

2016-04-11

Linearly typed dyadic group sessions for building multiparty sessions

H. Xi, H. Wu. 2016. "Linearly Typed Dyadic Group Sessions for Building Multiparty Sessions." <https://arxiv.org/abs/1604.04501>
<https://hdl.handle.net/2144/45001>

"Downloaded from OpenBU. Boston University's institutional repository."

Linearly Typed Dyadic Group Sessions for Building Multiparty Sessions

Hongwei Xi Hanwen Wu

Boston University
{hwx,hww}@cs.bu.edu

Abstract

Traditionally, each party in a (dyadic or multiparty) session implements exactly one role specified in the type of the session. We refer to this kind of session as an individual session (i-session). As a generalization of i-session, a group session (g-session) is one in which each party may implement a group of roles based on one channel. In particular, each of the two parties involved in a dyadic g-session implements either a group of roles or its complement. In this paper, we present a formalization of g-sessions in a multi-threaded lambda-calculus (MTLC) equipped with a linear type system, establishing for the MTLC both type preservation and global progress. As this formulated MTLC can be readily embedded into ATS, a full-fledged language with a functional programming core that supports both dependent types (of DML-style) and linear types, we obtain a direct implementation of linearly typed g-sessions in ATS. The primary contribution of the paper lies in both of the identification of g-sessions as a fundamental building block for multiparty sessions and the theoretical development in support of this identification.

1. Introduction

In broad terms, a session is a sequence of interactions between two or more concurrently running programs (often referred to as parties), and a session type is a form of type for specifying (or classifying) sessions. Traditionally, each party in a session implements exactly one role in the session type assigned to the session. For instance, each of the two parties in a dyadic session implements either the role of a client or the role of a server. Let us suppose that there are more than two roles in a session type (e.g., seller, buyer 1, and buyer 2). Conceptually, we can assign this session type to a session in which one party may implement a group of roles. For instance, there may be two parties in the session such that one implements the role of seller and the other implements both of the roles of buyer 1 and buyer 2. We coin the name *g-session* (for group session) to refer to a session in which a party may implement multiple roles. In contrast, a session is referred to as an *i-session* (for individual session) if each party in the session implements exactly one role. Therefore, an i-session is just a special case of g-session where each involved group is a singleton. As far as we can tell, this form

```

fun P() = let
  val () =
    channel_send(CH, I1, 0, 1) // send to Q
  val () =
    channel_send(CH, I2, 0, 1) // send to Q
  val b0 = channel_recv(CH, 1, 0) // recv from Q
  val () = channel_close(CH) // close the P-end of CH
in b0 end (* end of [P] *)

fun Q() = let
  val i1 =
    channel_recv(CH, 0, 1) // recv from P
  val i2 =
    channel_recv(CH, 0, 1) // recv from P
  val () =
    channel_send(CH, i1 < i2, 1, 0) // send to P
  val () = channel_close(CH) // close the Q-end of CH
in () end (* end of [Q] *)

```

Figure 1. Some pseudo code in ML-like syntax

of generalization from (dyadic) i-sessions to (dyadic) g-sessions is novel.

As an example (for clarifying basic concepts), let us assume that a dyadic session consists of two running programs (parties) P and Q that are connected with a bidirectional channel. From the perspective of P, the channel (that is, the endpoint at P's side) may be specified by a term sequence of the following form:

$$\text{snd}(\text{int}) :: \text{snd}(\text{int}) :: \text{rcv}(\text{bool}) :: \text{nil}$$

which means that an integer is to be sent, another integer is to be sent, a boolean is to be received, and finally the channel is to be closed. Clearly, from the perspective of Q, the channel (that is, the endpoint at Q's side) should be specified by the following term sequence:

$$\text{rcv}(\text{int}) :: \text{rcv}(\text{int}) :: \text{snd}(\text{bool}) :: \text{nil}$$

which means precisely the dual of what the previous term sequence does. We may think of P as a client who sends two integers to the server Q and then receives from Q either true or false depending on whether or not the first sent integer is less than the second one. A simple but crucial observation is that the above two term sequences can be unified as follows:

$$\text{msg}(0, 1, \text{int}) :: \text{msg}(0, 1, \text{int}) :: \text{msg}(1, 0, \text{bool}) :: \text{nil}$$

where 0 and 1 refer to the two roles implemented by P and Q, respectively. Given a type T , $\text{msg}(i, j, T)$ means a value of the type T is transferred from the party implementing role i to the one implementing role j , where both i and j range over 0 and 1.

In Figure 1, we present some pseudo code showing a plausible way to implement the programs P and Q. Please note that the functions P and Q, though written together here, can be written

in separate contexts. We use `CH` to refer to a channel available in the surrounding context of the code and `I1` and `I2` for two integers; the functions `channel_send` and `channel_recv` are for sending and receiving data via a given channel, and `channel_close` for closing a given channel.

Let us now sketch a way to make the above pseudo code type-check. Given an integer i and a session type S , let $\mathbf{chan}(i, S)$ be the type for a channel of role i , that is, a channel held by a party for implementing role i . We can assign the following type to `channel_send`:

$$(!\mathbf{chan}(i, \text{msg}(i, j, T) :: S) \gg \mathbf{chan}(i, S), \mathbf{int}(i), \mathbf{int}(j), T) \rightarrow \mathbf{1}$$

where $i \neq j$ is assumed, and $\mathbf{int}(i)$ and $\mathbf{int}(j)$ are singleton types for integers equal to i and j , respectively, and T and S stand for a type and a session type, respectively. Basically, this type¹ means that calling `channel_send` on a channel of the type $\mathbf{chan}(i, \text{msg}(i, j, T) :: S)$, integer i , integer j and a value of the type T returns a unit while *changing* the type of the channel to $\mathbf{chan}(i, S)$. Clearly, \mathbf{chan} is required to be a linear type constructor for this to make sense. As can be expected, the type assigned to `channel_recv` should be of the following form:

$$(!\mathbf{chan}(j, \text{msg}(i, j, T) :: S) \gg \mathbf{chan}(j, S), \mathbf{int}(i), \mathbf{int}(j)) \rightarrow T$$

where $i \neq j$ is assumed. This type essentially indicates that calling `channel_recv` on a channel of the type $\mathbf{chan}(j, \text{msg}(i, j, T) :: S)$, integer i and integer j returns a value of the type T while *changing* the type of the channel to $\mathbf{chan}(j, S)$. As for `channel_close`, it is assigned the following type:

$$(\mathbf{chan}(i, \text{nil})) \rightarrow \mathbf{1}$$

indicating that calling `channel_close` on a channel consumes the channel (so that the channel is no longer available for use).

Given an integer i (representing a role) and a session type S , the type $\mathbf{chan}(i, S)$ for single-role channels can be naturally transitioned into one of the form $\mathbf{chan}(G, S)$ for multirole channels, where G stands for a finite set of integers (representing roles). The fundamental issue to be addressed in this transition is to figure out a consistent interpretation for each term $\text{msg}(i, j, T)$ by a party based on the group of roles it implements. Assume there exists a fixed set of n roles ranging from 0 to $n - 1$ for some $n \geq 2$. For each G , we use \bar{G} for the complement of G , which consists of all of the natural numbers less than n that are not in G . We have the following four scenarios for interpreting $\text{msg}(i, j, T)$ based on the group G of roles implemented by a party:

- Assume $i \in G$ and $j \in G$. Then $\text{msg}(i, j, T)$ is interpreted as an internal message, and it is ignored.
- Assume $i \in G$ and $j \notin G$. Then $\text{msg}(i, j, T)$ is interpreted as sending a value of the type T by the party implementing G to the party implementing \bar{G} .
- Assume $i \notin G$ and $j \in G$. Then $\text{msg}(i, j, T)$ is interpreted as receiving a value of the type T by the party implementing G from the party implementing \bar{G} .
- Assume $i \notin G$ and $j \notin G$. Then $\text{msg}(i, j, T)$ is interpreted as an external message, and it is ignored.

With this interpretation, `channel_send` can be assigned the following type:

$$(!\mathbf{chan}(G, \text{msg}(i, j, T) :: S) \gg \mathbf{chan}(G, S), \mathbf{int}(i), \mathbf{int}(j), T) \rightarrow \mathbf{1}$$

where $i \in G$ and $j \notin G$ is assumed; `channel_recv` can be assigned the following type:

$$(!\mathbf{chan}(G, \text{msg}(i, j, T) :: S) \gg \mathbf{chan}(G, S), \mathbf{int}(i), \mathbf{int}(j)) \rightarrow T$$

¹Strictly speaking, this type should be referred to as a type schema as it contains occurrences of meta-variables.

where $i \notin G$ and $j \in G$ is assumed. As for `channel_close`, the following type is assigned:

$$(\mathbf{chan}(G, \text{nil})) \rightarrow \mathbf{1}$$

While transitioning single-role channels into multirole channels may seem mostly intuitive, there are surprises. In particular, we have the following result for justifying the use of multirole channels as a building block for implementing multiparty sessions (that involve more than 2 parties):

THEOREM 1.1. *Assume that ch_0 and ch_1 are two multirole channels (held by a party belonging to two sessions) of the types $\mathbf{chan}(G_0, S)$ and $\mathbf{chan}(G_1, S)$, respectively, where G_0 and G_1 are disjoint. Then there is a generic method for building a multirole channel ch_2 of the type $\mathbf{chan}(G_0 \cap G_1, S)$ such that each message received on one of ch_0 , ch_1 and ch_2 can be forwarded onto one of the other two in a type-correct manner, where ch_2' refers to the dual of ch_2 .*

The significance of Theorem 1.1 will be elaborated later on. Intuitively, this theorem justifies some form of “wiring” to allow two existing channels to be connected to provide the behavior of new channel (related to them in some way), enabling a multiparty session to be built based on dyadic g-sessions.

ATS [5, 27] is a full-fledged language with a functional programming core based on ML that supports both dependent types (of DML-style [28, 29]) and linear types. Its highly expressive type system makes it largely straightforward to implement session types in ATS (e.g., based on the outline given above) if our concern is primarily about type-correctness. For instance, there have already been implementations of session types in Haskell (e.g., [17, 19]) and elsewhere that offer type-correctness. However, mere type-correctness is inadequate. We are to establish formally the property that concurrency based on session types (formulated in this paper) can never result in deadlocking, which is often referred to as *global progress*. There have been many formalizations of session types in the literature (e.g., [4, 8, 11, 12, 23, 24, 26]). Often the dynamics formulated in a formalization of session types is based on π -calculus [16] or its variants/likes. We instead use multi-threaded λ -calculus (MTLC) as a basis for formalizing session types as such a formalization is particularly suitable for guiding implementation (due to its being less abstract and more operational).

The rest of the paper is organized as follows. In Section 2, we formulate a multi-threaded λ -calculus MTLC_0 equipped with a simple linear type system, setting up the basic machinery for further development. We then extend MTLC_0 to MTLC_{ch} in Section 3 with support for session types and establish both type preservation and global progress for MTLC_{ch} . In Section 4, we establish a key theorem needed for building multiparty sessions based on dyadic sessions. We present a few commonly used constructors for session types in Section 5 and then briefly mention the implementation of a classic example of 3-party sessions in Section 6. We also mention some key steps taken in both of our implementations of session-typed channels in ATS and in Erlang in Section 7. Lastly, we discuss some closely related work in Section 8 and then conclude.

The primary contribution of the paper lies in both of the identification of g-sessions as a fundamental building block for multiparty sessions and the theoretical development in support of this identification. We consider the formulation and proof of Theorem 1.1 a particularly important part of this contribution.

