The Complexity of Boundedness for Guarded Logics

Michael Benedikt\textsuperscript{1}, Balder ten Cate\textsuperscript{2}, Thomas Colcombet\textsuperscript{3}, Michael Vanden Boom\textsuperscript{1}

\textsuperscript{1}University of Oxford \hspace{1cm} \textsuperscript{2}LogicBlox and UC Santa Cruz \hspace{1cm} \textsuperscript{3}Université Paris Diderot

LICS 2015
Kyoto, Japan
Least fixpoint

Consider $\psi(y, Y)$ positive in $Y$ (of arity $m = |y|$).

For all structures $\mathfrak{A}$, the formula $\psi$ induces a monotone operation

$$
\mathcal{P}(A^m) \longrightarrow \mathcal{P}(A^m)
$$

$$
V \longmapsto \psi_{\mathfrak{A}}(V) := \{ a \in A^m : \mathfrak{A}, a, V \models \psi \}
$$

$\Rightarrow$ there is a unique least fixpoint $[\text{Lfp}_{Y,y}.\psi(y, Y)]_{\mathfrak{A}} := \bigcup_{\alpha} \psi_{\mathfrak{A}}^{\alpha}

\psi_{\mathfrak{A}}^0 := \emptyset

\psi_{\mathfrak{A}}^{\alpha+1} := \psi_{\mathfrak{A}}(\psi_{\mathfrak{A}}^\alpha)

\psi_{\mathfrak{A}}^\lambda := \bigcup_{\alpha<\lambda} \psi_{\mathfrak{A}}^{\alpha}$
## Boundedness problem

### Boundedness problem for $\mathcal{L}$

**Input:** $\psi(y, Y) \in \mathcal{L}$ positive in $Y$

**Question:** is there $n \in \mathbb{N}$ s.t. for all structures $\mathcal{A}$, $\psi^*_n(\mathcal{A}) = \psi^{n+1}(\mathcal{A})$? (i.e. the least fixpoint is always reached within $n$ iterations)
Boundedness problem

Boundedness problem for \( \mathcal{L} \)

**Input:** \( \psi(y, Y) \in \mathcal{L} \) positive in \( Y \)

**Question:** is there \( n \in \mathbb{N} \) s.t. for all structures \( \mathfrak{A} \), \( \psi_{\mathfrak{A}}^n = \psi_{\mathfrak{A}}^{n+1} \)?
(i.e. the least fixpoint is always reached within \( n \) iterations)

\[
\psi_1(xy, Y) := Rxy \lor \exists z (Rxz \land Yzy)
\]
Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(y, Y) \in \mathcal{L}$ positive in $Y$

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathcal{A}$, $\psi^n_{\mathcal{A}} = \psi^{n+1}_{\mathcal{A}}$? (i.e. the least fixpoint is always reached within $n$ iterations)

\[
\psi_1(xy, Y) := Rxy \lor \exists z (Rxz \land Yzy)
\] unbounded
Boundedness problem

Boundedness problem for $\mathcal{L}$

**Input:** $\psi(y, Y) \in \mathcal{L}$ positive in $Y$

**Question:** is there $n \in \mathbb{N}$ s.t. for all structures $\mathcal{A}$, $\psi^n_{\mathcal{A}} = \psi^{n+1}_{\mathcal{A}}$? (i.e. the least fixpoint is always reached within $n$ iterations)

$$\psi_1(xy, Y) := Rxy \lor \exists z (Rxz \land Yzy)$$

$[\text{lfp}_{Y,xy}.\psi_1] \equiv \text{transitive closure of } R$

unbounded
Boundedness problem

Boundedness problem for \( \mathcal{L} \)

**Input:** \( \psi(y, Y) \in \mathcal{L} \) positive in \( Y \)

**Question:** is there \( n \in \mathbb{N} \) s.t. for all structures \( \mathcal{A} \), \( \psi^n_{\mathcal{A}} = \psi^{n+1}_{\mathcal{A}} \)?

