A Note on DL-Lite with Boolean Role Inclusions

Roman Kontchakov¹, Vladislav Ryzhikov¹, Frank Wolter², and Michael Zakharyaschev¹

> ¹ Birkbeck, University of London, U.K. ² University of Liverpool, U.K.

Abstract. We discuss the complexity of reasoning and ontology-mediated query answering with the logics from the *DL-Lite* family extended by various types (Horn, Krom and Boolean) of role inclusions. We compare the expressive power of those logics with 1- and 2-variable fragments of first-order logic. Our most interesting findings show that binary disjunctions on roles do not change the complexity of satisfiability if disjunction is allowed on concepts, while still causing undecidability of UCQ answering.

1 Introduction

Concepts (unary predicates) and roles (binary predicates) are usually treated quite differently in description logics (DLs). For example, the basic expressive DL \mathcal{ALC} allows statements about concepts in the form of arbitrary Boolean formulas but disallows any statements about roles. A notable exception is DL-Lite_{\mathcal{R}} [9], also denoted DL-Lite^{\mathcal{H}} in the classification of [1], where concept and role inclusions take the form of binary Horn (aka core) formulas $\vartheta_1 \sqsubseteq \vartheta_2$ or $\vartheta_1 \sqcap \vartheta_2 \sqsubseteq \bot$, where the ϑ_i are either both concepts or both roles. On the other hand, DL-Lite_{\mathcal{R},\sqcap} [10], aka DL-Lite^{$\mathcal{H}}$ </sup>_{horm} [1], allows arbitrary Horn concept inclusions $\vartheta_1 \sqcap \cdots \sqcap \vartheta_k \sqsubseteq \vartheta_{k+1}$ but only core role inclusions.

To explore what happens if concepts and roles are treated equally, we introduce a few new members to the DL-Lite family (which could be qualified as distant relatives or illegitimate children) and denote the logics in the extended 'clan' by DL-Lite^r_c, where the parameters $r, c \in \{bool, horn, krom, core\}$ define the allowed structure of role and, respectively, concept inclusions. We present our observations on the complexity of checking satisfiability of DL-Lite $_{c}^{r}$ knowledge bases and answering (unions of) conjunctive queries mediated by DL-Lite $_{c}^{r}$ ontologies. It turns out that the resulting languages provide a new way of classifying some fundamental fragments of firstorder logic (FO). Consider first *DL-Lite^{bool}*, which is easily seen to be contained in the two-variable fragment FO₂ of FO. Using the fact that $\mathcal{ALC}^{\top,id,\cap,\neg,-}$, the extension of ALCI with Boolean operations on roles and the identity role, has exactly the same expressive power as FO₂ [16], we show that DL- $Lite_{bool}^{bool}$ almost captures FO₂: it corresponds to $ALC^{\top,\cap,\neg,\neg}$, the language $ALC^{\top,id,\cap,\neg,\neg}$ without the identity role. DL-Litebool inherits from these languages NEXPTIME-completeness of satisfiability and undecidability of CQ-answering [17, 15]. Note that DL-Lite^{bool} also almost captures the class of linear existential disjunctive rules with two variables [8]. On the other hand, we show that satisfiability of DL-Lite^r_c knowledge bases with $r \in \{horn, krom\}$

Table 1. Combined complexity of satisfiability checking (all bounds are tight). For the languages $DL-Lite_r^r$ in the grey cells, the complexity is the same as for $DL-Lite_r^r$ as concept inclusions in c can be expressed by means of role inclusions in r.

CIs RIs	Bool	guarded Bool	Horn	Krom	core
Bool Horn Krom core	NEXPTIME (Th. 3)	EXPTIME (Th. 3)	NP (Th. 6) P (Th. 6) P (Th. 6)	NP (Th. 4) NP (Th. 4) NL (Th. 4)	NP [1] P [10] NL [1] NL [9]

and $c \in \{bool, horn, krom, core\}$ can be encoded in propositional logic (or in the onevariable fragment of FO), with the complexity of satisfiability checking ranging from NL to P and NP; see Table 1. Our most interesting findings demonstrate that binary disjunctions on roles do not change the complexity of satisfiability if disjunction is allowed on concepts, while still causing undecidability of UCQ answering. Thus, satisfiability of $DL-Lite_{krom}^{krom}$ knowledge bases is NL-complete, while answering UCQs mediated by $DL-Lite_{krom}^{krom}$ ontologies is undecidable.

The logics introduced above leave a huge gap between those having Boolean role inclusions and those that only admit Horn or Krom role inclusions. To 'complete' the picture, we add DL-Lite_c^{g-bool}, the fragment of DL-Lite_c^{bool} in which Boolean role inclusions are guarded. It turns out that DL-Lite_c^{g-bool} approximates the two-variable guarded fragment GF₂ of FO, from which it inherits EXPTIME-completeness of satisfiability and 2EXPTIME-completeness of UCQ-answering.

Our motivation to investigate the complexity of the family of languages DL-Lite^r_c with $r, c \in \{bool, horn, krom, core\}$ stems from the observation that in the context of answering queries mediated by temporal DL-Lite ontologies [3, 2] in some cases admitting the same type of concept and role inclusions does not affect the data complexity of query answering, while in others the effect is dramatic.

2 DL-Lite with Complex Role Inclusions

Let a_0, a_1, \ldots be individual names, A_0, A_1, \ldots concept names, and P_0, P_1, \ldots role names. We define roles S and basic concepts B as usual in DL-Lite:

 $S ::= P_i \mid P_i^-, \qquad B ::= A_i \mid \exists S.$

A concept or role inclusion (CI or, respectively, RI) takes the form

$$\vartheta_1 \sqcap \dots \sqcap \vartheta_k \sqsubseteq \vartheta_{k+1} \sqcup \dots \sqcup \vartheta_{k+m}, \qquad k, m \ge 0, \tag{1}$$

where the ϑ_i are all basic concepts or, respectively, roles. As usual, we denote the empty \sqcap by \top and the empty \sqcup by \bot . When it does not matter whether we talk about concepts or roles, we refer to the ϑ_i as *terms*. A *TBox* \mathcal{T} and an *RBox* \mathcal{R} are finite sets of concept and, respectively, role inclusions; their union $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$ is called an *ontology*.

