Expressive Cardinality Constraints on $\mathcal{ALCSCC}$ Concepts

Franz Baader*
Theoretical Computer Science, TU Dresden
Dresden, Germany
franz.baader@tu-dresden.de

ABSTRACT

In two previous publications we have, on the one hand, extended the description logic (DL) $\mathcal{ALCQ}$ by more expressive number restrictions using numerical and set constraints expressed in the quantifier-free fragment of Boolean Algebra with Presburger Arithmetic (QFBAPA). The resulting DL was called $\mathcal{ALCSCC}$. On the other hand, we have extended the terminological formalism of the well-known description logic $\mathcal{ALC}$ from concept inclusions (CIs) to more general cardinality constraints expressed in QFBAPA, which we called extended cardinality constraints. Here, we combine the two extensions, i.e., we consider extended cardinality constraints on $\mathcal{ALCSCC}$ concepts. We show that this does not increase the complexity of reasoning, which is NExpTime-complete both for extended cardinality constraints in $\mathcal{ALC}$ and $\mathcal{ALCSCC}$. The same is true for a restricted version of such cardinality constraints, where the complexity of reasoning decreases to ExpTime, not just for $\mathcal{ALC}$, but also for $\mathcal{ALCSCC}$.

CCS CONCEPTS

• **Theory of computation → Description logics**.

KEYWORDS

Description Logic, Number Restrictions, Cardinality Restrictions, QFBAPA, Complexity

ACM Reference Format:
Franz Baader. 2019. Expressive Cardinality Constraints on $\mathcal{ALCSCC}$ Concepts. In *The 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19), April 8–12, 2019, Limassol, Cyprus.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3297280.3297390

1 INTRODUCTION

Description Logics (DLs) [3] are a well-investigated family of logic-based knowledge representation languages, which are frequently used to formalize ontologies for application domains such as biology and medicine [7]. To define the important notions of such an application domain as formal concepts, DLs state necessary and sufficient conditions for an individual to belong to a concept. These conditions can be Boolean combinations of atomic properties required for the individual (expressed by concept names) or properties that refer to relationships with other individuals and their properties (expressed as role restrictions). Using an example from [4], the concept of a motor vehicle can be formalized by the concept description *Vehicle* $\sqcap$ $\exists$*part.Motor*, which uses the concept names *Vehicle* and *Motor* and the role name *part* as well as the concept constructors conjunction ($\sqcap$) and existential restriction ($\exists r.C$). The concept inclusion (CI) *Motor-vehicle* $\sqsubseteq$ *Vehicle* $\sqcap$ $\exists$*part.Motor* can then be used to state that every motor vehicle needs to belong to this concept description.

Numerical constraints on the number of role successors (so-called number restrictions) have been used early on in DLs [5, 8, 9]. For example, using number restrictions, motorcycles can be constrained to being motor vehicles with exactly two wheels:

$$\mathit{Motorcycle} \sqsubseteq \mathit{Motor\text{-}vehicle} \sqcap (\leqslant 2\, \mathit{part.Wheel}) \sqcap (\geqslant 2\, \mathit{part.Wheel}).$$

The exact complexity of reasoning in $\mathcal{ALCQ}$, the DL that has all Boolean operations and number restrictions of the form $(\leqslant n\, r.C)$ and $(\geqslant n\, r.C)$ as concept constructors, was determined by Stephan Tobies [12, 14]: it is PSpace-complete without CIs and ExpTime-complete w.r.t. CIs, independently of whether the numbers occurring in the number restrictions are encoded in unary or binary.

The classical number restrictions available in $\mathcal{ALCQ}$ can only be used to compare the number of role successors of an individual with a fixed natural number. They cannot relate numbers of different kinds of role successors to each other. This would, e.g., be required to state that the number of cylinders of a motor coincides with the number of spark plugs in this motor, without fixing what this number actually is. To overcome this deficit, we have extended $\mathcal{ALCQ}$ by allowing the statement of constraints on role successors that are more general than the number restrictions of $\mathcal{ALCQ}$ [1]. To formulate these constraints, we have used the quantifier-free fragment of Boolean Algebra with Presburger Arithmetic (QF-BAPA) [10], in which one can express Boolean combinations of set constraints and numerical constraints on the cardinalities of sets. In the resulting logic $\mathcal{ALCSCC}$, the above constraint regarding cylinders and spark plugs can be expressed using a cardinality constraint on the role successors: $\mathit{Motor} \sqsubseteq \mathit{succ}(|\mathit{part} \cap \mathit{Cylinder}| = |\mathit{part} \cap \mathit{SparkPlug}|)$. In general, such a *succ*-expression considers the set of all role successors of a given individual, and requires certain subsets to satisfy the stated QFBAPA constraints. In our example, the cardinality of the set of *part*-successors that belong to the concept *Cylinder* must be the same as of the set of *part*-successors that belong to the concept *SparkPlug*.

In [1] it was shown that this extension of the expressive power of $\mathcal{ALCQ}$ does not increase the complexity of reasoning: it is still PSpace-complete without CIs and ExpTime-complete w.r.t. CIs. While the PSpace result also follows from previous work on modal logics with Presburger constraints [6], the ExpTime result was new.

*Partially supported by DFG within the Research Unit 1513 Hybris.