(i.e. the least fixpoint is always reached within \( n \) iterations)

\[
\psi_1(xy, Y) := Rxy \lor \exists z (Rxz \land Yzy)
\]

unbounded

\[
[\text{Ifp}_{y,xy} \cdot \psi_1] \equiv \text{transitive closure of } R
\]

\[
\psi_2(xy, Y) := Rxy \lor \exists z (Yzy)
\]
Boundedness problem

Boundedness problem for $\mathcal{L}$

**Input:** $\psi(y, Y) \in \mathcal{L}$ positive in $Y$

**Question:** is there $n \in \mathbb{N}$ s.t. for all structures $\mathcal{A}$, $\psi_\mathcal{A}^n = \psi_\mathcal{A}^{n+1}$? (i.e. the least fixpoint is always reached within $n$ iterations)

\[
\psi_1(xy, Y) := Rxy \lor \exists z (Rxz \land Yzy)
\]

\[
[lfp_{Y,xy} \cdot \psi_1] \equiv \text{transitive closure of } R
\]

\[
\psi_2(xy, Y) := Rxy \lor \exists z (Yzy)
\]

\[
[lfp_{Y,xy} \cdot \psi_2](xy) \equiv Rxy \lor \exists z (Rzy)
\]
Boundedness problem

Boundedness problem for $\mathcal{L}$

**Input:** $\psi(y, Y) \in \mathcal{L}$ positive in $Y$

**Question:** is there $n \in \mathbb{N}$ s.t. for all structures $\mathcal{A}$, $\psi_{\mathcal{A}}^n = \psi_{\mathcal{A}}^{n+1}$?

(i.e. the least fixpoint is always reached within $n$ iterations)

\[
\psi_1(xy, Y) := Rxy \lor \exists z (Rxz \land Yzy)
\]

$[\text{Ifp}_{Y,xy}\psi_1] \equiv \text{transitive closure of } R$

\[
\psi_2(xy, Y) := Rxy \lor \exists z (Yzy)
\]

$[\text{Ifp}_{Y,xy}\psi_2](xy) \equiv Rxy \lor \exists z (Rzy)$
Some prior results

Boundedness is **undecidable** for

- binary predicate in positive existential FO (i.e. Datalog)
  [Hillebrand, Kanellakis, Mairson, Vardi ’95]

- monadic predicate in existential FO with inequalities
  [Gaifman, Mairson, Sagiv, Vardi ’87]

- monadic predicate in FO²
  [Kolaitis, Otto ’98]
Some prior results

Boundedness is **undecidable** for

- binary predicate in positive existential FO (i.e. Datalog)
  [Hillebrand, Kanellakis, Mairson, Vardi '95]
- monadic predicate in existential FO with inequalities
  [Gaifman, Mairson, Sagiv, Vardi '87]
- monadic predicate in $\text{FO}^2$
  [Kolaitis, Otto '98]

Boundedness is **decidable** for

- monadic predicate in positive existential FO (i.e. monadic Datalog)
  [Cosmadakis, Gaifman, Kanellakis, Vardi '88]
  $2\text{EXPTIME}$
- monadic predicate in modal logic
  [Otto '99]
  $\text{EXPTIME}$
- predicates in “guarded logics”
  [Blumensath, Otto, Weyer '14]
  [Bárány, ten Cate, Otto '12]
  non-elementary upper bound
Some prior results

Boundedness is **undecidable** for

- binary predicate in positive existential FO (i.e. Datalog)
  [Hillebrand, Kanellakis, Mairson, Vardi ‘95]

- monadic predicate in existential FO with inequalities
  [Gaifman, Mairson, Sagiv, Vardi ‘87]

- monadic predicate in FO$^2$
  [Kolaitis, Otto ‘98]

Boundedness is **decidable** for

- monadic predicate in positive existential FO (i.e. monadic Datalog)
  [Cosmadakis, Gaifman, Kanellakis, Vardi ‘88]
  \[2\text{EXPTIME}\]

- monadic predicate in modal logic
  [Otto ‘99]
  \[\text{EXPTIME}\]

- predicates in “guarded logics”
  [Blumensath, Otto, Weyer ‘14]
  [Bárány, ten Cate, Otto ‘12]
  non-elementary upper bound

