

Designing Asynchronous Multiparty Protocols with Crash-Stop Failures

Adam D. Barwell¹ Ping Hou² Nobuko Yoshida² Fangyi Zhou^{2,3}

¹University of St Andrews

²University of Oxford

³Imperial College London

Stardust Meeting, Glasgow

18 December 2023



Imperial College
London

Concurrent and Communicating Systems

- » Are **ubiquitous**
 - Mobile phones, desktops, distributed systems, the internet, &c.
- » Can be difficult to get **right** unaided
 - Communications mismatch, deadlocks, livelocks, &c.

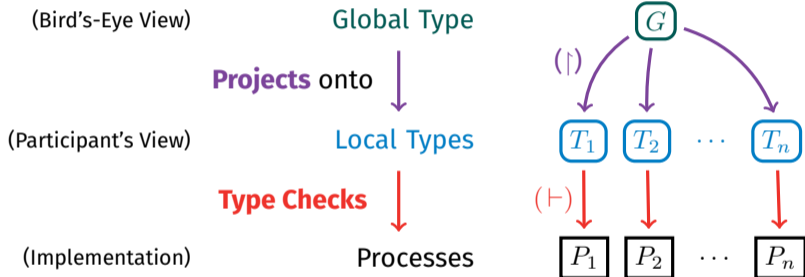
Concurrent and Communicating Systems

- » Are ubiquitous
 - Mobile phones, desktops, distributed systems, the internet, &c.
- » Can be difficult to get right **unaided**
 - Communications mismatch, deadlocks, livelocks, &c.
- » Some assistance: **Session Types**

(Multiparty) Session Types

- » Formally describe **communications behaviour** between two or more systems
- » Communications behaviour is **enforced statically**

Top-Down Multiparty Session Types



Top-Down Multiparty Session Types – Advantages

- » **Correct-by-construction** behavioural properties
 - Communication safety
 - Deadlock-freedom
 - Liveness

- » **Generate** protocol-conforming code
 - Library support in Scala, Rust, Haskell, Erlang, OCaml, &c.

One Caveat: Fault-Tolerance, or Lack Thereof

- » Traditional MPST systems assume a **perfect world**
 - No process failures
 - No message loss, duplication, or corruption



One Caveat: Fault-Tolerance, or Lack Thereof

- » Traditional MPST systems assume a **perfect world**
 - No process failures
 - No message loss, duplication, or corruption
- » Distributed systems can, and will, fail
 - Reality is not so convenient...



One Caveat: Fault-Tolerance, or Lack Thereof

- » Traditional MPST systems assume a **perfect world**
 - No process failures
 - No message loss, duplication, or corruption
- » Distributed systems can, and will, fail
 - Reality is not so convenient...
- » Such MPST systems cannot reason about failures
 - How do participants handle crashes?
 - Do behavioural guarantees still hold in the presence of crashes?



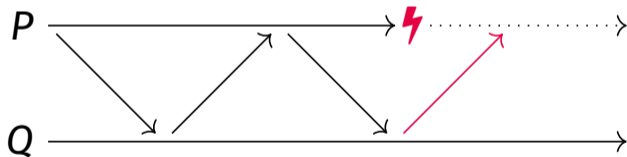
Asynchronous Multiparty Protocols with Crash-Stop Failures

We present an **asynchronous top-down** MPST theory with **crash-stop failures**.

1. Behavioural Guarantees
2. Optional Reliability Assumptions
3. Code generation toolchain, TEATRINO

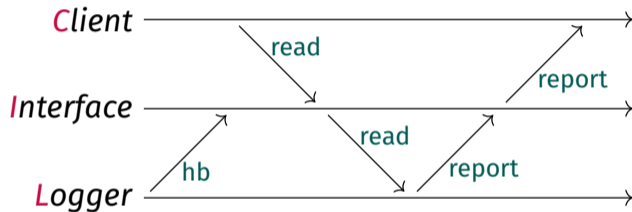
Crash-Stop Failures

- » Processes can **crash arbitrarily**
- » Crashed processes **make no progress** and **do not recover**
- » Communications channels deliver messages in order and without losses