SAC '19, April 8–12, 2019, Limassol, Cyprus

ACM ISBN 978-1-4503-5933-7/19/04…$15.00
https://doi.org/10.1145/3297280.3297390

Whereas number restrictions are local in the sense that they consider role successors of an individual under consideration (e.g. the wheels that are part of a particular motor vehicle), cardinality restrictions on concepts (CRs) [2, 13] are global, i.e., they consider all individuals in an interpretation. For example, the cardinality restriction ($\leqslant 45000000\,(Car \sqcap \exists registered\text{-}in.German\text{-}district)$) states that at most 45 million cars are registered all over Germany. Such cardinality restrictions can express CIs ($C \sqsubseteq D$ is equivalent to $(\leqslant 0\,(C \sqcap \neg D))$), but are considerably more expressive. In particular, they increase the complexity of reasoning: for the DLs $\mathcal{ALC}$ and $\mathcal{ALCQ}$, consistency w.r.t. CIs is ExpTime-complete [11, 14], but consistency w.r.t. CRs is NExpTime-complete if the numbers occurring in the CRs are assumed to be encoded in binary [13]. With unary coding of numbers, consistency stays ExpTime-complete even w.r.t. CRs [13], but the above example considering 45 million cars clearly shows that unary coding is not appropriate if numbers with large values are employed.

Just like classical number restrictions, CRs can only relate the cardinality of a concept to a fixed number. In [4], we have introduced and investigate more general constraints on the cardinalities of concepts, which we called extended cardinality constraints. The main idea was again to use QFBAPA to formulate and combine these constraints. An example of a constraint expressible this way, but not expressible using CRs is $2 \cdot |Car \sqcap \exists registered\text{-}in.German\text{-}district \sqcap \exists fuel.Diesel| \leq |Car \sqcap \exists registered\text{-}in.German\text{-}district \sqcap \exists fuel.Petrol|$, which states that, in Germany, cars running on petrol outnumber cars running on diesel by a factor of at least two. In [4] it is shown that, in the DL $\mathcal{ALC}$, the complexity of reasoning w.r.t. extended cardinality constraints (NExpTime for binary coding of numbers) is the same as for reasoning w.r.t. CRs. In addition, the paper introduces a restricted version of this formalism, which can express CIs, but not CRs, and shows that this way the complexity can be lowered to ExpTime. The NExpTime upper bound for the general case actually also follows from the NExpTime upper bound in [15] for a more expressive logic with n-ary relations and function symbols, but the ExpTime result for the restricted case was new.

The results on extended cardinality constraints in [4] were restricted to concepts of the DL $\mathcal{ALC}$. It was not even clear whether the complexity upper bounds also hold for $\mathcal{ALCQ}$, let alone the considerably more expressive DL $\mathcal{ALCSCC}$. In the present paper, we combine the work in [1] and [4] by considering extended cardinality constraints in $\mathcal{ALCSCC}$. This turned out to be non-trivial since the local cardinality constraints of $\mathcal{ALCSCC}$ may interact with the global ones in the extended cardinality constraints. Nevertheless, we are able to show that the complexity results (NExpTime-complete in general, and ExpTime-complete in the restricted case) hold not only for $\mathcal{ALC}$, but also for $\mathcal{ALCSCC}$.

2 PRELIMINARIES

Before defining $\mathcal{ALCSCC}$ and extended cardinality constraints, we must introduce QFBAPA, on which both are based. More details on this logic can be found in [10].

The logic QFBAPA. In this logic one can build *set terms* by applying Boolean operations (intersection $\cap$, union $\cup$, and complement $\cdot^c$) to set variables as well as the constants $\emptyset$ and $\mathcal{U}$. Set terms s, t can then be used to state inclusion and equality constraints ($s = t, s \subseteq t$) between sets. *Presburger Arithmetic (PA) expressions* are built from integer variables, integer constants, and set cardinalities $|s|$ using addition as well as multiplication with an integer constant. They can be used to form numerical constraints of the form $k = \ell, k < \ell, N \operatorname{dvd} \ell$, where k, ℓ are PA expressions, N is an integer constant, and dvd stands for divisibility. A *QFBAPA formula* is a Boolean combination of set and numerical constraints.

A *solution* σ of a QFBAPA formula ϕ assigns a finite set $\sigma(\mathcal{U})$ to $\mathcal{U}$, subsets of $\sigma(\mathcal{U})$ to set variables, and integers to integer variables such that ϕ is satisfied by this assignment. The evaluation of set terms, PA expressions, and set and numerical constraints w.r.t. σ is defined in the obvious way. For example, σ satisfies the numerical constraint $|s \cup t| = |s| + |t|$ for set variables s, t if the cardinality of the union of the sets $\sigma(s)$ and $\sigma(t)$ is the same as the sum of the cardinalities of these sets. Note that this is the case iff $\sigma(s)$ and $\sigma(t)$ are disjoint, which we could also have expressed using the set constraint $s \cap t \subseteq \emptyset$. A QFBAPA formula ϕ is *satisfiable* if it has a solution. As shown in [10], satisfiability of QFBAPA formulae is an NP-complete problem.

