## Verified Time-Aware Stream Processing

Rafael Castro G. Silva

rasi@di.ku.dk

Department of Computer Science University of Copenhagen

02/11/2023



# What is this PhD/Status seminar about?

- Distributed Systems
  - Stream processing frameworks
    - Dataflow models
      - Time-Aware Computations
- Formal Methods
  - Verification using proof assistants
    - Isabelle proofs
    - Verified and executable code
- Formalization of Time-Aware Stream Processing

# What is this PhD/Status seminar about?

- Distributed Systems
  - Stream processing frameworks
    - Dataflow models
    - Time-Aware Computations
- Formal Methods
  - Verification using proof assistants
    - Isabelle proofs
    - Verified and executable code
- Formalization of Time-Aware Stream Processing

# What is this PhD/Status seminar about?

- Distributed Systems
  - Stream processing frameworks
    - Dataflow models
    - Time-Aware Computations
- Formal Methods
  - Verification using proof assistants
    - Isabelle proofs
      - Verified and executable code
- Formalization of Time-Aware Stream Processing

#### Contents

- Introduction
- Preliminaries
- Lazy Lists Processors
- Time-Aware Operators
- Case Study
- Next Steps

## Introduction

#### Dataflow Models

- Stream Processing
- Dataflow Model
- Time-Aware Computations
  - Explain watermarks
- Bugs in Stream Processing

## **Preliminaries**

# Isabelle/HOL

- HOL
- Isabelle/HOL

• Datatypes and Codatatypes

```
codatatype (lset: 'a) //list = Inull: LNil | LCons (lhd: 'a) (ltl: 'a //list)
for map: Imap where ltl LNil = LNil
```

- Examples:
  - LNil
  - LCons 1 (LCons 2 (LCons 3 LNil))
  - LCons 0 (LCons 0 (LCons 0 (...)))
- Induction principle assuming membership in the lazy list
- Coinductive principle for lazy list equality:
  - Show that there is a pair of goggles that makes them to look the same, which implies that:
    - The first lazy list if empty iff second is
    - They have the same head
    - Their tail looks the same

Datatypes and Codatatypes

```
codatatype (lset: 'a) //ist = Inull: LNil | LCons (lhd: 'a) (ltl: 'a //ist)
for map: Imap where ltl LNil = LNil
```

- Examples:
  - LNil
  - LCons 1 (LCons 2 (LCons 3 LNil))
  - LCons 0 (LCons 0 (...)))
- Induction principle assuming membership in the lazy list
- Coinductive principle for lazy list equality:
  - Show that there is a pair of goggles that makes them to look the same, which implies that:
    - The first lazy list if empty iff second is
    - They have the same head
    - Their tail looks the same

Datatypes and Codatatypes

```
codatatype (lset: 'a) //ist = Inull: LNil | LCons (lhd: 'a) (ltl: 'a //ist)
for map: Imap where ltl LNil = LNil
```

- Examples:
  - LNil
  - LCons 1 (LCons 2 (LCons 3 LNil))
  - LCons 0 (LCons 0 (LCons 0 (...)))
- Induction principle assuming membership in the lazy list
- Coinductive principle for lazy list equality:
  - Show that there is a pair of goggles that makes them to look the same, which must imply that:
    - The first lazy list if empty iff second is
    - They have the same head
    - Their tail looks the same

Datatypes and Codatatypes

```
\label{eq:codatatype} \begin{array}{l} \mathbf{codatatype} \ (\mathsf{lset}: 'a) \ \mathit{llist} = \mathsf{Inull}: \ \mathsf{LNil} \ | \ \mathsf{LCons} \ (\mathsf{lhd}: 'a) \ (\mathsf{ltl}: 'a \ \mathit{llist}) \\ \mathbf{for} \ \mathsf{map}: \ \mathsf{lmap} \ \mathbf{where} \ \mathsf{ltl} \ \mathsf{LNil} = \mathsf{LNil} \end{array}
```

