Programming and Proving with Distributed Protocols

Ilya Sergey



James R. Wilcox Zachary Tatlock





- Liveness: $\Box(Q(\text{state of } n_1) \Rightarrow \Diamond(R(\text{state of } n_2)))$
- Refinement: P ("implementation") \leq P' ("specification")



• Invariants: $Inv(s_0) \land \forall s s', Inv(s) \land transfer(P, s, s') \Rightarrow Inv(s')$

Verified Distributed Systems

- Invariants
- Liveness
- Refinement















- Asynchronous updates
- Message reordering
- Packet loss
- Node crashes
- Network partitions
- Reconfiguration
- Byzantine faults



- Invariants
- Liveness
- Refinement
- Composition
 (*aka* Reusability)

Composition in Distributed Systems

Modular program verification

Horizontal System Decomposition

Inter-Protocol Dependencies







Composition: A way to make proofs harder (Lamport, 1997)

When distracting *language features* are removed and the underlying *mathematics* is revealed, compositional reasoning is seen to be of little use.





"language features"

$\left\{ P\right\} c \left\{ Q\right\}$ precondition postcondition

If the initial state satisfies P, then, after cterminates, the final state satisfies Q.

Working Example: Cloud Compute System

Cloud Compute



Cloud Compute: Server





Cloud Compute: Server

while True: (from, n) < - recv



send (n, factors(n)) to from

Cloud Compute: Server





Cloud Compute: Client



Cloud Compute: Client

- send 21 to server
- (_, ans) <- recv from server</pre>
- **assert** ans $== \{3, 7\}$

ver **cv from** server {3, 7}

Protocols



Protocols



Protocols

State: abstract state of each node + all ever sent messages Transitions: allowed sends and receives





Cloud Compute State



DISTRIBUTED STATE OF THE PROTOCOL CC

Send-Transitions

$ au_{S}$	Requires (m, to)
	$n \in \overline{C} \land to \in \overline{S} \land$
sreq	$n \mapsto rs \land m = (Re)$
	$args \in dom(f)$
	$n \in \overline{S} \land f(args) =$
sresp	$n \mapsto (to, args) \uplus rs$
	m = (Resp, v, args)



Send-Transitions



sreq ((Req, args1), s)



	Ensures
$(Req, args) \land$	$n\mapsto (to, args) \uplus rs$
$v = v \land rs \land rs \land rs$	$n \mapsto rs$



Send-Transitions







	Ensures
\land (Req, <i>args</i>) \land	$n\mapsto (to, args) \uplus rs$
$v = v \land v \land rs \land vs)$	$n \mapsto rs$

[(**Req**, args₁), **from**: *c*₁, **to**: *s*]



*C*₂

Receive-Transitions

$ au_r$	Requires $(m, from)$
rreq	$n \in \overline{S} \And n \mapsto rs \And$
	m = (Req, args)
	$n \in \overline{C}$
rresp	$n \mapsto (from, args) \cup$
	m = (Resp, ans, an)



Receive-Transitions

$ au_r$	Requires $(m, from)$		Ensures
rreq	$n \in \overline{S} \And n \mapsto rs \And m = (Req, args)$		$n \mapsto (from, args) \cup rs$
rresp	$n \in \overline{C} \qquad \& \\ n \mapsto (from, args) \cup rs \& \\ m = (Resp, ans, args) \end{cases}$	&z &z	$n \mapsto rs$

[(**Req**, args₁), **from**: *c*₁, **to**: *s*]

{ (s, args1) }



Receive-Transitions

$ au_r$	Requires (m, from)	Ensures
rreq	$n \in \overline{S} \And n \mapsto rs \And m = (Req, args)$	$n \mapsto (from, args) \cup rs$
rresp	$n \in \overline{C} \qquad \& x \\ n \mapsto (from, args) \cup rs \& x \\ m = (Resp, ans, args)$	$n \mapsto rs$

