Uniqueness typing for resource management in message-passing concurrency

EDSKO DEVRIES, Computer Science, Department of Trinity College Dublin, College Green, Dublin 2 Ireland.
E-mail: edsko.de.vries@cs.tcd.ie

ADRIAN FRANCALANZA, Information and Communication Technology, University of Malta, MSD 2080, Malta.
E-mail: adrian.francalanza@um.edu.mt

MATTHEW HENNESSY, Department of Computer Science, Trinity College Dublin, College Green, Dublin 2, Ireland.
E-mail: matthew.hennessy@cs.tcd.ie

Abstract

We view channels as the main form of resources in a message-passing programming paradigm. These channels need to be carefully managed in settings where resources are scarce. To study this problem, we extend the pi-calculus with primitives for channel allocation and deallocation and allow channels to be reused to communicate values of different types. Inevitably, the added expressiveness increases the possibilities for runtime errors. We define a substructural type system, which combines uniqueness typing and affine typing to reject these ill-behaved programs.

Keywords: Message-passing concurrency, type systems, resource management

1 Introduction

Message-passing concurrency is a programming paradigm whereby shared memory is prohibited and process interaction is limited to explicit message communication. This concurrency paradigm forms the basis for a number of process calculi such as the pi-calculus [22] and has been adopted by programming languages such as the actor-based language Erlang [3].

Message-passing concurrency often abstracts away from resource management and programs written at this abstraction level exhibit poor resource awareness. In this article, we study ways of improving this shortcoming. Specifically, we develop a statically typed extension of the pi-calculus in which resources, i.e. channels, can be reused at varying types and resources can be safely deallocated when they are no longer required.

Idiomatic pi-calculus processes are often characterized by wasteful use-once-throw-away channels [21, 22]. Consider the following two pi-calculus process definitions

\[
\text{timeSrv}(\text{getTime}) \triangleq \text{rec } x \in c X. \text{getTime}?x.x!(\text{hr, min}).X \\
\text{dateSrv}(\text{getDate}) \triangleq \text{rec } x \in c X. \text{getDate}?x.x!(\text{year, mon, day}).X
\]

\text{timeSrv} defines a server that repeatedly waits on a channel named \text{getTime} to dynamically receive a channel name, represented by the bound variable \(x\), and then replies with the current time on \(x\).
2. Uniqueness typing for message-passing concurrency

DATESRV is a similar service which returns the current date. An idiomatic pi-calculus client definition is

\[ \text{CLIENT}_0 \triangleq \text{new } \text{ret}_1, \text{getTime}_1!\text{ret}_1 \parallel \text{getTime}_2!\text{ret}_1 \parallel \text{ret}_1?(y_{\text{hr}}, y_{\text{min}}).\text{ret}_1?(y'_{\text{hr}}, y'_{\text{min}}).\text{getClient}_2.\text{getDate}_1\text{ret}_2.\text{ret}_2?(z_{\text{year}}, z_{\text{mon}}, z_{\text{day}}). P \]

CLIENT\_0 first sends a request to two separate time servers in order to obtain an accurate measurement. It does not matter in which order the servers reply, but once they have both replied the client continues to send a request to a date server before continuing as P. It uses two distinct channels ret\_1 and ret\_2 as return channels to query the time and date servers; these return channels are scoped (private) to preclude interference from other clients concurrently querying the servers.

From a resource management perspective, it makes pragmatic sense to try and reduce the number of channels used for the interactions between this client and the two servers. In particular, since the servers use the reply channel once and do not keep any reference to this channel after use, the client can economize on the resources required for these interactions and use one channel to communicate sequentially with both the time servers and the date servers.

\[ \text{CLIENT}_1 \triangleq \text{new } \text{getTime}_1!\text{ret} \parallel \text{getTime}_2!\text{ret} \parallel \text{ret}(y_{\text{hr}}, y_{\text{min}}).\text{getTime}_1!\text{ret}(y'_{\text{hr}}, y'_{\text{min}}).\text{getTime}_2!\text{ret}(z_{\text{year}}, z_{\text{mon}}, z_{\text{day}}). P \]

From a typing perspective, this reuse of the same channel entails strong update on the channel: that is, reuse of a channel to communicate values of different types. Strong update must be carefully controlled; for instance, an attempt to use one channel to communicate with a time server and date server in parallel is unsafe:

\[ \text{CLIENT}_{\text{err}} \triangleq \text{new } \text{getTime}_1!\text{ret}(y_{\text{hr}}, y_{\text{min}}).P_1 \parallel \text{getTime}_2!\text{ret}(z_{\text{year}}, z_{\text{mon}}, z_{\text{day}}). P_2 \]

Standard pi-calculus type systems accept only CLIENT\_0 and rule out both CLIENT\_1 and CLIENT\_err. However, CLIENT\_1 is safe because, apart from the fact that the time servers use the return channel once, the communication with the date server happens strictly after the communication with the time servers.

Adequate resource management also requires precise descriptions of when resources are allocated and existing ones are disposed. The standard scoping construct of the \( \pi \) calculus new \( c \). \( P \) is unfit for this purpose as the extrusion rule means that, for instance, channel ret\_2 is allocated before or after the communication with the time server in CLIENT\_0. Moreover, the \( \pi \) calculus does not offer an explicit channel deallocation construct, so that the point of deallocation of channels (garbage collection) is likewise unknown. We address this by introducing an explicit allocation construct alloc \( c \). \( P \)

When the allocation is executed, a new channel \( c \) is created at runtime and the alloc \( c \). \( P \) changes to new \( c \). \( P[c/x] \). Dually, we also extend the calculus with an explicit deallocation operator free \( c \). \( P \)

For example:

\[ \text{CLIENT}_2 \triangleq \text{alloc } x.\text{getTime}_1!x?(y_{\text{hr}}, y_{\text{min}}).\text{getTime}_1!x?(y'_{\text{hr}}, y'_{\text{min}}).\text{getTime}_2!x?(z_{\text{year}}, z_{\text{mon}}, z_{\text{day}}).\text{free } x. P \]

Inevitably, the added expressiveness of this extended pi-calculus increases the possibilities for runtime errors such as in CLIENT\_err\_2 below. In this client, potential interleavings of the subprocesses getTime\_1!x?(y_{\text{hr}}, y_{\text{min}}). P_1 and getTime\_2!x?(z_{\text{year}}, z_{\text{mon}}, z_{\text{day}}). free x. P_2 can lead to both value mismatch during communication (since the subprocesses are using channel x to communicate values...
Uniqueness typing for message-passing concurrency

of different types) and also to a premature deallocation of this channel by the second subprocess while it is still in use by the first subprocess.

\[ \text{CLIENT} \triangleq \text{alloc} \circ x.(\text{getTime}!:x.(y_{hr},y_{min}).P_1 \parallel \text{getDate}!:x.(c_{year},c_{mon},c_{day}).x \in x.P_2) \]

We define a type system which rejects processes that are unsafe; the type system combines uniqueness typing [5] and affine typing [21], while permitting value coercion across these modes through subtyping. Uniqueness typing gives us *global guarantees*, simplifying local reasoning when typing both strong updates and safe deallocations. Uniqueness typing can be seen as dual to affine typing [15], and we make essential use of this duality to keep track of uniqueness across channel-passing parallel processes.

The rest of the article is structured as follows. Section 2 introduces our message-passing language with explicit channel allocation and deallocation. Section 3 describes our type system and outlines how examples are typechecked. Section 4 then details the proof of soundness for our type system. Finally, Section 5 discusses related work and Section 6 concludes with directions for future work.

## 2 The Resource Pi-Calculus

Figure 1 shows the syntax for the resource pi-calculus. The language is the standard pi-calculus extended with primitives for explicit channel allocation and deallocation; moreover, channel scoping records whether a channel is allocated (⊤) or deallocated (⊥). The syntax assumes two separate denumerable sets of channel names \(c, d \in \text{NAME}\) and variables \(x, y \in \text{VAR}\), and lets identifiers \(u, v \in \text{NAME} \cup \text{VAR}\) range over both. The input and channel allocation constructs are binders for variables \(\vec{x}\) and \(x\) respectively, whereas scoping is a binder for names (i.e. \(c\)). The syntax also assumes a denumerable set of process variables \(X, Y \in \text{PVAR}\) which are bound by the recursion construct.

Channels are stateful (allocated, \(⊤\), or deallocated, \(⊥\)) and process semantics is defined over systems, \((M, P)\) where \(M \in \Sigma : \text{CHAN} \rightarrow \{⊤, ⊥\}\) describes the state of the free channels in \(P\), and stateful scoping \(\text{new} : c : s.P\) describes the state of scoped channels. A tuple \((M, P)\) is a system whenever \(\text{fn}(P) \subseteq \text{dom}(M)\) and is denoted as \(M \triangleright P\). We say that a system \(M \triangleright P\) is closed whenever \(P\) does not have any free variables. An example system would be

\[ \text{c:⊤}(\text{new} : d : \text{⊤}.c) \]

whereby, apart from the visible state relating to channel \(c\), \(M = c:⊤\), the system also describes internal state through the stateful scoped channel in the process part \(P = \text{new} : d : \text{⊤}.c\).