**our contribution:**

elementary upper bound
(or better)
Guarded logics

\[ \exists x(\text{GF}(xy) \land \psi(xy)) \]
\[ \forall x(\text{GF}(xy) \rightarrow \psi(xy)) \]

[Andréka, van Benthem, Németi ’95-’98]

constrain quantification
Guarded logics

constrain quantification

\[ \exists x(G(xy) \land \psi(xy)) \]
\[ \forall x(G(xy) \rightarrow \psi(xy)) \]

[Andréka, van Benthem, Németi ‘95–’98]

constrain negation

\[ \exists x(\psi(xy)) \]
\[ \neg \psi(x) \]

[ten Cate, Segoufin ‘11]
Guarded logics

- Constrain quantification
  \[ \exists x (G(xy) \land \psi(xy)) \]
  \[ \forall x (G(xy) \rightarrow \psi(xy)) \]
  [Andréka, van Benthem, Németi ‘95-’98]

- Constrain negation
  \[ \exists x (\psi(xy)) \]
  \[ G(xy) \land \neg \psi(xy) \]
  [ten Cate, Segoufin ‘11]
  [Bárány, ten Cate, Segoufin ‘11]
Guarded logics

constrain quantification

\[ \exists x (G(xy) \land \psi(xy)) \]
\[ \forall x (G(xy) \rightarrow \psi(xy)) \]

[Andréka, van Benthen, Németi ’95–’98]

constrain negation

\[ \exists x (\psi(xy)) \]
\[ G(xy) \land \neg \psi(xy) \]

[ten Cate, Segoufin ’11]
[Bárány, ten Cate, Segoufin ’11]
Guarded logics

Guarded logics are **expressive**. For instance, GNFP captures:

- mu-calculus, even with backwards modalities;
- positive existential FO (i.e. unions of conjunctive queries);
- description logics including $\text{ALC}, \text{ALCHIO}, \text{ELI}$;
- monadic Datalog.
Guarded logics

Guarded logics are **expressive**. For instance, GNFP captures:

- mu-calculus, even with backwards modalities;
- positive existential FO (i.e. unions of conjunctive queries);
- description logics including \(\mathcal{ALC}, \mathcal{ALCHIO}, \mathcal{ELI}\);
- monadic Datalog.

Guarded logics have many nice model theoretic properties.

- GF, UNF, and GNF have **finite models**.
- GFP, UNFP, and GNFP have **tree-like models**
  (models of bounded tree-width).
Guarded logics

Guarded logics are **expressive**. For instance, GNFP captures:

- mu-calculus, even with backwards modalities;
- positive existential FO (i.e. unions of conjunctive queries);
- description logics including $\mathcal{ALC}, \mathcal{ALCHIO}, \mathcal{ELI}$;
- monadic Datalog.

Guarded logics have many nice model theoretic properties.

- GF, UNF, and GNF have **finite models**.
- GFP, UNFP, and GNFP have **tree-like models** (models of bounded tree-width).

Guarded logics have nice **computational properties**.

- Satisfiability is decidable, and is 2EXPTIME-complete (even EXPTIME-complete for fixed-width GFP).
Boundedness for guarded logics

(We say $\psi(x)$ is answer-guarded if it is of the form $G(x) \land \psi'(x)$.)

**Corollary** to tree-like model property

For $\psi$ in GFP or answer-guarded GNFP:
$\psi$ is bounded over all structures iff $\psi$ is bounded over **tree-like structures**.
Boundedness for guarded logics

(We say $\psi(x)$ is **answer-guarded** if it is of the form $G(x) \land \psi'(x)$.)

**Corollary** to tree-like model property

For $\psi$ in GFP or answer-guarded GNFP:
$\psi$ is bounded over all structures iff $\psi$ is bounded over **tree-like structures**.

$\Rightarrow$ amenable to techniques using **tree automata**
Boundedness for guarded logics

(We say $\psi(x)$ is **answer-guarded** if it is of the form $G(x) \land \psi'(x)$.)

**Corollary** to tree-like model property

For $\psi$ in GFP or answer-guarded GNFP:

$\psi$ is bounded over all structures iff $\psi$ is bounded over **tree-like structures**.