[(**Req**, args₁), **from**: *c*₁, **to**: *s*]

{ (s, args1) }



{ (c₁, args₁) }

Composition in Distributed Systems

Modular Program Verification

Horizontal System Decomposition

• Inter-Protocol Dependencies



From Protocols to Hoare Specs

$\left\{ P \right\} c \left\{ Q \right\}$

From Protocols to Hoare Specs

 $tr \in \mathcal{P} \to Pre_{tr}$

letrec server loop = let ans = factor(args) in server loop() in server loop()

Send-transitions

$$\tau_s$$
Requires (m, to) $n \in \overline{C} \land to \in \overline{S} \land$ $sreq$ $n \in \overline{C} \land to \in \overline{S} \land$ $n \mapsto rs \land m = (\mathbb{R})$ $args \in \operatorname{dom}(f)(\mathbb{C})$ $n \in \overline{S} \land f(args) =$ $sresp$ $n \mapsto (to, args) \uplus rs$ $m = (\operatorname{Resp}, v, args)$

Receive-transitions

$ au_r$	Requires $(m, from)$	Ensures
rreq	$n \in \overline{S} \And n \mapsto rs \And m = (\operatorname{Req}, \operatorname{args}) \text{(b)}$	$n\mapsto (f\!rom,args) \uplus rs$
rresp	$n \in \overline{C} \And$ $n \mapsto (from, args) \uplus rs \And$ $m = (Resp, ans, args)$	$n\mapsto rs$

	Ensures		
eq, $args$) \land	$n \mapsto (to, args) \uplus rs$		
$= v \land$	$n \mapsto rs$		

$Inv_1(s) \triangleq \forall m \in s.MS, m = \langle from, to, -, (Req, args) \rangle$ $\Rightarrow args \in dom(factor)$

Inductive Invariant Inv₁

A rule for Invariant Strengthening

 $Inv_1(s) \triangleq \forall m \in s.MS, m = \langle from, to, -, (Req, args) \rangle$ $\Rightarrow args \in dom(factor)$

$\Gamma; \langle \ell \mapsto \mathcal{P}_{\ell} \uplus W, H \rangle \stackrel{n}{\vdash} c: \{P\}\{Q\} \qquad I \text{ is inductive } wrt. \ \mathcal{P}_{\ell} \qquad \mathcal{I} \triangleq \forall s, \text{this } s \Rightarrow I(s)$ $\Gamma; \langle \ell \mapsto \mathsf{WithInv}(\mathcal{P}_{\ell}, I) \uplus W, H \rangle \stackrel{n}{\vdash} c : \{ P \land \mathcal{I} \} \{ Q \land \mathcal{I} \}$ A "protocol combinator"

 $Inv_1(s) \triangleq \forall m \in s.MS, m = \langle from, to, -, (Req, args) \rangle$ $\Rightarrow args \in dom(factor)$

WithInv(CC, Inv₁) $CCI_1 \vdash letrec$ server loop = (from, args) ← blocking receive(); let ans = factor(args) in send_{sresp} ((Resp, ans, args), from); server loop() in server loop(): {this is a server \land Brs, this \Rightarrow rs} { False }

More Implementations for Cheap

A Batching Server

letrec receive_batch $(k : nat) \triangleq$ if k = k' + 1then fargs \leftarrow receive_req ();
 rest \leftarrow receive_batch k';
 return fargs :: rest
else return []

CCl₁ ⊢ batch_server(5) :
{ this is a server ∧ ∃rs, this → rs }
{ False }

letrec send_batch $(rs : [(Node, [nat])]) \triangleq$ if rs = (from, args) :: rs'then let v = f(args) in send[sresp, ℓ]((Resp, v, args), from); send_batch rs'else return ()