(Part of) A Simple Distributed Logger

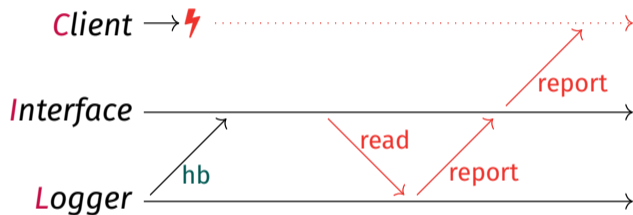
A **C**lient requests the accumulated logs from a distributed **L**ogger via an **I**nterface process.


$$G = L \rightarrow I : hb . C \rightarrow I : read . I \rightarrow L : read . L \rightarrow I : report(log) . I \rightarrow C : report(log) . end$$

The full version of the protocol (and other variants) can be found in the paper/artefact.

(Part of) A Simple Distributed Logger

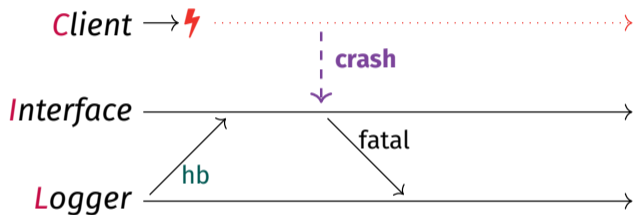
A **C**lient requests the accumulated logs from a distributed **L**ogger via an **I**nterface process.



$G = L \rightarrow I : hb. ??$

(Part of) A Simple Distributed Logger

A **C**lient requests the accumulated logs from a distributed **L**ogger via an **I**nterface process.


$$G = L \rightarrow I : hb . C \rightarrow I : \left\{ \begin{array}{l} \text{read } . I \rightarrow L : \text{read} . L \rightarrow I : \text{report}(\text{log}) . I \rightarrow C : \text{report}(\text{log}) . \text{end} \\ \text{crash } . I \rightarrow L : \text{fatal} . \text{end} \end{array} \right\}$$

↑ Reserved label; represents **crash detection**

Global Type Syntax

| | | | |
|-----------|-------|---|-----------------------|
| B | $::=$ | <code>int</code> <code>bool</code> <code>real</code> <code>unit</code> ... | Basic types |
| G | $::=$ | <code>p</code> → <code>q</code> † : { <code>m_i</code> (<code>B_i</code>). <code>G_i</code> } _{$i \in I$} | Transmission |
| | | <code>p</code> † \rightsquigarrow <code>q</code> : <code>j</code> { <code>m_i</code> (<code>B_i</code>). <code>G_i</code> } _{$i \in I$} ($j \in I$) | Transmission en route |
| | | μ <code>t</code> . <code>G</code> | Recursion |
| | | <code>t</code> | Type variable |
| | | <code>end</code> | Termination |
| \dagger | $::=$ | <code>.</code> <code>⚡</code> | Crash annotation |

- » **Runtime** types are not available to the user, only via reduction.
- » **crash** is a reserved label indicating crash-handling branches.

Crashing Clients

Participant **C** Crashes.

$$G = L \rightarrow I : hb. C \rightarrow I : \left\{ \begin{array}{l} \text{read.I} \rightarrow L : \text{read.L} \rightarrow I : \text{report(log).I} \rightarrow C : \text{report(log).end} \\ \text{crash.I} \rightarrow L : \text{fatal.end} \end{array} \right\}$$

Crashing Clients

Participant **C** Crashes.

$$G = L \rightarrow I : \text{hb}. C \rightarrow I : \left\{ \begin{array}{l} \text{read}. I \rightarrow L : \text{read}. L \rightarrow I : \text{report}(\text{log}). I \rightarrow C : \text{report}(\text{log}). \text{end} \\ \text{crash}. I \rightarrow L : \text{fatal}. \text{end} \end{array} \right\}$$

