How to Stay CALM
While Seeing What is not There?

Tom J Ameloot\textsuperscript{1}, Bas Ketsman\textsuperscript{1}, Frank Neven\textsuperscript{1}, Daniel Zinn\textsuperscript{2}

\textsuperscript{1} Hasselt University & Transnational University of Limburg
\textsuperscript{2} LogicBlox, Inc

DBDBD 2014
Introduction

- **Cloud-computing**: Setting with asynchronous communication via messages which can be arbitrarily delayed but not lost
- **CALM-conjecture**: No coordination = Monotonicity

\[\text{[Hellerstein, 2010]}\]

(CALM = Consistency And Logical Monotonicity)
Monotonicity

Definition
A query $Q$ is monotone if $Q(I) \subseteq Q(I \cup J)$ for all database instances $I$ and $J$.

Notation
$\mathcal{M}$: class of monotone queries

Example
- $Q_{\Delta}$: Select triangles in a graph $\in \mathcal{M}$
- $Q_{<}$: Select open triangles in a graph $\notin \mathcal{M}$
CALM by Example

\( Q_\Delta: \) select all triangles \( \in \mathcal{M} \)

Input instance

Algorithm
- broadcast all data
- periodically output local triangles

No coordination + Eventually consistent
CALM by Example

$Q_\leq$: select all open triangles $\not\in M$

Open triangle or fact not yet arrived??

Input instance

Requires global coordination
No-coordination = Monotonicity

[Ameloot, Neven, Van den Bussche, 2011]: TRUE
  for a setting where nodes have no information about the distribution of facts

[Zinn, Green, Ludäscher, 2012]: FALSE
  for settings where nodes have information about the distribution of facts

TRUE when also refining monotonicity
Overview

1. CALM

2. CALM Revision 1

3. CALM Revision 2
Relational Transducer Networks

[Ameloot, Neven, Van den Bussche, 2011]

- Network $\mathcal{N} = \{x, y, u, z\}$
- Transducer $\Pi$
- Messages can be arbitrarily delayed but never get lost

Semantics defined in terms of runs over a transition system
Definition

A transducer $\Pi$ computes a query $Q$ if

- for all networks $\mathcal{N}$,
- for all databases $I$,
- for all horizontal distributions $H$, and
- for every run of $\Pi$, $\text{out}(\Pi) = Q(I)$.

Consistency requirement

Network independent

Data distribution independent
Coordination-free Algorithms

[Ameloot, Neven, Van den Bussche, 2011]

**Definition**

Π is coordination-free if for all inputs I there is a distribution on which Π computes \( Q(I) \) without having to do communication.

**Goal:** separate data-communication from coordination-communication
Example: Ideal Distribution

$Q_\Delta$: select all triangles

Input instance

Algorithm

- (broadcast all data)
- periodically output local triangles

No communication required
A query has a coordination-free and eventually consistent execution strategy iff the query is monotone.

Theorem
\[ \mathcal{F}_0 = \mathcal{M} \]

Definition
\[ \mathcal{F}_0 = \text{set of queries which are distributedly computed by coordination-free transducers} \]
Overview

1. CALM

2. CALM Revision 1

3. CALM Revision 2
Policy-aware Transducers

Input instance

knows about missing fact

“Distribution Policy”
Definition
A distribution policy \( P \) for \( \sigma \) and \( \mathcal{N} \) is a total function from \( \text{facts}(\sigma) \) to the power set of \( \mathcal{N} \).

Definition
A policy-aware transducer is a transducer with access to \( P \) restricted to its active domain.

Definition
\( \mathcal{F}_1 = \) set of queries which are distributedly computed by policy-aware coordination-free transducers.
Domain-distinct-monotonicity

Definition
A fact $f$ is **domain distinct** from instance $I$ when $\text{dom}(f) \not\subseteq \text{dom}(I)$.

Example
Domain-distinct-monotonicity

**Definition**
An instance $J$ is domain distinct from instance $I$ when every fact $f \in J$ is domain distinct from $I$.