We classify ontologies according to the form of their concept and role inclusions. Let $r, c \in \{bool, horn, krom, core\}$. We denote by DL-Lite^r_c the description logic whose ontologies contain concept and role inclusions of the form (1) satisfying the following constraints for c and r, respectively:

(horn) $m \le 1$, (krom) $k + m \le 2$, (core) $m \le 1$ and $k + m \le 2$, (bool) $k, m \ge 0$.

Note that all of our logics allow (disjointness) inclusions of the form $\vartheta_1 \sqcap \vartheta_2 \sqsubseteq \bot$.

When defining our *DL-Lite* logics, we treat concept and role inclusions in a uniform way, which is usually not the case in standard DLs. In particular, the *DL-Lite* logics [9, 1] only allow *core* role inclusions of the form $S_1 \sqsubseteq S_2$ and $S_1 \sqcap S_2 \sqsubseteq \bot$. Also, standard *DL-Lite* logics often do not admit CIs and RIs with \top on the left-hand side, which are mostly harmless and do not affect the complexity of reasoning and query answering. Modulo the latter insignificant difference, the logic *DL-Lite*^{core}_{core} is known under the names *DL-Lite*_R [9] and *DL-Lite*^H_{core} [1], where the subscript *core* refers to concept inclusions and the superscript \mathcal{H} stands for 'role hierarchies', *DL-Lite*^{core}_{horn} goes by the names *DL-Lite*_{R, \partial} [10] and *DL-Lite*^H_{horn} [1], *DL-Lite*^{core}_{krom} is *DL-Lite*^H_{krom} and *DL-Lite*^{bool} = *DL-Lite*^H_{Dool} [1]. Note also that the logic *DL-Lite*^{core}_{krom} is *DL-Lite*^P_{bool}, we disallow role inclusions of the form $\top \sqsubseteq S_1 \sqcup S_2$ and, in particular, the inclusions $\top \sqsubseteq P$ saying that P is the *universal* role. Finally, in the logic *DL-Lite*^{g-bool}_{bool}, we disallow role inclusions of the form $\top \sqsubseteq S_1 \sqcup \cdots \sqcup S_n$ for $n \ge 1$ (g stands for 'guarded'). While *DL-Lite*^{bool}_{bool} and *DL-Lite*^{g-bool}_{bool} turn out to behave radically different, for *DL-Lite*^{horn}_{horn} and *DL-Lite*^{core}_{core} the restriction to guarded RIs (i.e., excluding universal roles) would have no effect on the problems we consider.

As usual, an *ABox*, \mathcal{A} , is a finite set of assertions of the form $A_i(a_j)$ and $P_i(a_j, a_k)$. A *DL-Lite*^{*r*}_{*c*} knowledge base (KB, for short) is a pair $\mathcal{K} = (\mathcal{O}, \mathcal{A})$, where \mathcal{O} is a *DL-Lite*^{*r*}_{*c*} ontology and \mathcal{A} an ABox. *Interpretations* are also standard, taking the form $\mathcal{I} = (\Delta^{\mathcal{I}}, \mathcal{I})$, where $\Delta^{\mathcal{I}} \neq \emptyset$, $a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, $A_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and $P_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. We call \mathcal{I} a *model* of a KB \mathcal{K} and write $\mathcal{I} \models \mathcal{K}$ if all of the assertions in \mathcal{A} and inclusions in \mathcal{O} are true in \mathcal{I} . If a KB has a model, it is called *consistent* or *satisfiable*. The computational (combined) complexity of satisfiability checking for *DL-Lite*^{*r*}_{*c*} KBs is one of our concerns in this paper.

The other is the (combined and data) complexity of answering conjunctive queries (CQs) and unions of CQs (UCQs) mediated by DL-Lite^r_c ontologies. A CQ is a first-order (FO) formula of the form $q(\vec{x}) = \exists \vec{y} \varphi(\vec{x}, \vec{y})$, where φ is a conjunction of unary or binary atoms $Q(\vec{z})$ with $\vec{z} \subseteq \vec{x} \cup \vec{y}$. A UCQ is a disjunction of CQs with the same answer variables \vec{x} . A DL-Lite^r_c ontology-mediated query (OMQ) is a pair $Q = (\mathcal{O}, q(\vec{x}))$, where \mathcal{O} is a DL-Lite^r_c ontology and q a (U)CQ.

Given an ABox \mathcal{A} , we denote by $\operatorname{ind}(\mathcal{A})$ the set of individual names that occur in \mathcal{A} . A tuple \vec{a} in $\operatorname{ind}(\mathcal{A})$ is a *certain answer* to an OMQ $\mathbf{Q} = (\mathcal{O}, \mathbf{q}(\vec{x}))$ if \vec{x} and \vec{a} are of the same length and $\mathcal{I} \models \mathbf{q}(\vec{a})$, for every model \mathcal{I} of $(\mathcal{O}, \mathcal{A})$; in this case we write $\mathcal{O}, \mathcal{A} \models \mathbf{q}(\vec{a})$. If the set \vec{x} of answer variables is empty (that is, \mathbf{q} is *Boolean*), a *certain answer* to \mathbf{Q} over \mathcal{A} is 'yes' if $\mathcal{I} \models \mathbf{q}$, for every model \mathcal{I} of $(\mathcal{O}, \mathcal{A})$, and 'no' otherwise. The *OMQ* answering problem for *DL-Lite*^{*r*}_{*c*} is to check, given an OMQ $Q = (\mathcal{O}, q(\vec{x}))$ with a *DL-Lite*^{*r*}_{*c*} ontology \mathcal{O} , an ABox \mathcal{A} and a tuple \vec{a} in ind(\mathcal{A}), whether \vec{a} is a certain answer to Q over \mathcal{A} .

