

Number Restrictions on Transitive Roles in Description Logics with Nominals

Víctor Gutiérrez-Basulto

Cardiff University, UK
GutierrezBasultoV@cardiff.ac.uk

Yazmín Ibáñez-García

TU Wien, Austria
yazmin.garcia@tuwien.ac.at

Jean Christoph Jung

Universität Bremen, Germany
jeanjung@informatik.uni-bremen.de

Abstract

We study description logics (DLs) supporting number restrictions on transitive roles. We first take a look at SOQ and SON with binary and unary coding of numbers, and provide algorithms for the satisfiability problem and tight complexity bounds ranging from EXPTIME to NEXPTIME. We then show that by allowing for counting only up to one (functionality), inverse roles and role inclusions can be added without losing decidability. We finally investigate DLs of the *DL-Lite*-family, and show that, in the presence of role inclusions, the *core* fragment becomes undecidable.

1 Introduction

Description Logics (DLs) are a successful family of logic-based knowledge representation formalisms. The relevance of DLs comes from the fact that they are arguably the most popular language for the formulation of ontologies. For instance, they provide the logical basis of the web ontology language OWL 2, the medical ontology SNOMED CT, and the NCI thesaurus. One of the main reasons for the take-up of DLs is that, in general, they provide a good trade-off between expressivity and computational complexity. Unfortunately, in some cases this is not easy to ensure, e.g., the unrestricted interaction of (*qualified*) *number restrictions* and *transitive roles* tends to destroy this good balance; in many cases, leading to undecidability. On the other hand, support of these features is required, e.g., for ontological modeling in the biomedical domain (Rector and Rogers 2006; Kazakov, Sattler, and Zolin 2007; Stevens et al. 2007). For instance, in the classification of proteins (Wolstencroft et al. 2005), certain classes of proteins are defined in terms of their composition: *If a protein contains at least n_1 X_1 -components ... and at least n_k X_k components, then it belongs to class B* . Moreover, these definitions require modeling of parthood, which is intended to be a *transitive relation*. Hence there is need of a clear understanding of the decidability frontier for DLs supporting these features.

With this in mind, in the last 15 years, the DL community has developed a vast amount of research on the complexity of reasoning in the presence of transitive roles and number restrictions, see, (Horrocks, Sattler, and Tobies 2000; Kazakov, Sattler, and Zolin 2007; Schröder and Pattinson 2008;

Kaminski and Smolka 2010) and references therein. In particular, it has been shown that the extensions of \mathcal{SN} (\mathcal{ALC} enriched with transitive roles and *unqualified* counting) with role inclusions (\mathcal{SHN}) or inverse roles (\mathcal{SLN}) are undecidable (Horrocks, Sattler, and Tobies 2000; Kazakov, Sattler, and Zolin 2007). These negative results are, intuitively, explained by the interaction of these two constructors, i.e., by the possibility of counting over transitive roles. In order to regain decidability, different restrictions on their interaction have been proposed, e.g., to completely disallow number restrictions on transitive roles, or impose certain restrictions on the transitive roles occurring in role inclusions (Horrocks, Sattler, and Tobies 2000; Kazakov, Sattler, and Zolin 2007). On the positive side, it was shown that if role inclusions and inverse roles are not present, as in \mathcal{SN} , \mathcal{SQ} , \mathcal{SOQ} , decidability is then regained (Kazakov, Sattler, and Zolin 2007; Kaminski and Smolka 2010). However, no (elementary) complexity bounds are obtained from these results. Interestingly, if we have inverse roles or role inclusions, but only *functionality* (counting up to one) is allowed the panorama is less clear. The only known result is that satisfiability relative to \mathcal{SLF} -TBoxes is decidable in 2EXPTIME (Tendera 2005), but decidability of, e.g., \mathcal{SHLF} , \mathcal{SHOLF} remains an open problem.

The main contribution of this paper is to establish a complete picture of the complexity of the problem of concept satisfiability relative to TBoxes in DLs supporting counting over transitive roles, by resolving the aforementioned open problems. Moreover, for all considered DLs including nominals, our upper bound results transfer to knowledge base satisfiability.

Our investigation starts (Section 3) with the DL \mathcal{SOQ} , allowing for qualified counting and nominals. As mentioned above, decidability was shown by Kaminski and Smolka [2010], and NEXPTIME-hardness is inherited from graded modal logic (Kazakov and Pratt-Hartmann 2009).¹ However, the exact computational complexity of \mathcal{SOQ} was unknown. We close here this gap, by providing a NEXPTIME upper bound. To this aim, we use a two-step approach. First, we provide a decomposition of \mathcal{SOQ} models, permitting us to ‘independently reason’ about the different (transitive) roles. In a second step, carefully adapting a tech-

¹Graded modalities correspond to qualified number restrictions.

nique developed by Kazakov and Pratt-Hartmann [2009] in the context of graded modal logic, we show a small (that is, exponential) model property of each member of the decomposition, which lifts to \mathcal{SOQ} and thus leads to the desired NEXPTIME upper bound.

As the next step (Section 4), we turn our attention to \mathcal{SON} , the restriction of \mathcal{SOQ} to unqualified number restrictions. In particular, we are interested in understanding the impact of the coding of numbers on the computational complexity. We first show that with unary coding, satisfiability in \mathcal{SON} is EXPTIME-complete, and therefore easier than in \mathcal{SOQ} . We devise a type-elimination procedure that exploits the unary coding by the observation that certain witnesses are of only polynomial size, and can thus be all enumerated. We then show that with binary coding, the complexity of satisfiability jumps to NEXPTIME-complete. In fact, the lower bound holds already for concept satisfiability of \mathcal{SN} concepts over a single transitive role (no TBox). It is interesting to note that when only *non-transitive* roles are allowed in number restrictions the coding has no impact on the computational complexity, that is, regardless of the coding of numbers, satisfiability in \mathcal{SON} and \mathcal{SOQ} is EXPTIME-complete (Calvanese, Eiter, and Ortiz 2009).

We then take a look (Section 5) at the case when only functionality is allowed. We show that in this case inverse roles, role inclusions and nominals can be added without losing decidability. In particular, we show that satisfiability in \mathcal{SHLF} and \mathcal{SHOLF} is EXPTIME- and NEXPTIME-complete, respectively, and hence not harder than when one cannot impose number restrictions on transitive roles.