Figure 2 defines contexts over systems where \(C[M \triangleright P]\) denotes the application of a context \(C\) to a system \(M \triangleright P\). In the case where a context scopes a name \(c\), the definition extracts the state relating
4 Uniqueness typing for message-passing concurrency

Contexts

\[
\begin{align*}
\mathcal{C} & := \mathbf{[-]} \mid \mathcal{C} \parallel \mathcal{P} \mid \mathcal{P} \parallel \mathcal{C} \mid \text{new } c.\mathcal{C}
\end{align*}
\]

\[
\begin{align*}
[M \triangleright \mathcal{P}] \quad & \xymatrix{\mathcal{T}} \quad [M \triangleright \mathcal{P}] \parallel Q \quad \xymatrix{\mathcal{T} \parallel [M \triangleright \mathcal{P}] \parallel Q} & [M \triangleright \mathcal{P}] \parallel Q' \parallel [M' \triangleright \mathcal{P}'] \parallel Q \quad \xymatrix{[M \triangleright \mathcal{P}] \parallel Q \parallel [M' \triangleright \mathcal{P}'] \parallel Q} \quad \text{if } [M \triangleright \mathcal{P}] = M' \triangleright \mathcal{P}'
\end{align*}
\]

\[
\begin{align*}
\text{new } c.\mathcal{C} \triangleright \mathcal{P} \quad \xymatrix{\mathcal{T} \parallel \text{new } c.\mathcal{P}'} & [\text{new } c.\mathcal{C} \triangleright \mathcal{P} \parallel Q] \quad \xymatrix{\text{new } c.\mathcal{C} \parallel [\text{new } c.\mathcal{C} \triangleright \mathcal{P} \parallel Q]} \quad \text{if } [\text{new } c.\mathcal{C} \triangleright \mathcal{P} \parallel Q] = [\text{new } c.\mathcal{C} \parallel \mathcal{P}'] = M', c : s > P'
\end{align*}
\]

Structural equivalence

\[
\begin{align*}
\text{sCom} & \quad P \parallel Q \equiv Q \parallel P & \text{sAss} & \quad P \parallel (Q \parallel R) \equiv (P \parallel Q) \parallel R
\end{align*}
\]

\[
\begin{align*}
\text{sNIl} & \quad P \parallel \text{nil} \equiv P & \text{sNnM} & \quad \text{new } c.\text{nil} \equiv \text{nil}
\end{align*}
\]

\[
\begin{align*}
\text{sSwp} & \quad \text{new } c.s(\text{new } d.s'.P) \equiv \text{new } d.s'(\text{new } c.s.P)
\end{align*}
\]

\[
\begin{align*}
\text{sExt} & \quad P \parallel \text{new } c.s Q \equiv \text{new } c.s(P \parallel Q) & c \not\in \text{fn } P
\end{align*}
\]

Reduction Rules

\[
\begin{align*}
R\text{Com} & \quad M(c) = \top & R\text{Rec} & \quad M \triangleright \text{rec } X.P \rightarrow M \triangleright P(X\ltrightarrow X[P])
\end{align*}
\]

\[
\begin{align*}
R\text{Then} & \quad \text{if } c = d \text{ then } P \rightarrow M \triangleright P & R\text{Else} & \quad \text{if } c = d \text{ then } P \rightarrow M \triangleright Q
\end{align*}
\]

\[
\begin{align*}
R\text{All} & \quad M \triangleright \text{alloc } c.P \rightarrow M \triangleright \text{new } c.\top.P(\parallel \top) & R\text{Free} & \quad M, c : \top \triangleright \text{free } c.P \rightarrow M, c : \downarrow \parallel P
\end{align*}
\]

\[
\begin{align*}
R\text{Str} & \quad P \equiv P' \quad M \triangleright P' \rightarrow M \triangleright Q' & & \equiv Q' \equiv Q \quad M \triangleright P \rightarrow M \triangleright Q
\end{align*}
\]

Error Reduction Rules

\[
\begin{align*}
\text{rAty} & \quad |\overline{d}| \parallel |\overline{e}| \quad M(c) = \bot & \text{rOut} & \quad M(c) = \bot \quad M \triangleright c?\overline{e}.Q \rightarrow M \triangleright c?\overline{e}.Q
\end{align*}
\]

\[
\begin{align*}
\text{rIn} & \quad M(c) = \bot \quad M \triangleright c\overline{e}.Q \rightarrow M \triangleright c\overline{e}.Q & \text{rStr} & \quad P \equiv Q \quad M \triangleright Q \rightarrow M \triangleright P
\end{align*}
\]

The reduction relation is defined as the least contextual relation over closed systems satisfying the rules in Figure 2. It relies on a quasi-standard pi-calculus structural equivalence relation over processes (\(\equiv\)) whereby the only notable difference is that redundant channel scoping can only be discarded when it is deallocated (sNnM); this allows us characterize ‘memory leakages’ during the computation, which take the form of \(\text{new } c.\top.\text{nil}\). System communication (rCom) requires the communicating channel to be allocated but does not place any constraint on the status of the channels communicated. Allocation (rAll) creates a private allocated channel and substitutes it for the bound variable of the allocation construct in the continuation; the condition \(c \not\in \text{fn } (M)\) ensures that \(c\) is
Uniqueness typing for message-passing concurrency

We can, therefore, express computations such as
\[ M \vdash \text{alloc}(x \mid x \cdot \text{nil} \parallel x \cdot ?y \cdot \text{free} \cdot P) \rightarrow M \vdash \text{newd} : \top \cdot (d \mid x \cdot \text{nil} \parallel d \mid ?y \cdot \text{free} \cdot P) \quad x \notin \text{fv}(P) \]
\[ \rightarrow M \vdash \text{newd} : \top \cdot (\text{free} \cdot P[c/y]) \]
\[ \rightarrow M \vdash \text{newd} : \bot \cdot (P[c/y]) \equiv M \vdash P[c/y] \]

Figure 2 also defines error reductions as the least contextual relation satisfying the rules for \( \text{err} \). These rules capture errors resulting from arity mismatch and attempts to communicate on deallocated channels. In particular, arity mismatch can come from the unconstrained use of strong updates such as in the case of \( \text{client}_{\text{err}} \) in Section 1.

**Example 2.1**

For \( Q = \text{timeSrv} \parallel \text{dateSrv} \), whenever \( M(\text{getTime}) = M(\text{getTime}) = \top \) we have the following successful reduction sequences:
\[ M \vdash \text{client}_2 \parallel Q \rightarrow * M \vdash P[\text{hr, min}/\text{hr, min}, \text{year, mon, day}/\text{year, mon, day}, \text{zyear}, \text{zmon}, \text{zday}] \parallel Q \]
\[ M \vdash \text{client}_{\text{err}} \parallel Q \rightarrow * M \vdash P[\text{hr, min}/\text{hr, min}, \text{year, mon, day}/\text{year, mon, day}, \text{zyear}, \text{zmon}, \text{zday}] \parallel Q \]

However, in the case of \( \text{client}_{\text{err}} \), we can also have premature deallocation errors—the second subprocess \( \text{getDate}!x. x?c[(\text{year}, \text{mon}, \text{day})], \text{free} \cdot P_2 \) may deallocate the shared allocated channel, \( c \) below, while the right sub-process is still using it.

\[ M \vdash \text{client}_{\text{err}} \parallel Q \rightarrow * \]
\[ M \vdash \text{newd} : \bot \cdot (c?[(\text{hr, min})], P_1 \parallel c!(\text{hr, min}) \cdot \text{timeSrv} \parallel P_2[\text{year, mon, day}/\text{year, mon, day}, \text{zyear}, \text{zmon}, \text{zday}]) \parallel \text{dateSrv}^{\text{err}} \]

We can also have arity mismatch errors caused by communication interference on the shared allocated channel \( c \):
\[ M \vdash \text{client}_{\text{err}} \parallel Q \rightarrow * M \vdash \text{newd} : \top \cdot (c?[(\text{hr, min})], P_1 \parallel c!(\text{year, mon, day}) \cdot \text{dateSrv} \parallel (c?[(\text{year}, \text{mon}, \text{zday})], \text{free} \cdot P_2 \parallel c!(\text{hr, min}) \cdot \text{timeSrv})^{\text{err}} \]

Strong updates, as in the case of \( \text{client}_2 \), should only be allowed when there is no other processes still using the channel.

**Example 2.2**

Client \( \text{client}_3 \) demonstrates channel reuse across processes. Rather than allocating a new channel, \( \text{client}_3 \) requests a currently unused channel from a heap of channels and returns the channel to the heap when it no longer needs it. A heap, in this case, is a designated channel \( \text{heap} \).

\[ \text{client}_3 = \text{heap}!x. \text{getTime}!x. x?((\text{hr, min}), \text{getDate}!x. x?c[(\text{year}, \text{mon}, \text{zday})], \text{heap}!x \cdot P) \]

This allows us to have multiple replicas of the clients while keeping the cost of the server interactions (in terms of additional channels required) constant. Of course, in the arrangement below, client–server interactions are sequentialized for each client, depending on who currently has access to channel \( c \).
6 Uniqueness typing for message-passing concurrency

held on heap. But this is necessary so as not to have any interference during the multiple reuses of channel c.

\[ M \triangleright CLIENT_3 \parallel CLIENT_3 \parallel \ldots \parallel CLIENT_3 \parallel \pi\text{new}:T.(heap!c) \parallel TIME\text{srv} \parallel DATE\text{srv} \]  

(1)

It is instructive to run through an execution of this system and analyse how access to channel c evolves throughout computation, as it turns out that exclusive access to such a channel is what enables a process to safely performs operations such as deallocations and strong updates. For instance in (1), after a client inputs on heap, it obtains access to c; as standard in the pi-calculus, this is denoted by the scope extrusion of channel c to the client and the system in (1) reduces to

\[ M \triangleright CLIENT_3 \parallel CLIENT_3 \parallel \ldots \parallel \pi\text{new}:T.(CLIENT'_3) \parallel TIME\text{srv} \parallel DATE\text{srv} \]  

(2)

At that point, only CLIENT’_3 is in a position to interact with the servers, while the other clients are blocked waiting for the next output on channel heap. After CLIENT’_3 queries TIME\text{srv} the scope of c is extruded again, this time to TIME\text{srv}, and from (2) we obtain

\[ M \triangleright CLIENT_3 \parallel CLIENT_3 \parallel \ldots \parallel \pi\text{new}:T.(CLIENT''_3 \parallel TIME\text{srv}) \parallel DATE\text{srv} \]  

(3)

After the interaction with TIME\text{srv} terminates the scope of c can be tightened again to

\[ M \triangleright CLIENT_3 \parallel CLIENT_3 \parallel \ldots \parallel \pi\text{new}:T.(CLIENT'''_3 \parallel TIME\text{srv}) \parallel DATE\text{srv} \]  

(4)

In similar fashion the scope of c extrudes to DATE\text{srv} and then contracts back again to the client so as to reach the following system from (4).