3 Relationship to Other Fragments of FO

Every DL-Lite^{bool}_{bool} ontology is easily seen to be equivalent to a sentence in the twovariable fragment of FO, and every DL-Lite^{g-bool}_{bool} ontology is equivalent to a sentence in the two-variable guarded fragment of FO. The converse directions do not hold. We show, however, that both languages can be equivalently captured by natural description logics with Boolean operators on roles (and no role inclusions). We give a concise description of the expressivity of both DL-Lite^{g-bool}_{bool} and DL-Lite^{g-bool}_{bool} ontologies within those more familiar description logics.

Denote by FO₂ the fragment of first-order logic with unary and binary relation symbols, the equality symbol, no function symbols, and the individual variables x and y only. FO₂ is well understood [13], in particular, it is known to have the finite model property, and the satisfiability problem for FO₂-formulas is NEXPTIME-complete. It is straightforward to show that DL- $Lite_{bool}^{bool}$ is a fragment of FO₂ in the sense that every DL- $Lite_{bool}^{bool}$ CI and RI is equivalent to a sentence in FO₂. Moreover, the translation is linear, and so the satisfiability problem for DL- $Lite_{bool}^{bool}$ KBs is in NEXPTIME. To answer the question how much of FO₂ is captured by DL- $Lite_{bool}^{bool}$, we consider another description logic in which roles have a similar status to concepts.

Denote by $\mathcal{ALC}^{id,\top,\cap,\neg,-}$ the DL whose roles are built according to the rule

$$S,S_1,S_2 \quad ::= \quad \top \quad | \quad \operatorname{id} \quad | \quad P_i \quad | \quad S_1 \sqcap S_2 \quad | \quad \neg S \quad | \quad S^-,$$

and whose concepts are built according to the rule

$$C, C_1, C_2 ::= \top \mid A_i \mid \exists S.C \mid C_1 \sqcap C_2 \mid \neg C,$$

where S is an $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ -role. An $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ -CI takes the form $C_1 \sqsubseteq C_2$, where C_1, C_2 are $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ -concepts. The semantics of $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ is as expected, with id interpreted in any interpretation \mathcal{I} as the relation $\{(d,d) \mid d \in \Delta^{\mathcal{I}}\}$. $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ was introduced in [16], various fragments had been considered before [15, 11]. It is easy to see that every $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ -CI is equivalent to a sentence in FO₂, and the translation is linear. The converse inclusion holds as well: every sentence in FO₂ is equivalent to an $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ -CI, but the known translation introduces an exponential blow-up [16]. In fact, despite having exactly the same expressive power, $\mathcal{ALC}^{\mathrm{id},\top,\cap,\neg,-}$ behaves differently from FO₂ in that satisfiability is still NExPTIMEcomplete but becomes, in contrast to FO₂, EXPTIME-complete if the signature of unary and binary relation symbols is fixed.

The relation symbols is fixed. The relationship between $\mathcal{ALC}^{id,\top,\cap,\neg,-}$ and DL- $Lite_{bool}^{bool}$ can be stated in a precise way: DL- $Lite_{bool}^{bool}$ is $\mathcal{ALC}^{id,\top,\cap,\neg,-}$ without the relation id. Observe that we clearly cannot express $\mathcal{ALC}^{id,\top,\cap,\neg,-}$ -CIs such as $A \sqsubseteq \exists \neg id.A$ in DL- $Lite_{bool}^{bool}$. Denote by $\mathcal{ALC}^{\top,\cap,\neg,-}$ the language $\mathcal{ALC}^{id,\top,\cap,\neg,-}$ without the relation id. We say that an ontology \mathcal{O} is a *model conservative extension* of an ontology \mathcal{O}' if $\mathcal{O} \models \mathcal{O}'$, the signature of \mathcal{O}' is contained in the signature of \mathcal{O} , and every model of \mathcal{O}' can be expanded to a model of \mathcal{O} by providing interpretations of the fresh symbols of \mathcal{O} and leaving the domain and the interpretation of the symbols in \mathcal{O}' unchanged.

Theorem 1. (*i*) For every DL-Lite^{bool} ontology, one can compute in linear time an equivalent $ALC^{\top,\cap,\neg,-}$ ontology.

(*ii*) For every $\mathcal{ALC}^{\top,\cap,\neg,-}$ ontology, one can compute in polynomial time a model conservative extension in DL-Lite^{bool}_{bool}.

Proof. Clearly, every CI in *DL-Lite*^{bool}_{bool} is also in $\mathcal{ALC}^{\top,\cap,\neg,-}$ ($\exists S$ abbreviates $\exists S.\top$). Every role inclusion $S_1 \sqcap \cdots \sqcap S_k \sqsubseteq S_{k+1} \sqcup \cdots \sqcup S_{k+m}$ in *DL-Lite*^{bool}_{bool} is equivalent to the $\mathcal{ALC}^{\top,\cap,\neg,-}$ -CI $\exists (S_1 \sqcap \cdots \sqcap S_k \sqcap \neg (S_{k+1} \sqcup \cdots \sqcup S_{k+m})).\top \sqsubseteq \bot$.