- Examples:
  - LNil
  - LCons 1 (LCons 2 (LCons 3 LNil))
  - LCons 0 (LCons 0 (LCons 0 (...)))
- Induction principle assuming membership in the lazy list
- Coinductive principle for lazy list equality:
  - Show that there is a pair of goggles that makes them to look the same, which must imply that:
    - The first lazy list if empty iff second is
    - They have the same head
    - Their tail looks the same

## Isabelle/HOL: Recursion and While Combinator

Recursion

```
fun lshift :: 'a list \Rightarrow 'a llist \Rightarrow 'a llist (infixr @@ 65) where lshift [] lxs = lxs | lshift (x \# xs) lxs = LCons x (lshift xs \ lxs)
```

While Combinator

```
definition while_option :: ('a \Rightarrow bool) \Rightarrow ('a \Rightarrow 'a) \Rightarrow 'a \Rightarrow 'a \text{ option where} while_option b \in S = \dots
```

- While rule for invariant reasoning (hoare-style):
  - There is something that holds before a step; that thing still holds after the step

- Corecursion is like recursion, but instead of always eventually reducing an argument it always eventually produces something
- Corec:

- Friendly function
  - Preserves productivity: it may consume at most one constructor to produce one constructor.

Coinduction up to congruence: Coinduction for Lazy list equality can be extended to compare an

- Corecursion is like recursion, but instead of always eventually reducing an argument it always eventually produces something
- Corec:

```
corec lapp :: 'a llist \Rightarrow 'a llist \Rightarrow 'a llist where lapp lxs lys = case lxs of LNil \Rightarrow lys | LCons x lxs' \Rightarrow LCons x (lapp lxs' lys)
```

- Friendly function
  - Preserves productivity: it may consume at most one constructor to produce one constructor.

```
friend_of_corec lshift where

xs @@ lxs = (case xs of)

[] \Rightarrow (case lxs of LNil \Rightarrow LNil | LCons x lxs' \Rightarrow LCons x lxs')

| x \# xs' \Rightarrow LCons x (xs' @@ lxs))

by (auto split: list.splits llist.splits) (transfer_prover)
```

• Coinduction up to congruence: Coinduction for Lazy list equality can be extended to compare an entire finite prefix through a congruence relation

- Corecursion is like recursion, but instead of always eventually reducing an argument it always eventually produces something
- Corec:

```
corec lapp :: 'a llist \Rightarrow 'a llist \Rightarrow 'a llist where lapp lxs lys = case lxs of LNil \Rightarrow lys | LCons x lxs' \Rightarrow LCons x (lapp lxs' lys)
```

- Friendly function
  - Preserves productivity: it may consume at most one constructor to produce one constructor.

• Coinduction up to congruence: Coinduction for Lazy list equality can be extended to compare an entire finite prefix through a congruence relation

- Corecursion is like recursion, but instead of always eventually reducing an argument it always eventually produces something
- Corec:

```
corec lapp :: 'a llist \Rightarrow 'a llist \Rightarrow 'a llist where lapp lxs lys = case lxs of LNil \Rightarrow lys | LCons x lxs' \Rightarrow LCons x (lapp lxs' lys)
```

- Friendly function
  - Preserves productivity: it may consume at most one constructor to produce one constructor.

```
friend_of_corec lshift where

xs @@ lxs = (case xs of \\ [] \Rightarrow (case lxs of LNil \Rightarrow LNil | LCons x lxs' \Rightarrow LCons x lxs') \\ | x#xs' \Rightarrow LCons x (xs' @@ lxs)) \\ by (auto split: list.splits llist.splits) (transfer_prover)

|concat lxs = case lxs of LNil \Rightarrow LNil | LCons xs lxs' \Rightarrow |shift xs (|concat lxs')|
```

• Coinduction up to congruence: Coinduction for Lazy list equality can be extended to compare an entire finite prefix through a congruence relation

## Isabelle/HOL: (Co)inductive Predicates

- Inductive predicate
  - Finite number of introduction rule applications

```
inductive in_llist :: 'a \Rightarrow 'a \text{ llist} \Rightarrow bool \text{ where}
In_llist: in_llist x \text{ (LCons } x \text{ lxs)}
| Next_llist: in_llist x \text{ lxs} \Rightarrow \text{in_llist } x \text{ (LCons } y \text{ lxs)}
in_llist 2 (LCons 1 (LCons (2 (...))))
```