$\downarrow \mathcal{R}$

$$G_1 = L \rightarrow I : \text{hb}. C^{\downarrow} \rightsquigarrow I : \left\{ \begin{array}{l} \text{read}. I \rightarrow L : \text{read}. L \rightarrow I : \text{report}(\text{log}). I \rightarrow C^{\downarrow} : \text{report}(\text{log}). \text{end} \\ \text{crash}. I \rightarrow L : \text{fatal}. \text{end} \end{array} \right\}$$

Crashing Clients

Participant C Crashes.

$$G = L \rightarrow I : \text{hb}. C \rightarrow I : \left\{ \begin{array}{l} \text{read}. I \rightarrow L : \text{read}. L \rightarrow I : \text{report}(\log). I \rightarrow C : \text{report}(\log). \text{end} \\ \text{crash}. I \rightarrow L : \text{fatal}. \text{end} \end{array} \right\}$$

p is unreliable

p occurs in G

G is anything but a μ -term

$$p \notin \mathcal{R}$$

$$p \in \text{roles}(G)$$

$$G \neq \mu t. G'$$

[GR- ζ]

$\downarrow \mathcal{R}$

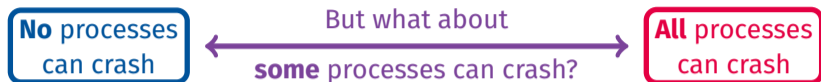
$$\langle \mathcal{C}; G \rangle \xrightarrow{P \zeta \mathcal{R}} \langle \mathcal{C} \cup \{p\}; G \zeta p \rangle$$

↑ Set of crashed roles

↑ 'Remove' p from G

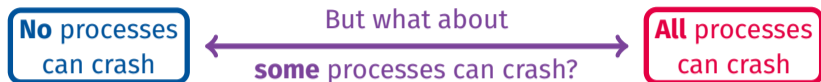
$$G_1 = L \rightarrow I : \text{hb}. C \zeta \rightsquigarrow I : \left\{ \begin{array}{l} \text{read}. I \rightarrow L : \text{read}. L \rightarrow I : \text{report}(\log). I \rightarrow C \zeta : \text{report}(\log). \text{end} \\ \text{crash}. I \rightarrow L : \text{fatal}. \text{end} \end{array} \right\}$$

But Some are More Reliable than Others



- » Some participants may represent services that can be **assumed reliable**

But Some are More Reliable than Others



- » Some participants may represent services that can be **assumed reliable**

Optional Reliability Assumptions

- » Specific participants can be marked as **reliable**
- » Reliable participants **will not crash**
- » Reduction rules, &c. conform to these assumptions

Crashing Clients

Participant **I** Detects

$$G_1 = L \rightarrow I : hb.C^{\downarrow} \rightsquigarrow I : \left\{ \begin{array}{l} \text{read.I} \rightarrow L : \text{read.L} \rightarrow I : \text{report}(\log).I \rightarrow C^{\downarrow} : \text{report}(\log).end \\ \text{crash.I} \rightarrow L : \text{fatal}.end \end{array} \right\}$$

\downarrow^*
 \mathcal{R}

Crashing Clients

Participant **I** Detects

$$G_1 = L \rightarrow I : hb . C^{\downarrow} \rightsquigarrow I : \left\{ \begin{array}{l} \text{read} . I \rightarrow L : \text{read} . L \rightarrow I : \text{report}(\log) . I \rightarrow C^{\downarrow} : \text{report}(\log) . \text{end} \\ \text{crash} . I \rightarrow L : \text{fatal} . \text{end} \end{array} \right\}$$

$\downarrow^*_{\mathcal{R}}$

$$G_4 = C^{\downarrow} \rightsquigarrow I : \left\{ \begin{array}{l} \text{read} . I \rightarrow L : \text{read} . L \rightarrow I : \text{report}(\log) . I \rightarrow C^{\downarrow} : \text{report}(\log) . \text{end} \\ \text{crash} . I \rightarrow L : \text{fatal} . \text{end} \end{array} \right\}$$