\[ M \triangleright CLIENT_3 \parallel CLIENT_3 \parallel \ldots \parallel \pi\text{new}:T.(heap!c.P') \parallel TIME\text{srv} \parallel DATE\text{srv} \]  

(5)

At this point, the client heap!c.P’ explicitly transfers exclusive access to channel c to some other client.

\[ M \triangleright CLIENT_3 \parallel \pi\text{new}:T.(CLIENT'_3) \parallel \ldots \parallel P' \parallel TIME\text{srv} \parallel DATE\text{srv} \]  

(6)

The following variants of client CLIENT’_3 are unsafe and lead to errors in the above computational sequence:

\[ \text{CLIENT}_3^{err1} \triangleq \text{heap}!x, \text{getTime}!x.x?yhr, \text{ymin} \cdot \text{getTime}!x.x?y\text{year}, \text{zmon} \cdot \text{zday}.\text{heap}!x.x!P \]

\[ \text{CLIENT}_3^{err2} \triangleq \text{heap}!x, \text{getTime}!x.x?yhr, \text{ymin} \cdot \text{getTime}!x.x?y\text{year}, \text{zmon} \cdot \text{zday}.\text{free} x.\text{heap}!x.x \]

The erroneous client CLIENT’_3^{err1} highlights the fact that, for the above system to work as intended, it is crucial that the client transfers exclusive access to channel c on heap. If, instead, the client attempts to use this channel after transfer, it will lead to an interference that can result in a runtime error as shown earlier in Example 2.1. Although the communication of deallocated channels is permitted in resource pi-calculus, the second erroneous client CLIENT’_3^{err2} shows how this can indirectly lead to a premature deallocation when this deallocated channel is then transferred to other clients.
3 Type system

We define a type system to statically approximate the class of safe resource pi-calculus systems. The type-level concept that corresponds to ‘exclusive access’ is uniqueness: when a process is typed in a typing environment in which a channel has a unique type, then that process is guaranteed to have exclusive access to that channel. Strong update and deallocation are therefore safe for unique channels.

The type system is an adaptation of uniqueness type systems for lambda calculi [5, 9], but unlike the latter it allows uniqueness to be temporarily violated. As we saw in Example 2.1, a client might have exclusive access to a channel, then use this channel to communicate with a server, and then regain exclusive access to this channel, provided that the server does not use the channel anymore.

We must, therefore, generalize uniqueness typing to introduce a type that records that a channel may not be unique now, but it will be after some number \( i \) of communication steps (‘unique-after-\( i \)’); moreover, we need to combine uniqueness typing with affine types to limit how often a channel can be used (for instance, we might limit the time server to use a channel at most once).

3.1 The type language

The core type of our system is the channel type, \([\overrightarrow{T}]^n\), consisting of an \( n \)-ary tuple of types describing the values carried over the channel, \( \overrightarrow{T} \), and an attribute \( a \) which can take one of three forms:

- A channel of type \([\overrightarrow{T}]^{\omega}\) is an unrestricted channel: it can be used arbitrary often, and provides no information on how many other processes have access to the channel. Such type assumptions correspond to type assumptions of the form \([\overrightarrow{T}]\) in standard (non-substructural) type systems.
- A channel of type \([\overrightarrow{T}]^1\) is affine, and comes with an obligation: it can be used at most once.
- A channel of type \([\overrightarrow{T}]^{(\star, i)}\) is unique-after-\( i \) communication actions, comes with a guarantee: a process typed using this assumption will have exclusive access to the channel after \( i \) communication actions. We abbreviate the type \([\overrightarrow{T}]^{(\star, 0)}\) of channels that are unique now to \([\overrightarrow{T}]^{\star}\) and refer to it as simply a unique channel.

We give a number of examples illustrating the use of these types with reference to the examples from Section 2.

(1) Unrestricted channels can be used to describe channels such as \( \text{getTime} \) and \( \text{getDate} \) on which replicated servers receive requests.

(2) Affine types can be used to limit how often servers can use a return channel sent to them by a client. For instance, we can give the type \([\overrightarrow{T}]^{\star}\) to \( \text{getTime} \) and \( \text{getDate} \).

(3) A channel which has just been allocated will have a unique type. We can also use uniqueness to ensure that the heap really carries channels that are unused, by giving the heap channel the type \([\overrightarrow{T}]^{\star}\).

(4) A channel which is unique-after-1 can be used by the client to record that after it has communicated with a server it will once again recover exclusive access to the channel.

3.2 Typing environments

Processes will be typed under a typing environment \( \Gamma \), which is a multiset of pairs of identifiers and types. Since the type system is substructural, typing assumptions can be used only once in a typing
Uniqueness typing for message-passing concurrency

8 Type Structure

\[ T ::= [T] \quad \text{(channel type)} \]  
\[ \text{proc (process)} \]  
\[ a ::= \omega \quad \text{(affine)} \]  
\[ \cdot i \quad \text{(unrestricted)} \]  
\[ (\text{unique after } i \text{ steps}, i \in \mathbb{N}) \]

Type Splitting

\[ \frac{[T] b = [T] b \circ [T] b}{\text{pUNR}} \quad \frac{\text{proc} = \text{proc } \circ \text{proc}}{\text{pPROC}} \quad \frac{[T] \cdot (\cdot i) = [T] \circ [T] \cdot (\cdot i + 1)}{\text{pUNQ}} \]

Subtyping

\[ \frac{(\cdot i) \varphi \cdot (\cdot i + 1)}{\text{sINDEX}} \quad \frac{\omega \varphi \omega}{\text{sUNQ}} \quad \frac{a \varphi a_2}{\text{sAFF}} \quad \frac{[T] \varphi [T] b}{\text{sTyp}} \]

Structural rules \((\varphi)\) is the least reflexive transitive relation satisfying

\[ T = T_1 \circ T_2 \quad \Gamma, u : T \prec \Gamma, u : T_1, u : T_2 \quad \text{tCON} \]
\[ T = T_1 \circ T_2 \quad \Gamma, u : T_1, u : T_2 \prec \Gamma, u : T \quad \text{tJOIN} \]
\[ \Gamma, u : T_1 \prec \Gamma, u : T_2 \quad \text{tWEAK} \]
\[ \Gamma, u : T_1 \prec \Gamma, u : T_2 \quad \text{tSUB} \]
\[ \Gamma, u : [T_1] \cdot (\cdot i) \prec \Gamma, u : [T_2] \cdot (\cdot i) \quad \text{tREV} \]

Figure 3. Type language and operations.

derivation. For instance, when typing two parallel processes under some environment \(\Gamma\), assumptions in \(\Gamma\) can be used to type the left process or the right, but never both.

However, we define a number of structural rules, shown in Figure 3, which allow to manipulate typing environments. The simplest is weakening (tWeak), which allows to disregard typing assumptions (the type system does not guarantee that a channel will be used). We discuss the other rules separately, below.

3.2.1 Splitting and joining

Although typing assumptions can be used only once, typing assumptions can be split (logically, contracted) under certain conditions (tCON). We write \(T = T_1 \circ T_2\) if an assumption \(x : T\) can be split as two assumptions \(x : T_1\) and \(x : T_2\).

For instance, after allocation CLIENT_2 is allowed to split the unique assumption \(x : [\overrightarrow{T}] \cdot (\cdot i)\) as two separate assumptions \(x : [\overrightarrow{T}] \cdot (\cdot i - 1)\) and \(x : [\overrightarrow{T}] 1\). This allows the client to send \(x\) to the server at type [\overrightarrow{T}] while retaining the assumption \(x : [\overrightarrow{T}] \cdot (\cdot i)\) for its own use.

More generally, an assumption \(c : [\overrightarrow{T}] \cdot (\cdot i)\) can be split as two assumptions \(c : [\overrightarrow{T}] \cdot (\cdot i - 1)\) and \(c : [\overrightarrow{T}] 1\). Unrestricted assumptions can be split as two unrestricted channels, so that unrestricted channels can be used arbitrary often. Finally, affine assumptions cannot be split at all, so that affine channels can be used only once.

Joining (tJOIN) is simply the dual of splitting; it is not strictly necessary to define the type system, but makes the technical development a lot cleaner.
3.2.2 Revision

After the client from Example 2.1 communicates with the time server over channel $x$, it regains exclusive access to that channel. It is therefore safe to the client to reuse the same channel to communicate with the date server, even though the date server sends values of a different type. This strong update is safe only for unique channels, and is embodied by rule $\tau_{\text{Rev}}$.

3.2.3 Subtyping

Since uniqueness is a guarantee that a channel $c$ will be unique after a number of uses, it is safe to weaken this guarantee: a channel at type $\langle \vec{T}, \cdot, u \rangle$ can also be used at type $\langle \vec{T}, \cdot, u+1 \rangle$ or even at $\langle \vec{T}, \omega \rangle$. For example, after getting a unique channel from the heap a process might choose to disregard the uniqueness information about this channel and use it in an unrestricted fashion instead.

Conversely, affinity is an obligation forcing an upper bound on the number of uses. Thus, it is also safe to regard a unique channel or an unrestricted channel as an affine channel. For instance, a client might send an unrestricted channel to a server expecting an affine channel.

We therefore have the following subtyping chain ($\tau_{\text{Sub}}$):

$$\langle \vec{T}, \bullet \rangle \prec_s \langle \vec{T}, \bullet, 1 \rangle \prec_s \langle \vec{T}, \bullet, 2 \rangle \prec_s \cdots \langle \vec{T}, \omega \rangle \prec_s \langle \vec{T}, 1 \rangle$$

Through a combination of subtyping and splitting, we can split a unique-after-$i$ assumption into two unrestricted assumptions, $\langle \vec{T}, \bullet, i \rangle \prec_s \langle \vec{T}, \omega \rangle = \langle \vec{T}, \omega \rangle \circ \langle \vec{T}, \omega \rangle$, and an unrestricted assumption into an unrestricted assumption and an affine assumption, $\langle \vec{T}, \omega \rangle = \langle \vec{T}, \omega \rangle \circ (\langle \vec{T}, \omega \rangle \prec_s \langle \vec{T}, 1 \rangle)$.