The main tool used in [10] to show that satisfiability in QFBAPA is in NP is a "sparse solution" lemma (see Lemma 1 below), which will also turn out to be useful for showing one of our complexity upper bounds. Assume that ϕ is a QFBAPA formula containing the set variables $X_1, \ldots, X_k$. A *Venn region* is of the form

$$X_1^{c_1} \cap \ldots \cap X_k^{c_k},$$

where c_i is either empty or c for $i = 1, \ldots, k$. It is shown in [10] that, given ϕ, one can easily compute a number N whose value is polynomial in the size of ϕ such that the following holds: ϕ is satisfiable iff it has a solution in which $\leq N$ Venn regions are interpreted by non-empty sets. Taking a closer look at how this result is proved in [10], one can actually strengthen it (see [1] for a proof).

Lemma 1. *For every QFBAPA formula ϕ, one can compute in polynomial time a number N whose value is polynomial in the size of ϕ such that the following holds for every solution σ of ϕ: there is a solution σ' of ϕ such that*

- $|\{v \mid v \text{ Venn region and } \sigma'(v) \neq \emptyset\}| \leq N$, *and*
- $\{v \mid v \text{ Venn region and } \sigma'(v) \neq \emptyset\} \subseteq \{v \mid v \text{ Venn region and } \sigma(v) \neq \emptyset\}$.

The DL $\mathcal{ALCSCC}$. In the following, we recall syntax and semantics of $\mathcal{ALCSCC}$ (a more detailed introduction can be found in [1]). Basically, the DL $\mathcal{ALCSCC}$ has all Boolean operations ($\sqcap$, $\sqcup$, $\neg$) as concept constructors and can state constraints on role successors using the expressiveness of QFBAPA.

To be more precise, a *successor constraint* is either a set constraint or a cardinality constraint. As in QFBAPA, *set constraints* are inclusion constraints or equality constraints ($s = t$, $s \subseteq t$) between set terms s, t, but now s, t use role names and $\mathcal{ALCSCC}$ concept descriptions in place of set variables. For example, the $\mathcal{ALCSCC}$ concept description $Male \sqcap succ(child \cap Female \subseteq Rich)$ describes all male individuals that have only rich daughters (i.e., female children), where *Male*, *Female* are concept names and *child* is a role name. Similarly, *cardinality constraint* are of the form $k = \ell, k < \ell, N \operatorname{dvd} \ell$, but in the PA expressions k, ℓ there are no integer variables, but only set cardinalities $|s|$. For example, the $\mathcal{ALCSCC}$ concept description $Female \sqcap succ(|child \cap Female| = |child \cap Male|)$ describes all female individuals that have exactly as many sons as daughters. Of course, successor constraints can also be nested, as in the $\mathcal{ALCSCC}$ concept description $succ(|child \cap succ(child \subseteq Female)| = |child \cap succ(child \subseteq Male)|)$, which describes all individuals having as many children that have only daughters as they have children having only sons.

Given disjoint sets N_C and N_R of concept and role names, respectively, an *interpretation* of N_C and N_R consists of a *finite*[1] non-empty set $\Delta^\mathcal{I}$ and a mapping $\cdot^\mathcal{I}$ that maps every concept name $A \in N_C$ to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$ and every role name $r \in N_R$ to a binary relation $r^\mathcal{I}$ over $\Delta^\mathcal{I}$. Given an individual $d \in \Delta^\mathcal{I}$ and a role name $r \in N_R$, we define $r^\mathcal{I}(d) := \{e \in \Delta^\mathcal{I} \mid (d, e) \in r^\mathcal{I}\}$ (r-successors) and $ars^\mathcal{I}(d) := \bigcup_{r \in N_R} r^\mathcal{I}(d)$ (all role successors).

As usual, the interpretation function $\cdot^\mathcal{I}$ is inductively extended to $\mathcal{ALCSCC}$ concept descriptions by interpreting $\sqcap$, $\sqcup$, and $\neg$ respectively as intersection, union and complement. Successor constraints are evaluated according to the semantics of QFBAPA. To be more precise, to determine whether $d \in succ(c)^\mathcal{I}$ or not, roles r occurring in c are evaluated as $r^\mathcal{I}(d)$ (i.e., the set of r-successors of d) and concept descriptions D as $D^\mathcal{I} \cap ars^\mathcal{I}(d)$ (i.e., the set of role successors of d that belong to D).[2] Then $d \in succ(c)^\mathcal{I}$ iff the valuation obtained this way is a solution of the QFBAPA formula c (see [1] for a more detailed definition of the semantics of $\mathcal{ALCSCC}$).

Note that $\mathcal{ALCSCC}$ contains the well-known DLs $\mathcal{ALC}$ and $\mathcal{ALCQ}$ as sub-logics. For example, the value restriction $\forall r.C$ of $\mathcal{ALC}$ can be expressed as $succ(r \subseteq C)$ and the number restriction $\geqslant n\, r.C$ as $succ(|r \cap C| \geq n)$.