Lightweight DLs of the *DL-Lite* family allowing for number restrictions on transitive roles have not been considered yet; indeed, only counting over non-transitive roles has been studied in *DL-Lite* (Artale et al. 2009). In the last part of the paper (Section 6), we initiate this study. In particular, we complement known undecidability results by considering a light sub-Boolean DL with unqualified existential restrictions and show that the *core* fragment of *DL-Lite*, with role inclusions, allowing for number restrictions on transitive roles is undecidable.

Missing proofs are available at www.informatik.uni-bremen.de/tdki/research/papers/GIJ17.pdf.

2 Preliminaries

Syntax. We introduce the DL \mathcal{SHOIQ} (Hollunder and Baader 1991), which extends the classical DL \mathcal{ALC} with transitivity declarations on roles (\mathcal{S}), role inclusion axioms (\mathcal{H}), nominals (\mathcal{O}), inverses (\mathcal{I}), and qualified number restrictions (\mathcal{Q}). We consider a vocabulary consisting of countably infinite disjoint sets of *concept names* N_C , *role names* N_R and *individual names* N_I . The syntax of \mathcal{SHOIQ} -concepts C, D is given by the following grammar:

$$C, D ::= A \mid \neg C \mid C \sqcap D \mid \{o\} \mid \exists r.C \mid (\sim n r C)$$

where $A \in N_C$, $o \in N_I$, $r \in \{s, s^- \mid s \in N_R\}$ is a *role*, \sim is a comparison operator \leq or \geq , and n is a number (given in binary, unless stated otherwise). Roles of the form r^- are called *inverse roles*, concepts of the form

$\{o\}$, $\exists r.C$, $(\leq n r C)$, $(\geq n r C)$ are called, respectively, *nominals*, *existential restrictions*, *at most-restrictions* and *at least-restrictions*. We identify r^- with $s \in N_R$ if $r = s^-$, and use standard abbreviations \top , \perp , $C \sqcup D$, $\forall r.C$, and $C \rightarrow D$.

A \mathcal{SHOIQ} -TBox (*ontology*) \mathcal{T} is a finite set of *concept inclusions* (CIs) $C \sqsubseteq D$, *transitivity declarations* $\text{Tra}(r)$ and *role inclusions* (RIs) $r \sqsubseteq s$, where C, D are \mathcal{SHOIQ} -concepts and r, s roles. We use $\text{CN}(\mathcal{T})$, $\text{Rol}(\mathcal{T})$ and $\text{Nom}(\mathcal{T})$ to denote, respectively, the set of *all concept names, roles and nominals occurring in \mathcal{T}* . Wlog. we assume that if $\text{Tra}(r) \in \mathcal{T}$ then $\text{Tra}(r^-) \in \mathcal{T}$. Indeed, by the semantics, if a role is transitive, so is its inverse.

Semantics. As usual, the semantics is defined in terms of interpretations. An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty *domain* $\Delta^{\mathcal{I}}$ and an *interpretation function* $\cdot^{\mathcal{I}}$ mapping concept names to subsets of the domain and role names to binary relations over the domain. We define, mutually recursive, the set $r_{\mathcal{I}}(d, C) = \{e \in C^{\mathcal{I}} \mid (d, e) \in r^{\mathcal{I}}\}$ of r -successors of d satisfying C , and the interpretation of complex concepts $C^{\mathcal{I}}$ by taking

$$\begin{aligned} (r^-)^{\mathcal{I}} &= \{(e, d) \mid (d, e) \in r^{\mathcal{I}}\}; \\ (\neg C)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}; \\ (C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}}; \\ \{o\}^{\mathcal{I}} &= \{o^{\mathcal{I}}\}; \\ (\exists r.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} \mid \exists e \in C^{\mathcal{I}} \text{ with } (d, e) \in r^{\mathcal{I}}\}; \\ (\sim n r C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} \mid |r_{\mathcal{I}}(d, C)| \sim n\}. \end{aligned}$$

The satisfaction relation \models is defined standardly:

$$\begin{aligned} \mathcal{I} \models C \sqsubseteq D &\text{ iff } C^{\mathcal{I}} \subseteq D^{\mathcal{I}}; \\ \mathcal{I} \models r \sqsubseteq s &\text{ iff } r^{\mathcal{I}} \subseteq s^{\mathcal{I}}; \\ \mathcal{I} \models \text{Tra}(r) &\text{ iff } r^{\mathcal{I}} \text{ is transitive.} \end{aligned}$$

An interpretation \mathcal{I} is a *model* of a TBox \mathcal{T} , denoted $\mathcal{I} \models \mathcal{T}$, if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{T}$. A concept C is *satisfiable relative to a TBox \mathcal{T}* if there is a model \mathcal{I} of \mathcal{T} such that $C^{\mathcal{I}} \neq \emptyset$.

Reasoning Problem. We are interested in the problem of *concept satisfiability*, that is, given a TBox \mathcal{T} and a concept C , we want to determine whether C is satisfiable relative to \mathcal{T} . We restrict our attention to the case when $C = A \in N_C$ because C is satisfiable relative to \mathcal{T} iff A_C is satisfiable relative to $\mathcal{T} \cup \{A_C \sqsubseteq C\}$ for any fresh concept name A_C .

Please note that in the presence of nominals our upper bounds transfer to the problem of *knowledge base satisfiability*; indeed, so-called *ABox assertions* can be internalized in the TBox using nominals (Baader et al. 2003).

Fragments. We consider the following fragments:

- \mathcal{SOQ} is obtained from \mathcal{SHOIQ} by disallowing role inclusions and inverse roles.
- \mathcal{SON} is obtained from \mathcal{SOQ} by supporting only *unqualified* number restrictions (indicated by letter \mathcal{N}) of the form $(\sim n r \top)$, which we usually abbreviate as $(\sim n r)$.
- \mathcal{SHOLF} is obtained from \mathcal{SHOIQ} by supporting only *local functionality* constraints (indicated by letter \mathcal{F}) of the form $(\leq 1 r)$.