3.2.4 Consistency

The type systems allow for compositional reasoning with respect to uniqueness (we prove formally in Section 4.2). For instance, the clients in Examples 2.1 and 2.2 can be shown to be to be safe using only the typing assumptions assigned to them, and without requiring the full code relating to the servers they interact with.

Such compositional reasoning is safe only if the total type assumptions used to analyse the entire system are consistent among themselves. For example, an environment of the form

$$c : [\langle \vec{T}, \bullet \rangle, c : [\vec{T}] \rangle$$

allows to type a system in which one process deallocates channel $c$ while another, parallel, process attempts to communicate on it. This is clearly unsound.

Clearly an environment with at most one assumption per channel (a partial function) must be consistent, but we have already seen that we sometimes need to split assumptions. We define an environment to be consistent if it can be obtained from a partial function by applying any of the structural rules we just defined:

\textbf{Definition 3.1 (Consistency)}

A typing environment $\Gamma$ is consistent if there exists a partial function $\Gamma'$ such that $\Gamma' \leq \Gamma$.

\footnote{Since we do not consider channel input/output-modalities in this article, channels are invariant with respect to their object types, $\vec{T}$.}
Uniqueness typing for message-passing concurrency

This definition is justified by the soundness theorem, which says that if a system can be typed under a consistent environment, it will not produce any runtime errors. For now, observe that the structural rules have been carefully defined so that if a consistent environment contains an assumption $c : [\overrightarrow{T}]^{(i, 1)}$ then it cannot contain any unrestricted assumptions about $c$, and the number of affine assumptions about $c$ must be less than or equal to $i$ (Lemma 4.1).

3.3 The typing relation

The typing relation over (open and closed) processes is the least relation defined by the rules in Figure 4. It takes the usual shape, $\Gamma \vdash P$, which reads as ‘$P$ is well-typed under the typing environment $\Gamma$’. It is extended to systems so that $\Gamma \vdash M \triangleright P$ is well-typed if $\Gamma$ types $P$ and $\Gamma$ only contains assumptions about channels that are allocated in $M$.

We do not a priori make the assumption that the type environment is consistent (Section 3.5), but soundness of the type system will only be stated with respect to consistent environments.

The rule for parallel (tPar) requires to divide the assumptions in the typing environment between the left and right process. Restriction introduces a new typing assumption into the typing environment, provided that the channel is allocated (tRst1, tRst2). As expected, allocation introduces unique channels (tAll), and only unique channels can be deallocated (tFree). We discuss the remaining rules separately, below.

3.3.1 Input and output

In the client–server example, after the client has communicated with the server over channel $x$, the server is no longer allowed to use $x$, and the client is guaranteed that it once again has exclusive access to the channel. From a typing perspective, the server had an affine permission $x : [\overrightarrow{T}]_{1}$ and lost that permission after the communication; the client had a permission $x : [\overrightarrow{T}]^{(i, 1)}$ which became $x : [\overrightarrow{T}]_{i}$ after the communication.

This is expressed in the type system as an operation $\Gamma, c : [\overrightarrow{T}]_{a - 1}$ on typing environments (defined in Figure 4); it computes the typing environment after a communication action on $c$: affine permission are removed, unique-after-$(i + 1)$ permissions become unique-after-$i$ and unrestricted permissions are unaffected.

This operation is used in the rules for input and output (tOut and tIn). Moreover, tOut requires separate typing assumptions for each of the channels that are sent. The attributes on these channels are not decremented, because no communication action has been performed on them; instead, the corresponding assumptions are handed over to the parallel process receiving the message. If the sending process wants to use any of these channels in its continuation ($P$) it must split the corresponding assumptions first.

3.3.2 Conditionals

As is standard in substructural (and standard) type systems, the rule for conditions RIf types both branches under the same typing environment, as only one will be executed. It also requires that the process must have a typing assumption for the two channels it is comparing.

We claimed that after a communication with the server over channel $x$ the client regains exclusive access to that channel. At the type level, the server looses its affine permission for $x$ after the communication, as we saw above. The side condition on the rule for conditionals implies that this
means that the server cannot even use \( x \) in a comparison anymore, thus guaranteeing the client true exclusive access to the channel.

### 3.3.3 Recursion

Recursive processes must be typed in an environment that contains only unrestricted channels (\( \text{tRec} \)). This is reasonable since recursion can be used to define processes with an unbounded number of parallel uses of some channel. Nevertheless, it is not as serious a restriction as it may seem, as recursive processes can still send, receive and allocate unique channels. For instance, the following process models an ‘infinite heap’ that keeps allocating new channels and sends them across a channel \( \text{heap} : [\![ T ]\!] \):

\[
\text{INFHEAP} \triangleq \text{rec} X . \text{alloc} X . \text{heap}! x . X
\]

### 3.4 Examples

The systems \( \text{client}_i \parallel \text{timesrv} \parallel \text{datesrv} \) for \( i \in \{0, 1, 2\} \) can all be typed in our type system, whereas \( \text{client}_{\text{err}} \) is rejected because type splitting enforces a common object type (cf. \( \text{pUnr} \), \( \text{pUnq} \) in Figure 4.) The derivation below outlines how \( \text{client}_2 \) from the introduction is typed, where we recall

\[
\begin{align*}
\Gamma, c : T \vdash P & \quad \text{TRst1} \\
\Gamma, c : T \vdash P & \quad \text{TRst2} \\
\end{align*}
\]
Uniqueness typing for message-passing concurrency

that $x$ is not free in $P$. It assumes an environment $\Gamma = getTime :: [[T_1, T_2]^1\omega]$, $getDate :: [[T_3, T_4, T_5]^1\omega]$ and proceeds as follows:

\[
\Gamma, x : [T_3, T_4, T_5]*, \gamma hr : T_1, \gamma min : T_2 \vdash getDate! x . x ? (z_{year}, z_{mon}, z_{day}) . \text{free} x . P \quad \text{T STR}
\]

\[
\Gamma, x : [T_1, T_2]^*[\omega], x : [T_1, T_2]^1 \vdash getDate! x . x ? (z_{year}, z_{mon}, z_{day}) . \text{free} x . P \quad \text{T IN}
\]

\[
\Gamma, x : [T_1, T_2]^*[\omega], y : [T_1, T_2]^1 \vdash getDate! x . x ? (z_{year}, z_{mon}, z_{day}) . \text{free} x . P \quad \text{T IN}
\]

Rule $\text{TA LL}$ assigns the unique type $[T_1, T_2]^\bullet$ to variable $x$. Using $\text{T STR}$ with $\text{T CON}$, this unique assumption is split in two using $\text{P UNQ}$. Rule $\text{T OUT}$ uses the affine assumption for $x$ for the output argument and the unique-after-1 assumption to type the continuation. Rule $\text{T IN}$ restores the uniqueness of $x$ for the continuation of the input after decrementing the uniqueness index, at which point $\text{T REV}$ is applied to the environment through $\text{T STR}$ to change the object type of $x$ from pairs of values (for time) to triples (for dates). The pattern of applying $\text{T STR}$ with $\text{T CON}$, $\text{T OUT}$ and $\text{T IN}$ repeats, at which point $x$ is unique again and can be safely deallocated by $\text{T FREE}$.

Through the type environment $\Gamma$ below, we can type system (1) from Example 2.2 as well.

\[
\Gamma = \text{getTime} :: [[T_1, T_2]^1\omega], \text{getDate} :: [[T_3, T_4, T_5]^1\omega], \text{heap} :: [[\bar{T}]^\omega]
\]

Since the type for channel $\text{heap}$ is unrestricted, we can split this assumption for every occurrence of $\text{CLIENT}_3$ in (1). When typechecking each instance of this client, we obtain a unique channel assumption as soon as we apply $\text{T IN}$ for the input on channel $\text{heap}$. The rest of the type derivation then proceeds in similar fashion to that of $\text{CLIENT}_2$ discussed above. It is, however, instructive to see why $\text{CLIENT}_3^{\omega\omega}$ and $\text{CLIENT}_3^{\omega2}$, also from Example 2.2, cannot be typed. Through the application of $\text{T STR}$ with $\text{T CON}$, $\text{T OUT}$ and $\text{T IN}$ we reach the two type judgements below for $\text{CLIENT}_3^{\omega\omega}$ and $\text{CLIENT}_3^{\omega2}$, respectively:

\[
\Gamma, x : [T_3, T_4, T_5]*, \gamma yr : T_1, \gamma min : T_2, z_{year} : T_3, z_{mon} : T_4, z_{day} : T_5 \vdash \text{heap}! x . x \cdot P \quad (7)
\]

\[
\Gamma, x : [T_3, T_4, T_5]*, \gamma yr : T_1, \gamma min : T_2, z_{year} : T_3, z_{mon} : T_4, z_{day} : T_5 \vdash \text{free} x . \text{heap}! x . P \quad (8)
\]

In the case of type derivation (7), even though we can revise the object type of the type assumption for $x$ from $T_3, T_4, T_5$ to $\bar{T}$, so as to enable the typing of the output of $x$ on channel $\text{heap}$, this would leave us with no remaining type assumptions for channel $x$, precluding the typing of the remaining process $x! . P$. Similarly, in the case of type derivation (8), typing the deallocation of $x$ would consume all the type assumptions relating to $x$, which rules out any possible typing for the remaining process $\text{heap}! x . P$.