2. MTLC_0 with Linear Types

We first present a multi-threaded lambda-calculus MTLC_0 equipped with a simple linear type system, setting up the basic machinery for further development. The dynamic semantics of MTLC_0 can essentially be seen as an abstract form of evaluation of multi-threaded programs.

expr.	e	$::=$	$x \mid f \mid rc \mid c(\vec{e}) \mid$ $\langle \rangle \mid \langle e_1, e_2 \rangle \mid \mathbf{fst}(e) \mid \mathbf{snd}(e) \mid$ $\mathbf{let} \langle x_1, x_2 \rangle = e_1 \mathbf{in} e_2 \mathbf{end} \mid$ $\mathbf{lam} x. e \mid \mathbf{app}(e_1, e_2) \mid \mathbf{fix} f. v$
values	v	$::=$	$x \mid rc \mid cc(\vec{v}) \mid \langle \rangle \mid \langle v_1, v_2 \rangle \mid \mathbf{lam} x. e$
types	T	$::=$	$\delta \mid \mathbf{1} \mid T_1 * T_2 \mid \hat{T}_1 \rightarrow_i \hat{T}_2$
viewtypes	\hat{T}	$::=$	$\hat{\delta} \mid T \mid \hat{T}_1 \otimes \hat{T}_2 \mid \hat{T}_1 \rightarrow_i \hat{T}_2$
int. expr. ctx.	Γ	$::=$	$\emptyset \mid \Gamma, xf : T$
lin. expr. ctx.	Δ	$::=$	$\emptyset \mid \Delta, x : \hat{T}$

Figure 2. Some syntax for MTLC_0

Some syntax of MTLC_0 is given in Figure 2. We use x for a lam-variable and f for a fix-variable, and xf for either a lam-variable or a fix-variable. Note that a lam-variable is considered a value but a fix-variable is not. We use rc for constant resources and c for constants, which include both constant functions cf and constant constructors cc . We treat resources in MTLC_0 abstractly and will later introduce communication channels as a concrete form of resources. The meaning of various standard forms of expressions in MTLC_0 should be intuitively clear. We may refer to a closed expression (containing no free variables) as a *program*.

We use T and \hat{T} for (non-linear) types and (linear) viewtypes, respectively, and refer \hat{T} to as a true viewtype if it is a viewtype but not a type. We use δ and $\hat{\delta}$ for base types and base viewtypes, respectively. For instance, **bool** is the base type for booleans and **int** for integers. We also assume the availability of integer constants when forming types. For instance, we may have a type $\mathbf{int}(i)$ for each integer constant i , which can only be assigned to a (dynamic) value equal to integer i .

For a simplified presentation, we do not introduce any concrete base viewtypes in MTLC_0 . We assume a signature SIG for assigning a viewtype to each constant resource rc and a constant type (c-type) schema of the form $(\hat{T}_1, \dots, \hat{T}_n) \Rightarrow \hat{T}$ to each constant. For instance, we may have a constant function *iadd* of the following c-type schema:

$$(\mathbf{int}(i), \mathbf{int}(j)) \rightarrow \mathbf{int}(i + j)$$

where i and j are meta-variables ranging over integer constants; each occurrence of *iadd* in a program is given a c-type that is an instance of the c-type schema assigned to *iadd*.

Note that a type is always considered a viewtype. Let \hat{T}_1 and \hat{T}_2 be two viewtypes. The type constructor \otimes is based on multiplicative conjunction in linear logic. Intuitively, if a resource is assigned the viewtype $\hat{T}_1 \otimes \hat{T}_2$, then the resource is a conjunction of two resources of viewtypes \hat{T}_1 and \hat{T}_2 . The type constructor \rightarrow_i is essentially based on linear implication \multimap in linear logic. Given a function of the viewtype $\hat{T}_1 \rightarrow_i \hat{T}_2$ and a value of the viewtype \hat{T}_1 , applying the function to the value yields a result of the viewtype \hat{T}_2 while the function itself is consumed. If the function is of the type $\hat{T}_1 \rightarrow_i \hat{T}_2$, then applying the function does not consume it. The subscript i in \rightarrow_i is often dropped, that is, \rightarrow is assumed to be \rightarrow_i by default. The meaning of various forms of types and viewtypes is to be made clear and precise when the rules are presented for assigning viewtypes to expressions in MTLC_0 .

There is a special constant function *thread.create* in MTLC_0 for thread creation, which is assigned the following interesting c-type:

$$\mathbf{thread.create} : (\mathbf{1} \rightarrow_i \mathbf{1}) \Rightarrow \mathbf{1}$$

A function of the type $\mathbf{1} \rightarrow_i \mathbf{1}$ is a procedure that takes no arguments and returns no result (when its evaluation terminates). Given that $\mathbf{1} \rightarrow_i \mathbf{1}$ is a true viewtype, a procedure of this type may contain resources and thus must be called exactly once. The operational semantics of *thread.create* is to be formally defined later.

A variety of mappings, finite or infinite, are to be introduced in the rest of the presentation. We use $[\]$ for the empty mapping and

$\rho(rc)$	$=$	$\{rc\}$
$\rho(c(e_1, \dots, e_n))$	$=$	$\rho(e_1) \uplus \dots \uplus \rho(e_n)$
$\rho(xf)$	$=$	\emptyset
$\rho(\langle \rangle)$	$=$	\emptyset
$\rho(\langle e_1, e_2 \rangle)$	$=$	$\rho(e_1) \uplus \rho(e_2)$
$\rho(\mathbf{fst}(e))$	$=$	$\rho(e)$
$\rho(\mathbf{snd}(e))$	$=$	$\rho(e)$
$\rho(\mathbf{if}(e_0, e_1, e_2))$	$=$	$\rho(e_0) \uplus \rho(e_1)$
$\rho(\mathbf{let} \langle x_1, x_2 \rangle = e_1 \mathbf{in} e_2 \mathbf{end})$	$=$	$\rho(e_1) \uplus \rho(e_2)$
$\rho(\mathbf{lam} x. e)$	$=$	$\rho(e)$
$\rho(\mathbf{app}(e_1, e_2))$	$=$	$\rho(e_1) \uplus \rho(e_2)$
$\rho(\mathbf{fix} f. v)$	$=$	$\rho(v)$

Figure 3. The definition of $\rho(\cdot)$

$[i_1, \dots, i_n \mapsto o_1, \dots, o_n]$ for the finite mapping that maps i_k to o_k for $1 \leq k \leq n$. Given a mapping m , we write $\mathbf{dom}(m)$ for the domain of m . If $i \notin \mathbf{dom}(m)$, we use $m[i \mapsto o]$ for the mapping that extends m with a link from i to o . If $i \in \mathbf{dom}(m)$, we use $m \setminus i$ for the mapping obtained from removing the link from i to $m(i)$ in m , and $m[i := o]$ for $(m \setminus i)[i \mapsto o]$, that is, the mapping obtained from replacing the link from i to $m(i)$ in m with another link from i to o .

We define a function $\rho(\cdot)$ in Figure 3 to compute the multiset (that is, bag) of constant resources in a given expression. Note that \uplus denotes the multiset union. In the type system of MTLC_0 , it is to be guaranteed that $\rho(e_1)$ equals $\rho(e_2)$ whenever an expression of the form $\mathbf{if}(e_0, e_1, e_2)$ is constructed, and this justifies $\rho(\mathbf{if}(e_0, e_1, e_2))$ being defined as $\rho(e_0) \uplus \rho(e_1)$.

We use R to range over finite multisets of resources. Therefore, R can also be regarded as a mapping from resources to natural numbers: $R(rc) = n$ means that there are n occurrences of rc in R . It is clear that we may not combine resources arbitrarily. For instance, we may want to exclude the combination of one resource stating integer 0 at a location L and another one stating integer 1 at the same location. We fix an abstract collection \mathbf{RES} of finite multisets of resources and assume the following:

- $\emptyset \in \mathbf{RES}$.
- For any R_1 and R_2 , $R_2 \in \mathbf{RES}$ if $R_1 \in \mathbf{RES}$ and $R_2 \subseteq R_1$, where \subseteq is the subset relation on multisets.

We say that R is a valid multiset of resources if $R \in \mathbf{RES}$ holds.

In order to formalize threads, we introduce a notion of *pools*. Conceptually, a pool is just a collection of programs (that is, closed expressions). We use Π for pools, which are formally defined as finite mappings from thread ids (represented as natural numbers) to (closed) expressions in MTLC_0 such that 0 is always in the domain of such mappings. Given a pool Π and $tid \in \mathbf{dom}(\Pi)$, we refer to $\Pi(tid)$ as a thread in Π whose id equals tid . In particular, we refer to $\Pi(0)$ as the main thread in Π . The definition of $\rho(\cdot)$ is extended as follows to compute the multiset of resources in a given pool:

$$\rho(\Pi) = \uplus_{tid \in \mathbf{dom}(\Pi)} \rho(\Pi(tid))$$

We are to define a relation on pools in Section 2.2 to simulate multi-threaded program execution.

2.1 Static Semantics

We present typing rules for MTLC_0 in this section. It is required that each variable occur at most once in an intuitionistic (linear) expression context Γ (Δ), and thus Γ (Δ) can be regarded as a finite mapping. Given Γ_1 and Γ_2 such that $\mathbf{dom}(\Gamma_1) \cap \mathbf{dom}(\Gamma_2) = \emptyset$, we write (Γ_1, Γ_2) for the union of Γ_1 and Γ_2 . The same notation also applies to linear expression contexts (Δ). Given an intuitionistic expression context Γ and a linear expression context Δ , we can form a combined expression context $(\Gamma; \Delta)$ if $\mathbf{dom}(\Gamma) \cap \mathbf{dom}(\Delta) = \emptyset$.