Conversely, assume an $\mathcal{ALC}^{\top,\cap,\neg,-}$ ontology \mathcal{O} is given. Using standard normalisation one can construct in linear time an $\mathcal{ALC}^{\top,\cap,\neg,-}$ ontology with CIs of the form

$$A \sqsubseteq \forall S.B, \quad \forall S.B \sqsubseteq A, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad A \sqsubseteq \neg B, \quad \neg A \sqsubseteq B,$$

where A, B, A_1, A_2 range over concept names and \top , which is a model conservative extension of \mathcal{O} . Now we can replace any CI $A \sqsubseteq \forall S.B$ by

$$S \sqsubseteq Q \sqcup R, \quad \exists Q^- \sqsubseteq B, \quad A \sqsubseteq \neg \exists R,$$

where Q and R are fresh role names, and any CI $\forall S.B \sqsubseteq A$ by

$$\neg A \sqsubseteq \exists R, \quad R \sqsubseteq S, \quad \exists R^- \sqsubseteq \neg B,$$

where R is a fresh role name. It remains to observe that we can replace the Boolean role S by a role name and DL-Lite^{bool}_{bool} role inclusions to obtain a model conservative extension.

The *two-variable guarded fragment* of FO, denoted GF_2 , is the fragment of FO₂ defined as follows:

- every atomic formula in FO_2 is in GF_2 ;
- GF₂ is closed under \wedge and \neg ;
- if $\vec{v} \in \{x, y, xy\}$, ψ is in GF₂, and G is in atomic formula in FO₂ containing all free variables in ψ , then $\exists \vec{v} (G \land \psi)$ is in GF₂.

It is not difficult to see that GF_2 corresponds to the fragment $\mathcal{ALC}_{guarded}^{id,\top,\cap,\neg,-}$ of the DL $\mathcal{ALC}^{id,\top,\cap,\neg,-}$ in which every role expression is either \top or contains a guard: a top-level conjunct that is either id, a role name, or the inverse of a role name. The fragment DL-Lite $_{bool}^{gbool}$ clearly lies within $\mathcal{ALC}_{guarded}^{\cap,\neg,-}$, the fragment of $\mathcal{ALC}_{guarded}^{id,\top,\cap,\neg,-}$ without the roles id and \top . In fact, similarly to the proof of Theorem 1, one can show the following:

Theorem 2. (*i*) For every DL-Lite^{g-bool}_{bool} ontology, one can compute in linear time an equivalent $\mathcal{ALC}_{guarded}^{\cap,\neg,-}$ ontology. (*ii*) For every $\mathcal{ALC}_{guarded}^{\cap,\neg,-}$ ontology, one can compute in polynomial time a model conservative extension in DL-Lite^{g-bool}_{bool}.

Complexity of Satisfiability 4

It is known [9,1] that satisfiability checking is NP-complete for DL-Lite^{core}_{bool} KBs, Pcomplete for *DL-Lite^{core}*_{horn} KBs, and NL-complete for *DL-Lite^{core}*_{krom} and *DL-Lite^{core}*_{core} KBs.

Theorem 3. Satisfiability checking is NEXPTIME-complete for DL-Lite^{bool}_{bool} KBs and EXPTIME-complete for DL-Lite^{g-bool}_{bool} KBs.

Proof. For DL-Litebool, NEXPTIME-completeness follows from Theorem 1 and [15, Theorem 14]. For DL-Lite^{g-bool}_{bool}, the EXPTIME upper bound follows from the fact that satisfiability of DL-Lite^{g-bool}_{bool} KBs is reducible to satisfiability of guarded FO-formulas with at most binary predicates, which is known to be in EXPTIME [12]. The matching lower bound can be established using the encoding of $A \sqsubseteq \forall S.B$ given in the proof of Theorem 1 and the proof of [4, Theorem 3.27].

Now we consider ontologies with Krom and Horn role inclusions. For an ontology \mathcal{O} , let $role^{\pm}(\mathcal{O}) = \{P, P^- \mid P \text{ is a role name in } \mathcal{O}\}$. We also set $(P^-)^- = P$. Suppose $\mathcal{J} = (\Delta^{\mathcal{J}}, \mathcal{J})$ is an interpretation. Denote by $t^{\mathcal{J}}(x)$ the concept type of $x \in \Delta^{\mathcal{J}}$ in \mathcal{J} , which comprises all B with $x \in B^{\mathcal{J}}$ and all $\neg B$ with $x \notin B^{\mathcal{J}}$, for basic concepts B of the form A, for concept names in \mathcal{O} , and $\exists S$, for $S \in role^{\pm}(\mathcal{O})$. Similarly, we denote by $r^{\mathcal{J}}(x,y)$ the *role type* of $(x,y) \in \Delta^{\mathcal{J}} \times \Delta^{\mathcal{J}}$ in \mathcal{J} , which comprises all S with $(x,y) \in S^{\mathcal{J}}$ and all $\neg S$ with $(x,y) \notin S^{\mathcal{J}}$, for $S \in role^{\pm}(\mathcal{O})$.

Lemma 1. For every satisfiable DL-Lite^{krom}_{bool} KB $\mathcal{K} = (\mathcal{O}, \mathcal{A})$, there exists a model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ of \mathcal{K} such that

- $-\Delta^{\mathcal{I}} = \operatorname{ind}(\mathcal{A}) \cup \{w_{S}^{i} \mid S \in \operatorname{role}^{\pm}(\mathcal{O}) \text{ and } 0 \leq i < 3\},\$
- $if a \in (\exists S)^{\mathcal{I}}, then (a, w_S^0) \in S^{\mathcal{I}}, for any a \in ind(\mathcal{A}) and any S \in role^{\pm}(\mathcal{O}),$ $if w_R^i \in (\exists S)^{\mathcal{I}}, then (w_R^i, w_S^{i\oplus 1}) \in S^{\mathcal{I}}, for any R, S \in role^{\pm}(\mathcal{O}) and 0 \le i < 3,$

where \oplus denotes addition modulo 3. In particular, DL-Lite^{krom}_{bool} enjoys the linear model property: $|\Delta^{\mathcal{I}}| = |\mathsf{ind}(\mathcal{A})| + 3|\mathsf{role}^{\pm}(\mathcal{O})|.$

Proof. Suppose *DL-Lite*^{krom}_{bool} KB $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ is satisfied in $\mathcal{J} = (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$. We construct the required model \mathcal{I} as follows. For a role literal L (that is, a role or its negation), we denote by cl(L) the set of all role literals L' such that $\mathcal{O} \models L \sqsubseteq L'$.