- Coinductive predicate
  - Infinite number of introduction rule applications

```
coinductive | prefix :: 'a | llist \Rightarrow 'a | llist \Rightarrow bool where | LNil_|prefix: | prefix LNil | lys | LCons_|prefix: | prefix | lxs | lys \Rightarrow | prefix (LCons \times lxs) (LCons \times lys) | | prefix (LCons 1 (LCons (2 (...)))) |
```

## Isabelle/HOL: (Co)inductive Predicates

- Inductive predicate
  - Finite number of introduction rule applications

```
inductive in_llist :: 'a \Rightarrow 'a \text{ llist} \Rightarrow bool \text{ where}
In_llist: in_llist x \text{ (LCons } x \text{ lxs)}
| Next_llist: in_llist x \text{ lxs} \Rightarrow \text{in_llist } x \text{ (LCons } y \text{ lxs)}
in_llist 2 (LCons 1 (LCons (2 (...))))
```

- Coinductive predicate
  - Infinite number of introduction rule applications

## Isabelle/HOL: (Co)inductive Predicates

- Inductive predicate
  - Finite number of introduction rule applications

```
inductive in_llist :: 'a \Rightarrow 'a \text{ llist} \Rightarrow bool \text{ where}
In_llist: in_llist x \text{ (LCons } x \text{ lxs)}
| Next_llist: in_llist x \text{ lxs} \Rightarrow \text{in_llist } x \text{ (LCons } y \text{ lxs)}
in_llist 2 (LCons 1 (LCons (2 (...))))
```

- Coinductive predicate
  - Infinite number of introduction rule applications

# Lazy Lists Processors

#### Operator formalization

- Operator as a codatatype
  - Taking 'i as the input type, and 'o as the output type: codatatype ('o, 'i) op = Logic (apply: ('i  $\Rightarrow$  ('o, 'i) op  $\times$  'o list))
  - Infinite trees: applying the selector apply "walks" a branch of the tree

#### Operator formalization

- Operator as a codatatype
  - Taking 'i as the input type, and 'o as the output type: codatatype ('o, 'i) op = Logic (apply: ('i  $\Rightarrow$  ('o, 'i) op  $\times$  'o list))
  - Infinite trees: applying the selector apply "walks" a branch of the tree

#### Execution formalization

• Produce function: applies the logic (co)recursively throughout a lazy list

```
definition \operatorname{produce}_1' op lxs = \operatorname{while\_option} (\lambda(op, lxs). \neg \operatorname{Inull} lxs \land \operatorname{snd} (\operatorname{apply} op (\operatorname{Ihd} lxs)) = []) (\lambda(op, lxs). (\operatorname{fst} (\operatorname{apply} op (\operatorname{Ihd} lxs)), \operatorname{Itl} lxs)) (op, lxs) definition \operatorname{produce}_1 op lxs = (\operatorname{case} \operatorname{produce}_1' op \, lxs \, \operatorname{of} \, \operatorname{None} \Rightarrow \operatorname{None} |\operatorname{Some} (op', lxs') \Rightarrow \operatorname{if} \operatorname{Inull} \, lxs' \, \operatorname{then} \, \operatorname{None} \, \operatorname{else} \operatorname{let} (op'', out) = \operatorname{apply} op' (\operatorname{Ihd} \, lxs') \, \operatorname{in} \, \operatorname{Some} (op'', \operatorname{hd} \, out, \operatorname{tl} \, out, \operatorname{Itl} \, lxs')) corec produce \operatorname{where} \operatorname{produce} \, op \, lxs = (\operatorname{case} \, \operatorname{produce}_1 \, op \, lxs \, \operatorname{of} \, \operatorname{None} \Rightarrow \operatorname{LNil} |\operatorname{Some} (op', x, xs, lxs') \Rightarrow \operatorname{LCons} x \, (xs \, @@ \, \operatorname{produce} \, op' \, lxs'))
```