$$\begin{array}{c}
\frac{\text{SIG} \models rc : \hat{\delta}}{\Gamma; \emptyset \vdash rc : \hat{\delta}} \text{ (ty-res)} \\
\frac{\text{SIG} \models c : (\hat{T}_1, \dots, \hat{T}_n) \Rightarrow \hat{T} \quad \Gamma; \Delta_i \vdash e_i : \hat{T}_i \text{ for } 1 \leq i \leq n}{\Gamma; \Delta_1, \dots, \Delta_n \vdash c(e_1, \dots, e_n) : \hat{T}} \text{ (ty-cst)} \\
\frac{}{(\Gamma, xf : T; \emptyset) \vdash xf : T} \text{ (ty-var-i)} \\
\frac{}{(\Gamma; \emptyset, x : \hat{T}) \vdash x : \hat{T}} \text{ (ty-var-l)} \\
\frac{\Gamma; \Delta_0 \vdash e_0 : \mathbf{bool} \quad \Gamma; \Delta \vdash e_1 : \hat{T} \quad \Gamma; \Delta \vdash e_2 : \hat{T} \quad \rho(e_1) = \rho(e_2)}{\Gamma; \Delta_0, \Delta \vdash \mathbf{if}(e_0, e_1, e_2) : \hat{T}} \text{ (ty-if)} \\
\frac{}{\Gamma; \emptyset \vdash \langle \rangle : \mathbf{1}} \text{ (ty-unit)} \\
\frac{\Gamma; \Delta_1 \vdash e_1 : T_1 \quad \Gamma; \Delta_2 \vdash e_2 : T_2}{\Gamma; \Delta_1, \Delta_2 \vdash \langle e_1, e_2 \rangle : T_1 * T_2} \text{ (ty-tup-i)} \\
\frac{\Gamma; \Delta \vdash e : T_1 * T_2}{\Gamma; \Delta \vdash \mathbf{fst}(e) : T_1} \text{ (ty-fst)} \quad \frac{\Gamma; \Delta \vdash e : T_1 * T_2}{\Gamma; \Delta \vdash \mathbf{snd}(e) : T_2} \text{ (ty-snd)} \\
\frac{\Gamma; \Delta_1 \vdash e_1 : \hat{T}_1 \quad \Gamma; \Delta_2 \vdash e_2 : \hat{T}_2}{\Gamma; \Delta_1, \Delta_2 \vdash \langle e_1, e_2 \rangle : \hat{T}_1 \otimes \hat{T}_2} \text{ (ty-tup-l)} \\
\frac{\Gamma; \Delta_1 \vdash e_1 : \hat{T}_1 \otimes \hat{T}_2 \quad \Gamma; \Delta_2, x_1 : \hat{T}_1, x_2 : \hat{T}_2 \vdash e_2 : \hat{T}}{\Gamma; \Delta_1, \Delta_2 \vdash \mathbf{let} \langle x_1, x_2 \rangle = e_1 \text{ in } e_2 \text{ end} : \hat{T}} \text{ (ty-tup-l-elim)} \\
\frac{(\Gamma; \Delta), x : \hat{T}_1 \vdash e : \hat{T}_2}{\Gamma; \Delta \vdash \mathbf{lam} x. e : \hat{T}_1 \rightarrow \hat{T}_2} \text{ (ty-lam-l)} \\
\frac{\Gamma; \Delta_1 \vdash e_1 : \hat{T}_1 \rightarrow \hat{T}_2 \quad \Gamma; \Delta_2 \vdash e_2 : \hat{T}_1}{\Gamma; \Delta_1, \Delta_2 \vdash \mathbf{app}(e_1, e_2) : \hat{T}_2} \text{ (ty-app-l)} \\
\frac{(\Gamma; \emptyset), x : \hat{T}_1 \vdash e : \hat{T}_2 \quad \rho(e) = \emptyset}{\Gamma; \emptyset \vdash \mathbf{lam} x. e : \hat{T}_1 \rightarrow \hat{T}_2} \text{ (ty-lam-i)} \\
\frac{\Gamma; \Delta_1 \vdash e_1 : \hat{T}_1 \rightarrow \hat{T}_2 \quad \Gamma; \Delta_2 \vdash e_2 : \hat{T}_1}{\Gamma; \Delta_1, \Delta_2 \vdash \mathbf{app}(e_1, e_2) : \hat{T}_2} \text{ (ty-app-i)} \\
\frac{\Gamma, f : T; \emptyset \vdash v : T}{\Gamma; \emptyset \vdash \mathbf{fix} f. v : T} \text{ (ty-fix)} \\
\frac{(\emptyset; \emptyset) \vdash \Pi(0) : \hat{T} \quad (\emptyset; \emptyset) \vdash \Pi(\mathit{rid}) : \mathbf{1} \text{ for each } 0 < \mathit{tid} \in \mathbf{dom}(\Pi)}{\vdash \Pi : \hat{T}} \text{ (ty-pool)}
\end{array}$$

Figure 4. The typing rules for MTLC_0

Given $(\Gamma; \Delta)$, we may write $(\Gamma; \Delta), x : \hat{T}$ for either $(\Gamma; \Delta, x : \hat{T})$ or $(\Gamma, x : \hat{T}; \Delta)$ (if \hat{T} is actually a type).

A typing judgment in MTLC_0 is of the form $(\Gamma; \Delta) \vdash e : \hat{T}$, meaning that e can be assigned the viewtype \hat{T} under $(\Gamma; \Delta)$. The typing rules for MTLC_0 are listed in Figure 4. In the rule **(ty-cst)**, the following judgment requires that the c-type be an instance of the c-type schema assigned to c in SIG :

$$\text{SIG} \models c : (\hat{T}_1, \dots, \hat{T}_n) \Rightarrow \hat{T}$$

For the constant function iadd mentioned previously, the following judgment is valid:

$$\text{SIG} \models \mathit{iadd} : (\mathbf{int}(0), \mathbf{int}(1)) \Rightarrow \mathbf{int}(0 + 1)$$

and the following judgment is valid as well:

$$\text{SIG} \models \mathit{iadd} : (\mathbf{int}(2), \mathbf{int}(2)) \Rightarrow \mathbf{int}(2 + 2)$$

By inspecting the typing rules in Figure 4, we can readily see that a closed value cannot contain any resources if the value itself can be assigned a type (rather than a linear type). More formally, we have the following proposition:

PROPOSITION 2.1. *If $(\emptyset; \emptyset) \vdash v : T$ is derivable, then $\rho(v) = \emptyset$.*

This proposition plays a fundamental role in MTLC_0 : The rules in Figure 4 are actually so formulated in order to make it hold.

The following lemma, which is often referred to as *Lemma of Canonical Forms*, relates the form of a value to its type:

LEMMA 2.2. *Assume that $(\emptyset; \emptyset) \vdash v : \hat{T}$ is derivable.*

- If $\hat{T} = \delta$, then v is of the form $cc(v_1, \dots, v_n)$.
- If $\hat{T} = \hat{\delta}$, then v is of the form rc or $cc(v_1, \dots, v_n)$.
- If $\hat{T} = \mathbf{1}$, then v is $\langle \rangle$.
- If $\hat{T} = T_1 * T_2$ or $\hat{T} = \hat{T}_1 \otimes \hat{T}_2$, then v is of the form $\langle v_1, v_2 \rangle$.
- If $\hat{T} = \hat{T}_1 \rightarrow_i \hat{T}_2$ or $\hat{T} = \hat{T}_1 \rightarrow_i \hat{T}_2$, then v is of the form $\mathbf{lam} x. e$.

Proof By an inspection of the rules in Figure 4. ■

We use θ for substitution on variables xf :

$$\theta ::= [] \mid \theta[x \mapsto v] \mid \theta[f \mapsto e]$$

For each θ , we define the multiset $\rho(\theta)$ of resources in θ as follows:

$$\rho(\theta) = \uplus_{xf \in \mathbf{dom}(\theta)} \rho(\theta(xf))$$

Given an expression e , we use $e[\theta]$ for the result of applying θ to e , which is defined in a standard manner. We write $(\Gamma_1; \Delta_1) \vdash \theta : (\Gamma_2; \Delta_2)$ to mean that

- $\mathbf{dom}(\theta) = \mathbf{dom}(\Gamma_2) \cup \mathbf{dom}(\Delta_2)$, and
- $(\Gamma_1; \emptyset) \vdash \theta(xf) : \Gamma_2(xf)$ is derivable for each $xf \in \Gamma_2$, and
- there exists a linear expression context $\Delta_{1,x}$ for each $x \in \mathbf{dom}(\Delta_2)$ such that $(\Gamma_1; \Delta_{1,x}) \vdash \theta(x) : \Delta_2(x)$ is derivable, and
- $\Delta_1 = \bigcup_{x \in \mathbf{dom}(\Delta_2)} \Delta_{1,x}$

The following lemma, which is often referred to as *Substitution Lemma*, is needed to establish the soundness of the type system of MTLC_0 :

LEMMA 2.3. *Assume $(\Gamma_1; \Delta_1) \vdash \theta : (\Gamma_2; \Delta_2)$ and $(\Gamma_2; \Delta_2) \vdash e : \hat{T}$. Then $(\Gamma_1; \Delta_1) \vdash e[\theta] : \hat{T}$ is derivable and $\rho(e[\theta]) = \rho(e) \uplus \rho(\theta)$.*

Proof By induction on the derivation of $(\Gamma_2; \Delta_2) \vdash e : \hat{T}$. ■

2.2 Dynamic Semantics

We present evaluation rules for MTLC_0 in this section. The evaluation contexts in MTLC_0 are defined below:

$$\begin{array}{l}
\text{eval. ctx. } E ::= \\
[] \mid c(\vec{v}, E, \vec{e}) \mid \mathbf{if}(E, e_1, e_2) \mid \\
\langle E, e \rangle \mid \langle v, E \rangle \mid \mathbf{let} \langle x_1, x_2 \rangle = E \text{ in } e \text{ end} \mid \\
\mathbf{fst}(E) \mid \mathbf{snd}(E) \mid \mathbf{app}(E, e) \mid \mathbf{app}(v, E)
\end{array}$$

Given an evaluation context E and an expression e , we use $E[e]$ for the expression obtained from replacing the only hole $[]$ in E with e .

DEFINITION 2.4. *We define pure redexes and their reducts as follows.*

- $\mathbf{if}(\mathit{true}, e_1, e_2)$ is a pure redex whose reduct is e_1 .
- $\mathbf{if}(\mathit{false}, e_1, e_2)$ is a pure redex whose reduct is e_2 .
- $\mathbf{let} \langle x_1, x_2 \rangle = \langle v_1, v_2 \rangle \text{ in } e \text{ end}$ is a pure redex whose reduct is $e[x_1, x_2 \mapsto v_1, v_2]$.
- $\mathbf{fst}(\langle v_1, v_2 \rangle)$ is a pure redex whose reduct is v_1 .
- $\mathbf{snd}(\langle v_1, v_2 \rangle)$ is a pure redex whose reduct is v_2 .
- $\mathbf{app}(\mathbf{lam} x. e, v)$ is a pure redex whose reduct is $e[x \mapsto v]$.
- $\mathbf{fix} f. v$ is a pure redex whose reduct is $v[f \mapsto \mathbf{fix} f. v]$.

Evaluating calls to constant functions is of particular importance in MTLC_0 . Assume that cf is a constant function of arity n . The expression $cf(v_1, \dots, v_n)$ is an *ad-hoc* redex if cf is defined at v_1, \dots, v_n , and any value of $cf(v_1, \dots, v_n)$ is a reduct of $cf(v_1, \dots, v_n)$. For instance, $1 + 1$ is an ad hoc redex and 2 is its sole reduct. In contrast, $1 + \mathit{true}$ is not a redex as it is undefined. We can

even have non-deterministic constant functions. For instance, we may assume that the ad-hoc redex $\text{randbit}()$ can evaluate to both 0 and 1.

Let e be a well-typed expression of the form $cf(v_1, \dots, v_n)$ and $\rho(e) \subseteq R$ holds for some valid R (that is, $R \in \mathbf{RES}$). We always assume that there exists a redex v in M TLC_0 for $cf(v_1, \dots, v_n)$ such that $(R \setminus \rho(e)) \uplus \rho(v) \in \mathbf{RES}$. By doing so, we are able to give a presentation with much less clutter.

DEFINITION 2.5. *Given expressions e_1 and e_2 , we write $e_1 \rightarrow e_2$ if $e_1 = E[e]$ and $e_2 = E[e']$ for some E, e and e' such that e' is a reduct of e , and we may say that e_1 evaluates or reduces to e_2 purely if e is a pure redex.*

Note that resources may be generated as well as consumed when ad-hoc reductions occur. This is an essential issue of great importance in any linear type system designed to support practical programming.