$\Rightarrow$ amenable to techniques using **tree automata**

**Logic-automata connection** utilized in Blumensath et al. ’14 but only yields non-elementary complexity since their proof goes via MSO.
Boundedness for guarded logics

(We say $\psi(x)$ is **answer-guarded** if it is of the form $G(x) \land \psi'(x)$.)

**Corollary** to tree-like model property

For $\psi$ in GFP or answer-guarded GNFP:
$\psi$ is bounded over all structures iff $\psi$ is bounded over **tree-like structures**.

$\Rightarrow$ amenable to techniques using **tree automata**

**Logic-automata connection** utilized in Blumensath et al. ’14 but only yields non-elementary complexity since their proof goes via MSO.

**Our strategy**: construct automata for boundedness problem directly.
Boundedness for guarded logics

(We say $\psi(x)$ is answer-guarded if it is of the form $G(x) \land \psi'(x)$.)

**Corollary** to tree-like model property

For $\psi$ in GFP or answer-guarded GNFP:
$\psi$ is bounded over all structures iff $\psi$ is bounded over tree-like structures.

$\Rightarrow$ amenable to techniques using tree automata

Logic-automata connection utilized in Blumensath et al. ’14 but only yields non-elementary complexity since their proof goes via MSO.

Our strategy: construct cost automata for boundedness problem directly.
Cost automata

Cost automaton $\mathcal{A}$
classical automaton + finite set of counters with operations $i$, $r$, and $\varepsilon$
Cost automata

Cost automaton $\mathcal{A}$
classical automaton + finite set of counters with operations $i, r, \text{ and } \varepsilon$

Semantics $\llbracket \mathcal{A} \rrbracket : \text{trees} \rightarrow \mathbb{N} \cup \{\infty\}$

$\llbracket \mathcal{A} \rrbracket(t) := \min \{n : \exists \text{ run } \rho \text{ of } \mathcal{A} \text{ on } t \text{ such that}$
\hspace{1cm} $\rho$ satisfies the acceptance condition and keeps counters below $n\}$
Cost automata

Cost automaton $\mathcal{A}$
classical automaton + finite set of counters with operations $\top, \cdot, \text{ and } \varepsilon$

Semantics $\llbracket \mathcal{A} \rrbracket : \text{trees} \rightarrow \mathbb{N} \cup \{\infty\}$

$\llbracket \mathcal{A} \rrbracket (t) := \min \{n : \exists \text{ run } \rho \text{ of } \mathcal{A} \text{ on } t \text{ such that }$

$\rho \text{ satisfies the acceptance condition and}$

keeps counters below $n\}$

Theorem

For all $\psi \in \text{GNFP}[\sigma]$, we can construct a 2-way cost automaton $\mathcal{A}_\psi$ such that

$\psi$ is bounded

iff $\exists \ n \in \mathbb{N}$ such that $\forall$ trees $t$, $\llbracket \mathcal{A}_\psi \rrbracket (t) \leq n$. 
Boundedness problem for cost automata

**Input:** cost automaton $\mathcal{B}$

**Question:** is there $n \in \mathbb{N}$ such that for all trees $t$, $\|\mathcal{B}\|(t) \leq n$?
Boundedness problem for cost automata

**Input:** cost automaton $\mathcal{B}$

**Question:** is there $n \in \mathbb{N}$ such that for all trees $t$, $\llbracket \mathcal{B} \rrbracket(t) \leq n$?

Decidability of boundedness is not known in general for cost automata over infinite trees...
Boundedness for cost automata

Boundedness problem for cost automata

**Input:** cost automaton \( \mathcal{B} \)

**Question:** is there \( n \in \mathbb{N} \) such that for all trees \( t \), \( \left\lfloor \mathcal{B} \right\rfloor(t) \leq n \)?

Decidability of boundedness is not known in general for cost automata over infinite trees...

...but we are interested in special types of cost automata:
1 counter that is only incremented or left unchanged (never reset).