For each role S in \mathcal{O} with $S^{\mathcal{J}} \neq \emptyset$, we pick $w_S \in \Delta^{\mathcal{J}}$ with $\exists S^- \in t^{\mathcal{J}}(w_S)$; for S with $S^{\mathcal{J}} = \emptyset$, we pick an arbitrary w_S . Without loss of generality, we can assume that all the selected w_S are pairwise distinct. The set comprising three copies w_S^0, w_S^1, w_S^2 of all those w_S together with $ind(\mathcal{A})$ is denoted by $\Delta^{\mathcal{I}}$ (this technique with three copies is similar to the one used in the proof of [7, Proposition 8.1.4] establishing the finite model property of FO₂). Define a function f by taking f(a) = a, for all $a \in ind(\mathcal{A})$, and $f(w_S^i) = w_S$, for all S and i. We then set $t^{\mathcal{I}}(w) = t^{\mathcal{J}}(f(w))$, for all $w \in \Delta^{\mathcal{J}}$.

We now show how to define $r^{\mathcal{I}}(w, w')$ for $w, w' \in \Delta^{\mathcal{I}}$. The following three cases need consideration.

- If $w = a, w' = w_S^0$ and $\exists S \in \boldsymbol{t}^{\mathcal{J}}(w)$, then we define $\boldsymbol{r}^{\mathcal{I}}(w,w')$ by first taking cl(S). Suppose now that $\top \sqsubseteq S_j \sqcup S'_j$, for $1 \le j \le n$, are all disjunctions in \mathcal{O} such that S_j , S'_j and their negations are not in cl(S). As \mathcal{J} is a model of \mathcal{O} , for each j, either S_i or S'_i must be in $r^{\mathcal{J}}(f(w), f(w'))$. Suppose for definiteness that it is S_i . Then we add $cl(S_i)$ to $r^{\mathcal{I}}(w, w')$. Using the main property of Krom formulas (if a contradiction is derivable, then it is derivable from two literals), we can show that the resulting $r^{\mathcal{I}}(w, w')$ is consistent with \mathcal{O} and respects $t^{\mathcal{I}}(w)$ and $t^{\mathcal{I}}(w')$ in the sense that $R \in r^{\mathcal{I}}(w, w')$ implies $\exists R \in t^{\mathcal{I}}(w)$ and $\exists R^{-} \in t^{\mathcal{I}}(w')$. - If $w = w_{R}^{i}, w' = w_{S}^{i \oplus 1}$ and $\exists S \in t^{\mathcal{J}}(w)$, then we define $r^{\mathcal{I}}(w, w')$ as above. - For any other w, w' not covered above (in particular, if $w = a, w' = w_{S}^{i}$ and either
- i > 0 or $\exists S \notin t^{\mathcal{J}}(a)$), we set $r^{\mathcal{I}}(w, w') = r^{\mathcal{J}}(f(w), f(w'))$.

It is readily checked that the constructed concept types $t^{\mathcal{I}}(w)$ and role types $r^{\mathcal{I}}(w, w')$, for $w, w' \in \Delta^{\mathcal{I}}$, define a model of \mathcal{K} .

The existence of a model specified in Lemma 1 can be encoded by an essentially propositional formula $\varphi_{\mathcal{K}}$ with atomic propositions of the form $B^{\dagger}(x)$, for $x \in \Delta^{\mathcal{I}}$ and basic concepts B from \mathcal{O} , saying that B holds on x in \mathcal{I} , and $P^{\dagger}(x, x')$, for a role name P in \mathcal{O} , saying that P holds on (x, x') in \mathcal{I} . The formula $\varphi_{\mathcal{K}}$ is a conjunction of the following sentences, for all $x, x' \in \Delta^{\mathcal{I}}$:

$$\begin{split} B_1^{\dagger}(x) \wedge \cdots \wedge B_k^{\dagger}(x) &\to B_{k+1}^{\dagger}(x) \vee \cdots \vee B_{k+m}^{\dagger}(x), \\ & \text{for each CI } B_1 \sqcap \cdots \sqcap B_k \sqsubseteq B_{k+1} \sqcup \cdots \sqcup B_{k+m} \text{ in } \mathcal{O}; \\ S_1^{\dagger}(x,x') \wedge \cdots \wedge S_k^{\dagger}(x,x') &\to S_{k+1}^{\dagger}(x,x') \vee \cdots \vee S_{k+m}^{\dagger}(x,x'), \\ & \text{for each RI } S_1 \sqcap \cdots \sqcap S_k \sqsubseteq S_{k+1} \sqcup \cdots \sqcup S_{k+m} \text{ in } \mathcal{O}, \\ A^{\dagger}(a), \quad \text{for each } A(a) \in \mathcal{A}, \quad \text{and} \quad P^{\dagger}(a,b), \quad \text{for each } P(a,b) \in \mathcal{A}, \\ (\exists S)^{\dagger}(a) \to S^{\dagger}(a,w_S^0) \quad \text{and} \quad (\exists S)^{\dagger}(w_R^i) \to S^{\dagger}(x,w_S^{i\oplus 1}), \quad \text{ for each } S \text{ and } i, \\ S^{\dagger}(x,x') \to (\exists S)^{\dagger}(x), \quad \text{ for each } S. \end{split}$$

where $(P_i^-)^{\dagger}(x, x') = P_i^{\dagger}(x', x)$. Note that, if \mathcal{K} is a *DL-Lite*_{krom}^{krom} KB, then $\varphi_{\mathcal{K}}$ is a Krom formula, which can be constructed by a logspace transducer. It can now be straightforwardly shown that:

Lemma 2. A DL-Lite^{krom}_{bool} KB \mathcal{K} is satisfiable iff $\varphi_{\mathcal{K}}$ is satisfiable.