DEFINITION 2.6. *Given pools Π_1 and Π_2 , the relation $\Pi_1 \rightarrow \Pi_2$ is defined according to the following rules:*

$$\frac{e_1 \rightarrow e_2}{\Pi[tid \mapsto e_1] \rightarrow \Pi[tid \mapsto e_2]} \text{ (PR0)}$$

$$\frac{\Pi(tid_0) = E[\text{thread_create}(\text{lam } x. e)]}{\Pi \rightarrow \Pi[tid_0 := E[\langle \rangle]][tid \mapsto \text{app}(\text{lam } x. e, \langle \rangle)]} \text{ (PR1)}$$

$$\frac{tid > 0}{\Pi[tid \mapsto \langle \rangle] \rightarrow \Pi} \text{ (PR2)}$$

If a pool Π_1 evaluates to another pool Π_2 by the rule (PR0), then one program in Π_1 evaluates to its counterpart in Π_2 and the rest stay the same; if by the rule (PR1), then a fresh program is created; if by the rule (PR2), then a program (that is not the main program) is eliminated.

From this point on, we always (implicitly) assume that $\rho(\Pi) \in \mathbf{RES}$ holds whenever Π is well-typed. The soundness of the type system of M TLC_0 rests upon the following two theorems:

THEOREM 2.7. *(Subject Reduction on Pools) Assume that $\vdash \Pi_1 : \hat{T}$ is derivable and $\Pi_1 \rightarrow \Pi_2$ holds for some Π_2 satisfying $\rho(\Pi_2) \in \mathbf{RES}$. Then $\vdash \Pi_2 : \hat{T}$ is also derivable.*

Proof By structural induction on the derivation of $\vdash \Pi_1 : \hat{T}$. Note that Lemma 2.3 is needed. ■

THEOREM 2.8. *(Progress Property on Pools) Assume that $\vdash \Pi_1 : \hat{T}$ is derivable. Then we have the following possibilities:*

- Π_1 is a singleton mapping $[0 \mapsto v]$ for some v , or
- $\Pi_1 \rightarrow \Pi_2$ holds for some Π_2 such that $\rho(\Pi_2) \in \mathbf{RES}$.

Proof By structural induction on the derivation of $\vdash \Pi_1 : \hat{T}$. Note that Lemma 2.2 is needed. Essentially, we can readily show that $\Pi_1(tid)$ for any $tid \in \text{dom}(\Pi_1)$ is either a value or of the form $E[e]$ for some evaluation context E and redex e . If $\Pi_1(tid)$ is a value for some $tid > 0$, then this value must be $\langle \rangle$. So the rule (PR2) can be used to reduce Π_1 . If $\Pi_1(tid)$ is of the form $E[e]$ for some redex e , then the rule (PR0) can be used to reduce Π_1 . ■

By combining Theorem 2.7 and Theorem 2.8, we immediately conclude that the evaluation of a well-typed pool either leads to a pool that itself is a singleton mapping of the form $[0 \mapsto v]$ for some value v , or it goes on forever. In other words, M TLC_0 is type-sound.

3. Extending M TLC_0 with Channels

There is no support for communication between threads in M TLC_0 , making M TLC_0 uninteresting as a multi-threaded language. We extend M TLC_0 to M TLC_{ch} with support for synchronous communication channels in this section. Supporting asynchronous communication channels is certainly possible but would result in a more

involved theoretical development. We do support both synchronous and asynchronous session-typed communication channels in practice, though. In order to assign types to channels, we introduce session types as follows:

$$S ::= \text{nil} \mid \text{msg}(i, j) :: S$$

An empty session is specified by nil . Given integers i and j (representing roles), the precise meaning of the term $\text{msg}(i, j)$ is to be given later, which depends on the group of the roles implemented by a party. Intuitively speaking, this term refers to transferring a message of some kind from a party implementing the role i (and possibly others) to another one implementing the role j (and possibly others). Please notice that $\text{msg}(i, j)$ is written instead of $\text{msg}(i, j, \hat{T})$ for some viewtype \hat{T} . The omission of \hat{T} is solely for the purpose of a simplified presentation as the primary focus is on communications between parties in a session (rather than values transferred during communications).

Let us assume the availability of finite sets of integers for forming types. As another step towards a simplified presentation, we fix a set of roles $0, 1, \dots, nrole - 1$ for some natural number $nrole \geq 2$ and require that the roles mentioned in every session type belong to this set. Given a group G of roles and a session type S , we can form a linear base viewtype $\mathbf{chan}(G, S)$ for channels that are often referred to as G -channels; the party in a session that holds a G -channel is supposed to implement all of the roles in G .

The function chan.create for creating a channel is assigned the following c-type schema:

$$\text{chan.create} : (\mathbf{chan}(G, S) \rightarrow_1 \mathbf{1}) \Rightarrow \mathbf{chan}(\overline{G}, S)$$

where $G \neq \emptyset$ and $\overline{G} \neq \emptyset$ is assumed. Given a linear function of the type $\mathbf{chan}(G, S) \rightarrow_1 \mathbf{1}$ for some session type S , chan.create essentially creates two properly connected channels of the types $\mathbf{chan}(G, S)$ and $\mathbf{chan}(\overline{G}, S)$, and then starts a thread for evaluating the call that applies the function to the channel of the type $\mathbf{chan}(G, S)$ and then returns the other channel of the type $\mathbf{chan}(\overline{G}, S)$. The newly created two channels share the same id.

The function for sending onto a channel is given the following type schema:

$$\text{send} : (\mathbf{chan}(G, \text{msg}(i, j) :: S)) \Rightarrow \mathbf{chan}(G, S)$$

where $i \in G$ and $j \notin G$ is assumed. The function for receiving from a channel is given the following type schema:

$$\text{recv} : (\mathbf{chan}(G, \text{msg}(i, j) :: S)) \Rightarrow \mathbf{chan}(G, S)$$

where $i \notin G$ and $j \in G$ is assumed. The function for skipping an internal or external message is given the following type schema:

$$\text{skip} : (\mathbf{chan}(G, \text{msg}(i, j) :: S)) \Rightarrow \mathbf{chan}(G, S)$$

where either $i \in G$ and $j \in G$ is assumed or $i \notin G$ and $j \notin G$ is assumed. The function for closing a channel is given the following type schema:

$$\text{close} : (\mathbf{chan}(G, S)) \Rightarrow \mathbf{1}$$

Note that send and recv correspond to the functions `channel.send` and `channel.recv` mentioned in Section 1, respectively, and close corresponds to `channel.close`.

In M TLC_{ch} , there are resource constants ch_n^+ and ch_n^- referring to positive and negative channels, respectively, where n ranges over natural numbers. For each n , ch_n^+ and ch_n^- are dual to each other and their channel ids are n . We use ch^+ and ch^- for positive and negative channels, and ch for both. If ch^+ and ch^- appear in the same context, it is assumed (unless specified otherwise) that they refer to ch_n^+ and ch_n^- for the same id n .

Given a group G of roles, a G -channel is positive if $0 \in G$ and it is negative if $0 \notin G$. Note that calling chan.create creates a positive channel and a negative channel of the same id; one is passed to a newly created thread while the other is returned to the caller.

There are no new typing rules in MTLC_{ch} over MTLC_0 . Given G and S , we say that the type $\mathbf{chan}(G, S)$ matches the type $\mathbf{chan}(\overline{G}, S)$ and vice versa. In any type derivation of $\Pi : \hat{T}$ satisfying $\rho(\Pi) \in \mathbf{RES}$, the type assigned to a positive channel ch^+ is always required to match the one assigned to the corresponding negative channel ch^- of the same channel id. For evaluating pools in MTLC_0 , we have the following additional rules in MTLC_{ch} :

$$\frac{\Pi(tid_0) = E[\mathbf{chan_create}(\mathbf{lam} \ x. e)]}{\Pi \rightarrow \Pi[tid_0 := E[ch^-]] [tid \mapsto \mathbf{app}(\mathbf{lam} \ x. e, ch^+)]} \quad (\text{PR3})$$

$$\frac{\Pi(tid_1) = E_1[\mathbf{send}(ch^+)] \quad \Pi(tid_2) = E_2[\mathbf{recv}(ch^-)]}{\Pi \rightarrow \Pi[tid_1 := E_1[ch^+]] [tid_2 := E_2[ch^-]]} \quad (\text{PR4-send})$$

$$\frac{\Pi(tid_1) = E_1[\mathbf{recv}(ch^+)] \quad \Pi(tid_2) = E_2[\mathbf{send}(ch^-)]}{\Pi \rightarrow \Pi[tid_1 := E_1[ch^+]] [tid_2 := E_2[ch^-]]} \quad (\text{PR4-recv})$$

$$\frac{\Pi(tid_1) = E_1[\mathbf{skip}(ch^+)] \quad \Pi(tid_2) = E_2[\mathbf{skip}(ch^-)]}{\Pi \rightarrow \Pi[tid_1 := E_1[ch^+]] [tid_2 := E_2[ch^-]]} \quad (\text{PR4-skip})$$

$$\frac{\Pi(tid_1) = E_1[\mathbf{close}(ch^+)] \quad \Pi(tid_2) = E_2[\mathbf{close}(ch^-)]}{\Pi \rightarrow \Pi[tid_1 := E_1[\langle \rangle]] [tid_2 := E_2[\langle \rangle]]} \quad (\text{PR4-close})$$

For instance, the rule PR4-send states: If a program in a pool is of the form $E_1[\mathbf{send}(ch^+)]$ and another of the form $E_2[\mathbf{recv}(ch^-)]$, then this pool can be reduced to another pool by replacing these two programs with $E_1[ch^+]$ and $E_2[ch^-]$, respectively.

While Theorem 2.7 (Subject Reduction) can be readily established for MTLC_{ch} , Theorem 2.8 (Progress) requires some special treatment due to the presence of session-typed primitive functions $\mathbf{chan_create}$, \mathbf{send} , \mathbf{recv} , \mathbf{skip} , and \mathbf{close} .

A partial (ad-hoc) redex in MTLC_{ch} is of one of the following forms: $\mathbf{send}(ch)$, $\mathbf{recv}(ch)$, $\mathbf{skip}(ch)$, and $\mathbf{close}(ch)$. We say that $\mathbf{send}(ch^+)$ and $\mathbf{recv}(ch^-)$ match, and $\mathbf{recv}(ch^+)$ and $\mathbf{send}(ch^-)$ match, and $\mathbf{skip}(ch^+)$ and $\mathbf{skip}(ch^-)$ match, and $\mathbf{close}(ch^+)$ and $\mathbf{close}(ch^-)$ match. We can immediately prove in MTLC_{ch} that each well-typed program is either a value or of the form $E[e]$ for some evaluation context E and expression e that is either a redex or a partial redex. We refer to an expression as a *blocked* one if it is of the form $E[e]$ for some partial redex e . We say two blocked expressions $E_1[e_1]$ and $E_2[e_2]$ match if e_1 and e_2 are matching partial redexes. Clearly, a pool containing two matching blocked expressions can be reduced according to one of the rules PR4-send, PR4-recv, PR4-skip, and PR4-close.

Intuitively, a pool Π is deadlocked if $\Pi(tid)$ for $tid \in \mathbf{dom}(\Pi)$ are all blocked expressions but there are no matching ones among them, or if $\Pi(0)$ is a value and $\Pi(tid)$ for positive $tid \in \mathbf{dom}(\Pi)$ are all blocked expressions but there are no matching ones among them. The following lemma states that a well-typed pool in MTLC_{ch} can never be deadlocked:

LEMMA 3.1. (Deadlock-Freedom) *Let Π be a well-typed pool in MTLC_{ch} such that $\Pi(0)$ is either a value containing no channels or a blocked expression and $\Pi(tid)$ for each positive $tid \in \mathbf{dom}(\Pi)$ is a blocked expression. If Π is obtained from evaluating an initial pool containing no channels, then there exist two thread ids tid_1 and tid_2 such that $\Pi(tid_1)$ and $\Pi(tid_2)$ are matching blocked expressions.*

Note that it is entirely possible to encounter a scenario where the main thread in a pool returns a value containing a channel while another thread is waiting for something to be sent on the channel. Technically, we do not classify this scenario as a deadlocked one. There are many forms of values that contain channels. For instance, such a value can be a channel itself, or a closure-function containing a channel in its environment, or a compound value like a tuple that contains a channel as one part of it, etc. Clearly, any value containing a channel can only be assigned a true viewtype.