A Memoising Server

letrec memo_server $(mmap : map) \triangleq$ $(from, args) \leftarrow \text{receive}_{req}();$ let ans = lookup mmap inif $ans \neq \perp then send[sresp, \ell]((Resp, ans, args), from);$ memo_server mmap

else let ans = f(args) in

 $send[sresp, \ell](m, (Resp, ans, args));$ let mmap' = update(mmap, args, ans) in memo_server mmap'

 $CC_1 \vdash memo server({})$: { this is a server ∧ ∃rs, this → rs } { False }

A Client Implementation

- $CCI_1 \vdash fun$ compute factor (arg, serv) = send_{sreq} ((Req, args), serv); r ← receive resp();
 - return r
 - { **serv** is a server ∧ $arg \in dom(factor) \land$
 - this $\rightarrow \emptyset$

$\{ res = factor(arg) \land this \rightarrow \emptyset \}$

Cannot conclude res = factor(args).

$CCI_1 \vdash receive resp():$ { this \mapsto {(serv, arg) } this $\mapsto \emptyset$

Send-transitions

$$\tau_s$$
Requires (m, to) $n \in \overline{C} \land to \in \overline{S} \land$ $sreq$ $n \mapsto rs \land m = (\mathbb{R}$ $args \in \mathsf{dom}(f)$ $n \in \overline{S} \land f(args) =$ $sresp$ $n \mapsto (to, args) \uplus rs$ $m = (\mathbb{Resp}, v, args)$

Receive-transitions

$ au_r$	Requires $(m, from)$	Ensures
rreq	$n \in \overline{S} \And n \mapsto rs \And$ $m = (Req, args)$	$n \mapsto (from, args) \uplus rs$
rresp	$n \in \overline{C} \& x$ $n \mapsto (from, args) \uplus rs \& x$ $m = (Resp, ans, args) (C)$	$n\mapsto rs$

	Ensures		
$eq, args) \land$	$n \mapsto (to, args) \uplus rs$		
$v \wedge (b)$ (a)	$n \mapsto rs$		

$Inv_2(s) \triangleq \forall m \in s.MS, m = <-, -, -, (Resp, ans, args) >$ \Rightarrow factor(args) = ans)

Inductive Invariant Inv2

With $Inv(CCI_1, Inv_2)$ $CCI_2 \vdash fun compute_fa$ send_{sreq} ((Req $r \leftarrow receive$ return r : { serv is a server ∧ $arg \in dom(factor) \land$ this $\rightarrow \emptyset$

 $Inv_2(s) \triangleq \forall m \in s.MS, m = \langle -, -, -, (Resp, ans, args) \rangle$ \Rightarrow **factor**(*args*) = *ans*)

{res = factor(arg) \land this $\Rightarrow \emptyset$ }

Composition in Distributed Systems

Modular Program Verification

Inter-Protocol Dependencies

Horizontal System Decomposition

- $\{s | s = S_1\}$
- send(msg);
 doStuff();

 T1
 - $\{s | s = S_2\}$
- m <- receive(c);
 doMoreStuff();
 </pre>
 T2
 - $\{s | s = S_3\}$

 $\{s | s = S_1\}$ send(msg);
doStuff();

T1 $\{s \mid s = S_2\}$ m <- receive(c);
doMoreStuff();
</pre>
T2 $\{s \mid s = S_3\}$ doOtherStuff(); $\{s \mid s = S_3\}$

"frame"

 $\{s \mid s = S_1 \oplus T_6\}$ send(msg);
doStuff();
T1 $\{ \mathbf{S} \mid \mathbf{S} = \mathbf{S}_2 \oplus \mathbf{T}_6 \}$ m <- receive(c);
doMoreStuff();
</pre>
T2 $\{s \mid s = S_3 \oplus T_6\}$ doOtherStuff(); $T_3 \\ \{s \mid s = S_3 \oplus T_7\}$

A Delegating Server

delegating server (n') in delegating server(server)

letrec delegating server (n': Node) = (from, args) ← blocking receive(); let ans = [compute factor(args)] in send_{sresp} ((Resp, ans, args), from);