3 SOQ

We start by devising an algorithm for concept satisfiability relative to \mathcal{SOQ} -TBoxes, yielding a tight NEXPTIME upper bound. The matching lower bound follows from the fact that satisfiability in the graded modal logic over transitive frames, **GrK4**, is NEXPTIME-complete (Kazakov and Pratt-Hartmann 2009).

We assume that the input TBox is in the following normal form. Let $\mathcal{C}_{\text{Bool}}$ be the set of \mathcal{SOQ} -concept descriptions that are obtained without using the constructors $(\sim n r C)$ and $\exists r.C$ (we treat $\exists r.C$ as $(\geq 1 r C)$). We say that a TBox is in normal form if all concept inclusions are of the shape

$$C \sqsubseteq D \quad \text{or} \quad C \sqsubseteq (\sim n r D),$$

for $C, D \in \mathcal{C}_{\text{Bool}}$. We show that, by introducing fresh concept names, every TBox can be transformed in polynomial time into a satisfiability-equivalent TBox in normal form.

In order to obtain the desired NEXPTIME upper bound it clearly suffices to show a *small model property*, that is, whenever A is satisfiable relative to \mathcal{T} , then there is a model of exponential size, since we can then simply “guess” the model. To this end, we will first characterize concept satisfiability in terms of the existence of a *quasimodel*, which is a decomposition of a model of \mathcal{SOQ} -TBoxes into components that interpret only a single role name. This is in line with viewing \mathcal{SOQ} as a fusion logic. Note that decompositions of fusion logics have been studied (Baader et al. 2002), but so far nominals were not considered. Nominals impose the additional difficulty that the models of a \mathcal{SOQ} -TBox are not closed under union.

Let \mathcal{T} be the input TBox. For $r \in \mathbb{N}_{\mathbb{R}}$, we define $\mathcal{T}_r := \mathcal{T} \setminus \{C \sqsubseteq (\sim n r' D) \mid r' \neq r\}$. Intuitively, \mathcal{T}_r reflects \mathcal{T} on the single role r . Given two interpretations \mathcal{I}, \mathcal{J} , we say that $d \in \Delta^{\mathcal{I}}, d' \in \Delta^{\mathcal{J}}$ are *Boolean equivalent* iff for all $C \in \mathcal{C}_{\text{Bool}}$ we have $d \in C^{\mathcal{I}}$ iff $d' \in C^{\mathcal{J}}$.

We are now in a position to define the intended decomposition. A *quasimodel* for \mathcal{T} is a finite collection of interpretations $\Omega = \{\mathcal{I}_r \mid r \in \text{Rol}(\mathcal{T})\}$ such that the following two conditions are satisfied:

(qm1) $\mathcal{I}_r \models \mathcal{T}_r$ for each role name $r \in \text{Rol}(\mathcal{T})$;

(qm2) for all role names r, s and $d \in \Delta^{\mathcal{I}_r}$ there exists a $d' \in \Delta^{\mathcal{I}_s}$ such that d and d' are Boolean equivalent.

Intuitively, **(qm1)** captures the TBox relative to a single role name r , and **(qm2)** ensures that the components \mathcal{I}_r can be combined into one model. A *quasimodel* for A and \mathcal{T} is a quasimodel Ω for \mathcal{T} such that $A^{\mathcal{I}_r} \neq \emptyset$ for some (equivalently: every) interpretation $\mathcal{I}_r \in \Omega$. Note that, by **(qm1)**, we can assume that each \mathcal{I}_r interprets only role r as (possibly) non-empty. Thus, quasimodels provide a suitable decomposition of models. The *size* of a quasimodel is the sum of the domain sizes of all interpretations in the quasimodel.

The following lemma provides the characterization of satisfiability and additionally relates the size of the quasimodel to the size of a model.

Lemma 1. *A is satisfiable relative to \mathcal{T} iff there is a quasimodel for A and \mathcal{T} . Moreover, if there is a quasimodel of size κ for A and \mathcal{T} , there is a model of size $\leq \kappa$ for A and \mathcal{T} .*

It remains to restrict the sizes of quasimodels.

Lemma 2. *If there is a quasimodel for A and \mathcal{T} , there is a quasimodel for A and \mathcal{T} of exponential size.*

We give some intuitions on the proof here. Let Ω be a quasimodel for A and \mathcal{T} and $\mathcal{I}_r \in \Omega$. If r is a non-transitive role, it has been already shown that \mathcal{I}_r can be replaced by an exponentially sized interpretation \mathcal{I}'_r , preserving **(qm2)** (Lutz et al. 2005, Corollary 4.3). Therefore, we concentrate on the case when r is a transitive role, that is, $\text{Tra}(r) \in \mathcal{T}$.

First observe that we can assume that \mathcal{I}_r has at most exponentially many connected components, more precisely, $2^{|X|}$, where $X = \text{CN}(\mathcal{T}) \cup \text{Nom}(\mathcal{T})$. To see this, fix for every subset $Y \subseteq X$ a domain element d_Y such that $Y = \{C \in X \mid d_Y \in C^{\mathcal{I}_r}\}$, if such an element exists. It should be clear that the restriction \mathcal{I}' of \mathcal{I}_r to domain

$$\Delta^{\mathcal{I}'} = \{d \in \Delta^{\mathcal{I}_r} \mid \exists d_Y : (d_Y, d) \in (r^{\mathcal{I}_r})^*\}$$

still satisfies **(qm1)** and **(qm2)**, and additionally has at most $2^{|X|}$ connected components (each rooted at some d_Y).

Finally, to show Lemma 2, we carefully adapt a technique by Kazakov and Pratt-Hartmann [2009] showing the finite model property of **GrK4**, to prove that every connected component of \mathcal{I}_r can be assumed to be of exponential size. Crucially, we have to take care that, when domain elements are removed, we keep the witnesses d_Y for Condition **(qm2)**. See the appendix for a full proof.

Lemma 1 and 2 yield the small (that is, exponential) model property for \mathcal{SOQ} which, as argued above implies:

Theorem 1. *Concept satisfiability relative to \mathcal{SOQ} -TBoxes is NEXPTIME-complete.*

4 SON

In this section, we study the complexity of concept satisfiability relative to \mathcal{SON} -TBoxes, with both unary and binary coding of numbers. Note that for \mathcal{SOQ} the coding of numbers does not make any difference on the computational complexity because the NEXPTIME-hardness proof for satisfiability in **GrK4** only uses numbers that are at most 1.