Extended cardinality constraints. Basically, extended cardinality constraints are QFBAPA formulae where $\mathcal{ALCSCC}$ concept descriptions are used in place of set variables. However, since inclusion constraints $s \subseteq t$ (and thus equality constraints) can be expressed as cardinality constraints $|s \cap t^c| \leq 0$, we do not explicitly allow the use of set constraints. In addition, since $\mathcal{ALCSCC}$ has all Boolean operations on concepts, we do not need Boolean operations on set terms. Consequently, we define *extended cardinality constraints on $\mathcal{ALCSCC}$ concepts* as follows:

- *$\mathcal{ALCSCC}$ cardinality terms* are built from integer constants and concept cardinalities $|C|$ for $\mathcal{ALCSCC}$ concepts C using addition and multiplication with integer constants;
- *extended $\mathcal{ALCSCC}$ cardinality constraints* are of the form $k = \ell, k < \ell, N \text{ dvd } \ell$, where k, ℓ are $\mathcal{ALCSCC}$ cardinality terms and N is an integer constant;
- an *extended $\mathcal{ALCSCC}$ cardinality box (ECBox)* is a Boolean combination of extended $\mathcal{ALCSCC}$ cardinality constraints.

When defining the semantics of ECBoxes, we again restrict the attention to *finite* interpretations to ensure that cardinalities of concepts are always well-defined non-negative integers. Given a finite interpretation $\mathcal{I}$, concept cardinalities are interpreted in the obvious way, i.e., $|C|^\mathcal{I} := |C^\mathcal{I}|$. Addition and multiplication in cardinality terms are interpreted as the usual addition and multiplication operations on integers, and the same is the case for the comparison operators $=$, $<$, and divisibility dvd. This yields the semantics of extended $\mathcal{ALCSCC}$ cardinality constraints and thus also of their Boolean combinations.

Since $\mathcal{ALC}$ is a sub-logic of $\mathcal{ALCSCC}$, extended cardinality constraints on $\mathcal{ALC}$ concepts are a special case of extended cardinality constraints on $\mathcal{ALCSCC}$ concepts. In [4] it was shown that consistency of $\mathcal{ALC}$ ECBoxes is a NExpTime-complete problem if the numbers occurring in the ECBoxes are encoded in binary. The hardness result obviously transfers to the super-logic $\mathcal{ALCSCC}$.

Proposition 2. *Consistency of ECBoxes with numbers encoded in binary is NExpTime-hard in $\mathcal{ALCSCC}$.*

In the next section, we show that the NExpTime upper bound holds not only for $\mathcal{ALC}$, but also for $\mathcal{ALCSCC}$.

3 CONSISTENCY OF $\mathcal{ALCSCC}$ ECBOXES

In the following we consider an $\mathcal{ALCSCC}$ ECBox $\mathcal{E}$ and show how to test $\mathcal{E}$ for consistency by reducing this problem to the problem of testing satisfiability of QFBAPA formulae. Since the reduction is exponential and satisfiability in QFBAPA is in NP, this yields a NExpTime upper bound for consistency of $\mathcal{ALCSCC}$ ECBoxes. Such an ExpTime-reduction to satisfiability in QFBAPA has already been used in the consistency procedure for $\mathcal{ALC}$ ECBoxes in [4]. However, for $\mathcal{ALCSCC}$ it needs to be extended considerably. In addition to encoding the global cardinality constraints in $\mathcal{E}$ using the notion of a type, we also need to take care of the local constraints expressed by successor constraints within $\mathcal{ALCSCC}$, and to link both kinds of constraints. This makes both the reduction and the proof of its correctness considerably more complicated.

Given a set of concept descriptions $\mathcal{M}$, the *type* of an individual in an interpretation consists of the elements of $\mathcal{M}$ to which the individual belongs. Such a type t can also be seen as a concept description C_t, which is the conjunction of all the elements of t. We assume in the following, that $\mathcal{M}$ consists of *all subdescriptions* of the concept descriptions occurring in $\mathcal{E}$ as well as the *negations of these subdescriptions.*

Definition 3. *A subset t of $\mathcal{M}$ is a* type *for $\mathcal{E}$ if it satisfies the following properties for all concept descriptions C, D:*

(1) *if $\neg C \in \mathcal{M}$, then either C or $\neg C$ belongs to t;*
(2) *if $C \sqcap D \in \mathcal{M}$, then $C \sqcap D \in t$ iff $C \in t$ and $D \in t$;*
(3) *if $C \sqcup D \in \mathcal{M}$, then $C \sqcup D \in t$ iff $C \in t$ or $D \in t$.*

We denote the set of all types for $\mathcal{E}$ with $types(\mathcal{E})$. Given an interpretation $\mathcal{I}$ and an individual $d \in \Delta^\mathcal{I}$, the type *of d is the set*

$$t_\mathcal{I}(d) := \{C \in \mathcal{M} \mid d \in C^\mathcal{I}\}.$$

It is easy to show that the type of an individual really satisfies the conditions stated in the definition of a type. Due to Condition 1 in the definition of types, concept descriptions induced by different types are disjoint, and all concept descriptions in $\mathcal{M}$ can be obtained as the disjoint union of the concept descriptions induced by the types containing them. In particular, we have

$$|C^\mathcal{I}| = \sum_{t \text{ type with } C \in t} |C_t^\mathcal{I}| \quad \text{for all finite interpretations } \mathcal{I}.$$

We transform the ECBox $\mathcal{E}$ into a QFBAPA formula $\phi_\mathcal{E}$ by introducing an integer variable v_t for every type t, stating that these