As a consequence, and using the fact that $DL-Lite_{horn}^{krom}$ can express $DL-Lite_{bool}^{krom}$ (as Krom RIs can simulate Krom CIs, and the latter can express the complements of concepts), we obtain the following complexity results:

Theorem 4. Satisfiability checking is NP-complete for DL-Lite^{krom}_{bool} and DL-Lite^{krom}_{horn} KBs, and NL-complete for DL-Lite^{krom}_{krom} KBs.

Next, we consider DL-Lite^{horn} and DL-Lite^{horn} KBs. Note that universal roles defined by RIs such as $\top \sqsubseteq S$ can be trivially omitted from KBs, so in what follows we assume that our KBs do not contain universal roles. The next lemma is proved by a more or less standard unravelling construction. We need an unravelled infinite forest-shaped model rather than a finite one similar to that in Lemma 1 in order to obtain Theorem 12 on the data complexity of OMQ answering with DL-Litehorn

Lemma 3. Let $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ be a satisfiable DL-Lite^{horn}_{bool} KB. Then there is a model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ such that

- $-\Delta^{\mathcal{I}} \subseteq \operatorname{ind}(\mathcal{A}) \cup \{aw \mid a \in \operatorname{ind}(\mathcal{A}), w \text{ is a word over } w_S, \text{ for } S \in \operatorname{role}^{\pm}(\mathcal{O})\},\$
- $t^{\mathcal{I}}(\overline{w_1}w_S) = t^{\mathcal{I}}(\overline{w_2}w_S)$, for any w_1w_S and w_2w_S in $\widetilde{\Delta}^{\mathcal{I}}$.
- if w_1 and w_2 are not both in ind(A), then $(w_1, w_2) \in R^{\mathcal{I}}$ iff there is $\exists S \in t^{\mathcal{I}}(w_1)$ such that $R \in r_S$, S is minimal with this property w.r.t. \sqsubseteq , and $w_2 = w_1 w_S$, where $r_S = \{R \mid \mathcal{O} \models S \sqsubseteq R\}$.

If \mathcal{K} is a DL-Lite^{horn} KB, then $B \in t^{\mathcal{I}}(ww_S)$ iff $\mathcal{O} \models \exists S^- \sqsubseteq B$, for any basic concept B.

Proof. First, we can show that if \mathcal{K} is satisfied in a model \mathcal{I}_0 , then it is satisfied in a model \mathcal{J} with minimal role types where, for all $w, w' \in \Delta^{\mathcal{J}}$ with $\{w, w'\} \not\subseteq \operatorname{ind}(\mathcal{A})$, we have

$$\boldsymbol{r}^{\mathcal{J}}(w,w') = \boldsymbol{r}_{S}, \text{ for some role } S.$$

Indeed, suppose (w, w') violates this property and $w' \notin \operatorname{ind}(\mathcal{A})$. Let $r^{\mathcal{I}_0}(w, w') = \{S_1, \ldots, S_k\}$. Then we replace w' by k copies w'_1, \ldots, w'_k with the same concept types as w', and connect w to each w'_i by the roles in r_{S_i} only. In the resulting interpretation, the pairs (w, w'_i) are as required, but the w'_i may not satisfy CIs in \mathcal{O} due to missing witnesses for existential concepts $\exists R$. To fix each of these w'_i , we create k-1 copies w_{ij} of w for $j \neq i$ (again, the w_{ij} belong to the same concepts as w) and connect each w_{ij} to w'_i by all roles in r_{S_j} . It can be seen that, in the resulting interpretation, the pairs (w_{ij}, w'_i) satisfy the required property, but now the w_{ij} may not satisfy CIs in \mathcal{O} (again, due to some $\exists R$). To fix these new elements, we create copies of w', and so on.

Second, for each role name P such that $P^{\mathcal{J}} \not\subseteq \operatorname{ind}(\mathcal{A}) \times \operatorname{ind}(\mathcal{A})$, we can fix two 'witnesses' by arbitrarily choosing w_P and w_{P^-} such that $\{w_P, w_{P^-}\} \not\subseteq \operatorname{ind}(\mathcal{A})$ and $r^{\mathcal{J}}(w_{P^-}, w_P) = r_P$. Now we can unravel \mathcal{J} into the required forest-shaped model \mathcal{I} of \mathcal{K} by using only witnesses of the form w_P and w_{P^-} .

Using Lemma 3, we now define translations of *DL-Lite* KBs with Horn RIs into universal sentences in FO₁ (FO-sentences with one variable), Horn-FO₁ and Krom-FO₁, the satisfiability problems for which are known to be NP-, P- and NL-complete [7].

Theorem 5. (i) DL-Lite^{horn} satisfiability is poly-time-reducible to FO_1 -satisfiability. (ii) DL-Lite^{horn} satisfiability is poly-time-reducible to Horn-FO_1-satisfiability. (iii) DL-Lite^{horn} satisfiability is poly-time-reducible to Krom-FO_1-satisfiability.