produce<sub>1</sub> has an induction principle based on the while invariant rule

#### Execution formalization

• Produce function: applies the logic (co)recursively throughout a lazy list

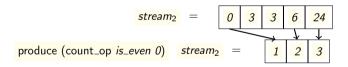
```
definition \operatorname{produce}_1' op lxs = \operatorname{while\_option} (\lambda(op, lxs). \neg \operatorname{Inull}\ lxs \land \operatorname{snd}\ (\operatorname{apply}\ op\ (\operatorname{Ihd}\ lxs)) = []) (\lambda(op, lxs). (fst (\operatorname{apply}\ op\ (\operatorname{Ihd}\ lxs)), \operatorname{Itl}\ lxs)) (op, lxs) definition \operatorname{produce}_1 op lxs = (\operatorname{case}\ \operatorname{produce}_1' op lxs of \operatorname{None} \Rightarrow \operatorname{None} |\operatorname{Some}\ (op', lxs') \Rightarrow \operatorname{if}\ \operatorname{Inull}\ lxs' then \operatorname{None}\ \operatorname{else} \operatorname{let}\ (op'', out) = \operatorname{apply}\ op'\ (\operatorname{Ihd}\ lxs') in \operatorname{Some}\ (op'', \operatorname{hd}\ out, \operatorname{tl}\ lxs')) corec produce \operatorname{where} \operatorname{produce}\ op\ lxs = (\operatorname{case}\ \operatorname{produce}_1\ op\ lxs\ of\ \operatorname{None} \Rightarrow \operatorname{LNil} |\operatorname{Some}\ (op', x, xs, lxs') \Rightarrow \operatorname{LCons}\ x\ (xs\ @@\ \operatorname{produce}\ op'\ lxs'))
```

produce<sub>1</sub> has an induction principle based on the while invariant rule

#### Operators: Count

#### • Example:

corec count\_op where count\_op P n = Logic ( $\lambda e$ . if P e then (count\_op P (n + 1), [n+1]) else (count\_op P n, []))



## Sequential Composition

 Sequential composition: take the output of the first operator and give it as input to the second operator.

```
definition fproduce op \ xs = \text{fold} \ (\lambda e \ (op, out).
let (op', out') = \text{apply } op \ e \ \text{in} \ (op', out @ out')) \ xs \ (op, [])
corec comp_op where
comp_op op_1 \ op_2 = \text{Logic} \ (\lambda ev.
let (op_1', out) = \text{apply } op_1 \ ev; \ (op_2', out') = \text{fproduce } op_2 \ out \ \text{in} \ (\text{comp\_op } op_1' \ op_2', out'))
```

Correctness:

```
produce (comp_op op_1 op_2) lxs = produce <math>op_2 (produce op_1 lxs)
```

- Proof: coinduction principle for lazy list equality and produce<sub>1</sub> induction principle
  - Generalization: we must be able to reason about elements in arbitrary positions

```
corec skip_op where skip_op op \ n = \text{Logic} \ (\lambda ev. \ \text{let} \ (op', out) = \text{apply} \ op \ ev \ \text{in} if length out < n then (skip_op op' \ (n - \text{length} \ out), [] else (op', \text{drop} \ n \ out))
```

Correctness: Coinduction up to congruence for lazy list equality

Correctness:

```
produce (comp_op op_1 op_2) lxs = produce <math>op_2 (produce op_1 lxs)
```

- Proof: coinduction principle for lazy list equality and <a href="produce1">produce1</a> induction principle
  - Generalization: we must be able to reason about elements in arbitrary positions

```
corec skip_op where skip_op op \ n = \text{Logic} \ (\lambda ev. \ \text{let} \ (op', \ out) = \text{apply} \ op \ ev \ \text{in} if length out < n then (\text{skip\_op} \ op' \ (n - \text{length} \ out), \ []) else (op', \ drop \ n \ out))
```

Correctness: Coinduction up to congruence for lazy list equality

Correctness:

```
produce (comp_op op_1 op_2) lxs = produce <math>op_2 (produce op_1 lxs)
```