[1] In [1] a weaker restriction is used, where only the number of role successors of all individuals must be finite. Here we use the stronger restrictions that $\Delta^\mathcal{I}$ is finite since we need this restriction in the presence of extended cardinality constraints [4].
[2] Note that, by induction, the sets $D^\mathcal{I}$ are well-defined.

variables have a non-negative value, and then replacing every concept cardinality $|C|$ in $\mathcal{E}$ by the sum of the corresponding type variables, i.e.,

$$|C| \text{ is replaced by } \sum_{t \text{ type with } C \in t} v_t.$$

A model $\mathcal{I}$ of $\mathcal{E}$ then yields a solution of $\phi_\mathcal{E}$ as follows: if we define $\sigma(v_t) := |C_t^\mathcal{I}|$ for all types t, then σ is a solution of $\phi_\mathcal{E}$.

However, not every solution of $\phi_\mathcal{E}$ is induced by a model of $\mathcal{E}$ in this way. For example, let $\mathcal{E} := |\operatorname{succ}(|A \cap r| \geq 5)| \geq 1 \wedge |A| \leq 3$. In this case, $\mathcal{M}$ consists of the concept descriptions $A, \neg A, \operatorname{succ}(|A \cap r| \geq 5), \neg \operatorname{succ}(|A \cap r| \geq 5)$, and there are four types:

$$t_1 := \{A, \operatorname{succ}(|A \cap r| \geq 5)\}, \quad t_2 := \{\neg A, \operatorname{succ}(|A \cap r| \geq 5)\},$$
$$t_3 := \{A, \neg \operatorname{succ}(|A \cap r| \geq 5)\}, \quad t_4 := \{\neg A, \neg \operatorname{succ}(|A \cap r| \geq 5)\}.$$

The ECBox $\mathcal{E}$ is now translated into the QFBAPA formula $\phi_\mathcal{E}$ by replacing $|\operatorname{succ}(|A \cap r| \geq 5)|$ with $v_{t_1} + v_{t_2}$ and $|A|$ with $v_{t_1} + v_{t_3}$ and adding the information that $v_{t_1}, v_{t_2}, v_{t_3}, v_{t_4}$ have values ≥ 0, i.e., $\phi_\mathcal{E}$ is the following formula:

$$v_{t_1} \geq 0 \wedge v_{t_2} \geq 0 \wedge v_{t_3} \geq 0 \wedge v_{t_4} \geq 0 \wedge v_{t_1} + v_{t_2} \geq 1 \wedge v_{t_1} + v_{t_3} \leq 3.$$

If we set $\sigma(v_{t_1}) = \sigma(v_{t_3}) = \sigma(v_{t_4}) = 0$ and $\sigma(v_{t_2}) = 1$, then σ is a solution of $\phi_\mathcal{E}$. However, $\mathcal{E}$ does not have a model. In fact, $\mathcal{E}$ requires that there is an element belonging to the concept $\operatorname{succ}(|A \cap r| \geq 5)$, which implies that this element must have at least 5 distinct r-successors belonging to A. However, this is prohibited by $\mathcal{E}$ since the second conjunct states that globally (i.e., in the whole model) there are at most 3 elements belonging to A.

This example shows that we must take elements required by successor constraints into account. Basically, if a type t is *realized* in the sense that the variable v_t receives a value > 0, then we must ensure that also types required by the successor constraints in t are realized with the right multiplicity. In our example, the types t_1 and t_2 require that an element belonging to them has at least 5 distinct r-successors belonging to A. As shown in [1], we can express such constraints again by QFBAPA formulae. Given a type t, the (possibly negated) successor constraints occurring in t induce a QFBAPA formula ψ_t, in which the concepts C and roles r occurring in these constraints induce set variables X_C^t and X_r^t. In addition, to take into account the semantics of $\mathcal{ALCSCC}$, we conjoin

$$\alpha_t := (\mathcal{U}^t = \bigcup_{r \in N_R} X_r^t) \wedge \bigwedge_{C \in \mathcal{M}} X_C^t \subseteq \mathcal{U}^t$$

to the formula obtained from the successor constraints and replace $\mathcal{U}$ in these constraints by $\mathcal{U}^t$. In our example, we have

$$\psi_{t_1} := |X_A^{t_1} \cap X_r^{t_1}| \geq 5 \wedge \alpha_{t_1}, \quad \psi_{t_2} := |X_A^{t_2} \cap X_r^{t_2}| \geq 5 \wedge \alpha_{t_2},$$
$$\psi_{t_3} := \neg(|X_A^{t_3} \cap X_r^{t_3}| \geq 5) \wedge \alpha_{t_3}, \quad \psi_{t_4} := \neg(|X_A^{t_4} \cap X_r^{t_4}| \geq 5) \wedge \alpha_{t_4}.$$

In principle, the formula ψ_t constrains the "local" role successors of an individual of type t. It remains to link these local constraints with the global ones in case type t is populated. Basically, we need to ensure that our solutions of the local constraints do not assume that a concept is populated by more individuals than the solution of the global constraints allows. This can be expressed by the following QFBAPA formula:

$$\gamma_t := \bigwedge_{t' \in \operatorname{types}(\mathcal{E})} |\bigcap_{C \in t'} X_C^t| \leq v_{t'}. \tag{1}$$

In our example, the formulae γ_{t_i} (for $1 \leq i \leq 4$) in particular contain the conjuncts

$v_{t_1} \geq |X_A^{t_i} \cap X_C^{t_i}|$ and $v_{t_3} \geq |X_A^{t_i} \cap X_{\neg C}^{t_i}|$ where $C = \operatorname{succ}(|A \cap r| \geq 5)$.