Crashing Clients

Participant **I** Detects

$$G_4 = C^{\not{I}} \rightsquigarrow I: \left\{ \begin{array}{l} \text{read.I} \rightarrow \text{L:read.L} \rightarrow \text{I:report(log).I} \rightarrow C^{\not{I}} : \text{report(log).end} \\ \text{crash.I} \rightarrow \text{L:fatal.end} \end{array} \right\}$$

$\downarrow \mathcal{R}$

$$G_5 = I \rightarrow \text{L:fatal.end}$$

Crashing Clients

Participant **I** Detects

$$G_4 = \mathbf{C}^{\not\downarrow} \rightsquigarrow \mathbf{I}: \left\{ \begin{array}{l} \text{read. I} \rightarrow \mathbf{L}: \text{read. L} \rightarrow \mathbf{I}: \text{report}(\text{log}). \mathbf{I} \rightarrow \mathbf{C}^{\not\downarrow}: \text{report}(\text{log}). \text{end} \\ \text{crash. I} \rightarrow \mathbf{L}: \text{fatal. end} \end{array} \right\}$$

$$\downarrow_{\mathcal{R}} \quad \frac{j \in I \quad m_j = \text{crash}}{\langle \mathbf{C}; \mathbf{p}^{\not\downarrow} \rightsquigarrow \mathbf{q}: j \{m_i(B_i). G'_i\}_{i \in I} \rangle \xrightarrow{\mathbf{q} \odot \mathbf{p}}_{\mathcal{R}} \langle \mathbf{C}; G'_j \rangle} \quad [\text{GR-}\odot]$$

$$G_5 = \mathbf{I} \rightarrow \mathbf{L}: \text{fatal. end}$$

Local Type Syntax

| | | | |
|--------|-------|--|-----------------|
| B | $::=$ | <code>int</code> <code>bool</code> <code>real</code> <code>unit</code> ... | Basic types |
| S, T | $::=$ | <code>p</code> { $m_i(B_i).T_i$ }_{ $i \in I$ } | External choice |
| | | <code>p</code> \oplus { $m_i(B_i).T_i$ }_{ $i \in I$ } | Internal choice |
| | | $\mu \mathbf{t}.T$ | Recursion |
| | | <code>t</code> | Type variable |
| | | <code>end</code> | Termination |
| | | <code>stop</code> | Crash |

- » The `runtime stop` type is the type of crashed participants
- » `crash` is a reserved label indicating crash-handling branches

Type Checking Distributed Logging with Local Types

$$T_C = I \oplus \text{read}. I \& \text{report}(\text{log}). \text{end}$$

$$P_C = I ! \text{read}. I ? \text{report}(x). \mathbf{0}$$

$$T_I = L \& \text{hb}. C \& \left\{ \begin{array}{l} \text{read}. L \oplus \text{read}. L \& \text{report}(\text{log}). C \oplus \text{report}(\text{log}). \text{end} \\ \text{crash}. L \oplus \text{fatal}. \text{end} \end{array} \right\}$$

$$P_I = L ? \text{hb}. \sum \left\{ \begin{array}{l} C ? \text{read}. L ! \text{read}. L ? \text{report}(x). C ! \text{report}(x). \mathbf{0} \\ C ? \text{crash}. L ! \text{fatal}. \mathbf{0} \end{array} \right\}$$

We largely elide the session π -calculus here, but full details can be found in the paper.

Local Type Context Reduction

$$\begin{array}{l} T_{\mathbf{c}} = \mathbf{I} \oplus \text{read}.\mathbf{I} \& \text{report}(\text{log}).\text{end} \\ P_{\mathbf{c}} = \mathbf{I} ! \text{read}.\mathbf{I} ? \text{report}(x).\mathbf{0} \end{array} \quad \rightarrow \quad \begin{array}{l} T_{\mathbf{c}} = \text{stop} \\ P_{\mathbf{c}} = \downarrow \end{array}$$

- » Rules are largely standard
 - Crashing, crash-detection, &c. rules are novel
- » Queues handle asynchronous message passing
 - Queues are made unavailable when the corresponding participant has crashed

Projecting Global Types into Local Types

Relates G and T for a given p .