**Example**

$I$

$J$

$\times$
Domain-distinct-monotonicity

**Definition**
A query $Q$ is **domain-distinct-monotone** if $Q(I) \subseteq Q(I \cup J)$ for all $I$ and $J$ for which $J$ is domain distinct from $I$.

**Notation**
$\mathcal{M}_{\text{distinct}}$: class of **domain-distinct-monotone** queries

**Remark**
$\mathcal{M}_{\text{distinct}}$: class of queries preserved under extensions
**Domain-distinct-monotonicity**

**Example**

Select open triangles in graph $\in M_{distinct}$.

Not domain-distinct from $I$
Revised CALM-conjecture

A query has a coordination-free and eventually consistent execution strategy under distribution policies iff the query is domain-distinct-monotone

Theorem
\[ \mathcal{F}_1 = \mathcal{M}_{\text{distinct}} \]

Definition
\[ \mathcal{F}_1 = \text{set of queries which are distributedly computed by policy-aware coordination-free transducers} \]
Proof of $\mathcal{M}_{\text{distinct}} \subseteq \mathcal{F}_1$

- **Monotonicity:** $Q(J) \subseteq Q(I)$ for every $J \subseteq I$
- **Domain-distinct-monotonicity:**

Let $I$ be an instance, $C \subseteq \text{adom}(I)$.

**Induced instance:** $I|_C = \{f \in I \mid \text{adom}(f) \subseteq C\}$

By domain-distinct-monotonicity: $Q(I|_C) \subseteq Q(I)$
Proof of $\mathcal{M}_{\text{distinct}} \subseteq \mathcal{F}_1$

- $\mathcal{F}_1$ setting:

Let $I$ be an instance, $C \subseteq \text{adom}(I)$.

$C$ is complete at node $x$ when $x$ knows for every fact $f$ with $\text{adom}(f) \subseteq C$ whether $f \in I$ or $f \notin I$.

complete set = instance based on complete $C$

= induced instance of $I$ based on $C$

Algorithm

- broadcast all present and deduced absent facts
- Evaluate query on complete sets
Overview

1. CALM

2. CALM Revision 1

3. CALM Revision 2
Domain-guided Policies

Input instance

“Distribution Policy”
Domain-guided Policies

[Zinn, Green, Ludäscher, 2012]

Definition

\( \mathcal{F}_2 = \) queries which are distributively computed under domain-guided distribution policies by policy-aware coordination-free transducers.
Domain-disjoint-monotonicity

**Definition**
An instance $J$ is domain disjoint from instance $I$ when $\text{adom}(I) \cap \text{adom}(J) = \emptyset$.

**Example**

![Diagram showing instances I, J, and J']
Domain-disjoint-monotonicity

Definition
A query $Q$ is domain-disjoint-monotone if $Q(I) \subseteq Q(I \cup J)$ for all $I$ and $J$ for which $J$ is domain disjoint from $I$.

Notation
$M_{disjoint}$: class of domain-disjoint-monotone queries
Revised CALM-conjecture

A query has a coordination-free and eventually consistent execution strategy under domain-guided distribution policies iff the query is domain-disjoint-monotone

Theorem
\[ \mathcal{F}_2 = \mathcal{M}_{\text{disjoint}} \]

Definition
\[ \mathcal{F}_2 = \text{queries which are distributedly computed under domain-guided distribution policies by policy-aware coordination-free transducers.} \]
Summary

\[ \mathcal{F}_0 \subseteq \mathcal{M} \subseteq \text{wlLOG}(\neq) \subseteq \text{Datalog}(\neq) \]

\[ \mathcal{F}_1 \subseteq \mathcal{M}_{\text{distinct}} \subseteq \text{SP-wILOG} \subseteq \text{SP-Datalog} \]

\[ \mathcal{F}_2 \subseteq \mathcal{M}_{\text{disjoint}} \subseteq \text{semicon-wILOG} \subseteq \text{semicon-Datalog} \]

- Coordination
- Monotonicity
- Datalog + value invention
- Datalog
References

Tom J Ameloot, B. K., Frank Neven and Daniel Zinn. Weaker Forms of Monotonicity for Declarative Networking: a more fine-grained answer to the CALM-conjecture. In PODS 2014.