As a channel can be sent from one thread to another one, establishing Lemma 3.1 is conceptually challenging. The following technical approach to addressing the challenge is adopted from

some existing work on dyadic sessions types (for i-sessions) [30]. One may want skip the rest of this section when reading the paper for the first time.

Let us use \mathcal{M} for sets of (positive and negative) channels and M for a finite non-empty collection (that is, multiset) of such sets. We say that \mathcal{M} is *regular* if the sets in \mathcal{M} are pairwise disjoint and each pair of channels ch^+ and ch^- are either both included in the multiset union $\biguplus(\mathcal{M})$ of all the sets in \mathcal{M} or both excluded from it. Of course, $\biguplus(\mathcal{M})$ is the same as the set union $\bigcup(\mathcal{M})$ as the sets in \mathcal{M} are pairwise disjoint.

Let \mathcal{M} be a regular collection of channel sets. We say that \mathcal{M} *DF-reduces* to \mathcal{M}' via ch^+ if there exist M_1 and M_2 in \mathcal{M} such that $ch^+ \in M_1$ and $ch^- \in M_2$ and $\mathcal{M}' = (\mathcal{M} \setminus \{M_1, M_2\}) \cup \{M_{12}\}$, where $M_{12} = (M_1 \cup M_2) \setminus \{ch^+, ch^-\}$. We say that \mathcal{M} *DF-reduces* to \mathcal{M}' if \mathcal{M} *DF-reduces* to \mathcal{M}' via some ch^+ . We may write $\mathcal{M} \rightsquigarrow \mathcal{M}'$ to mean that \mathcal{M} *DF-reduces* to \mathcal{M}' . We say that \mathcal{M} is *DF-normal* if there is no \mathcal{M}' such that $\mathcal{M} \rightsquigarrow \mathcal{M}'$ holds.

PROPOSITION 3.2. *Let \mathcal{M} be a regular collection of channel sets. If \mathcal{M} is DF-normal, then each set in \mathcal{M} consists of an indefinite number of channel pairs ch^+ and ch^- . In other words, for each M in a DF-normal \mathcal{M} , a channel ch^+ is in M if and only if its dual ch^- is also in M .*

Proof The proposition immediately follows from the definition of DF-reduction \rightsquigarrow . ■

DEFINITION 3.3. *A regular collection \mathcal{M} of channel sets is DF-reducible if either (1) each set in \mathcal{M} is empty or (2) \mathcal{M} is not DF-normal and \mathcal{M}' is DF-reducible whenever $\mathcal{M} \rightsquigarrow \mathcal{M}'$ holds.*

We say that a channel set M is self-looping if it contains both ch^+ and ch^- for some ch^+ . Obviously, a regular collection \mathcal{M} of channel sets is not DF-reducible if there is a self-looping M in \mathcal{M} .

PROPOSITION 3.4. *Let \mathcal{M} be a regular collection of channel sets. If \mathcal{M} is DF-reducible and $\mathcal{M}' = \mathcal{M} \setminus \{\emptyset\}$, then \mathcal{M}' is also DF-reducible.*

Proof Straightforwardly. ■

PROPOSITION 3.5. *Let \mathcal{M} be a regular collection of channel sets. If $\mathcal{M} \rightsquigarrow \mathcal{M}'$ and \mathcal{M}' is DF-reducible, then \mathcal{M} is also DF-reducible.*

Proof Clearly, $\mathcal{M} \rightsquigarrow \mathcal{M}'$ via some ch^+ . Assume $\mathcal{M} \rightsquigarrow \mathcal{M}_1$ via ch_1^+ for some M_1 and ch_1^+ . If ch^+ and ch_1^+ are the same, then M_1 is DF-reducible as it is the same as \mathcal{M}' . Otherwise, it can be readily verified that there exists M'_1 such that $M_1 \rightsquigarrow M'_1$ via ch^+ and $\mathcal{M}' \rightsquigarrow M'_1$ via ch_1^+ . Clearly, the latter implies M'_1 being DF-reducible. Note that the size of M_1 is strictly less than that of \mathcal{M} . By induction hypothesis on M_1 , we have M_1 being DF-reducible. By definition, \mathcal{M} is DF-reducible. ■

PROPOSITION 3.6. *Let \mathcal{M} be a regular collection of channel sets that is DF-reducible. If M_1 and M_2 in \mathcal{M} contain ch^+ and ch^- , respectively, then $\mathcal{M}' = (\mathcal{M} \setminus \{M_1, M_2\}) \cup \{M'_1, M'_2\}$ is also DF-reducible, where $M'_1 = M_1 \setminus \{ch^+\}$ and $M'_2 = M_2 \setminus \{ch^-\}$.*

Proof The proposition follows from a straightforward induction on the size of the set union $\bigcup(\mathcal{M})$. ■

LEMMA 3.7. *Let \mathcal{M} be a regular collection of n channel sets M_1, \dots, M_n for some $n \geq 1$. If the union $\bigcup(\mathcal{M}) = M_1 \cup \dots \cup M_n$ contains at least n channel pairs $(ch_1^+, ch_1^-), \dots, (ch_n^+, ch_n^-)$, then \mathcal{M} is not DF-reducible.*

Proof By induction on n . If $n = 1$, then \mathcal{M} is not DF-reducible as M_1 is self-looping. Assume $n > 1$. If either M_1 or M_2 is self-looping, then \mathcal{M} is not DF-reducible. Otherwise, we may assume

that $ch_1^+ \in M_1$ and $ch_1^- \in M_2$ without loss of generality. Then \mathcal{M} DF-reduces to \mathcal{M}' via ch_1^+ for some \mathcal{M}' containing $n - 1$ channel sets. Note that $\bigcup(\mathcal{M}')$ contains at least $n - 1$ channel pairs $(ch_2^+, ch_2^-), \dots, (ch_n^+, ch_n^-)$. By induction hypothesis, \mathcal{M}' is not DF-reducible. So \mathcal{M} is not DF-reducible, either. ■

Given an expression e in $\text{MTLC}_{\text{ch}}^-$, we use $\rho_{CH}(e)$ for the set of channels contained in e . Given a pool Π in MTLC_{ch} , we use $\mathcal{R}_{CH}(\Pi)$ for the collection of $\rho_{CH}(\Pi(tid))$, where tid ranges over $\text{dom}(\Pi)$.

LEMMA 3.8. *If $\mathcal{R}_{CH}(\Pi)$ is DF-reducible and Π evaluates to Π' , then $\mathcal{R}_{CH}(\Pi')$ is also DF-reducible.*

Proof Note that $\mathcal{R}_{CH}(\Pi)$ and $\mathcal{R}_{CH}(\Pi')$ are the same unless Π evaluates to Π' according to one of the rules PR3, PR4-send, PR4-recv, PR4-skip, and PR4-close.

- For the rule PR3: We have $\mathcal{R}_{CH}(\Pi') \rightsquigarrow \mathcal{R}_{CH}(\Pi)$ via the newly introduced channel ch^+ . By Proposition 3.5, $\mathcal{R}_{CH}(\Pi')$ is DF-reducible.
- For the rule PR4-send: Let ch^+ be the channel on which a value is sent when Π evaluates to Π' . Note that this value can itself be a channel or contain a channel. We have $\mathcal{R}_{CH}(\Pi) \rightsquigarrow \mathcal{M}$ via ch^+ for some \mathcal{M} . So \mathcal{M} is DF-reducible by definition. Clearly, $\mathcal{R}_{CH}(\Pi') \rightsquigarrow \mathcal{M}$ via ch^+ as well. By Proposition 3.6, $\mathcal{R}_{CH}(\Pi')$ is DF-reducible.
- For the rule PR4-recv: This case is similar to the previous one.
- For the rule PR4-skip: This case is trivial as $\mathcal{R}_{CH}(\Pi)$ and $\mathcal{R}_{CH}(\Pi')$ are the same.
- For the rule PR4-close: We have that $\mathcal{R}_{CH}(\Pi')$ is DF-reducible by Proposition 3.6.

In order to fully appreciate the argument made in the case of PR4-send, one needs to imagine a scenario where a channel is actually transferred from one thread into another. While this scenario does not happen here due to the simplified version of type schemas assigned to *send* and *recv*, one can find the essential details in the original paper on DF-reducibility [30]. ■

We are now ready to give a proof for Lemma 3.1:

Proof Note that any channel, either positive or negative, can appear at most once in $\mathcal{R}_{CH}(\Pi)$, and a channel ch^+ appears in $\mathcal{R}_{CH}(\Pi)$ if and only if its dual ch^- also appears in $\mathcal{R}_{CH}(\Pi)$. In addition, any positive channel ch^+ being assigned a type of the form $\mathbf{chan}(G, S)$ in the type derivation of Π for some session type S mandates that its dual ch^- be assigned the type of the form $\mathbf{chan}(\bar{G}, S)$.

Assume that $\Pi(tid)$ is a blocked expression for each $tid \in \text{dom}(\Pi)$. If the partial redex in $\Pi(tid_1)$ involves a positive channel ch^+ while the partial redex in $\Pi(tid_2)$ involves its dual ch^- , then these two partial redexes must match. This is due to Π being well-typed. In other words, the ids of the channels involved in the partial redexes of $\Pi(tid)$ for $tid \in \text{dom}(\Pi)$ are all distinct. This simply implies that there are n channel pairs (ch^+, ch^-) in $\bigcup(\mathcal{R}_{CH}(\Pi))$ for some n greater than or equal to the size of Π . By Lemma 3.7, $\mathcal{R}_{CH}(\Pi)$ is not reducible. On the other hand, $\mathcal{R}_{CH}(\Pi)$ is reducible by Lemma 3.8 as Π_0 evaluates to Π (in many steps) and $\mathcal{R}_{CH}(\Pi_0)$ (containing only sets that are empty) is reducible. This contradiction indicates that there exist tid_1 and tid_2 such that $\Pi(tid_1)$ and $\Pi(tid_2)$ are matching blocked expressions. Therefore Π evaluates to Π' for some pool Π' according to one of the rules PR4-clos, PR4-send, and PR4-recv.

With Proposition 3.4, the case can be handled similarly where $\Pi(0)$ is a value containing no channels and $\Pi(tid)$ is a blocked expression for each positive $tid \in \text{dom}(\Pi)$. ■

Please assume for the moment that we would like to add into MTLC_{ch} a function *chan2_create* of the following type schema:

$$((\mathbf{chan}(G_1, S_1), \mathbf{chan}(G_2, S_2)) \rightarrow_1 \mathbf{1}) \Rightarrow (\mathbf{chan}(\bar{G}_1, S_1), \mathbf{chan}(\bar{G}_2, S_2)))$$

One may think of *chan2_create* as a “reasonable” generalization of *chan_create* that creates in a single call two channels instead of one. Unfortunately, adding *chan2_create* into MTLC_{ch} can potentially cause a deadlock. For instance, we can easily imagine a scenario where the first of the two channels (ch_1^-, ch_2^-) returned from a call to *chan2_create* is used to send the second to the newly created thread by the call, making it possible for that thread to cause a deadlock by waiting for a value to be sent on ch_2^+ . Clearly, Lemma 3.8 is invalidated if *chan2_create* is added.