Finally, system (1) in Example 2.2 can be safely extended with processes such as $\text{CLIENT}_4$ which uses unique channels obtained from the heap in unrestricted fashion. Our type system accepts $\text{CLIENT}_4$ by applying subtyping from unique to unrestricted on the channel $x$ obtained from $\text{heap}$.

\[
\text{CLIENT}_4 \triangleq \text{heap}? x . x \in X . (\text{getTime}! x . x ? (y_{hr}, y_{min}) . P \parallel X)
\]
3.5 (In)consistency

Due to our substructural treatment of assumption, when two processes both need a typing assumption relating to the same channel \( c \), \( \text{tPar} \) from Figure 4 forces us to have two separate assumptions in the typing environment. As we discussed already, these two assumptions need not be identical and can be derived from a single assumption; for example, an assumption \( c : [\cdot]^{\bullet} \) can be split as two assumptions

\[
c : [\cdot]^{\bullet}, c : [\cdot]^{1}
\]

As mentioned, soundness of the type system is stated only with respect to consistent environments. However, communication in the pi-calculus provides synchronization points: when a \( u \) is constructed and can be split as two assumptions

\[
a : [\cdot]^{\bullet}, u : [\cdot]^{\cdot}, x : [\cdot]^{1} \vdash \text{freeu} \parallel a!x
\]

\[
a : [\cdot]^{\cdot}, u : [\cdot]^{\cdot}, x : [\cdot]^{1} \vdash u() \cdot \text{freeu} \parallel a!x
\]

\[
a : [\cdot]^{\cdot}, u : [\cdot]^{\cdot}, x : [\cdot]^{1} \vdash a!u \parallel a?x.u().\text{freeu} \parallel a!x
\]

\[
a : [\cdot]^{\cdot}, u : [\cdot]^{\cdot}, x : [\cdot]^{1} \vdash a!u \parallel a?x.u().\text{freeu} \parallel a!x
\]

This is a valid typing derivation, and moreover the typing environment used at every step is consistent. But now consider what happens after this process takes a reduction step:

\[
u : [\cdot]^{\cdot}, u : [\cdot]^{\cdot} \vdash \text{freeu} \parallel a!u
\]

\[
u : [\cdot]^{\cdot}, u : [\cdot]^{\cdot} \vdash u() \cdot \text{freeu} \parallel a!u
\]

The continuation of this process looks suspicious as it attempts to free \( u \) while simultaneously sending it on \( a \). Indeed, \( \text{freeu} \parallel a!u \) can only be typed in an inconsistent environment \( u : [\cdot]^{\cdot}, u : [\cdot]^{\cdot} \). Nevertheless, the fact that this process is typeable is not a violation of type safety. The assumption \( u : [\cdot]^{\cdot} \) tells us that there are no processes that output on \( u \) so that the input on \( u \) is blocked: the continuation of the process will never execute.

Thus, when an environment

\[
\Gamma, c : [\overline{T}]^u, c : [\overline{T}]^{u'}
\]

(e.g. \( u : [\cdot]^{\cdot}, u : [\cdot]^{1} \)) is consistent, it may be the case that

\[
\Gamma, c : [\overline{T}]^{u-1}, c : [\overline{T}]^{u'}
\]

(e.g. \( u : [\cdot]^{\cdot}, u : [\cdot]^{1} \)) is inconsistent: this means that the tails of input or output processes may have to be typed under inconsistent environments, even when the larger process is typed in a consistent environment. However, communication in the pi-calculus provides synchronization points: when
Uniqueness typing for message-passing concurrency

Two processes communicate, both will start executing their continuations. It is an important property of our typing environments that if \( \Gamma, c : \overrightarrow{T}^n, c : \overrightarrow{T}^n \) is consistent, then

\[ \Gamma, c : \overrightarrow{T}^n, c : \overrightarrow{T}^n \]

will also be consistent (Lemma 4.1). This is crucial when proving subject reduction, which says that if a system is typeable in a consistent environment, it will remain typeable in a consistent environment.

Note that allowing for inconsistent environments during type checking is not unusual for these kinds of type systems; for instance, in the session type system with subtyping of Gay and Hole a very similar issue arises [13, Section 5.2] (they use the word ‘unbalanced’ instead of inconsistent).

4 Soundness

We show that our type system is sound: well-typed programs do not go wrong. The presentation of this proof proceeds as follows: Section 4.1 outlines some preliminary properties of consistent environments and the typing relation, and states the necessary substitution lemmas; Section 4.2 details the subterm typing lemma, which turns out to be the most involved aspect of the soundness proof; Section 4.3 finally states and proves the standard safety and subject reduction theorems: well-typed programs do not have runtime errors and remain well-typed when they reduce.

4.1 Preliminaries

We first state some properties of consistent environments.

**Lemma 4.1 (Consistent environments)**

Let \( \Gamma \) be a consistent environment. Then:

1. If \( \Gamma = \Gamma', c : \overrightarrow{T}^n, c : \overrightarrow{T}^n \) then \( \overrightarrow{T} = \overrightarrow{T}' \).
2. If \( \Gamma = \Gamma', c : \overrightarrow{T}^n \) then \( c \notin \text{dom}(\Gamma') \).
3. If \( \Gamma \preceq \Gamma', c : \overrightarrow{T}^n \wedge \exists \Gamma'' : \overrightarrow{T}^n, c : \overrightarrow{T}^n \) such that \( \Gamma = \Gamma'', c : \overrightarrow{T}^n \) and \( \Gamma'' \preceq \Gamma' \).
4. If \( c \in \text{dom}(\Gamma) \) and \( \Gamma' \preceq \Gamma \) then \( c \in \text{dom}(\Gamma') \).
5. If \( \Gamma = \Gamma_1, \Gamma_2 \) then both \( \Gamma_1 \) and \( \Gamma_2 \) are consistent.
6. If \( \Gamma \) contains only unrestricted assumptions then \( \Gamma \preceq \Delta, \Gamma \).
7. If \( \Gamma = \Gamma', u : \overrightarrow{T}^m, u : \overrightarrow{T}^p \) then

\[ \Gamma', u : \overrightarrow{T}^m, u : \overrightarrow{T}^p \preceq \Gamma', u : \overrightarrow{T}^m - 1, u : \overrightarrow{T}^p - 1 \]

**Proof.** Properties (1)–(6) are easily verified. For the proof of (7), we do case analysis on \( a_1 \) and \( a_2 \). If \( a_1, a_2 \) are affine or unrestricted, the lemma follows from weakening (unrestricted assumptions are not affected by the decrement operation). The only interesting case is where \( a_1 = (\cdot, i) \) and \( a_2 = 1 \) (or vice versa), in which case it follows from \( \top JOIN \).

One of the advantages of phrasing the structural rules of the type system using an auxiliary transitive relation \( \preceq \) is that we can conveniently state inversion principles. For example,

**Lemma 4.2 (Inversion for output)**
Uniqueness typing for message-passing concurrency

If \( \Gamma \vdash e \bar{d}.P \) then
\[
\Gamma \preceq \Gamma', u: \overrightarrow{T}, d:T
\]
\( \Gamma', u: \overrightarrow{T} \bar{d}-1 \vdash P \).

The inversion lemmas for the other constructs are similar.

As usual, we need two substitution lemmas; however, the typing rules have been set up in such a way that these are easy to prove.

**Lemma 4.3 (Process substitution)**

If \( \Gamma_1, \overrightarrow{X} : \text{proc} \vdash P \) and \( \Gamma_2 \vdash Q \) then \( \Gamma_1, \Gamma_2 \vdash P[Q/\overrightarrow{X}] \).

**Proof.** By induction on \( P \) followed by inversion on the typing relation. The only tricky case is the case for parallel, where we might duplicate the process. However, since the process is typed under an environment containing only unrestricted assumptions, this follows from Lemma 4.1(6).

**Lemma 4.4 (Identifier substitution)**

If \( \Gamma_1, \bar{x} : T \vdash P \), where the \( \bar{x} \) are pairwise disjoint and do not occur in the domain of \( \Gamma \), then \( \Gamma, u: \overrightarrow{T} \vdash Q[\bar{u}/\bar{x}] \).

**Proof.** This is a simple renaming of variables throughout the typing derivation.

**Lemma 4.5 (Preservation of types under structural equivalence)**

If \( \Gamma \vdash P \) and \( P \equiv P' \) then \( \Gamma \vdash P' \).

**Proof.** Since the typing relation is not sensitive to the order of the assumptions in the typing environment, reordering parallel processes and extrusion do not affect typing. Removing or adding \( \text{nil} \) processes and adding or removing a restriction around the \( \text{nil} \) process do not affect typing because \( \text{nil} \) can be typed in any environment. Finally, alpha-renaming bound names does not affect typing because those bound names are not in the (original) typing environment.

**4.2 Subterm typing lemma**

The preservation theorem (Section 4.3) states that if a system \( M \Rightarrow P \) is typed under some environment \( \Gamma \), and \( M \Rightarrow P \) takes a step to \( M' \Rightarrow P' \), then this new system is typeable under some typing environment \( \Gamma' \) such that such that \( \Gamma \preceq \Gamma' \). An important case in the proof of this theorem is the case for contexts; i.e. part of the system takes a step, while the rest of the system remains the same. This case is dealt with by the subterm typing lemma, which we state and prove in this section. The subterm typing lemma is important, because it shows that the type system is compositional: in particular, if \( P \) and \( Q \) are two parallel processes in some system, and \( P \) deallocates a channel, then this will not affect \( Q \).

Unfortunately, the subterm typing lemma is also a little technical to state and non-trivial to prove. We will, therefore, give the proof in detail, after we first give a number of small definitions and lemmas that we will require in the proof.

**Definition 4.6 (Restriction of a typing environment to a resource environment)**

\[
\Gamma|_M = \{ d:T \mid d:T \in \Gamma \land M(d) = \top \}
\]

**Definition 4.7 (Applying a context to a typing environment)**
16 Uniqueness typing for message-passing concurrency

We define an operation $C(\Gamma)$ which removes all assumptions about channels that are restricted in $C$ from $\Gamma$:

$$[[\Gamma]] = \Gamma$$

$$(\pi \in w, C)(\Gamma) = C(\Gamma) \setminus c$$

$$(C \parallel P)(\Gamma) = C(\Gamma)$$

$$P \parallel C(\Gamma) = C(\Gamma)$$

**Definition 4.8 (Subenvironment)**

$M'$ is a subenvironment of $M$ with respect to a typing environment $\Gamma$, denoted as $M \preceq_\Gamma M'$, if

1. $\text{dom}(M') = \text{dom}(M)$;
2. If $M(c) = \top$ and $M'(c) = \bot$ then $c : [\top] \in \Gamma$.

Intuitively, $M \preceq_\Gamma M'$ if we might end up in state $M'$ by running a process typed under $\Gamma$ in state $M$. This is formalized by the following lemma:

**Lemma 4.9 (Subenvironments and Reduction)**

$\Gamma \vdash M \rightarrow P$ and $M \rightarrow P \rightarrow M' \rightarrow P'$ then $M \preceq_\Gamma M'$

We will also require the following result which relates subenvironments and the operation on typing environments that we defined above.