We first show that with unary coding, concept satisfiability relative to \mathcal{SON} -TBoxes is EXPTIME-complete, and thus easier than in \mathcal{SOQ} . We then show that with binary coding, the complexity of concept satisfiability relative to \mathcal{SON} -TBoxes coincides with that relative to \mathcal{SOQ} -TBoxes. In particular, we show that the latter holds already for \mathcal{SN} .

4.1 Unary Coding of Numbers

We focus on providing an EXPTIME algorithm for concept satisfiability relative to \mathcal{SON} -TBoxes with unary coding of numbers. The lower bound is inherited from \mathcal{ACC} .

We proceed in two steps. First, we give a characterization of concept satisfiability, independent of the coding of numbers. This characterization is then the basis for a type elimination procedure which runs in exponential time, given the unary coding. The main challenge lies in the interplay between nominals and transitive roles. In fact, the algorithm is not purely type-based, but needs to make explicit what we

call the *nominal core* of an interpretation, which is the part of the model ‘close’ to the nominals.

Denote with $\text{cl}(\mathcal{T})$ the set of all sub-concepts appearing in \mathcal{T} , closed under single negations. A *type for* \mathcal{T} is a set $t \subseteq \text{cl}(\mathcal{T})$ satisfying

- $D \in t$ iff $\neg D \notin t$ for all $\neg D \in \text{cl}(\mathcal{T})$;
- $D \sqcap E \in t$ iff $\{D, E\} \subseteq t$, for all $D \sqcap E \in \text{cl}(\mathcal{T})$;
- $C \in t$ implies $D \in t$, for all $C \sqsubseteq D \in \mathcal{T}$.

Let $\text{tp}(\mathcal{T})$ be the set of all types for \mathcal{T} . Two types $t, t' \in \text{tp}(\mathcal{T})$ are called *r-compatible*, written $t \rightsquigarrow_r t'$, if

- $\{\neg D \mid \neg \exists r.D \in t\} \subseteq t'$, in case r is non-transitive, and
- $\{\neg D, \neg \exists r.D \mid \neg \exists r.D \in t\} \subseteq t'$, in case r is transitive.

Fix a role r and a type t , and let ℓ be maximal with $(\geq \ell r) \in t$, and u be minimal with $(\leq u r) \in t$.² Then, t is called *r-realizable in* $T \subseteq \text{tp}(\mathcal{T})$ if $\ell \leq u$ and there are $k \leq u$ types $t_1, \dots, t_k \in T$ with $t \rightsquigarrow_r t_i$, for all i , such that:

1. for each $\exists r.C \in t$, there is some i with $C \in t_i$;
2. if $k < \ell$, there is an i with $\{o\} \notin t_i$ for all $\{o\} \in \text{cl}(\mathcal{T})$.

Item 1 states the known realizability condition for \mathcal{ALC} . Item 2 captures the interplay of nominals and *at-least* restrictions; in particular, if $k < \ell$, we need one type that can be repeated as a successor, which cannot be a nominal type.

For transitive roles in combination with *at-most* restrictions $(\leq n r)$ and nominals, we need to make explicit how the at-most restrictions in a type are realized. In order to formalize this, denote with $\text{tp}_{\mathcal{I}}(d)$ the type $\{C \in \text{cl}(\mathcal{T}) \mid d \in C^{\mathcal{I}}\}$ of d in an interpretation \mathcal{I} , and say that $t =_r t'$ if $C \in t$ iff $C \in t'$ for all concepts $C \in \text{cl}(\mathcal{T})$ not of the form $(\sim n s)$, $\neg(\sim n s)$, $\exists s.A$, $\neg \exists s.A$, for $s \neq r$. Moreover, given a role name r and sets $T' \subseteq T \subseteq \text{tp}(\mathcal{T})$, we say that an interpretation \mathcal{I} is a *\leq -witness for* (r, T, T') if

- (i) for each $d \in \Delta^{\mathcal{I}}$, there is $t \in T$ with $\text{tp}_{\mathcal{I}}(d) =_r t$, and
- (ii) for every $t \in T'$ such that $(\leq n r) \in t$, there is some d with $\text{tp}_{\mathcal{I}}(d) =_r t$.

Intuitively, \mathcal{I} realizes (relative to r) only types from T , but at least those in T' . Using the notion of \leq -witness we give the following characterization of concept satisfiability, which also provides the starting point of our decision procedure.

Lemma 3. *A is satisfiable relative to \mathcal{T} iff there is a set $T \subseteq \text{tp}(\mathcal{T})$ with $A \in t$ for some $t \in T$ such that:*

- (E1) *for any $\{o\}$ in \mathcal{T} there is exactly one $t \in T$ with $\{o\} \in t$;*
- (E2) *every $t \in T$ is r -realizable in T , for each role r ;*
- (E3) *there is a \leq -witness for (r, T, T) , for any transitive r .*

Conditions (E1) and (E2) are straightforward; Condition (E3) requires, for each transitive role r , an interpretation realizing all types with an at-most restriction. Though condition (E3) is intuitive, it does not lend itself for implementation yet because \leq -witnesses are exponentially big in general.

²By convention, $\ell = 0$ and $u = \infty$, respectively, if no such concepts are in t .