Overall, we translate the $\mathcal{ALCSCC}$ ECBox $\mathcal{E}$ into the QFBAPA formula

$$\delta_\mathcal{E} := \phi_\mathcal{E} \wedge \bigwedge_{t \in \operatorname{types}(\mathcal{E})} v_t = 0 \vee (\psi_t \wedge \gamma_t).$$

Assume that σ is a solution of $\delta_\mathcal{E}$ for the ECBox $\mathcal{E}$ of our example. Then $\sigma(v_{t_1}) + \sigma(v_{t_2}) \geq 1$, which implies that $\sigma(v_{t_1}) > 0$ or $\sigma(v_{t_2}) > 0$. Let us assume first that $\sigma(v_{t_1}) > 0$. Then σ must satisfy ψ_{t_1} and γ_{t_1}. But then ψ_{t_1} yields that $|\sigma(X_A^{t_1})| \geq 5$. However, γ_{t_1} yields that $|\sigma(X_A^{t_1})| = |\sigma(X_A^{t_1} \cap X_C^{t_1})| + |\sigma(X_A^{t_1} \cap X_{\neg C}^{t_1})| \leq \sigma(v_{t_1}) + \sigma(v_{t_3}) \leq 3$, where the latter inequality holds because σ satisfies the conjunct $v_{t_1} + v_{t_3} \leq 3$ of $\phi_\mathcal{E}$. This yields a contradiction to our assumption that $\sigma(v_{t_1}) > 0$. However, then we must have $\sigma(v_{t_2}) > 0$, which also leads to a contradiction: $5 \leq |\sigma(X_A^{t_2})| \leq \sigma(v_{t_1}) + \sigma(v_{t_3}) \leq 3$. This shows that $\delta_\mathcal{E}$ does not have a solution in our example, which corresponds to the fact that $\mathcal{E}$ is actually inconsistent.

The next lemma shows that there is indeed a 1–1-relationship between solvability of $\delta_\mathcal{E}$ and consistency of $\mathcal{E}$.

Lemma 4. *The QFBAPA formula $\delta_\mathcal{E}$ is of size at most exponential in the size of $\mathcal{E}$, and it is satisfiable iff $\mathcal{E}$ is consistent.*

Proof. The at most exponential size of $\delta_\mathcal{E}$ is an easy consequence of the fact that there are at most exponentially many types t, and thus at most exponentially many variables v_t since the cardinality of $\mathcal{M}$ is linear in the size of $\mathcal{E}$. This implies that the size of $\phi_\mathcal{E}$ is at most exponential. For every type t, the size of ψ_t is polynomial in the size of $\mathcal{E}$ since the cardinality of $\mathcal{M}$ is linear in the size of $\mathcal{E}$ and the constraints in ψ_t not contained in α_t are obtained from successor constraints occurring in the concepts in $\mathcal{M}$. Finally, the size of γ_t is at most exponential in the size of $\mathcal{E}$ since there are at most $|\mathcal{M}|$ concepts C and at most exponentially many types t'.

Now, assume that the finite interpretation $\mathcal{I}$ is a model of $\mathcal{E}$. If we define σ as $\sigma(v_t) := |C_t^\mathcal{I}|$ for all types t, then we know that σ solves $\phi_\mathcal{E}$. Let t by a type such that $\sigma(v_t) \neq 0$. Then there is an individual $d \in \Delta^\mathcal{I}$ such that $d \in C_t^\mathcal{I}$. The semantics of $\mathcal{ALCSCC}$ then implies that we can extend σ to a solution of ψ_t by interpreting the set variables with superscript t using the role successors of d:

$$\sigma(X_r^t) := \{e \mid (d, e) \in r^\mathcal{I}\}, \quad \sigma(\mathcal{U}^t) := \textstyle\bigcup_{r \in N_R} \sigma(X_r^t), \text{ and}$$
$$\sigma(X_C^t) := C^\mathcal{I} \cap \sigma(\mathcal{U}^t).$$

This also yields a solution of γ_t since the following holds for all types t': $\bigcap_{C \in t'} \sigma(X_C^t) \subseteq \bigcap_{C \in t'} C^\mathcal{I} = C_{t'}^\mathcal{I}$ and thus $|\bigcap_{C \in t'} \sigma(X_C^t)| \leq |C_{t'}^\mathcal{I}| = \sigma(v_{t'})$.

If t is a type such that $\sigma(v_t) = 0$, then it is not necessary for σ to satisfy ψ_t and γ_t. We can thus extend σ to the set variables with superscript t in an arbitrary way, e.g. by interpreting all of them as the empty set. Overall, this show that we can use a model of $\mathcal{E}$ to define a solution σ of $\delta_\mathcal{E}$.