**Lemma 4.10**

Let $M_1 \preceq_\Gamma M_2$. Suppose $C[M_1 \triangleright P_1] = M_1' \triangleright P_1'$ and $C[M_2 \triangleright P_2] = M_2' \triangleright P_2'$ for some $P_1$ and $P_2$. Then $M_1' \preceq_{C(\Gamma)} M_2'$.

We can now state the subterm typing lemma:

**Lemma 4.11 (Subterm typing)**

Suppose $\Gamma_0^1 \vdash C[M_1 \triangleright P_1]$ ($\Gamma_0^1$ consistent). Then the following hold.

1. There exist some $\Gamma_0^2$ such that $\Gamma_0^2 \preceq_{C(\Gamma)^2} \Gamma_0^0$ and $\Gamma_0^2 \vdash M_1 \triangleright P_1$.
2. For all systems $M_2 \triangleright P_2$ where $M_1 \preceq_{\Gamma_0^1} M_2$ and for all environments $\Gamma_0^2$ where $\Gamma_0^0 \preceq_{\Gamma_0^1} \Gamma_0^2$,
   
   if $\Gamma_0^2 \vdash M_2 \triangleright P_2$ then there exist a $\Gamma_0^2$ such that $\Gamma_0^2 \preceq_{C(\Gamma)^2} \Gamma_0^0$ and $\Gamma_0^2 \vdash C[M_2 \triangleright P_2]$.

In words: if a system $M_1 \triangleright P_1$ is placed in some context $C$, and the whole thing is typeable under some environment $\Gamma_0^1$, then

1. There is some environment $\Gamma_0^0$ which types $M_1 \triangleright P_1$; the assumptions in this environment come from $\Gamma_0^1$, with the exception of the assumptions for channels restricted by $C$, and
2. If we replace $M_1 \triangleright P_1$ by $M_2 \triangleright P_2$, typed under an environment $\Gamma_0^2$ obtained by applying structural rules to $\Gamma_0^0$, then—provided there are unique permissions in $\Gamma_0^0$ for the channels that are allocated in $M_1$ but deallocated in $M_2$—the whole thing is typeable under an environment $\Gamma_0^2$, obtained by applying structural rules to $\Gamma_0^1$.

**Proof:** By induction on $C$. The base case ($C = []$) is trivial; remains to consider case for parallel and the case for restriction.
(1) Case $\mathcal{C} = \mathcal{C}' \parallel Q$. Let $\mathcal{C}[M_1 \triangleright P_1] = M'_1 \triangleright P'_1$. Then $\mathcal{C}[M_1 \triangleright P_1] = M'_1 \triangleright P'_1 \parallel Q$. By inversion we get

$$
\frac{\Gamma^c \vdash P'_1 \quad \Gamma^q \vdash Q \quad \Gamma^c \preceq \Gamma^c_1 \triangleq \Gamma^c \quad \text{dom}(\Gamma^c_1) \subseteq \text{alloc}(M'_1)}{\Gamma^c \vdash \mathcal{C}[M_1 \triangleright P_1] = M'_1 \triangleright P'_1 \parallel Q}
$$

(9)

It is not hard to see that

$$
\frac{\Gamma^c \vdash P'_1 \quad \text{dom}(\Gamma^c_1) \subseteq \text{alloc}(M'_1)}{\Gamma^c \vdash \mathcal{C}'[M_1 \triangleright P_1] = M'_1 \triangleright P'_1}
$$

so that by the induction hypothesis there exist a $\Gamma^q_2$ such that $\Gamma^c_2 \preceq \mathcal{C}'(\Gamma^q_2)$ and $\Gamma^c_2 \triangleright M_2 \triangleright P_2$. Choose $\Gamma^q = \Gamma^q_2$. Clearly $\Gamma^c_2 \preceq \Gamma^c(\Gamma^q_2) = \mathcal{C}(\Gamma^q_2)$.

Suppose we have a system $M_2 \triangleright P_2$ where $M_1 \preceq \Gamma^q_2 M_2$ and an environment $\Gamma^q_2$ where $\Gamma^q_2 \preceq \Gamma^q_2$, such that $\Gamma^q_2 \vdash M_2 \triangleright P_2$. By the induction hypothesis, there exist a $\Gamma^c_2$ such that $\Gamma^c_2 \preceq \Gamma^c_2$ and $\Gamma^c_2 \vdash \mathcal{C}'[M_2 \triangleright P_2]$.

Let $\mathcal{C}[M_2 \triangleright P_2] = M'_2 \triangleright P'_2 \parallel Q$ where $\mathcal{C}'[M_2 \triangleright P_2] = M'_2 \triangleright P'_2$. By inversion we have

$$
\frac{\mathcal{C}_2 \vdash P'_2 \quad \text{dom}(\Gamma^c_2) \subseteq \text{alloc}(M'_2)}{\Gamma^c_2 \vdash \mathcal{C}'[M_2 \triangleright P_2] = M'_2 \triangleright P'_2}
$$

(10)

By (9) we have $\Gamma^q = \Gamma^q_2 | M'_1$, and by Lemma 4.10 we have $M'_1 \preceq \mathcal{C}(\Gamma^q_2) | M'_2$. Hence, channels allocated in $M'_1$ but deallocated in $M'_2$ must have a unique permission in $\mathcal{C}(\Gamma^q_2)$. Since $\Gamma^c_2 \preceq \mathcal{C}(\Gamma^q_2)$ they must therefore have a unique permission in $\Gamma^c_2$. Finally, since $\Gamma^c_2 \preceq (\Gamma^c_2, \Gamma^q)$ and $\Gamma^c_2$ is assumed consistent, this means that there cannot be any assumptions about these channels in $\Gamma^q$. Hence $\Gamma^q = \Gamma^q_2 | M'_1 = \Gamma^q_2 | M'_2$. Finally, we have $\Gamma^c_2 = \Gamma^c_2 | M'_2$ by (10).

Choose $\Gamma^c_2 = \Gamma^c_1 | M'_2$ (clearly $\Gamma^c_1 \preceq \Gamma^c_2$). We have

$$
\Gamma^c_2 \preceq (\Gamma^c_1 \triangleright \Gamma^q_2) | M'_2 \preceq (\Gamma^c_2 \triangleright \Gamma^q_2) | M'_2 = \Gamma^c_2 \triangleright \Gamma^q
$$

The proof is completed by

$$
\frac{\mathcal{C}_2 \vdash P'_2 \quad \Gamma^q \vdash Q}{\mathcal{C}_2 \triangleright \Gamma^q \vdash P'_2 \parallel Q \quad \text{TPAR}}
$$

$$
\frac{\mathcal{C}_2 \triangleright \Gamma^q \vdash P'_2 \parallel Q \quad \text{dom}(\Gamma^c_2) \subseteq \text{alloc}(M'_2)}{\Gamma^c_2 \vdash \mathcal{C}[M_2 \triangleright P_2] = M'_2 \triangleright P'_2 \parallel Q \quad \text{TSTR}}
$$

$$
\frac{\Gamma^c_2 \vdash \mathcal{C}[M_2 \triangleright P_2] = M'_2 \triangleright P'_2 \parallel Q}{\Gamma^c_2 \vdash \mathcal{C}_2 \triangleright \Gamma^q \vdash P'_2 \parallel Q \quad \text{TSTR}}
$$

$$
\frac{\Gamma^c_2 \vdash \mathcal{C}[M_2 \triangleright P_2] = M'_2 \triangleright P'_2 \parallel Q}{\Gamma^c_2 \vdash \mathcal{C}_2 \triangleright \Gamma^q \vdash P'_2 \parallel Q \quad \text{TSTR}}
$$
18  Uniqueness typing for message-passing concurrency

(2) Case \( C \equiv \text{newc}.C' \). Let \( C'[M_1 \triangleright P_1] = M'_1, c : s_1 \triangleright P'_1 \). Then \( C[M_1 \triangleright P_1] = M'_1 \triangleright \text{newc}.s_1.P'_1 \).

We take cases on \( s_1 \).

(a) Case \( s_1 = T \). By inversion we get

\[
\begin{align*}
\Gamma^C_1 : c : T \vdash P_1' & \quad \Gamma^C_1 \subseteq \Gamma^C_1' \quad \text{dom}(\Gamma^C_1) \subseteq \text{alloc}(M'_1) \\
\Gamma^C_1 \vdash C[M_1 \triangleright P_1] = M'_1 \triangleright \text{newc}.T.P'_1
\end{align*}
\]

(11)

where \( c \notin \text{dom}(\Gamma^C_1) \subseteq \text{dom}(\Gamma^C_1') \) by Barendregt. It is not hard to see that

\[
\begin{align*}
\Gamma^C_1 : c : T \vdash P_1' & \quad \text{dom}(\Gamma^C_1, c : T) \subseteq \text{alloc}(M'_1, c : T) \\
\Gamma^C_1' : c : T \vdash C'[M_1 \triangleright P_1] = M'_1, c : T \triangleright P'_1
\end{align*}
\]

so that by the induction hypothesis, there exist \( \Gamma^H_1 \) such that \( \Gamma^H_1 : c : T \vdash C'(\Gamma^H_1) \) and \( \Gamma^H_1 \vdash M_1 \triangleright P_1 \). Choose \( \Gamma^H_1 = \Gamma^H_2 \). We have

\[
\Gamma^C_1 \subseteq (\Gamma^H_1 : c : T) \vdash C'(\Gamma^H_1) \vdash C(\Gamma^H_1)
\]

Suppose we have a system \( M_2 \triangleright P_2 \) where \( M_1 \preceq \Gamma^H_1 \) and an environment \( \Gamma^H_2 \) where \( \Gamma^H_2 \preceq \Gamma^H_1 \), such that \( \Gamma^H_2 \vdash M_2 \triangleright P_2 \). By the induction hypothesis, there exist \( \Gamma^H_2 \) such that \( (\Gamma^C_1, c : T) \preceq \Gamma^H_2 \) and \( \Gamma^H_2 \vdash C'[M_2 \triangleright P_2] \).