As the next step, we analyze the \leq -witnesses and give an equivalent condition (E3'). Fix a role name r . The *nominal core of an interpretation \mathcal{I} wrt. r* , written $\text{core}_r(\mathcal{I})$, is obtained from \mathcal{I} by restricting the domain to

$$\{o^{\mathcal{I}} \mid \{o\} \in \text{cl}(\mathcal{T})\} \cup \{d \mid (o^{\mathcal{I}}, d) \in r^{\mathcal{I}}, o^{\mathcal{I}} \in (\leq m r)^{\mathcal{I}}, (\leq m r) \in \text{cl}(\mathcal{T})\}$$

We prove that Lemma 3 remains correct if we use the following instead of (E3):

- (E3') *for each transitive r , there is an interpretation \mathcal{I}_r such that, for each $t \in T$ with $(\leq n r) \in t$, there is a \leq -witness \mathcal{I}_{rt} for $(r, T, \{t\})$ with $\text{core}_r(\mathcal{I}_{rt}) = \mathcal{I}_r$.*

Before we give the algorithm, observe that there are only exponentially many *maximal* sets $T \subseteq \text{tp}(\mathcal{T})$ satisfying (E1), that is, there is no set T' satisfying (E1) and $T \subsetneq T' \subseteq \text{cl}(\mathcal{T})$. Moreover, it is crucial to observe that the (domain) size of the nominal core of a \leq -witness (and in fact of any interpretation) is *polynomial in the size of \mathcal{T}* ; more precisely, its size is bounded by $\ell_1 \ell_2$, where ℓ_1 is the number of nominals in \mathcal{T} , and ℓ_2 is the largest number in \mathcal{T} ; so the \mathcal{I}_r in (E3') has only polynomial size. Finally, given such a (polynomial) \mathcal{I}_r and some t with $(\leq n r) \in t$, we can check in exponential time the existence of \mathcal{I}_{rt} , because we can just try all possible extensions of \mathcal{I}_r with n elements.

These arguments show that the following procedure runs in exponential time. For each maximal set $T \subseteq \text{cl}(\mathcal{T})$ satisfying (E1) and each possible combination of nominal cores (one for each role name), exhaustively remove types from T if they do not satisfy (E2) or (E3'). Accept if, in this way, a set \hat{T} is found which satisfies all conditions and there is $t \in \hat{T}$ with $A \in t$. Overall, this shows:

Theorem 2. *Concept satisfiability relative to \mathcal{SON} -TBoxes with unary coding of numbers is EXPTIME-complete.*

4.2 Binary Coding of Numbers

Now, we show that with binary coding, concept satisfiability relative to \mathcal{SON} -TBoxes becomes NEXPTIME-hard. The matching upper bound follows from Theorem 1 above. Note that the lower bound does not follow from (the proof of) NEXPTIME-hardness of satisfiability in **GrK4** because that relies on qualified number restrictions.

The NEXPTIME-hardness proof is by reduction of the problem of tiling a torus of exponential size (van Emde Boas 1997). For the reduction to work nominals are not required, that is, the lower bound already holds for \mathcal{SN} . Intuitively, in the reduction, we cope with the lack of qualified number restrictions by exploiting the fact that ‘big numbers’ can be used due to the binary coding.

We concentrate here on the most interesting part, the construction of an \mathcal{SN} -TBox \mathcal{T}_{tor} and a concept L_0 whose satisfiability characterize the $2^n \times 2^n$ -torus. To this aim, we use the following signature:

- concept names $X_0, \dots, X_{n-1}, Y_0, \dots, Y_{n-1}$ that serve to encode the (x, y) coordinates in the torus
- concept names L_0, \dots, L_{2n} that mark the levels of a binary tree,

- a transitive role r .

We start by enforcing that certain models of \mathcal{T}_{tor} contain a complete binary tree, called *torus-tree*, with $2n$ levels, where the 2^{2n} leaves of the torus-tree will represent the $2^n \times 2^n$ points in the torus. To this end, we include the transitivity statement $\text{Tra}(r)$ and the following CIs, for $0 \leq i < n$ in \mathcal{T}_{tor} :

$$\begin{aligned} L_i &\sqsubseteq \exists r.(X_i \sqcap L_{i+1}) \sqcap \exists r.(\neg X_i \sqcap L_{i+1}) \\ L_{n+i} &\sqsubseteq \exists r.(Y_i \sqcap L_{n+i+1}) \sqcap \exists r.(\neg Y_i \sqcap L_{n+i+1}). \end{aligned}$$

Moreover, we force the levels to be disjoint by adding the CI

$$L_i \sqsubseteq \neg L_j, \quad \text{for } 0 \leq i < j \leq 2n.$$

We now propagate the concepts X_i and Y_i down to level L_{2n} to encode numbers between 0 and $2^n - 1$ at the leaves of the torus-tree. Since r is a transitive role, the following concept inclusions, for all i with $0 \leq i < n$ suffice:

$$\begin{aligned} L_{i+1} \sqcap X_i &\sqsubseteq \forall r.(L_j \rightarrow X_i) && \text{for } i < j \leq 2n, \\ L_{i+1} \sqcap \neg X_i &\sqsubseteq \forall r.(L_j \rightarrow \neg X_i) && \text{for } i < j \leq 2n, \\ L_{n+i+1} \sqcap Y_i &\sqsubseteq \forall r.(L_j \rightarrow Y_i) && \text{for } n+i < j \leq 2n, \\ L_{n+i+1} \sqcap \neg Y_i &\sqsubseteq \forall r.(L_j \rightarrow \neg Y_i) && \text{for } n+i < j \leq 2n. \end{aligned}$$

Next, we introduce some required notation. Fix an interpretation \mathcal{I} . For each element $d \in \Delta^{\mathcal{I}}$, we define $\text{pos}(d)$ as the pair of integers

$$(\text{xpos}(d), \text{ypos}(d)) = (\sum_{0 \leq i < n} x_i \cdot 2^i, \sum_{0 \leq i < n} y_i \cdot 2^i),$$

where

$$x_i = \begin{cases} 0 & \text{if } d \notin X_i^{\mathcal{I}}, \\ 1 & \text{otherwise;} \end{cases} \quad y_i = \begin{cases} 0 & \text{if } d \notin Y_i^{\mathcal{I}}, \\ 1 & \text{otherwise.} \end{cases}$$

It should be clear that in any model \mathcal{I} of L_0 and the CIs defined so far, there are 2^{2n} elements which satisfy L_{2n} ; even more, for each pair of values $0 \leq i, j < 2^n$, there is an element $d_{ij} \in L_{2n}^{\mathcal{I}}$ such that $\text{pos}(d_{ij}) = (i, j)$. However, the elements d_{ij} are not necessarily connected in a particularly useful way; thus, we now relate elements at level $2n$ to their horizontal and vertical neighbors.