Composition in Distributed Systems

Modular Program Verification

Inter-Protocol Dependencies

 $r \leftarrow compute_factor(n);$

${r = factor(n)}$

r ← compute_factor(n); ps ← query_server(s)

r ← compute_factor(n);
ps ← query_server(s)
{r = factor(n) ∧ this ∉ ps}

${r = factor(n) \land this \notin ps}$

ps ← query_server(s)

 $r \leftarrow compute_factor(n);$

{True}

CloudComp + Inv (CCI)

Logical Hooks

 $G[CCI, QS, tr](m, S_{CCI}, S_{QS}) \triangleq$ $tr \in QS \land$ $m = perms(s_{cci})$

tr is send-response-to-enquiry ∧

 $r \leftarrow compute_factor(n);$ ps ← query_server(s) ${r = factor(n) \land this \notin ps}$

6[CCI, QS, tr]

$r \leftarrow compute_factor(n);$

 \bigwedge

{True} ps ← query_server(s) $\{\texttt{this} \notin \texttt{PS}\}$

Hooks and Framing

- Hooks allow to reuse complex protocol invariants for server (dependable) components (e.g., CCI);
- that cannot be "framed out";
- Can be more fine-grained: consider specific transitions.

Hook Footprint (e.g., CCI) determines necessary server protocol

Composition in Distributed Systems

Modular Program Verification

DISEL: **Distributed Separation Logic** $-\{P\} \ c \ \{Q\}$

https://github.com/DistributedComponents/disel

- Cloud Compute + Variations;
- Two-Phase Commit: Protocol, Invariants, Clients;
- Simple Blockchain Consensus protocol;
- Lease-based lock and distributed resource (WIP);
- Extraction and trusted shim implementation,

Compositional Reasoning about Distributed Systems

Separation of Programs and Protocols: Program Logics

- Separation of Invariant Proofs: Framing
- Separation of Inter-Protocol Dependencies: Hooks

Plenty of aspects to address in the future:

node crashes, reconfiguration, byzantine faults, protocol updates, authentication, per-node concurrency, dynamic network topologies, integrating automation tools, (Ivy, TLA+, CVC4)...

- $-\{P\}\ c\ \{Q\}$

Backup Slides

How is it different from (Multiparty) Session Types?

- Session types do not describe the state of nodes;
- Limited support for horizontal system composition.

• No way to express global system invariants (*e.g.*, consensus);

How is it different from proving program refinement?

- Our logic establishes a version of refinement by means of "programming with linearization points";
- Protocol transitions (send/receive) observable LPs.
- Information hiding by means of abstract predicates.

Verification Efforts

Protocol Invariants

Chapar

IronFleet

Verdi PSync

EventML

Mace

DistAlgo

System Correctness

Verification Efforts

	Protocol- implementation modularity	Modular program verification	Horizontal protocol composition
IronFleet	Yes	Sort of	No
Verdi	Νο	No	Νο
PSync	Νο	Νο	Νο
EventML	Νο	No	No

Protocol Framing with Hooks

FRAME $\Gamma; W \stackrel{n}{\vdash} c : \{P\}\{Q\}$ NotHooked(W, H) R is C-stable $\overline{\Gamma; W \uplus \langle C, H \rangle \stackrel{n}{\vdash} c : \{P * R\}\{Q * R\}}$