Proof. (i) Let $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ be a *DL-Lite*^{hom}_{bool} KB. We assume without loss of generality that together with RIs of the form $S_1 \sqcap \cdots \sqcap S_k \sqsubseteq S_{k+1}$ the TBox also contains $S_1^{-} \sqcap \cdots \sqcap S_k^{-} \sqsubseteq S_{k+1}^{-}$, and similarly for $S_1 \sqcap \cdots \sqcap S_k \sqsubseteq \bot$. For each concept \mathcal{A} , fix a unary predicate \mathcal{A} and, for each role name P, fix a binary predicate P, two unary predicates EP and EP^- , and two constants w_P and w_{P^-} . Intuitively, the two constants are representatives of the range and domain of the role, respectively (if it is non-empty); the unary predicates of the form \mathcal{A}, EP and EP^- encode the unary types; the binary predicate P encodes the binary types of pairs of ABox elements. For a concept C, let $C^{\dagger}(x)$ be

$$A(x)$$
 if $C = A$, $EP(x)$ if $C = \exists P$, $EP^{-}(x)$ if $C = \exists P^{-}$.

A CI $C_1 \sqcap \cdots \sqcap C_k \sqsubseteq C_{k+1} \sqcup \cdots \sqcup C_{k+m}$ is then naturally translated into

$$\forall x \left(C_1^{\dagger}(x) \wedge \dots \wedge C_k^{\dagger}(x) \to C_{k+1}^{\dagger}(x) \vee \dots \vee C_{k+m}^{\dagger}(x) \right).$$

We also include the following sentences, for all roles S:

$$\forall x \left((\exists S)^{\dagger}(x) \to (\exists S^{-})^{\dagger}(w_S) \right)$$

Since we assume that the binary types for (w_1, w_2) with $\{w_1, w_2\} \not\subseteq ind(\mathcal{A})$ are minimal, that is, generated by a single role, RIs are translated into the the following sentences:

$$\forall x \left((\exists S)^{\dagger}(x) \to (\exists R)^{\dagger}(x) \right), \quad \text{for all roles } S \text{ and } R \text{ with } \mathcal{O} \models S \sqsubseteq R; \\ \forall x \left((\exists S)^{\dagger}(x) \to \bot \right), \quad \text{for all roles } S \text{ with } \mathcal{O} \models S \sqsubseteq \bot.$$

Assertions of the form $A(a_i)$ are translated into $A(a_i)$ and of the form $P(a_i, a_j)$ into $P(a_i, a_j)$. We also add

$$P(a_i, a_j) \to (\exists P)^{\dagger}(a_i) \land (\exists P^-)^{\dagger}(a_j)$$

for all pairs $a_i, a_j \in ind(\mathcal{A})$ and all role names P, and

$$S_1(a_i, a_j) \wedge \dots \wedge S_k(a_i, a_j) \to S_{k+1}(a_i, a_j), \tag{2}$$

for all pairs $a_i, a_j \in ind(\mathcal{A})$ and RIs $S_1 \sqcap \cdots \sqcap S_k \sqsubseteq S_{k+1}$ and similarly for RIs $S_1 \sqcap \cdots \sqcap S_k \sqsubseteq \bot$, with \bot in the conclusion, where $P^-(a_i, a_j)$ is a shortcut for $P(a_j, a_i)$. It can easily be seen that \mathcal{K} is satisfiable iff the translation is consistent.

(*ii*) We observe that the translation above is in Horn-FO₁ if \mathcal{K} is in *DL-Lite*^{horn}_{horn}.

(*iii*) The translation in (*i*) for *DL-Lite*^{horn}_{krom} is not in Krom-FO₁. However, since the translation works in polynomial time, we can modify it so that instead of (2) it produces all $S(a_i, a_j)$ such that $\mathcal{R}, \mathcal{A} \models S(a_i, a_j)$ or \bot if $\mathcal{R}, \mathcal{A} \models \bot$, where \mathcal{R} is the RBox of \mathcal{O} .

As a consequence, we obtain the following complexity results:

Theorem 6. Satisfiability checking is NP-complete for DL-Lite^{horn}_{bool} KBs, and P-complete for DL-Lite^{horn}_{horn} and DL-Lite^{horn}_{krom} KBs.

5 The Complexity of (U)CQ-Answering

In this section, we give a brief survey of results on the combined and data complexity of answering OMQs with DL-Lite^r_c ontologies and (U)CQs, which can mostly be obtained in a rather straightforward way from existing results. We start with undecidability results, which show that the language DL-Lite^{bool}_{bool} behaves again similarly to the twovariable fragment of FO and that even in DL-Lite^{krom}_{krom} UCQ answering is undecidable. The following can be shown using [17, Theorem 1] and the encoding of $\forall R.C$ given in the proof of Theorem 1:

Theorem 7. Answering OMQs with DL-Lite^{bool} ontologies and CQs is undecidable for combined complexity.

The next result follows from [17, Theorem 3]:

Theorem 8. Answering OMQs with DL-Lite^{krom} $_{krom}$ ontologies and UCQs is undecidable for combined complexity.

It remains open whether answering OMQs with DL- $Lite_{krom}^{krom}$ ontologies and CQs is undecidable. The language DL- $Lite_{bool}^{g-bool}$ can be regarded as a fragment of the guarded fragment, GF, of FO. Query evaluation for GF was first investigated in [5], where it was proved that (U)CQ evaluation in GF is in 2EXPTIME for combined complexity. It follows directly from the results in [8] that a matching lower bound holds already for DL- $Lite_{bool}^{core}$. Thus, we obtain the following:

Theorem 9. Answering OMQs with DL-Lite^{g-bool} ontologies and (U)CQs can be done in 2EXPTIME for combined complexity. It is 2EXPTIME-hard for DL-Lite^{core}_{bool} ontologies.