Let \( C[M_2 \triangleright P_2] = M'_2, c : s_2 \triangleright P'_2 \) where \( C'[M_2 \triangleright P_2] = M'_2, c : s_2 \triangleright P'_2 \). By inversion we have

\[
\begin{align*}
\frac{\Gamma^H_2 \vdash P'_2 \quad \text{dom}(\Gamma^H_2) \subseteq \text{alloc}(M'_2, c : s_2) \quad \text{dom}(\Gamma^H_2) \subseteq \text{alloc}(M'_2)}{\Gamma^H_2 \vdash C'[M_2 \triangleright P_2] = M'_2, c : s_2 \triangleright P'_2}
\end{align*}
\]

(12)

Choose \( \Gamma^H_2 = \Gamma^H_2 | M'_2 \) (clearly \( \Gamma^H_2 \subseteq \Gamma^H_2 \)). Note that since \( \Gamma^C_1 \preceq \Gamma^H_2 \) by (11) we have \( \Gamma^C_2 \preceq \Gamma^C_2 | M_2 \), and moreover \( \Gamma^H_2 = \Gamma^H_2 | M_2, c : s_2 \) by (12).

We take cases on \( s_2 \).

(i) Case \( s_2 = T \). We have \( \Gamma^C_2, c : T \vdash (\Gamma^C_1 | M_2 : c : T) = (\Gamma^C_1' | M'_2 : c : T) \leq \Gamma^C_2' \).

The proof is completed by

\[
\begin{align*}
\frac{\Gamma^H_2 \vdash P'_2 \quad \text{STR}}{\Gamma^H_2, c : T \vdash P'_2} & \quad \frac{\Gamma^H_2 \vdash \text{newc}.T.P'_2 \quad \text{TRST1}}{\Gamma^H_2 \vdash \text{newc}.T.P'_2} \\
\frac{\Gamma^C_2 \vdash C[M_2 \triangleright P_2] = M'_2 \triangleright \text{newc}.T.P'_2 \quad \text{dom}(\Gamma^C_2) \subseteq \text{alloc}(M'_2) \quad \text{dom}(\Gamma^C_2) \subseteq \text{alloc}(M'_2)}{\Gamma^C_2 \vdash C[M_2 \triangleright P_2] = M'_2 \triangleright \text{newc}.T.P'_2}
\end{align*}
\]
We can now state and prove the main theorems of this article: type safety and safety reduction.

4.3 Safety and preservation

We can now state and prove the main theorems of this article: type safety and safety reduction.

**Theorem 4.12 (Type safety)**

If $\Gamma \vdash M \triangleright P$ ($\Gamma$ consistent) then $M \triangleright P \rightarrow^{\text{err}}$.
20 Uniqueness typing for message-passing concurrency

PROOF. By induction on $P \xrightarrow{\text{ext}}$ followed by inversion on $\Gamma \vdash M \triangleright P$. We do not show the entire proof here as it is reasonably straightforward. The case for $\text{eATY}$ relies on Lemma 4.1(1), the case for $\text{eSTR}$ relies on Lemma 4.5, and the case for contexts relies on Lemma 4.11.

THEOREM 4.13 (Subject Reduction)
If $\Gamma \vdash M \triangleright P$ ($\Gamma$ consistent) and $M \triangleright P \rightarrow M' \triangleright P'$ then there exists an $\Gamma'$ such that $\Gamma \leq \Gamma'$ and $\Gamma' \vdash M' \triangleright P'$.

PROOF. By induction on $M \triangleright P \rightarrow M' \triangleright P'$. We show only a number of the cases; of the omitted cases, $\text{rRec}$ relies on the process substitution lemma, $\text{rThen}$ and $\text{rElse}$ are trivial, $\text{rStr}$ follows from the induction hypothesis and preservation of types under structural equivalence (Lemma 4.5), and the case for contexts finally is dealt with by the subterm typing lemma (Section 4.2).

(1) Case $M(c) = \top$

$M \triangleright c?d. P \parallel c?\bar{x}. Q \rightarrow M \triangleright P \parallel Q[d/\bar{x}]$

By inversion on the typing relation we get $\text{dom}(\Gamma) \subseteq \text{alloc}(M)$ and

$\begin{align*}
\Gamma_1, c : [\bar{T}]^{\mu_1 - 1} \vdash P \\
\Gamma_p \leq \Gamma_1, c : [\bar{T}]^{\mu_1 - 1}, d : T \\
\Gamma_q \leq \Gamma_2, c : [\bar{T}]^{\mu_2 - 1}, x : ?T \vdash Q
\end{align*}$

$\Gamma \vdash c?d. P \parallel c?\bar{x}. Q$

where Lemma 4.1(1) allows us to conclude that the two assumptions $c : [\bar{T}]^{\mu_1}$ and $c : [\bar{T}]^{\mu_2}$, used to type the left and right process, respectively, must refer to the same object type $\bar{T}_1 = \bar{T} = \bar{T}_2$. By the identifier substitution lemma (Lemma 4.4), we have that

$\Gamma_2, c : [\bar{T}]^{\mu_2 - 1}, d : T \vdash Q[d/\bar{x}]$

Pick $\Gamma' = \Gamma$. Clearly $\text{dom}(\Gamma') \subseteq \text{alloc}(M)$. We construct the required type derivation as follows:

$\begin{align*}
\Gamma_1, c : [\bar{T}]^{\mu_1 - 1} \vdash P \\
\Gamma_2, c : [\bar{T}]^{\mu_2 - 1}, d : T \vdash Q[d/\bar{x}]
\end{align*}$

$\Gamma \vdash P \parallel Q[d/\bar{x}]$

where the side condition $(\ast)$ is

$\Gamma \leq (\Gamma_1, c : [\bar{T}]^{\mu_1 - 1}), (\Gamma_2, c : [\bar{T}]^{\mu_2 - 1}, d : T)$

and follows from Lemma 4.1(7), since environment are unordered.

(2) Case $M \triangleright \text{allocx} \cdot P \rightarrow M \triangleright \text{newc} : \exists x. P[c/x]$

By inversion we get $\text{dom}(\Gamma) \subseteq \text{alloc}(M)$ and

$\begin{align*}
\Gamma_1, x : [\bar{T}]^{\star} \vdash P \\
\Gamma \vdash \text{allocx} \cdot P
\end{align*}$

We have $c \notin \text{dom}(M) \supseteq \text{alloc}(M) \supseteq \text{dom}(\Gamma) \supseteq \text{dom}(\Gamma_1)$ and hence $c \notin \text{fn}(P)$. 

Uniqueness typing for message-passing concurrency

Choose $\Gamma' = \Gamma$. Clearly $\Gamma \leq \Gamma'$ and $\text{dom}(\Gamma') \supseteq \text{alloc}(M)$. The proof is completed by

\[
\frac{\Gamma_1, c : [T]^* \vdash P[c/x] \quad \Gamma_1', c : [T]^* \leq \Gamma_1, c : [T]^*}{\Gamma', c : [T]^* \vdash P[c/x]} \quad \text{TSTR}
\]

\[
\frac{\Gamma', c : [T]^* \vdash P[c/x]}{\Gamma' \vdash \text{new} : T; P[c/x]} \quad \text{tRST1}
\]

where the premise is established by the identifier substitution lemma (Lemma 4.4).

(3) Case $\mathcal{RFree}$

\[ M, c : T \triangleright \text{free} : c.P \rightarrow M, c : \bot \triangleright P \]

By inversion we get $\text{dom}(\Gamma) \subseteq \text{alloc}(M, c : \top)$ and

\[
\frac{\Gamma_1 \vdash P \quad \Gamma \leq \Gamma_1', u : [T]^*}{\Gamma' \vdash \text{free}u.P} \quad \text{tStr}
\]

By Lemma 4.1(3) there exist an $\Gamma_1$ such that $\Gamma = \Gamma_1, c : [T]^*$ and $\Gamma_1 \leq \Gamma_1'$. Choose $\Gamma' = \Gamma_1$. Clearly $\Gamma \leq \Gamma'$. Moreover $\text{dom}(\Gamma') \subseteq \text{alloc}(M, c : \bot)$ since $c \notin \text{dom}(\Gamma')$ by Lemma 4.1(2). The proof is completed by

\[
\frac{\Gamma_1 \vdash P \quad \Gamma \leq \Gamma_1'}{\Gamma' \vdash P} \quad \text{TSTR}
\]

5 Related work

The literature on using substructural logics to support destructive or strong updates is huge and we can give but a brief overview here. More in-depth discussions can be found in [9, 24].

Resources and pi-calculus: resource usage in a pi-calculus extension is studied in [26] but it differs from our work in many respects. For a start, their scoping construct assumes an allocation semantics while we tease scoping and allocation apart as separate constructs. The resource reclamation construct in [26] is at a higher level of abstraction than $\text{free} : c.P$, and acts more like a ‘resource finalizer’ leading to garbage collection. Resource reclamation is implicit in [26], permitting different garbage collection policies for the same program, whereas in the resource pi-calculus resource reclamation is explicit and fixed for every program. The main difference, however, concerns the aim of the type systems: our type system ensures safe channel deallocation and reuse; the type system in [26] statically determines an upper bound for the number of resources used by a process and does not use substructural typing.

Linearity versus Uniqueness: in the absence of subtyping, affine typing and uniqueness typing coincide but when subtyping is introduced they can be considered dual [15]. For linear typing, the subtyping relation allows coercing non-linear assumptions into a linear assumptions, i.e. $!U \rightarrow U$, but for uniqueness typing, the subtyping relation permits coercing unique assumptions into non-unique assumptions. Correspondingly, the interpretation is different: linearity (respectively affinity) is a local obligation that a channel must be used exactly (respectively at most) once, while uniqueness is a global guarantee that no other processes have access to the channel.
Uniqueness typing for message-passing concurrency

Combining both subtyping relations as we have done in this article appears to be novel. The usefulness of the subtyping relation for affine or linear typing is well-known (e.g. see [19]); subtyping for unique assumptions allows to ‘forget’ the uniqueness guarantee; CLIENT4 above shows one scenario where this might be useful.