To this aim, we will use *glueing points*. More precisely, for every $d, d' \in L_{2n}^{\mathcal{I}}$ with $\text{pos}(d) = (x, y)$ and $\text{pos}(d') = (x \oplus_{2^n} 1, y)$,³ we enforce an element $g \in H^{\mathcal{I}}$ such that $(d, g) \in r^{\mathcal{I}}$ and $(d', g) \in r^{\mathcal{I}}$ and $\text{pos}(g) = (x \oplus_{2^n} 1, y)$, and similar for the y -coordinate. This is illustrated in Figure 1, where glueing points are depicted as \circ and labelled with H and V for horizontal and vertical, respectively.

To facilitate this task, we define the following concepts, for $0 \leq i \leq n-1$ and $i < j \leq n-1$:

$$\begin{aligned} X_i^* &\equiv \neg X_i \sqcap \prod_{0 \leq k \leq i-1} X_k; & X_i^+ &\equiv X_i \sqcap \prod_{0 \leq k \leq i-1} \neg X_k; \\ X_n^* &\equiv \prod_{0 \leq k \leq n-1} X_k; & X_n^+ &\equiv \prod_{0 \leq k \leq n-1} \neg X_k; \\ X_i^{\rightarrow} &\sqsubseteq (X_j \rightarrow \forall r.X_j) \sqcap (\neg X_j \rightarrow \forall r.\neg X_j); \end{aligned}$$

³ \oplus_k denotes the addition modulo k .

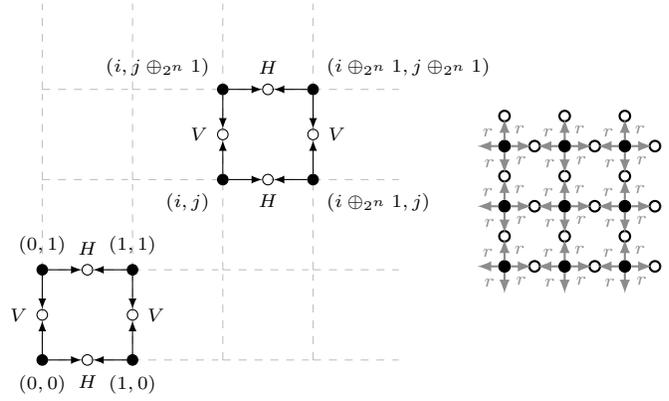


Figure 1: Glueing the points in the torus

and analogous concepts Y_i^* , Y_i^+ , and Y_i^{\rightarrow} . Observe that for every interpretation \mathcal{I} and $d \in \Delta^{\mathcal{I}}$, there is a *single* i such that $d \in (X_i^+)^{\mathcal{I}}$, and similarly for X_i^* . Moreover, for every $d, e \in \Delta^{\mathcal{I}}$ with $\text{xpos}(e) = \text{xpos}(d) \oplus_{2^n} 1$, we have that, if $d \in (X_i^*)^{\mathcal{I}}$, then $e \in (X_i^+)^{\mathcal{I}}$ and $d \in (X_j^{\rightarrow})^{\mathcal{I}}$ iff $e \in (X_j^{\rightarrow})^{\mathcal{I}}$, for all $i < j \leq n$; and similarly for ypos . With this in mind, we enforce for every element d in level $2n$ with $\text{pos}(d) = (x, y)$ four r -successors d_h^+ , d_h^- , d_v^+ , d_v^- such that

- $d_h^+ \in H^{\mathcal{I}}$, $\text{xpos}(d_h^+) = x \oplus_{2^n} 1$, $\text{ypos}(d_h^+) = y$,
- $d_h^- \in H^{\mathcal{I}}$, $\text{xpos}(d_h^-) = x$, $\text{ypos}(d_h^-) = y$,
- $d_v^+ \in V^{\mathcal{I}}$, $\text{xpos}(d_v^+) = x$, $\text{ypos}(d_v^+) = y \oplus_{2^n} 1$, and
- $d_v^- \in V^{\mathcal{I}}$, $\text{xpos}(d_v^-) = x$, $\text{ypos}(d_v^-) = y$,

using the following concept inclusions, for $0 \leq i \leq n$ and $0 \leq j \leq n-1$:

$$\begin{aligned} L_{2n} \sqcap X_i^* &\sqsubseteq X_i^{\rightarrow} \sqcap \exists r.(H \sqcap X_i^+) \sqcap \exists r.(H \sqcap X_i^*) \\ L_{2n} \sqcap Y_i^* &\sqsubseteq Y_i^{\rightarrow} \sqcap \exists r.(V \sqcap Y_i^+) \sqcap \exists r.(V \sqcap Y_i^*) \\ L_{2n} \sqcap X_j &\sqsubseteq \forall r.(V \rightarrow X_j) \\ L_{2n} \sqcap \neg X_j &\sqsubseteq \forall r.(V \rightarrow \neg X_j) \\ L_{2n} \sqcap Y_j &\sqsubseteq \forall r.(H \rightarrow Y_j) \\ L_{2n} \sqcap \neg Y_j &\sqsubseteq \forall r.(H \rightarrow \neg Y_j) \end{aligned}$$

Moreover, we make sure that the glueing points are fresh, and that horizontal and vertical are disjoint by adding:

$$H \sqsubseteq \neg V \quad \text{and} \quad H \sqcup V \sqsubseteq \neg L_i, \text{ for } 0 \leq i \leq 2n.$$

It remains to *identify* the introduced glueing points as indicated in Figure 1. For this, we use the (unqualified) number restrictions. In particular, we add the concept inclusion

$$L_0 \sqsubseteq (\leq k r),$$

with $k = (2^{2n+1} - 2) + 2^{2n}$. To justify the choice of k , note that, without the glueing points, the intended model of L_0 has 2^i elements in every level i , that is, $2^{2n+1} - 1$ elements overall, and hence L_0 has $2^{2n+1} - 2$ successors. As we want to have a single glueing point for every (i, j) with $0 \leq i, j < n$, we need to restrict the number of glueing points to 2^{2n} .

This finishes the definition of \mathcal{T}_{tor} . It is formally shown in the appendix that \mathcal{T}_{tor} properly defines the $2^n \times 2^n$ -torus.

Having this, it is standard to reduce the tiling problem on the torus, by using the elements in level $2n$ as the tiles and the glueing points to communicate between neighboring tiles.