Conversely, assume that there is a solution σ of $\delta_\mathcal{E}$. Let

$$T_\sigma := \{t \mid t \text{ type with } \sigma(v_t) \neq 0\}$$

be the types that are realized by σ. We now define a finite interpretation $\mathcal{I}$ and show that it is a model of $\mathcal{E}$. The interpretation domain consists of copies of the realized types, where the number of copies is determined by σ:

$$\Delta^{\mathcal{I}} := \{(t,j) \mid t \in T_\sigma \text{ and } 1 \leq j \leq \sigma(v_t)\}.$$

For concept names A we define

$$A^{\mathcal{I}} := \{(t,j) \in \Delta^{\mathcal{I}} \mid A \in t\}.$$

Defining the interpretation of a role name in $\mathcal{I}$ is a bit more involved. Consider a type t with $\sigma(v_t) \neq 0$. Then σ satisfies ψ_t and γ_t. The solution of ψ_t yields a finite set $\sigma(\mathcal{U}^t)$ and interprets the set variables X_r^t and X_C^t as subsets of $\sigma(\mathcal{U}^t)$ such that ψ_t is satisfied. For every element e of $\sigma(\mathcal{U}^t)$ there is a unique type t_e such that $e \in \bigcap_{C \in t_e} \sigma(X_C^t)$. In addition, the fact that σ satisfies γ_t implies that, for all types t', we have $|\{e \in \sigma(\mathcal{U}^t) \mid t_e = t'\}| \leq \sigma(v_{t'})$. This show that there exists an injective mapping π_t of $\sigma(\mathcal{U}^t)$ into $\Delta^{\mathcal{I}}$ such that $\pi_t(e) = (t', j)$ implies that $t' = t_e$. Given a role name r, we now define

$$r^{\mathcal{I}} := \{((t,j),(t',j')) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid \exists e \in \sigma(\mathcal{U}^t) : \pi_t(e) = (t',j') \wedge e \in \sigma(X_r^t)\}.$$

Note that π_t is a bijection between $\sigma(\mathcal{U}^t)$ and the role successors of $(t,j) \in \Delta^{\mathcal{I}}$. To show this it is enough to prove that, for all $e \in \sigma(\mathcal{U}^t)$, there is a role $r \in N_R$ such that $e \in \sigma(X_r^t)$. This is an immediate consequence of the fact that σ satisfies α_t.

We want to show that the following holds for all types $t \in T_\sigma$ and $j, 1 \leq j \leq \sigma(v_t)$:

$$(t,j) \in C_t^{\mathcal{I}}. \tag{2}$$

Note that, due to the disjointness of the type concepts, this implies that (t,j) cannot be an element of $C_{t'}^{\mathcal{I}}$ for any type $t' \neq t$. As an easy consequence we obtain that $|C_t^{\mathcal{I}}| = \sigma(v_t)$ for all types t. Thus, the fact that σ solves $\phi_{\mathcal{E}}$ implies that $\mathcal{I}$ is a model of $\mathcal{E}$.

It remains to show (2). For this it is sufficient to show the following: for all concept descriptions $C \in \mathcal{M}$, all types $t \in T_\sigma$ and all $j, 1 \leq j \leq \sigma(v_t)$ we have

$$(t,j) \in C^{\mathcal{I}} \text{ iff } C \in t. \tag{3}$$

We show (3) by *induction on the structure of* C:

- Let $C = A$ for a concept name A. Then (3) is an immediate consequence of the definition of $A^{\mathcal{I}}$.
- Let $C = \neg D$. Then induction yields $(t,j) \in D^{\mathcal{I}}$ iff $D \in t$. By contraposition, this is the same as $(t,j) \notin D^{\mathcal{I}}$ iff $D \notin t$. By Condition 1 in the definition of types and the semantics of negation, this is in turn equivalent to $(t,j) \in (\neg D)^{\mathcal{I}}$ iff $\neg D \in t$.
- Let $C = D \sqcap E$. Then induction yields $(t,j) \in D^{\mathcal{I}}$ iff $D \in t$ and $(t,j) \in E^{\mathcal{I}}$ iff $E \in t$. From this, we obtain $(t,j) \in (D \sqcap E)^{\mathcal{I}}$ iff $D \sqcap E \in t$ using Condition 2 in the definition of types and the semantics of conjunction.
- The case where $C = D \sqcup E$ can be handled similarly, using Condition 3 in the definition of types and the semantics of disjunction.
- $C = \text{succ}(c)$ be a successor restriction. First, assume that $C \in t$. Then the translation c' of c using set variables with superscript t is a conjunct in ψ_t. Consequently, σ satisfies this translation c'. We know that the mapping π_t is a bijection between $\sigma(\mathcal{U}^t)$ and the set $ars^{\mathcal{I}}(t,j)$ of role successors of (t,j) in $\mathcal{I}$. Because of our definition of the interpretation of roles in $\mathcal{I}$, we know more precisely that π_t is also a bijection between $\sigma(X_r^t)$ and $r^{\mathcal{I}}(t,j)$. It is thus sufficient to show that π_t is a bijection between $\sigma(X_D^t)$ and $D^{\mathcal{I}} \cap ars^{\mathcal{I}}(t,j)$ for all concept descriptions D occurring in the constraint c. In fact, then π_t is an "isomorphism" between $\sigma(\mathcal{U}^t)$ and $ars^{\mathcal{I}}(t,j)$, and thus the fact that σ satisfies c' implies that $(t,j) \in \text{succ}(c)^{\mathcal{I}}$. By induction, we know that the equivalence (3) holds for the concept descriptions D occurring in c. Let $e \in \sigma(X_D^t)$ and $\pi_t(e) = (t', i')$. Consequently, we have $t' = t_e$, and thus $D \in t'$. By induction, we obtain $(t', i') \in D^{\mathcal{I}}$, and we already know that $(t', i') \in ars^{\mathcal{I}}(t,j)$. This shows that π_t is an injective mapping from $\sigma(X_D^t)$ into $D^{\mathcal{I}} \cap ars^{\mathcal{I}}(t,j)$. To show surjectivity, assume that $(t', i') \in D^{\mathcal{I}} \cap ars^{\mathcal{I}}(t,j)$. By induction $(t', i') \in D^{\mathcal{I}}$ yields $D \in t'$. In addition, $(t', i') \in ars^{\mathcal{I}}(t,j)$ implies that there is an $e \in \sigma(\mathcal{U}^t)$ such that $\pi_t(e) = (t', i')$. Consequently, $t' = t_e$, which implies that $e \in \sigma(X_D^t)$, and thus establishes he desired surjectivity result.
  Second, assume that $C \notin t$. Then $\neg\,\text{succ}(c) \in t$, and thus the translation $\neg c'$ of $\neg c$ using set variables with superscript t is a conjunct in ψ_t. We can now proceed as in the first case, but with $\neg c$ and $\neg c'$ in place of c and c'.