$$\begin{array}{c} \text{BIND} \\ \Gamma; W \stackrel{n}{\vdash} c_{1} : \{P\}\{Q \land \text{res} : \mathcal{T}\} \\ \overline{\Gamma, x : \mathcal{T}; W \vdash [x/\text{res}]c_{2} : \{Q\}\{R\} \ x \notin \text{FV}(R)} \\ \overline{\Gamma; W \stackrel{n}{\vdash} x \leftarrow c_{1}; c_{2} : \{P\}\{R\}} \\ \end{array} \\ \begin{array}{c} \text{LETREC} \\ \Gamma, x : \mathcal{T}, f : \langle W, \forall x : \mathcal{T}, \{P\}\{Q\}\rangle; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \overline{\Gamma; W \vdash n} c : \{T, f : \langle W, \forall x : \mathcal{T}, \{P\}\{Q\}\rangle; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \overline{\Gamma; W \vdash n} c : \{T, f : \langle W, \forall x : \mathcal{T}, \{P\}\{Q\}\rangle; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \overline{\Gamma; W \vdash n} c : \{T, f : \langle W, \forall x : \mathcal{T}, \{P\}\{Q\}\rangle; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \overline{\Gamma; W \vdash n} c : \{T, f : \langle W, \forall x : \mathcal{T}, \{P\}\{Q\}\rangle; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \hline \Gamma; W \stackrel{n}{\vdash} e : \{C, H\} \\ \overline{\Gamma; W \vdash n} c : \{T, f : \langle W, \forall x : \mathcal{T}, \{P\}\{Q\}\rangle; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \hline \Gamma; W \stackrel{n}{\vdash} e : \{P\}\{Q\} \\ \hline \Gamma; W \stackrel{n}{\vdash} c : \{P\}\{Q\} \\ \hline \Gamma; W \stackrel{n}{\vdash} c : \{P\}\{Q x R\} \\ \hline \Gamma; W \stackrel{n}{\vdash} c : \{P x R\}\{Q x R\} \\ \hline \Gamma; \langle \ell \mapsto \forall H \lnv(\mathcal{P}_{\ell}, I) \uplus W, H) \stackrel{n}{\vdash} c : \{P \land \mathcal{I}\}\{Q \land \mathcal{I}\} \\ \hline \end{array}$$

$$W = \langle C, H \rangle \qquad W \vDash s \qquad \ell \in \operatorname{dom}(C) \qquad \mathcal{P}_{\ell} = C(\ell) \qquad (MS, d) = s(\ell) \qquad \left\{ n, \ to \right\} \subseteq \operatorname{dom}(d)$$

$$\underline{\tau_s \in \mathcal{P}_{\ell}.T_s \qquad \tau_s.pre(n, to, \ m, d) \qquad \operatorname{HooksOk}(W, \tau_s, \ell, s, n, m, to) \qquad MS' = MS \uplus \langle n, to, \circ, (\tau_s.tag, m) \rangle$$

$$s \sim_W^n s[\ell \mapsto (MS', d[n \mapsto \tau_s.step(to, m, d(n))])]$$

$$W = \langle C, H \rangle \quad W \vDash s \quad \ell \in \operatorname{dom}(C) \quad \mathcal{P}_{\ell} = C(\ell)$$

$$\mathsf{m} = \langle from, n, \circ, (\tau_r.tag, m) \rangle \quad \{from, n\} \subseteq \operatorname{dom}(d) \quad \tau_r$$

$$s \quad \rightsquigarrow_W^n \quad s[\ell \mapsto (MS'', d)]$$

Network Semantics

 $(MS,d) = s(\ell) \qquad \tau_r \in \mathcal{P}_{\ell}.T_r \qquad MS = MS' \uplus \mathsf{m}$ $\tau_r.pre(\mathsf{m},d(n)) \qquad MS'' = MS' \uplus \langle from, n, \bullet, (\tau_r.tag,m) \rangle \qquad \text{RECV}$ $d[n \mapsto \tau_r.step(\mathbf{m}, d(n))])]$

Defs/Specs	Impl	Proofs	Build	
culator (§2)				
239	_	243	4.8	
192	43	153	8.6	
120	24	99	4.8	
75	7	49	2.4	
Commit (§4	4.1–§4	.3)		
465	-	231	3.9	
236	35	440	20	
163	24	198	11	
997	-	2113	36	
' y/TPC (§4.4)				
169	-	115	2.1	
326	18	707	22	
76	5	89	2.6	