We now sketch the results on data complexity. A coNP-upper bound for (U)CQ evaluation in GF for data complexity is proved in [5]. Thus, UCQ answering is in CONP for DL-Lite^{g-bool}. A fine-grained analysis of the data complexity at the TBox and even OMQ level can be obtained from [6, 14]. While the structure of the space of possible complexity of answering OMQs over GF ontologies and UCQs is wide open (it corresponds to the complexity of MMSNP₂ [6]) and an active area of research in the constraint satisfaction community, it has recently been proved in the extended version of [14] that there is a P/CONP dichotomy for OMQs over the *two-variable* fragment GF₂ of GF and UCQs (i.e., every such OMQ is either in P or CONP-complete). As DL-Lite^{g-bool} is clearly contained in GF₂, we obtain the following:

Theorem 10. Answering any OMQ with a DL-Lite^{*g*-bool}_{bool} ontology and a UCQ is either in P or CONP-complete for data complexity.

We conjecture that important properties such as datalog rewritability and first-order rewritability of such OMQs are decidable as well, but this remains open. In many applications of ontology-mediated query answering, only the TBox is known in advance, but not the relevant queries. In this case, investigating the complexity at the level of OMQs is not appropriate, but instead one is interested in an upper bound for the complexity of query evaluation for *all* OMQs based on the given TBox. As *DL-Lite*^{*g*-bool} lies within the fragment $uGF_2^-(1, =)$ of GF introduced in [14], we obtain the following analysis from that paper:

Theorem 11. For every ontology \mathcal{O} in DL-Lite^{*g*-bool}_{bool}, either every OMQ using \mathcal{O} and a UCQ is datalog-rewritable or there exists such an OMQ for which query answering is CONP-hard. Moreover, datalog-rewritability is decidable in NEXPTIME.

We note that a dichotomy between datalog-rewritable and CONP is quite rare. There are, for example, ALC TBoxes for which not all OMQs are datalog rewritable but still in P [14]. Finally, the following (probably folklore) theorem is proved using Lemma 3:

Theorem 12. Every OMQ with a DL-Lite^{horn}_{horn} ontology and a UCQ is FO-rewritable, and so answering it is in AC^0 for data complexity.

Acknowledgements

The work was supported by the EPSRC UK grants EP/S032282/1 and EP/S032207/1.

References

- Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The DL-Lite family and relations. Journal of Artificial Intelligence Research (JAIR) 36, 1–69 (2009)
- Artale, A., Kontchakov, R., Kovtunova, A., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: First-order rewritability of temporal ontology-mediated queries. In: Yang, Q., Wooldridge, M. (eds.) Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 2706–2712. AAAI Press (2015), http://ijcai.org/Abstract/15/383
- Artale, A., Kontchakov, R., Wolter, F., Zakharyaschev, M.: Temporal description logic for ontology-based data access. In: Rossi, F. (ed.) IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013. pp. 711–717. IJCAI/AAAI (2013), http://www.aaai.org/ocs/index.php/IJCAI/ IJCAI13/paper/view/6824
- Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press (2003)
- Bárány, V., Gottlob, G., Otto, M.: Querying the guarded fragment. Logical Methods in Computer Science 10(2) (2014), https://doi.org/10.2168/LMCS-10 (2:3) 2014
- Bienvenu, M., ten Cate, B., Lutz, C., Wolter, F.: Ontology-based data access: A study through disjunctive datalog, CSP, and MMSNP. ACM Transactions on Database Systems 39(4), 33:1–44 (2014)
- Börger, E., Grädel, E., Gurevich, Y.: The Classical Decision Problem. Perspectives in Mathematical Logic, Springer (1997)
- Bourhis, P., Manna, M., Morak, M., Pieris, A.: Guarded-based disjunctive tuple-generating dependencies. ACM Trans. Database Syst. 41(4), 27:1–27:45 (2016), https://doi. org/10.1145/2976736
- Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: the *DL-Lite* family. Journal of Automated Reasoning 39(3), 385–429 (2007)
- Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Data complexity of query answering in description logics. In: Doherty, P., Mylopoulos, J., Welty, C.A. (eds.) Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006. pp. 260–270. AAAI Press (2006), http://www.aaai.org/Library/KR/2006/kr06-028.php
- Gargov, G., Passy, S.: A note on boolean modal logic. In: Petkov, P. (ed.) Mathematical Logic. Springer, Boston, MA (1990)
- Grädel, E.: On the restraining power of guards. J. Symb. Log. 64(4), 1719–1742 (1999), https://doi.org/10.2307/2586808
- 13. Grädel, E., Kolaitis, P.G., Vardi, M.Y.: On the decision problem for two-variable first-order logic. Bulletin of Symbolic Logic 3(1), 53–69 (1997), http://www.math.ucla.edu/ \%7Easl/bsl/0301/0301-003.ps
- Hernich, A., Lutz, C., Papacchini, F., Wolter, F.: Dichotomies in ontology-mediated querying with the guarded fragment. In: Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, 2017. pp. 185–199 (2017), https://doi.org/10.1145/3034786.3056108

- Lutz, C., Sattler, U.: The complexity of reasoning with boolean modal logics. In: Wolter, F., Wansing, H., de Rijke, M., Zakharyaschev, M. (eds.) Advances in Modal Logic 3, papers from the third conference on "Advances in Modal logic," held in Leipzig, Germany, 4-7 October 2000. pp. 329–348. World Scientific (2000)
- Lutz, C., Sattler, U., Wolter, F.: Modal logic and the two-variable fragment. In: Fribourg, L. (ed.) Computer Science Logic, 15th International Workshop, CSL 2001. 10th Annual Conference of the EACSL, Paris, France, September 10-13, 2001, Proceedings. Lecture Notes in Computer Science, vol. 2142, pp. 247–261. Springer (2001), https://doi.org/10. 1007/3-540-44802-0_18
- Rosati, R.: The limits of querying ontologies. In: Database Theory ICDT 2007, 11th International Conference, Barcelona, Spain, January 10-12, 2007, Proceedings. pp. 164–178 (2007), https://doi.org/10.1007/11965893_12