$$(q \rightarrow r^\dagger : \{m_i(B_i).G_i\}_{i \in I}) \Vdash_{\mathcal{R}} p = \begin{cases} r \oplus \{m_i(B_i).(G_i \Vdash_{\mathcal{R}} p)\}_{i \in \{j \in I \mid m_j \neq \text{crash}\}} & \text{if } p = q \\ q \& \{m_i(B_i).(G_i \Vdash_{\mathcal{R}} p)\}_{i \in I} & \text{if } p = r, \text{ and } q \notin \mathcal{R} \text{ implies} \\ & \exists k \in I : m_k = \text{crash} \\ \prod_{i \in I} G_i \Vdash_{\mathcal{R}} p & \text{if } p \neq q, \text{ and } p \neq r \end{cases}$$

$$(q^\dagger \rightsquigarrow r : j \{m_i(B_i).G_i\}_{i \in I}) \Vdash_{\mathcal{R}} p = \begin{cases} G_j \Vdash_{\mathcal{R}} p & \text{if } p = q \\ q \& \{m_i(B_i).(G_i \Vdash_{\mathcal{R}} p)\}_{i \in I} & \text{if } p = r, \text{ and } q \notin \mathcal{R} \text{ implies} \\ & \exists k \in I : m_k = \text{crash} \\ \prod_{i \in I} G_i \Vdash_{\mathcal{R}} p & \text{if } p \neq q, \text{ and } p \neq r \end{cases}$$

$$(\mu \mathbf{t}.G) \Vdash_{\mathcal{R}} p = \begin{cases} \mu \mathbf{t}.(G \Vdash_{\mathcal{R}} p) & \text{if } p \in G \text{ or } \text{fv}(\mu \mathbf{t}.G) \neq \emptyset \\ \text{end} & \text{otherwise} \end{cases} \quad \begin{array}{l} \mathbf{t} \Vdash_{\mathcal{R}} p = \mathbf{t} \\ \text{end} \Vdash_{\mathcal{R}} p = \text{end} \end{array}$$

The (full) merging operator definition can be found in the paper.

Projection Begat Properties

1. Association of Global Type and Configuration (Soundness and Completeness)
 - Global and local types do not diverge when reducing

Formal definitions and proofs can be found in the paper.

Projection Begat Properties

1. Association of Global Type and Configuration (Soundness and Completeness)
 - Global and local types do not diverge when reducing
2. Configuration Safety
 - There are no label mismatches
 - Each receiver must be able to handle the potential crash of the (unreliable) sender

Formal definitions and proofs can be found in the paper.

Projection Begat Properties

1. Association of Global Type and Configuration (Soundness and Completeness)
 - Global and local types do not diverge when reducing
2. Configuration Safety
 - There are no label mismatches
 - Each receiver must be able to handle the potential crash of the (unreliable) sender
3. Deadlock-Freedom (Progress)
 - Local types are able to reduce until they terminate or crash

Projection Begat Properties

1. Association of Global Type and Configuration (Soundness and Completeness)
 - Global and local types do not diverge when reducing
2. Configuration Safety
 - There are no label mismatches
 - Each receiver must be able to handle the potential crash of the (unreliable) sender
3. Deadlock-Freedom (Progress)
 - Local types are able to reduce until they terminate or crash
4. Liveness
 - Every pending internal/external choice is eventually triggered (by message transmission or crash detection)

Asynchronous Multiparty Protocols with Crash-Stop Failures

We present an **asynchronous top-down** MPST theory with **crash-stop failures**.

- ✓ Behavioural Guarantees
- ✓ Optional Reliability Assumptions
- ⇒ Code generation toolchain, TEATRINO

TEATRINO



» SCRIBBLE-inspired **code generation toolchain**

- Consumes SCRIBBLE protocols
- Produces protocol-conforming SCALA code using the EFFPI concurrency library
- Generated code is executable

» **Implements** our (non-runtime) Global and Local Types, and Projection

» **Extends** both the EFFPI and SCRIBBLE syntaxes with crash-stop failures

The (Extended) SCRIBBLE Protocol Description Language

- » We support (a less sugary subset of) the version accepted by ν SCR¹
 - No support for `do`-notation – recursion is expressed via `rec` and `continue`
 - No support for auxiliary protocol definitions