The soundness of the type system of MTLC_{ch} rests upon the following two theorems (corresponding to Theorem 2.7 and Theorem 2.8):

THEOREM 3.9. (*Subject Reduction on Pools*) *Assume that $\vdash \Pi_1 : \hat{T}$ is derivable and $\Pi_1 \rightarrow \Pi_2$ such that $\rho(\Pi_2) \in \mathbf{RES}$. Then $\vdash \Pi_2 : \hat{T}$ is derivable.*

Proof The proof is essentially the same as the one for Theorem 2.7. The only additional part is for checking that the rules PR3, PR4-clos, PR4-send, and PR4-recv are all consistent with respect to the typing rules listed in Figure 4. ■

THEOREM 3.10. (*Progress Property on Pools*) *Assume that $\vdash \Pi_1 : \hat{T}$ is derivable and $\rho(\Pi_1)$ is valid. Also assume that $\rho(v)$ contains no channels for every value v of the type \hat{T} . Then we have the following possibilities:*

- Π_1 is a singleton mapping $[0 \mapsto v]$ for some v , or
- $\Pi_1 \rightarrow \Pi_2$ holds for some Π_2 such that $\rho(\Pi_2) \in \mathbf{RES}$.

Proof The proof follows the same structure as the one for Theorem 2.8. Lemma 3.1 is needed to handle the case where all of the threads (possibly excluding the main thread) in a pool consist of blocked expressions. ■

4. From Dyadic to Multiparty

In this section, we present an approach to building multiparty sessions based on dyadic g-sessions. With this approach, we give justification in support of the theoretical development in Section 2 and Section 3.

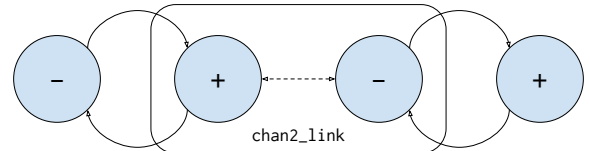


Figure 5. Illustrating *chan2_link*

4.1 Bidirectional Forwarding between Two Parties

Given two channels of dual types, there is a generic method for forwarding onto one channel each message received from the other channel and vice versa. In formalizations of session types that are directly based on linear logic (e.g., [1, 26]), this form of bidirectional forwarding of messages corresponds to the cut-elimination process in linear logic².

²It should be noted in this context that it is of no concern as to whether the cut-elimination process is terminating or not. Instead, the focus is solely on

THEOREM 4.1. Assume that ch_0 and ch_1 are two channels of the types $\mathbf{chan}(G, S)$ and $\mathbf{chan}(\overline{G}, S)$, respectively. For any party holding ch_0 and ch_1 , there is a generic method for forwarding onto ch_0 the messages received from ch_1 and vice versa (during the evaluation of a well-typed program). Let us use the name $chan2_link$ for a function of the following type that implements this generic method:

$$(\mathbf{chan}(G, S), \mathbf{chan}(\overline{G}, S)) \Rightarrow \mathbf{1}$$

Proof If S is \mathbf{nil} , then all that is needed is to call \underline{close} on both ch_0 and ch_1 . Assume that S is of the form $\mathbf{msg}(i, j) :: S_1$. Then we have the following 4 possibilities.

- Assume $i \in G$ and $j \in G$. Then $\mathbf{msg}(i, j)$ indicates an internal message for ch_0 and an external message for ch_1 . So this case is handled by calling \underline{skip} on ch_0 and ch_1 .
- Assume $i \in G$ and $j \in \overline{G}$. Then $\mathbf{msg}(i, j)$ indicates sending for ch_0 and receiving for ch_1 . So this case is handled by calling \underline{rcv} on ch_1 to receive a message and then calling \underline{send} on ch_0 to send the message.
- Assume $i \in \overline{G}$ and $j \in G$. This case is similar to the one where $i \in G$ and $j \in \overline{G}$.
- Assume $i \in \overline{G}$ and $j \in \overline{G}$. This case is similar to the one where $i \in G$ and $j \in G$.

After one of the above 4 possibilities is performed, a recursive call can be made on ch_0 and ch_1 to perform the rest of bidirectional forwarding. ■

An illustration of $chan2_link$ is given in Figure 5. We point out that Lemma 3.8 still holds after $chan2_link$ is added, and thus Lemma 3.1 still holds as well.

A dyadic i-session involves only two roles: 0 and 1. If we just study dyadic g-sessions corresponding to dyadic i-sessions, then a channel type is of the form $\mathbf{chan}(G, S)$ for either $G = \{0\}$ or $G = \{1\}$. In this context, it is unclear how Theorem 4.1 can be generalized. When more than two roles are involved, there turns out to be a natural generalization of Theorem 4.1, which we report in the next section.

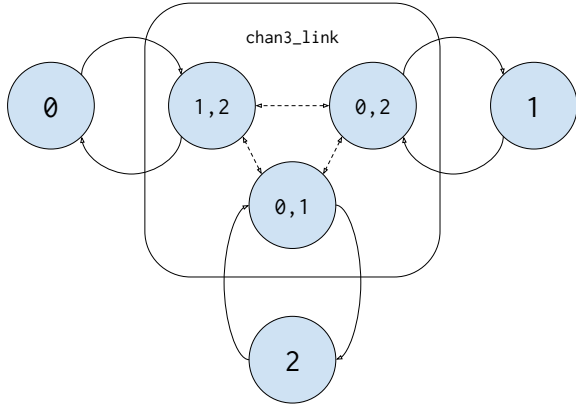


Figure 6. Illustrating $chan3_link$ involving 3 roles

4.2 Bidirectional Forwarding between Three Parties

The next theorem states the existence of a generic method for forwarding messages between three channels of certain types:

proving that each cut involving a compound formula can be reduced into one or more cuts (so as to make progress).

THEOREM 4.2. Assume that ch_0 and ch_1 are two channels of the types $\mathbf{chan}(G_0, S)$ and $\mathbf{chan}(G_1, S)$, respectively, where $\overline{G_0} \cap \overline{G_1} = \emptyset$ holds, and ch_2 is another channel of the type $\mathbf{chan}(G_2, S)$ for $G_2 = \overline{G_0} \cup \overline{G_1}$. Then there is a generic method for forwarding each message received from one of ch_0 , ch_1 and ch_2 onto one of the other two in a type-correct manner. Let us use the name $chan3_link$ for a function of the following type that implements this generic method:

$$(\mathbf{chan}(G_0, S), \mathbf{chan}(G_1, S), \mathbf{chan}(\overline{G_0} \cup \overline{G_1}, S)) \Rightarrow \mathbf{1}$$

Proof If S is \mathbf{nil} , then all that is needed is to call \underline{close} on each of ch_0 , ch_1 , and ch_2 . Assume S is of the form $\mathbf{msg}(i, j) :: S_1$ for some roles i and j . Note that $\overline{G_0}$, $\overline{G_1}$, and $\overline{G_2} = \overline{G_0} \cap \overline{G_1}$ are pairwise disjoint, and the union of these three equals the full set of roles. So we have 9 scenarios covering all of the possibilities of $i \in G$ and $j \in G'$ for G and G' ranging over $\overline{G_0}$, $\overline{G_1}$, and $\overline{G_2}$.

- Assume $i \in \overline{G_0}$ and $j \in \overline{G_0}$. Then $\mathbf{msg}(i, j)$ indicates an external message for ch_0 , an internal message for ch_1 , and an internal message for ch_2 . So this case is handled by calling \underline{skip} on each of ch_0 , ch_1 , and ch_2 .
- Assume $i \in \overline{G_0}$ and $j \in \overline{G_1}$. Then $\mathbf{msg}(i, j)$ indicates receiving for ch_0 , sending for ch_1 and an internal message for ch_2 . So this case is handled by calling \underline{rcv} on ch_0 to receive a message, and then calling \underline{send} on ch_1 to send the message, and then calling \underline{skip} on ch_2 .
- Assume $i \in \overline{G_0}$ and $j \in \overline{G_2}$. Then $\mathbf{msg}(i, j)$ indicates receiving for ch_0 , an internal message for ch_1 , and sending for ch_2 . So this case is handled by calling \underline{rcv} on ch_0 to receive a message, and then calling \underline{send} on ch_2 to send the message, and then calling \underline{skip} on ch_0 .
- Assume $i \in \overline{G_1}$ and $j \in \overline{G_0}$. The case is similar to the previous one where $i \in \overline{G_0}$ and $j \in \overline{G_1}$ holds.
- Assume $i \in \overline{G_1}$ and $j \in \overline{G_1}$. The case is similar to the previous one where $i \in \overline{G_0}$ and $j \in \overline{G_0}$ holds.
- Assume $i \in \overline{G_1}$ and $j \in \overline{G_2}$. The case is similar to the previous one where $i \in \overline{G_0}$ and $j \in \overline{G_2}$ holds.
- Assume $i \in \overline{G_2}$ and $j \in \overline{G_0}$. Then $\mathbf{msg}(i, j)$ indicates sending for ch_0 , an internal message for ch_1 , and receiving for ch_2 . So this case is handled by calling \underline{rcv} on ch_2 to receive a message, and then calling \underline{send} on ch_0 to send the message, and then calling \underline{skip} on ch_1 .
- Assume $i \in \overline{G_2}$ and $j \in \overline{G_1}$. The case is similar to the previous one where $i \in \overline{G_1}$ and $j \in \overline{G_2}$ holds.
- Assume $i \in \overline{G_2}$ and $j \in \overline{G_2}$. Then $\mathbf{msg}(i, j)$ indicates an internal message for ch_0 , an internal message for ch_1 , and an external message for ch_2 . So this case is handled by calling \underline{skip} on each of ch_0 , ch_1 , and ch_2 .

After one of the above 9 possibilities is performed, a recursive call can be made on ch_0 , ch_1 and ch_2 to perform the rest of bidirectional forwarding between these channels. ■

An illustration of $chan3_link$ involving 3 roles is given in Figure 6. We point out that Lemma 3.8 still holds after $chan3_link$ is added, and thus Lemma 3.1 still holds as well.

The following corollary (which is partly stated as Theorem 1.1 in Section 1) follows from Theorem 4.2 immediately:

COROLLARY 4.3. Assume that ch_0 and ch_1 are two channels of the types $\mathbf{chan}(G_0, S)$ and $\mathbf{chan}(G_1, S)$, respectively, where $\overline{G_0} \cap \overline{G_1} = \emptyset$ holds. Then there is a method for creating a channel ch_2 of the type $\mathbf{chan}(G_0 \cap G_1, S)$ such that the generic method for forwarding

messages as is stated in Theorem 4.2 applies to ch_0 , ch_1 and the dual ch'_0 of ch_0 . Let us use the name `chan2.link.create` for a function of the following type that implements the method for creating ch_2 based on ch_0 and ch_1 :

$$(\mathbf{chan}(G_0, S), \mathbf{chan}(G_1, S)) \Rightarrow \mathbf{chan}(G_0 \cap G_1, S)$$

Proof The channel ch_2 can simply be obtained by evaluating the following expression:

$$\mathit{chan_create}(\lambda x. \mathit{chan3_link}(ch_0, ch_1, x))$$

■

The very significance of Theorem 4.2 lies in its establishing a foundation for a type-based approach to building multiparty sessions based on dyadic g-sessions. Let us elaborate this point a bit further. Suppose that we start with one dyadic session where the two parties communicating via the dual channels ch_0 and ch'_0 and another dyadic session where the two parties communicating via the dual channels ch_1 and ch'_1 . If the party holding both ch'_0 and ch_1 (which means the party is shared between the two sessions) does bidirectional forwarding of messages between them according to Theorem 4.1 (that is, calling `chan2.link` on ch'_0 and ch_1), then a new session is created but it is still a dyadic one (where the communication is between ch_0 and ch'_1). It is impossible to build multiparty sessions (involving more than 2 parties) by simply relying on Theorem 4.1 if we can only start with dyadic ones. With Theorem 4.2, three dyadic sessions can be joined together by making a call to `chan3.link`, resulting in the creation of a 3-party session. In other words, `chan3.link` can be seen as a breakthrough. In order to build sessions involving more parties, we simply make more calls to `chan3.link`.