- Proof: coinduction principle for lazy list equality and <a href="produce1">produce1</a> induction principle
  - Generalization: we must be able to reason about elements in arbitrary positions

```
corec skip_op where skip_op op \ n = \text{Logic} \ (\lambda ev. \ \text{let} \ (op', out) = \text{apply} \ op \ ev \ \text{in} if length out < n then (skip_op op' \ (n - \text{length} \ out), \ []) else <math>(op', \text{drop} \ n \ out))
```

• Correctness: Coinduction up to congruence for lazy list equality

Correctness:

```
produce (comp_op op_1 op_2) lxs = produce <math>op_2 (produce op_1 lxs)
```

- Proof: coinduction principle for lazy list equality and <a href="produce1">produce1</a> induction principle
  - Generalization: we must be able to reason about elements in arbitrary positions

```
corec skip_op where skip_op op \ n = \text{Logic} \ (\lambda ev. \ \text{let} \ (op', out) = \text{apply} \ op \ ev \ \text{in} if length out < n \ \text{then} \ (\text{skip\_op} \ op' \ (n - \text{length} \ out), \ []) else (op', \ drop \ n \ out))
```

• Correctness: Coinduction up to congruence for lazy list equality

Time-Aware Operators

#### Time-Aware Streams

• Time-Aware lazy lists

```
datatype ('t::order, 'd) event = DT (tmp: 't) (data: 'd) | WM (wmk: 't)
```

- Generalization to partial orders
  - Cycles
  - Operators with multiple inputs

#### Time-Aware Streams

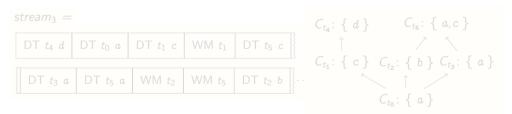
- Time-Aware lazy lists
  - datatype ('t::order, 'd) event = DT (tmp: 't) (data: 'd) | WM (wmk: 't)
- Generalization to partial orders
  - Cycles
  - Operators with multiple inputs

#### Monotone Time-Aware Streams

• Monotone: watermarks do not go back in time

```
coinductive monotone :: ('t::order, 'd) event llist \Rightarrow 't set \Rightarrow bool where LNil: monotone LNil W | LConsR: (\forall wm' \in W. \neg wm' \geq wm) \longrightarrow monotone lxs (\{wm\} \cup W) \longrightarrow monotone (LCons (WM wm) lxs) W | LConsL: (\forall wm \in W. \neg wm \geq t) \longrightarrow monotone lxs W \longrightarrow monotone (LCons (DT t d) lxs) W
```

- Up to congruence coinduction principle
- Example:



#### Monotone Time-Aware Streams

Monotone: watermarks do not go back in time

```
coinductive monotone :: ('t::order, 'd) event llist \Rightarrow 't set \Rightarrow bool where LNil: monotone LNil W | LConsR: (\forall wm' \in W. \neg wm' \geq wm) \longrightarrow monotone lxs (\{wm\} \cup W) \longrightarrow monotone (LCons (WM wm) lxs) W | LConsL: (\forall wm \in W. \neg wm \geq t) \longrightarrow monotone lxs W \longrightarrow monotone (LCons (DT t d) lxs) W
```

- Up to congruence coinduction principle
- Example

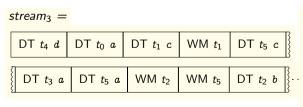


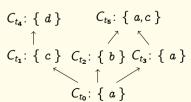
#### Monotone Time-Aware Streams

Monotone: watermarks do not go back in time

```
coinductive monotone :: ('t::order, 'd) event llist \Rightarrow 't set \Rightarrow bool where LNil: monotone LNil W | LConsR: (\forall wm' \in W. \neg wm' \geq wm) \longrightarrow monotone lxs (\{wm\} \cup W) \longrightarrow monotone (LCons (WM wm) lxs) W | LConsL: (\forall wm \in W. \neg wm \geq t) \longrightarrow monotone lxs W \longrightarrow monotone (LCons (DT t d) lxs) W
```

- Up to congruence coinduction principle
- Example:





- Productive: always eventually allows the production
  - Batching operators: accumulate data until its completion
  - Data is always eventually completed by some watermark

```
coinductive productive where
LFinite: Ifinite lxs \longrightarrow productive lxs
| EnvWM: \neg Ifinite lxs \longrightarrow (\exists u \in vimage WM (lset <math>lxs). u \ge t) \longrightarrow productive <math>lxs \longrightarrow productive (LCons (DT t d) lxs)
| EnvDT: \neg Ifinite lxs \longrightarrow productive lxs \longrightarrow productive (LCons (WM t) lxs)
```

- Productive: always eventually allows the production
  - Batching operators: accumulate data until its completion
  - Data is always eventually completed by some watermark

```
coinductive productive where
LFinite: Ifinite lxs \longrightarrow \text{productive } lxs
| EnvWM: \neg Ifinite lxs \longrightarrow (\exists u \in \text{vimage WM (lset } lxs). \ u \ge t) \longrightarrow \text{productive } lxs \longrightarrow \text{productive } (LCons (DT t d) \ lxs)
| EnvDT: \neg Ifinite lxs \longrightarrow \text{productive } lxs \longrightarrow \text{productive } (LCons (WM t) \ lxs)
```

- Productive: always eventually allows the production
  - Batching operators: accumulate data until its completion
  - Data is always eventually completed by some watermark

```
coinductive productive where
LFinite: Ifinite lxs \longrightarrow \text{productive } lxs
| EnvWM: \neg Ifinite lxs \longrightarrow (\exists u \in \text{vimage WM (lset } lxs). \ u \ge t) \longrightarrow \text{productive } lxs \longrightarrow \text{productive } (LCons (DT t d) \ lxs)
| EnvDT: \neg Ifinite lxs \longrightarrow \text{productive } lxs \longrightarrow \text{productive } (LCons (WM t) \ lxs)
```

- Productive: always eventually allows the production
  - Batching operators: accumulate data until its completion
  - Data is always eventually completed by some watermark

```
coinductive productive where
LFinite: Ifinite lxs \longrightarrow \text{productive } lxs
\mid \text{EnvWM:} \neg \text{ Ifinite } lxs \longrightarrow (\exists u \in \text{vimage WM (lset } lxs). \ u \geq t) \longrightarrow \text{productive } lxs \longrightarrow \text{productive (LCons (DT t d) } lxs)
\mid \text{EnvDT:} \neg \text{ Ifinite } lxs \longrightarrow \text{productive } lxs \longrightarrow \text{productive (LCons (WM t) } lxs)
```

### Building Blocks: Batch Operator

### Batch Operator: Soundness

#### Batch Operator: Completeness

- Uses soundness of batch\_op
- Proof by induction over n

```
mono_prod lxs\ W \longrightarrow (\exists i\ d.\ \text{enat}\ i < \text{llength}\ lxs \land \text{lnth}\ lxs\ i = \text{DT}\ t\ d \land n = \text{Suc}\ i) \lor n = 0 \land t \in \text{set\_t}\ buf \longrightarrow (\forall t' \in \text{set\_t}\ buf.\ \text{lfinite}\ lxs \lor \exists wm \ge t'\ .\ \text{WM}\ wm \in \text{lset}\ lxs) \longrightarrow \exists wm\ batch.\ \text{DT}\ wm\ batch \in \text{lset}\ (\text{produce}\ (\text{batch\_op}\ buf)\ lxs) \land t \in \text{set\_t}\ batch \lor (\forall k \in \{n\ ..< \text{the\_enat}\ (\text{llength}\ lxs)\}\ .\ \neg\ (\exists t' \ge t.\ \text{lnth}\ lxs\ k = \text{WM}\ t')) \land \text{lfinite}\ lxs}  (1)
```

### Batch Operator: Monotone

### Batch Operator: Productive

### Building Blocks: Incremental Operator

### Batch Operator: Soundness

### Batch Operator: Completeness

### Batch Operator: Monotone

### Batch Operator: Productive

### Compositional Reasoning

# Case Study

## Histogram

### Histogram: Soundness

### Histogram: Completeness

#### Histogram: Monotone

## Histogram: Productive

# Efficient Histogram

• Foo

#### Join

Join: Soundness

# Join: Completeness

Join: Monotone

Next Steps

## Next Steps

Questions, comments and suggestions