Extensions

1. `reliable` role declarations
2. Reserved `crash` label

¹<https://nuscr.dev/>

The (Extended) SCRIBBLE Protocol Description Language

$$G = L \rightarrow I : hb . C \rightarrow I : \left\{ \begin{array}{l} \text{read} . I \rightarrow L : \text{read} . L \rightarrow I : \text{report}(\text{log}) . I \rightarrow C : \text{report}(\text{log}) . \text{end} \\ \text{crash} . I \rightarrow L : \text{fatal} . \text{end} \end{array} \right\}$$

```
global protocol G(reliable role L, role C, reliable role I) {
  hb from L to I;
  choice at C {
    read from C to I;
    read from I to L;
    report(Log) from L to I;
    report(Log) from I to C;
  } or {
    crash from C to I;
    fatal from I to L;
  }
}
```

The (Extended) EFFPI Concurrency Library

- » Embedded Domain Specific Language for SCALA 3
- » Leverages type features to **represent local types directly in code**
 - Union types, match types, and dependent and polymorphic function types

Extensions

1. Support for crash-handling branches
 - New type-level receive construct: `InErr`
 - New value-level receive construct: `receiveErr`
2. Support for crash detection
 - Implemented using timeouts

The (Extended) EFFPI Concurrency Library

$$T_I = L \& \text{hb}. C \& \left\{ \begin{array}{l} \text{read}.L \oplus \text{read}.L \& \text{report}(\text{log}).C \oplus \text{report}(\text{log}).\text{end} \\ \text{crash}.L \oplus \text{fatal}.\text{end} \end{array} \right\}$$

```
type I[C0 <: InChan[Hb],
      C1 <: OutChan[Fatal],
      C2 <: InChan[Read],
      C3 <: InChan[Report],
      C4 <: OutChan[Report]] =
  In[C0, Hb, (X <: Hb) =>
    InErr[C2, Read,
      (Y <: Read) =>
        Out[C3, Read] >>: In[C4, Report, (Z <: Log) => Out[C5, Report]],
      (Err <: Throwable) => Out[C2, Fatal]
  ]]
```

Evaluating TEATRINO

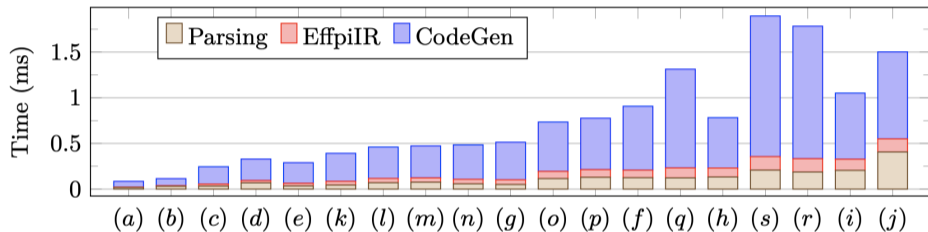
Expressiveness

- » Applied to examples from session type and distributed systems literature
- » Standard examples extended with crash-handling behaviour
- » Two patterns: **graceful failure** and **failover**

Feasibility

- » We give generation times for all of our examples
- » We report times for the three main generation phases:
 1. Parsing
 2. EffpiIR Generation
 3. Code Generation

Evaluating TEATRINO



- » Evaluated on 19 protocols taken from session type and distributed system literature
 - Code generation times all under 3 milliseconds

Summary

- » Asynchronous top-down MPST theory with crash-stop failures:
 - Support for fully-reliable to fully-unreliable protocols
 - Safety, deadlock-freedom, and liveness guarantees
- » TEATRINO: toolchain support for generating protocol-conformant SCALA code
- » Future work:
 - Investigate different crash and failure models (e.g. crash-recover, link failures)

Links

- » Full version: <https://arxiv.org/abs/2305.06238>
- » Artefact: <https://doi.org/10.5281/zenodo.7714132>
- » Artefact Source: <https://github.com/adbarwell/ECOOP23-Artefact>