Linearity in functional programming: in pure functional programming languages, data structures are always persistent and destructive updates are not supported: mapping a function \( f \) across a list \([x_1, \ldots, x_n]\) yields a new list \([f x_1, \ldots, f x_n]\), leaving the old list intact. However, destructive updates cannot always be avoided (e.g. when modelling system I/O [1]) and are sometimes required for efficiency (e.g. updating arrays). Substructural type systems can be used to support destructive update without losing referential transparency: destructive updates are only allowed on terms that are not shared. Both uniqueness typing [5] and linear typing have been used for this purpose, although even some proponents of linear typing agree that the match is not perfect [29, Section 3].

In functional languages with side effects, substructural type systems have been used to support strong (type changing) updates. For instance, Ahmed et al. have applied a linear type system to support ‘strong’ (type changing) updates to ML-style references [2] in a setting with no subtyping. It has been recognized early on that it is useful to allow the uniqueness of an object to be temporarily violated. In functional languages, this typically takes the form of a sequential construct that allows a unique object (such as an array) to be regarded as non-unique to allow multiple non-destructive accesses (such as multiple reads) after which the uniqueness is recovered again. Wadler’s \texttt{let!} construct [28] (or the equivalent Clean construct \#!) and observer types [23] both fall into this category, and this approach has also been adopted by some non-functional languages where it is sometimes called borrowing [8]. It is, however, non-trivial to extend this approach to a concurrent setting with a partial order over execution steps; our approach can be regarded as one preliminary attempt to do so.

Strong update in the presence of sharing: there is substantial research on type systems that allow strong update even in the presence of sharing; the work on alias types and Vault [11, 25, 31] and on CQual [12] are notable examples of this. These type systems do explicit alias analysis by reflecting memory locations at the type level through singleton types. This makes it possible to track within the type system that a strong (type changing) update to one variable changes the type of all its aliases. The interpretation of unique (or linear) in these systems is different: a unique reference (typically called a capability in this context) does not mean that there is only a single reference to the object, but rather that all its aliases are known. For non-unique reference not all aliases are known and so strong update is disallowed. These systems are developed for imperative languages. They are less useful for functional languages because they cannot guarantee referential transparency, and they appear to be even less useful for concurrent languages: even if we track the effect of a strong update on a shared object on all its aliases, this is only useful if we know \texttt{when} the update happens. In an inherently non-deterministic language such as the pi-calculus this is usually hard to know before execution.

Linearity in the pi-calculus: linear types for the pi-calculus were introduced by Kobayashi et al. [21] but do not employ any subtyping. Moreover, their system cannot be used as a basis for strong update or channel deallocation; although they split a bidirectional linear (‘unique’) channel into a linear input channel and a linear output channel (cf. Definition 2.3.1 for the type combination operator \((+\) \)) these parts are never ‘collected’ or ‘counted’. The more refined type splitting operation we use in this article, combined with the type decrement operation (which has no equivalent in their system) is key to make uniqueness useful for strong updates and deallocation. Our system can easily be extended to incorporate modalities but it does not rely on them; in our case, channel modalities are an orthogonal issue.
Fractional permissions and permission accounting: Boyland [7] was one of the first to consider splitting permissions into \textit{fractional} permissions, which allow linearity or uniqueness to be temporarily violated. Thus, strong update is possible only with a full permission, whereas only passive access is permitted with a ‘fraction’ of a permission. When all the fractions have been reassembled into one whole permission, strong update is once again possible.

Boyland’s suggestion has been taken up by Bornat \textit{et al.} [6], who introduce both fractional permissions and ‘counting’ permissions to separation logic. Despite of the fact that their model of concurrency is shared-memory, their mechanism of permission splitting and counting is surprisingly similar to our treatment of unique assumptions. However, while their resource reading of semaphores targets \textit{implicit} ownership-transfer, uniqueness typing allow us to reason about explicit ownership-transfer. Moreover, subtyping from unique to unrestricted types provides the flexibility of not counting assumptions whenever this is not required, simplifying reasoning for resources that are not deallocated or strongly updated.

Session types: session types [16] and types with CCS-like usage annotations [19] are used to describe channels which send objects of different types. However, these types give detailed prescriptions on how channels can be used, constraining them to follow specific protocols determined at compile-time; this makes modular typing difficult. For example, the \textit{heap} channel used by \texttt{CLIENT}_3 cannot be given a type without knowing all the processes that use the heap. Moreover, most session type systems describe the interaction between \textit{two} parties, and cannot describe the interaction of a client with two time servers, as in example \texttt{CLIENT}_1: multi-party session types [17] lift this restriction but at the expense of a significant increase of the complexity of the type system.

6 Conclusions and future work

We have extended ideas from process calculi, substructural logics and permission counting to define a type system for the pi-calculus extended with primitives for channel allocation and deallocation, where strong update and channel deallocation is safe for unique channels.

The purpose of our type system is not to ensure that every resource that is allocated will also be deallocated (i.e. the absence of memory leaks). This is difficult to track in a type system. For instance, consider

\[
\text{allocx}(c!d_1,d_1!\text{nil} || c!d_2,\text{nil} || c?y??.\text{freex})
\]

Statically, it is hard to determine whether the third parallel process will eventually execute the \texttt{free}x operation. This is due to the fact that it can non-deterministically react with either the first or second parallel process and, should it react with the second process, it will block at \texttt{d}_2?..\texttt{free}x. In order to reject this process as ill-typed, the type-system needs to detect potential deadlocks. This can be done [20], but requires a type system that is considerably more complicated than ours. We leave the responsibility to deallocate to the user, but guarantee that resources once deallocated will no longer be used.

One can envision various extensions of and variations on the type system we have presented. One obvious generalization is to consider a pair of attributes \langle u, a \rangle indicating when the channel will be
Uniqueness typing for message-passing concurrency

unique \( u \) and how often the channel can be used \( a \). This leads to an infinite space of attributes of the form:

\[
\langle \omega, \omega \rangle \ldots \langle 2, \omega \rangle \langle 1, \omega \rangle \langle 0, \omega \rangle \\
\vdots \quad \vdots \quad \vdots \\
\langle \omega, 2 \rangle \langle \omega, 1 \rangle \langle \omega, 0 \rangle \\
\langle 0, 0 \rangle \langle 0, 1 \rangle \langle 0, 2 \rangle \\
\cdots \cdots \cdots \cdots \cdots \cdots \\
\cdots \cdots \cdots \cdots \cdots \cdots \\
\cdots \cdots \cdots \cdots \cdots \cdots
\]

The highlighted pairs are the pairs that we represent in our type system: \( \langle \omega, \omega \rangle \) we denote simply as \( \omega \) (unrestricted), \( \langle u, \omega \rangle \) we represent as \( (\bullet, u) \) or just \( \bullet \) if \( u = 0 \) (unique), and \( \langle \omega, 1 \rangle \) we represent as \( 1 \) (affine). We could have taken \( \omega \) to be syntactic shorthand for \( (\bullet, \omega) \), which too would correspond to \( \langle \omega, \omega \rangle \) in this matrix; although this would technically simplify the type language, we felt it would be confusing to describe unrestricted channels as a special kind of unique channels (i.e. channels that will be unique after an infinite number of steps).

Of the entries in the matrix that we do not represent, attributes of the form \( \langle \omega, a \rangle \) for other values of \( a \) would certainly be of use and it should be straightforward to extend the type system to support them. However, entries of the form \( \langle u, a \rangle \) with \( u, a < \omega \) are less useful. If \( a < u \), we know the channel will be unique after \( u \) steps but we are not allowed to take \( u \) steps, so we can consider the channel to have attribute \( \langle \omega, a \rangle \) instead. Similarly, if \( u < a \) we will only be able to do a restricted number of actions on the channel after it becomes unique, but (by definition of uniqueness) no other process will benefit from this restriction, so we can consider the channel to have attribute \( \langle u, \omega \rangle \). Hence, the natural space of attributes is one-dimensional:

\[
\langle \omega, 1 \rangle \langle \omega, 2 \rangle \cdots \langle \omega, \omega \rangle \cdots \langle 2, \omega \rangle \langle 1, \omega \rangle \langle 0, \omega \rangle \\
(1 \quad \omega \quad \cdots (\bullet, 2) (\bullet, 1) \bullet )
\]

One useful extension would be that of input/output modalities, which blend easily with the affine/unique duality. Presently, when a server process splits a channel \( c :: [T]^* \) into one channel of type \( [T]^* \) and two channels of type \( [T]^1 \) to be given to two clients, the clients can potentially use this channel to communicate among themselves instead of the server. Modalities are a natural mechanism to preclude this from happening. For instance, we could generalize the splitting rule so that a unique channel \( c :: [T]^* \) can be split into two affine output channels of type \( [T]^1_{\text{out}} \) and an input channel of type \( [T]^1_{\text{in}} \) which will be unique after two inputs.

One might even envision two usage attributes, one for input and one for output, so that we can have channels that can be limited to be used for \( x \) inputs and \( y \) outputs, or dually channels that can are guaranteed to be unique after \( x \) inputs and \( y \) outputs. It is not clear at present whether this additional expressivity would be useful, however; if one wants to be more precise about communication protocols, session types [16] are probably a better choice.

We are currently investigating ways how uniqueness types can be used to refine existing equational theories, so as to be able to equate processes such as \( \text{CLIENT}_1 \) and \( \text{CLIENT}_0 \).
Uniqueness typing for message-passing concurrency

We have not designed a type checker for our type system. Type checking for substructural language without subtyping is well-understood (for instance, see the textbook [30]). Type checking for a substructural type system with subtyping is a less well-charted territory, but here too there is a lot of work, especially for functional languages [4, 5, 14, 18, 27] but also for imperative languages [10] and for the $\pi$-calculus [13].

Funding

Edsko de Vries and Matthew Hennessy are supported by SFI project [SFI 06 IN.1 1898].

References

26 Uniqueness typing for message-passing concurrency


Received October 2010, revised July 2011, accepted February 2012