4.3 A Sketch for Building a 3-Party Session

Building a session often requires explicit coordination between the involved parties during the phase of setting-up. In practice, designing and implementing coordination between 3 or more parties is generally considered a difficult issue. By building a multiparty session based on dyadic g-sessions, we only need to be concerned with two-party coordination, which is usually much easier to handle.

Given a group G of roles and a session type S , we introduce a type **service**(G, S) that can be assigned to a value representing some form of *persistent* service. With such a service, channels of the type $\mathbf{chan}(\overline{G}, S)$ can be created repeatedly. A built-in function `service.create` is assigned the following type for creating a service:

$$(\mathbf{chan}(G, S) \rightarrow_i \mathbf{1}) \Rightarrow \mathbf{service}(G, S)$$

In contrast with `chan.create` for creating a channel, `service.create` requires that its argument be a non-linear function (so that this function can be called repeatedly). A client may call the following function to obtain a channel to communicate with a server that provides the requested service:

$$\mathit{service_request} : (\mathbf{service}(G, S)) \Rightarrow \mathbf{chan}(\overline{G}, S)$$

Suppose we want to build a 3-party session involving 3 roles: 0, 1, and 2. We may assume that there are two services of the types **service**({0}, S) and **service**({1}, S) available to a party (planning to implement role 2); this party can call `service.request` on the two services (which are just two names) to obtain two channels ch_0 and ch_1 of the types $\mathbf{chan}(\{1, 2\}, S)$ and $\mathbf{chan}(\{0, 2\}, S)$, respectively; it then calls `chan2.link.create`(ch_0, ch_1) to obtain a channel ch_2 of the type $\mathbf{chan}(\{2\}, S)$ for communicating with two servers providing the requested services. Obviously, there are many other ways of building a multiparty session by passing around multirole channels for dyadic g-sessions.

5. More Constructors for Session Types

We present in this section a few additional constructors for session types plus an example of user-defined session type.

5.1 Branching

Given an integer i (representing a role) and two session types S_0 and S_1 , we can form a branching session type **choose**(i, S_0, S_1) that is given the following interpretation based a group G of roles:

- Assume $i \in G$. Then the session type **choose**(i, S_0, S_1) means for the party implementing G to send a message to the party implementing \overline{G} so as to inform the latter which of S_0 and S_1 is chosen for specifying the subsequent communication. In order for Theorem 4.2 to work out in the presence of **choose**, this message needs to be sent repeatedly, targeting a new role in \overline{G} each time. It needs to be sent n times if there are n roles in \overline{G} .
- Assume $i \notin G$. Then the session type **choose**(i, S_0, S_1) means for the party implementing G to wait for a message indicating which of S_0 and S_1 is chosen for specifying the subsequent communication. Note that this message arrives repeatedly for n times, where n is the cardinality of \overline{G} .

As can be expected, we have two functions of the following type schemas:

$$\begin{aligned} \mathit{chan_choose_l} & : (\mathbf{chan}(G, \mathbf{choose}(i, S_0, S_1))) \Rightarrow \mathbf{chan}(G, S_0) \\ \mathit{chan_choose_r} & : (\mathbf{chan}(G, \mathbf{choose}(i, S_0, S_1))) \Rightarrow \mathbf{chan}(G, S_1) \end{aligned}$$

where $i \in G$ is assumed. We have another function `chan.choose.tag` of the following type schema:

$$(\mathbf{chan}(G, \mathbf{choose}(i, S_0, S_1))) \Rightarrow \exists \sigma. (\mathbf{ctag}(S_0, S_1, \sigma), \mathbf{chan}(G, \sigma))$$

where **ctag** is a datatype with the following two constructors (which are essentially represented as 0 and 1):

$$\begin{aligned} \mathit{ctag_l} & : () \Rightarrow \mathbf{ctag}(S_0, S_1, S_0) \\ \mathit{ctag_r} & : () \Rightarrow \mathbf{ctag}(S_1, S_0, S_1) \end{aligned}$$

By performing pattern matching on the first component of the tuple returned by a call to `chan.choose.tag`, one can tell whether the σ is S_0 or S_1 . Note that the existential quantification is available in ATS but it is not part of MTL_{ch}. Hopefully, the idea should be clear to the reader.

5.2 Sequencing

Given two session types S_0 and S_1 , we can form another one of the form **append**(S_0, S_1), which intuitively means the standard sequencing of S_0 and S_1 . Also, we have a function `chan.append` of the following type schema for operating on a channel of some sequencing session type:

$$(\mathbf{chan}(G, \mathbf{append}(S_0, S_1)), \mathbf{chan}(G, S_0) \rightarrow_i \mathbf{chan}(G, \mathbf{nil})) \Rightarrow \mathbf{chan}(G, S_1)$$

When applied to a channel and a linear function, `chan.append` essentially calls the linear function on the channel to return another channel. There is a bit of cheating here because there is no type-based enforcement of the requirement that the channel returned by the linear function be the same as the one passed to it, which on the other hand can be readily achieved in ATS.

5.3 Repeating Indefinitely

Given an integer i (representing a role) and a session type S , we can form another session type of the form **repeat**(i, S), which is essentially defined recursively as follows:

$$\mathbf{repeat}(i, S) = \mathbf{choose}(i, \mathbf{nil}, \mathbf{repeat}(i, S))$$

Intuitively, **repeat**(i, S) means that the party implementing the role i determines whether a session specified by S should be repeated

```

datatype
  ssn_queue(a:vtype, i:int, int) =
  | queue_nil(a,i,0) of (nil)
  | {n:nat}
  | queue_eng(a,i,n) of msg(i, 0, a)::ssn_queue(a, n+1)
  | {n:pos}
  | queue_deq(a,i,n) of msg(0, i, a)::ssn_queue(a, n-1)

where ssn_queue(a:vtype, n:int) =
  choose(0, ssn_queue(a, 1, n), ssn_queue(a, 2, n))

```

Figure 7. A type for queue-sessions

```

typedef
  ssn_s0b1b2_fail = nil

typedef
  ssn_s0b1b2_succ =
  msg(B2,S0,proof) ::
  msg(S0,B2,receipt) :: nil

typedef
  ssn_s0b1b2 =
  msg(B1,S0,title) :: msg(S0,B1,price) ::
  msg(S0,B2,price) :: msg(B1,B2,price) ::
  choose(B2, ssn_s0b1b2_succ, ssn_s0b1b2_fail)

```

Figure 8. A classic 3-role session type

indefinitely. For instance, a list-session (between two parties) can be specified as follows:

```
repeat(0, msg(0, 1, T) :: nil)
```

which means the server (implementing the role 0) offers to the client (implementing the role 1) a list of values of the type T . Similarly, a colist-session (between two parties) can be specified as follows:

```
repeat(1, msg(0, 1, T) :: nil)
```

which means the client requests from the server a list of values of the type T .

5.4 User-Defined Session Types

We can readily make use of various advanced programming features in ATS for formulating types for sessions as well as implementing these sessions. As a concrete example, a type for queue-sessions written in the source syntax of ATS is given in Figure 7, which makes direct use of dependent types (of DML-style). Given the following explanation, we reasonably expect that the reader be able to make sense of this interesting example.

During a queue-session (specified by *ssn_queue*), there are one server(S0) and two clients (C1 and C2); the server chooses (based on some external information) which client is to be served in the next round; the chosen client can request the server to create an empty queue (*queue_nil*), enqueue an element into the current queue (*queue_eng*), and dequeue an element from the current non-empty queue (*queue_deq*). Given a type T and a natural number N , the type *ssn_queue*(T, N) means that the size of the underlying queue in the current queue-session is N . This example is largely based on a type for 2-party queue-sessions in SILL [10]. What is novel here is an added party plus the use of dependent types (of DML-style) to specify the size of the underlying queue in a queue-session. As a side note, one may be wondering why C1 can keep the correct account of queue length after C2 performs an operation. The very reason is that C2 announces to both S0 and C1 which operation C2 is to perform before it performs it.

6. An Example of 3-Party Session

Let us assume the availability of three roles: Seller(S0), Buyer 1(B1) and Buyer 2(B2). A description of the classic one-seller-and-two-buyers(S0B1B2) protocol due to (Honda et al. 2008) is essentially given as follows:

1. B1 sends a book title to S0.
2. S0 replies a quote to both Buyer 1 and Buyer 2.
3. B1 tells B2 how much B1 can contribute:
 - (a) Assume B2 can afford the remaining part:
 - i. B2 sends S0 a proof of payment.
 - ii. S0 sends B2 a receipt for the sale.
 - (b) Assume B2 cannot afford the remaining part:
 - i. B2 terminates.

This protocol is formally captured by the type *ssn_s0b1b2* given in Figure 8. For taking a peek at a running implementation of this example in ATS, please visit the following link:

<http://pastebin.com/JmZRukRi>

The steps involved in building a 3-party session are basically those outlined in Section 4.3.

7. Implementing Session-Typed Channels

As far as implementation is of the concern, there is very little that needs to be done regarding typechecking in order to support session-typed channels in ATS. The only considerably significant complication comes from the need for solving constraints generated during typechecking that may involve various common set operations (on groups of roles), which the current built-in constraint-solver for ATS cannot handle. Fortunately, we have an option to export such constraints for them to be solved with an external constraint-solver based on Z3 [6].

The first implementation of session-typed channels (based on shared memory) for use in ATS is done in ATS itself, which compiles to C, the primary compilation target for ATS. Another implementation of session-typed channels (based on processes) is done in Erlang. As the ML-like core of ATS can already be compiled into Erlang, we have now an option to construct distributed programs in ATS that may make use of session types and then translate these programs into Erlang code for execution, thus taking great advantage of the infrastructural support for distributed computing in Erlang.

We outline some key steps taken in both of the implementations. In particular, we briefly mention an approach to implementing *chan2_link* and *chan3_link* that completely removes the need for explicit forwarding of messages and thus the inefficiency associated with it.

Let us use *uch* to refer to a uni-directional channel that can be held by two parties; one party can only write to it while the other can only read from it. Suppose we want to build a pair of multirole channels ch^+ and ch^- for some groups G and \bar{G} of roles; we first create a matrix M of the dimension $nrole$ by $nrole$, where $nrole$ is the total number of available roles; for each $i \in G$ and $j \in \bar{G}$, we use $uch(i, j)$ and $uch(j, i)$ to refer to the two uni-directional channels stored in $M[i, j]$ and $M[j, i]$, respectively. Note that this matrix M is shared by both ch^+ and ch^- ; for ch^+ , $uch(i, j)$ and $uch(j, i)$ are used to send messages from role i to role j and receive messages sent from role j to role i on ch^- , respectively; for ch^- , it is precisely the opposite.