Theorem 3. *Concept satisfiability relative to SON-TBoxes with binary coding of numbers is NEXPTIME-complete.*

Note that, in presence of a single transitive role r , we can always rewrite the TBox \mathcal{T} as a concept $C_{\mathcal{T}}$: add a conjunct $(C \rightarrow D) \sqcap \forall r.(C \rightarrow D)$ for each $C \sqsubseteq D \in \mathcal{T}$; hence:

Corollary 1. *Given a transitive r , satisfiability of SON-concepts with binary coding is NEXPTIME-complete.*

5 The Case of Functionality

We next study *SHOIF* which allows for both inverse roles and role inclusions. Recall that these features lead to undecidability already with unqualified counting with numbers greater than 1 (Horrocks, Sattler, and Tobies 2000; Kazakov, Sattler, and Zolin 2007). However, it was open whether decidability could be attained by sticking to functionality. We answer positively this question, by reducing satisfiability in *SHOIF* to satisfiability in \mathcal{ALCHOL}_{sf} , the extension of \mathcal{ALCHOL} with *local reflexivity* concepts $\exists r.\text{self}$, whose semantics is given by

$$(\exists r.\text{self})^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid (d, d) \in r^{\mathcal{I}}\}.$$

Let \mathcal{T} be a *SHOIF*-TBox and recall the notation $\text{cl}(\mathcal{T})$ from Section 4.1. We obtain the TBox \mathcal{T}' from \mathcal{T} by taking $\mathcal{T}' = (\mathcal{T} \setminus \{\text{Tra}(r) \in \mathcal{T}\}) \cup \mathcal{T}''$, where \mathcal{T}'' is the set of the following CIs, for every $\text{Tra}(r) \in \mathcal{T}$ and $C \in \text{cl}(\mathcal{T})$:

$$\begin{aligned} \forall r.C \sqsubseteq \forall r.(\forall r.C), \\ (\leq 1 r) \sqsubseteq \forall r.(\forall r.\perp \sqcup \exists r.\text{self}). \end{aligned}$$

Concept inclusions of the first type have been used to mimic the behavior of transitive roles for example in (Tobies 2001); note also the similarity to the axiom $\Box p \rightarrow \Box \Box p$ characterizing transitive frames in modal logic (Chagrov and Zakharyashev 1997). Concept inclusions of the second type capture the interplay of transitivity and (local) functionality: if two elements d and e are r -connected, for some functional transitive r , e cannot have an r -successor other than itself.

The correctness of this reduction is established in the appendix. It is easy to see that \mathcal{T}' is an \mathcal{ALCHOL}_{sf} TBox and that it can be computed in polynomial time. Moreover, the reduction also works for *SHIF* yielding an \mathcal{ALCHI}_{sf} -TBox. It is known that concept satisfiability in \mathcal{ALCHI}_{sf} and \mathcal{ALCHOL}_{sf} can be checked in EXPTIME (Calvanese, Eiter, and Ortiz 2009) and NEXPTIME (Motik, Shearer, and Horrocks 2009), respectively. Matching lower bounds are inherited from \mathcal{ALC} and \mathcal{ALCFIO} , respectively (Baader et al. 2003; Lutz 2004).

Theorem 4. *Concept satisfiability is NEXPTIME-complete for SHOIF-TBoxes and EXPTIME-complete for SHIF.*

6 A Look at DL-Lite

We next show that supporting number restrictions on transitive roles in $DL\text{-Lite}_{core}^{\mathcal{H}LN}$ (Artale et al. 2009) leads to undecidability. This result strengthens the known undecidability result for *SHLN* in the sense that $DL\text{-Lite}_{core}^{\mathcal{H}LN}$ is a very

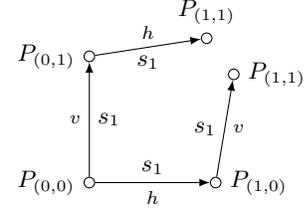


Figure 2: Grid Square

weak sub-Boolean logic without qualified existential restrictions. $DL\text{-Lite}_{core}^{\mathcal{H}LN}$ -concepts C are defined as

$$C ::= \perp \mid A \mid (\sim n r),$$

where $A \in \mathbb{N}_C$, r is a role and \sim is an arbitrary comparison. $DL\text{-Lite}_{core}^{\mathcal{H}LN}$ -TBoxes are defined as in Section 2, but CIs can only take the form: $C \sqsubseteq D$ or $C \sqcap D \sqsubseteq \perp$ with C, D $DL\text{-Lite}_{core}^{\mathcal{H}LN}$ -concepts.

The undecidability proof (cf. appendix) is by reduction of the halting problem of deterministic Turing machines. In the proof, RIs and counting over transitive roles are key for the construction of squares of a grid, and for ensuring that such grid is infinite. For instance, in Figure 2, if we declare (i) the transitive role s_1 as super-role of h and v , and (ii) that each element has at-most 3 s_1 -successors, then the two elements in $P_{(1,1)}$ are forced to be the same. Roughly, we then arrange a sequence of configurations (a computation) as a ‘two-dimensional’ grid of domain elements.

Theorem 5. *Concept satisfiability relative to $DL\text{-Lite}_{core}^{\mathcal{H}LN}$ -TBoxes is undecidable.*

Decidability is regained again for functionality. In particular, Theorem 4 yields a (tight) EXPTIME upper bound for $DL\text{-Lite}_{bool}^{\mathcal{H}LF}$, which is the fragment of *SHLF* allowing only for *unqualified* existential restrictions. Note that the upper bound holds for local functionality; indeed, in $DL\text{-Lite}$ functionality is normally meant to be *global*, which is weaker than the local one. The lower bound is inherited from $DL\text{-Lite}_{bool}^{\mathcal{H}LF}$ (Artale et al. 2009). We thus obtain:

Theorem 6. *Concept satisfiability relative to $DL\text{-Lite}_{bool}^{\mathcal{H}LF}$ -TBoxes is EXPTIME-complete.*

Further, if we drop role inclusions and consider global functionality, we show that, similar to Section 5, we can reduce satisfiability in $DL\text{-Lite}_{bool}^{\mathcal{S}F}$ to satisfiability in $DL\text{-Lite}_{bool}^{\mathcal{F},sf}$, extending $DL\text{-Lite}_{bool}^{\mathcal{F}}$ with local reflexivity concepts. To obtain the desired result, we first show:

Lemma 4. *Concept satisfiability relative to $DL\text{-Lite}_{bool}^{\mathcal{F},sf}$ -TBoxes is NP-complete.*

The lower bound is inherited from $DL\text{-Lite}_{bool}^{\mathcal{F}}$. The upper bound can be proved by extending the reduction from $DL\text{-Lite}_{bool}^{\mathcal{F}}$ to the one-variable fragment of first-order logic (Artale et al. 2009) so as to deal with local reflexivity. With Lemma 4 at hand, we obtain the following (where the lower bound is also inherited from $DL\text{-Lite}_{bool}^{\mathcal{F}}$):

Theorem 7. *Concept satisfiability relative to $DL\text{-Lite}_{bool}^{SF}$ TBoxes is NP-complete.*

Note that the reduction from $DL\text{-Lite}_{bool}^{SF}$ to $DL\text{-Lite}_{bool}^{F, sf}$ relies on the availability of disjunction. Hence we cannot lift the reduction to non-Boolean complete fragments.

7 Conclusions and Future Work

In this paper, we have made progress on the understanding of the computational complexity of DLs allowing for number restrictions on transitive roles. In particular, we have established a tight NEXPTIME upper bound for satisfiability in SOQ , and showed that in SON the coding of numbers plays a role on the computational complexity.

As the next step, we will look for ways to incorporate inverse roles and numbers greater than 1 without losing decidability. In this direction, we will investigate languages based on $DL\text{-Lite}$ without RIs. In $SOTQ$, decidability might be regained by admitting counting only over r or r^- , but not over both (Kazakov, Sattler, and Zolin 2007). We will also investigate ways to include some other forms of complex roles, such as role composition.

On the practical side, we are interested in developing a consequence-based calculus for our logics – a promising starting point is the recently proposed calculus for $SRIQ$, supporting number restrictions on non-transitive roles (Bate et al. 2016).

We will also study DLs supporting counting over transitive roles in the context of *ontology-based data access*. We want to understand the impact of these features on the problem of *conjunctive query answering*, in the case where transitive roles occur in the query. Moreover, we will consider conjunctive queries incorporating some type of counting, e.g., restricted versions of inequalities, such as local inequalities (Gutiérrez-Basulto et al. 2015).

Acknowledgments We thank anonymous reviewers for useful comments and suggestions. The first author was supported by a Marie Curie SIRCIW COFUND grant.

References

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyashev, M. 2009. The dl-lite family and relations. *J. Artif. Intell. Res. (JAIR)* 36:1–69.

Baader, F.; Lutz, C.; Sturm, H.; and Wolter, F. 2002. Fusions of description logics and abstract description systems. *J. Artif. Intell. Res. (JAIR)* 16:1–58.

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

Bate, A.; Motik, B.; Grau, B. C.; Simancik, F.; and Horrocks, I. 2016. Extending consequence-based reasoning to $SRIQ$. In *Proc. of KR-16*, 187–196.

Calvanese, D.; Eiter, T.; and Ortiz, M. 2009. Regular path queries in expressive description logics with nominals. In *Proc. of IJCAI-09*, 714–720.

Chagrov, A. V., and Zakharyashev, M. 1997. *Modal Logic*, volume 35 of *Oxford logic guides*. Oxford University Press.

Gutiérrez-Basulto, V.; Ibáñez-García, Y. A.; Kontchakov, R.; and Kostylev, E. V. 2015. Queries with negation and inequalities over lightweight ontologies. *J. Web Sem.* 35:184–202.

Hollunder, B., and Baader, F. 1991. Qualifying number restrictions in concept languages. In *Proc. of KR-91*, 335–346.

Horrocks, I.; Sattler, U.; and Tobies, S. 2000. Practical reasoning for very expressive description logics. *Logic Journal of the IGPL* 8(3):239–263.

Kaminski, M., and Smolka, G. 2010. Terminating tableaux for SOQ with number restrictions on transitive roles. In *Theoretical Computer Science - 6th IFIP TC Proceedings*, 213–228.

Kazakov, Y., and Pratt-Hartmann, I. 2009. A note on the complexity of the satisfiability problem for graded modal logics. In *Proc. of LICS-09*, 407–416.

Kazakov, Y.; Sattler, U.; and Zolin, E. 2007. How many legs do I have? Non-simple roles in number restrictions revisited. In *Proc. of LPAR-07*, 303–317.

Lutz, C.; Areces, C.; Horrocks, I.; and Sattler, U. 2005. Keys, nominals, and concrete domains. *J. Artif. Intell. Res. (JAIR)* 23:667–726.

Lutz, C. 2004. An improved NEXPTIME-hardness result for the description logic \mathcal{ALC} extended with inverse roles, nominals and counting. Technical Report LTCS-04-07, Chair for Automata Theory, TU Dresden, Germany.

Motik, B.; Shearer, R.; and Horrocks, I. 2009. Hypertableau reasoning for description logics. *J. Artif. Intell. Res. (JAIR)* 36:165–228.

Rector, A. L., and Rogers, J. 2006. Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN. In *Proc. of RW-06*, 197–231.

Schröder, L., and Pattinson, D. 2008. How many toes do I have? parthood and number restrictions in description logics. In *Proc. of KR-08*, 307–317.

Stevens, R.; Aranguren, M. E.; Wolstencroft, K.; Sattler, U.; Drummond, N.; Horridge, M.; and Rector, A. L. 2007. Using OWL to model biological knowledge. *International Journal of Man-Machine Studies* 65(7):583–594.

Tendera, L. 2005. Counting in the two variable guarded logic with transitivity. In *Proc. of STACS-05*, 83–96.

Tobies, S. 2001. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation*. Ph.D. Dissertation, RWTH Aachen.

van Emde Boas, P. 1997. The convenience of tilings. In *Complexity, Logic, and Recursion Theory*, 331–363.

Wolstencroft, K.; Brass, A.; Horrocks, I.; Lord, P. W.; Sattler, U.; Turi, D.; and Stevens, R. 2005. A little semantic web goes a long way in biology. In *Proc. of ISWC-05*, 786–800.