**Theorem**

For some special types of 2-way cost automata, the boundedness problem is decidable in elementary time.
<table>
<thead>
<tr>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundedness is decidable in <strong>elementary time</strong> for answer-guarded GNF and GF.</td>
</tr>
</tbody>
</table>
## Summary of results

<table>
<thead>
<tr>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundedness is decidable in <strong>elementary time</strong> for answer-guarded GNF and GF.</td>
</tr>
</tbody>
</table>

Using unpublished results of Colcombet, this can be improved to **2EXPTIME**, and elementary bound can be extended to answer-guarded GNFP and GFP.
Summary of results

**Theorem**

Boundedness is decidable in *elementary time* for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to $2\text{EXPTIME}$, and elementary bound can be extended to answer-guarded GNFP and GFP.

This yields elementary time algorithms for:

- deciding boundedness for some Datalog-like languages
### Theorem

Boundedness is decidable in **elementary time** for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to **2EXPTIME**, and elementary bound can be extended to answer-guarded GNFP and GFP.

This yields elementary time algorithms for:

- deciding boundedness for some Datalog-like languages
- deciding FO-rewritability of $[\text{Lfp}_y, y \cdot \psi](y)$ for $\psi$ in answer-guarded GNF or GF (using [Bárány, ten Cate, Otto ’12])
Summary of results

Theorem

Boundedness is decidable in **elementary time** for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to **2EXPTIME**, and elementary bound can be extended to answer-guarded GNFP and GFP.

This yields elementary time algorithms for:

- deciding boundedness for some Datalog-like languages
- deciding FO-rewritability of $[\text{Ifp}_y, y \cdot \psi](y)$ for $\psi$ in answer-guarded GNF or GF (using [Bárány, ten Cate, Otto ’12])
- deciding FO-rewritability of CQs over guarded and frontier-guarded TGDs (using [Bárány, Benedikt, ten Cate ’13])
Boundedness is decidable in **elementary time** for guarded logics.

**Contributions**

- General translation from GNFP to automata that can be used for satisfiability testing and boundedness questions.
- Finer analysis of complexity of some cost automata constructions.
Syntax of $\text{cGNFP}[\sigma]$

$$
\varphi ::= \cdots \mid [\text{lfp}^N_{Y,y} G(y) \land \varphi(y, Y, Z)](x) \quad \text{for } \varphi \text{ positive in } Y
$$

where $\text{lfp}^N$ operators only appear positively in the formula.
Syntax of cGNFP[σ]

φ ::= · · · | [\text{lfp}_N^Y.y.G(y) \land \varphi(y, Y, Z)](x) \quad \text{for } \varphi \text{ positive in } Y

where \text{lfp}_N operations only appear positively in the formula.

Semantics \llbracket \varphi \rrbracket : \sigma\text{-structures } \rightarrow \mathbb{N} \cup \{\infty\}

\llbracket \varphi \rrbracket(\mathcal{A}) := \min \left\{ n \in \mathbb{N} : \mathcal{A} \text{ satisfies } \varphi \text{ when } [\text{lfp}_N^Y.y.\psi] \text{ replaced by } \psi^n \right\}

where \quad \psi^0 := \bot \quad \text{and} \quad \psi^n := \psi[\psi^{n-1}/Y]
Bringing cost capabilities to guarded logics

Syntax of $\text{cGNFP}[\sigma]\n
\[ \phi ::= \cdots \mid [\text{lfp}^N_{Y,y}.G(y) \land \phi(y, Y, Z)](x) \quad \text{for } \phi \text{ positive in } Y \]

where $\text{lfp}^N$ operators only appear positively in the formula.

Semantics $\llbracket \phi \rrbracket : \sigma$-structures $\rightarrow \mathbb{N} \cup \{\infty\}$

$\llbracket \phi \rrbracket(\mathcal{A}) := \min\{n \in \mathbb{N} : \mathcal{A} \text{ satisfies } \phi \text{ when } [\text{lfp}^N_{Y,y}.\psi] \text{ replaced by } \psi^n\}$

where $\psi^0 := \bot$ and $\psi^n := \psi[\psi^{n-1}/Y]\n
Example

$\phi(y) := [\text{lfp}^N_{Y,y}.S y \lor \exists z(Ryz \land Yz)](y)$

$\llbracket \phi \rrbracket(\mathcal{A}, a) := \text{minimum length of } R\text{-chain to reach } S \text{ from } a$