If *chan2_link* is implemented by following Theorem 4.1 directly, then a call to *chan2_link* creates a thread/process for performing bidirectional forwarding of messages explicitly, and the created thread/process only terminates after no more forwarding is needed.

If we assume that a *uch* can be sent onto another *uch*, which can be readily supported in both ATS and Erlang, then a much more efficient approach to implementing *chan2.link* can be described as follows. Suppose we call *chan2.link* on two channels ch_0 and ch_1 of the types $\mathbf{chan}(G, S)$ and $\mathbf{chan}(\overline{G}, S)$; let M_0 and M_1 be the matrices in ch_0 and ch_1 for holding uni-directional channels in ch_0 and ch_1 , respectively; for each $i \in G$ and $j \in \overline{G}$, we send the *uch* in $M_1[i, j]$ onto the one in $M_0[i, j]$ and the *uch* in $M_0[j, i]$ onto the one in $M_1[j, i]$; if a *uch* is received on the one in $M_0[i, j]$ ($M_1[j, i]$), then the *uch* is put into $M_0[i, j]$ ($M_1[j, i]$) to replace the original one. It should be clear that *chan3.link* can be implemented similarly.

8. Related Work and Conclusion

Session types were introduced by Honda [11] and further extended subsequently [12, 21]. There have since been extensive theoretical studies on session types in the literature (e.g., [1, 4, 8, 15, 22, 24, 26]). Multiparty session types, as a generalization of (dyadic) session types, were introduced by Honda and others [13], together with the notion of global types, local types, projection and coherence. By introducing dyadic group sessions (g-sessions), we give a novel form of generalization going from dyadic sessions to dyadic g-sessions.

The notion of dyadic g-sessions is rooted in a very recent attempt to incorporate session types for dyadic sessions into ATS [30]. In an effort to formalize session types, two kinds of channel types $\mathbf{chpos}(S)$ and $\mathbf{chneg}(S)$ are introduced for positive and negative channels, respectively, directly leading to the discovery of the notion of single-role channels (as $\mathbf{chpos}(S)$ and $\mathbf{chneg}(S)$ can simply be translated into $\mathbf{chan}(0, S)$ and $\mathbf{chan}(1, S)$, respectively) and then the discovery of the key notion of multirole channels in this paper.

In [7, 18], a party can play multiple roles by holding channels belonging to multiple sessions. We see no direct relation between such a multirole party and a multirole channel. In [3], coherence is treated as a generalization of duality. In particular, the binary cut rule is extended to a multiparty cut rule. But this multiparty cut rule is not directly related to Theorem 4.2 as far as we can tell.

It is in general a challenging issue to establish deadlock-freedom for session-typed concurrency. There are variations of session types that introduce a partial order on time stamps [20] or a constraint on dependency graphs [2]. As for formulations of session types (e.g., [1, 26]) based on linear logic [9], the standard technique for cut-elimination is often employed to establish global progress (which implies deadlock-freedom). In MTLC_{ch} , there is no explicit tracking of cut-rule applications in the type derivation of a program. The notion of DF-reducibility (taken from [30]) is introduced in order to carry out cut-elimination in the absence of explicit tracking of cut-rule applications.

Probably, MTLC_{ch} is most closely related to SILL [23], a functional programming language that adopts via a contextual monad a computational interpretation of linear sequent calculus as session-typed processes. Unlike in MTLC_{ch} , the support for linear types in SILL is not direct and only monadic values (representing open process expressions) in SILL can be linear. In terms of theoretical development, the approach to establishing global progress in MTLC_{ch} is rooted in the one for SILL (though the latter does not apply directly).

Also, MTLC_{ch} is related to previous work on incorporating session types into a multi-threaded functional language [25], where a type safety theorem is established to ensure that the evaluation of a well-typed program can never lead to a so-called *faulty configuration*. However, this theorem does not imply global progress as a program that is not of faulty configuration can still deadlock. Also, we point out MTLC_{ch} is related to recent work on assigning an operational semantics to a variant of GV [15]. In particular, the approach based on DF-reducibility to establishing global progress

in MTLC_{ch} is analogous to the one taken to establish deadlock-freedom for this variant.

As for future work we are particularly interested in applying the notion of dyadic g-sessions to the design and formalization of a type system for some variant of π -calculus. Also, it should be both exciting and satisfying if a logic-based interpretation can be found for Theorem 4.2.

There are a variety of programming issues that need to be addressed in order to facilitate the use of session types in practice. Currently, session types are often represented as datatypes in ATS, and programming with such session types tends to involve writing a very significant amount of boilerplate code. In the presence of large and complex session types, writing such code can be tedious and error-prone. Naturally, we are interested in developing some meta-programming support for generating such code automatically. Also, we are in the process of designing and implementing session combinators (in a spirit similar to parsing combinators [14]) that can be conveniently called to assemble subsessions into a coherent whole.

References

- [1] Luís Caires and Frank Pfenning. Session types as intuitionistic linear propositions. In *CONCUR 2010 - Concurrency Theory, 21th International Conference, CONCUR 2010, Paris, France, August 31-September 3, 2010. Proceedings*, pages 222–236, 2010.
- [2] Marco Carbone and Søren Debois. A graphical approach to progress for structured communication in web services. In *Proceedings Third Interaction and Concurrency Experience: Guaranteed Interaction, ICE 2010, Amsterdam, The Netherlands, 10th of June 2010.*, pages 13–27, 2010.
- [3] Marco Carbone, Fabrizio Montesi, Carsten Schürmann, and Nobuko Yoshida. Multiparty Session Types as Coherence Proofs. *CONCUR*, pages 412–426, 2015.
- [4] Giuseppe Castagna, Mariangiola Dezani-Ciancaglini, Elena Giachino, and Luca Padovani. Foundations of session types. In *Proceedings of the 11th International ACM SIGPLAN Conference on Principles and Practice of Declarative Programming, September 7-9, 2009, Coimbra, Portugal*, pages 219–230, 2009.
- [5] Chiyen Chen and Hongwei Xi. Combining Programming with Theorem Proving. In *Proceedings of the Tenth ACM SIGPLAN International Conference on Functional Programming*, pages 66–77, Tallinn, Estonia, September 2005.
- [6] Leonardo Mendonça de Moura and Nikolaj Bjørner. Z3: an efficient SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems, 14th International Conference, TACAS 2008, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2008, Budapest, Hungary, March 29-April 6, 2008. Proceedings*, pages 337–340, 2008.
- [7] Pierre-malo Deniérou and Nobuko Yoshida. Dynamic multirole session types. *POPL*, pages 435–446, 2011.
- [8] Simon J. Gay and Vasco Thudichum Vasconcelos. Linear type theory for asynchronous session types. *J. Funct. Program.*, 20(1):19–50, 2010.
- [9] Jean-Yves Girard. Linear logic. *Theoretical Computer Science*, 50(1):1–101, 1987.
- [10] Dennis Griffith. SILL: A session-typed functional programming language, 2015. Available at: <https://github.com/ISANobody/sill>.
- [11] Kohei Honda. Types for dyadic interaction. In *CONCUR '93, 4th International Conference on Concurrency Theory, Hildesheim, Germany, August 23-26, 1993, Proceedings*, pages 509–523, 1993.
- [12] Kohei Honda, Vasco Vasconcelos, and Makoto Kubo. Language primitives and type discipline for structured communication-based programming. In *Programming Languages and Systems - ESOP '98, 7th European Symposium on Programming, Held as Part of the European Joint Conferences on the Theory and Practice of Software, ETAPS '98*,

- Lisbon, Portugal, March 28 - April 4, 1998, *Proceedings*, pages 122–138, 1998.
- [13] Kohei Honda, Nobuko Yoshida, and Marco Carbone. Multiparty asynchronous session types. In *Proceedings of the 35th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2008, San Francisco, California, USA, January 7-12, 2008*, pages 273–284, 2008.
- [14] Graham Hutton. Higher-order functions for parsing. *J. Funct. Program.*, 2(3):323–343, 1992.
- [15] Sam Lindley and J. Garrett Morris. A semantics for propositions as sessions. In *Programming Languages and Systems - 24th European Symposium on Programming, ESOP 2015, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2015, London, UK, April 11-18, 2015. Proceedings*, pages 560–584, 2015.
- [16] Robin Milner, J. Parrow, and David Walker. A calculus of processes, parts I and II. *Information and Computation*, 100:1–40 and 41–77, 1992.
- [17] Matthias Neubauer and Peter Thiemann. An implementation of session types. In *Practical Aspects of Declarative Languages, 6th International Symposium, PADL 2004, Dallas, TX, USA, June 18-19, 2004, Proceedings*, pages 56–70, 2004.
- [18] Romyana Neykova and Nobuko Yoshida. Multiparty Session Actors. *COORDINATION*, pages 131–146, 2014.
- [19] Riccardo Pucella and Jesse A. Tov. Haskell session types with (almost) no class. In *Proceedings of the 1st ACM SIGPLAN Symposium on Haskell, Haskell 2008, Victoria, BC, Canada, 25 September 2008*, pages 25–36, 2008.
- [20] Eijiro Sumii and Naoki Kobayashi. A generalized deadlock-free process calculus. *Electr. Notes Theor. Comput. Sci.*, 16(3):225–247, 1998.
- [21] Kaku Takeuchi, Kohei Honda, and Makoto Kubo. An interaction-based language and its typing system. In *PARLE '94: Parallel Architectures and Languages Europe, 6th International PARLE Conference, Athens, Greece, July 4-8, 1994, Proceedings*, pages 398–413, 1994.
- [22] Bernardo Toninho, Luís Caires, and Frank Pfenning. Dependent session types via intuitionistic linear type theory. In *Proceedings of the 13th International ACM SIGPLAN Conference on Principles and Practice of Declarative Programming, July 20-22, 2011, Odense, Denmark*, pages 161–172, 2011.
- [23] Bernardo Toninho, Luís Caires, and Frank Pfenning. Higher-order processes, functions, and sessions: A monadic integration. In *Programming Languages and Systems - 22nd European Symposium on Programming, ESOP 2013, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2013, Rome, Italy, March 16-24, 2013. Proceedings*, pages 350–369, 2013.
- [24] Vasco T. Vasconcelos. Fundamentals of session types. *Inf. Comput.*, 217:52–70, 2012.
- [25] Vasco Thudichum Vasconcelos, António Ravara, and Simon J. Gay. Session types for functional multithreading. In *CONCUR 2004 - Concurrency Theory, 15th International Conference, London, UK, August 31 - September 3, 2004, Proceedings*, pages 497–511, 2004.
- [26] Philip Wadler. Propositions as sessions. In *ACM SIGPLAN International Conference on Functional Programming, ICFP'12, Copenhagen, Denmark, September 9-15, 2012*, pages 273–286, 2012.
- [27] Hongwei Xi. Applied Type System (extended abstract). In *post-workshop Proceedings of TYPES 2003*, pages 394–408. Springer-Verlag LNCS 3085, 2004.
- [28] Hongwei Xi. Dependent ML: an approach to practical programming with dependent types. *Journal of Functional Programming*, 17(2):215–286, 2007.
- [29] Hongwei Xi and Frank Pfenning. Dependent Types in Practical Programming. In *Proceedings of 26th ACM SIGPLAN Symposium on Principles of Programming Languages*, pages 214–227, San Antonio, Texas, January 1999. ACM press.
- [30] Hongwei Xi, Zhiqiang Ren, Hanwen Wu, and William Blair. Session types in a linearly typed multi-threaded lambda-calculus, 2016. Available at: <http://arxiv.org/abs/1603.03727>.