This completes the proof of (3) and thus the proof of the lemma. □

Since satisfiability of QFBAPA formulae can be decided within NP even for binary coding of number [10], this lemma shows that consistency of $\mathcal{ALCSCC}$ ECBoxes can be decided within NExpTime. Together with the known NExpTime lower bound for consistency of $\mathcal{ALC}$ ECBoxes [4], this yields:

THEOREM 5. *Consistency of ECBoxes with numbers encoded in binary is NExpTime-complete in* $\mathcal{ALCSCC}$.

4 RESTRICTED CARDINALITY CONSTRAINTS

For $\mathcal{ALC}$, a restricted notion of ECBoxes was introduced in [4], and it was show that this restriction lowers the complexity of the consistency problem from NExpTime to ExpTime. We will show below that the same is true for $\mathcal{ALCSCC}$.

Restricted cardinality constraints on $\mathcal{ALCSCC}$ *concepts* are defined as follows:

- *restricted* $\mathcal{ALCSCC}$ *cardinality constraints* are of the form
  $$N_1|C_1| + \cdots + N_k|C_k| \leq N_{k+1}|C_{k+1}| + \cdots + N_{k+\ell}|C_{k+\ell}|,$$
  where C_i are $\mathcal{ALCSCC}$ concept descriptions and N_i are integer constants for $1 \leq i \leq k + \ell$;
- a *restricted* $\mathcal{ALCSCC}$ *cardinality box (RCBox)* is a conjunction of restricted $\mathcal{ALC}$ cardinality constraints.

Since ExpTime hardness already holds for restricted $\mathcal{ALC}$ cardinality boxes [4], we obtain the following complexity lower bounds for $\mathcal{ALCSCC}$. Actually, the hardness proof does not require large number, and thus ExpTime-hardness even holds for unary coding of numbers.

PROPOSITION 6. *Consistency of* $\mathcal{ALCSCC}$ *RCBoxes is ExpTime-hard, independently of whether numbers are encoded in unary or binary.*

We show the ExpTime upper bound for numbers encoded in binary using type elimination, where the notion of augmented type from [1] is used, and a second step for removing types is added to take care of the RCBox, similar to what is done in the type elimination procedure for $\mathcal{ALC}$ RCBoxes in [4].

The ExpTime upper bound for our procedure on the one hand depends on the following lemma, which applies in our setting due to the special form of RCBoxes. A proof of this lemma can be found in [4].

Lemma 7. *Let ϕ be a system of linear inequalities consisting of $A \cdot \boldsymbol{v} \geq \mathbf{0}$, $\boldsymbol{v} \geq \mathbf{0}$, and $B \cdot \boldsymbol{v} \geq \mathbf{1}$, where A, B are matrices of integer coefficients and $\boldsymbol{v}$ is the variable vector.*

1. *Deciding whether ϕ has a non-negative integer solution can be done in polynomial time.*
2. *The solutions of ϕ are closed under addition, and in particular we have the following: if $\{v_1, \ldots, v_k\}$ is a set of variables such that, for all v_i $(1 \leq i \leq k)$, ϕ has a solution σ_i in which v_i has a non-zero value, then there is a non-negative integer solution σ of ϕ such that $\sigma(v_i) \geq 1$ for all $i, 1 \leq i \leq k$.*

Another important ingredient of our ExpTime procedure are augmented types, which have been introduced in [1] to show that satisfiability in $\mathcal{ALCSCC}$ w.r.t. concept inclusions is in ExpTime.

Augmented types consider not just the concepts to which a single individual belongs, but also the Venn regions to which its role successors belong. Given a type t for $\mathcal{R}$, we consider the corresponding QFBAPA formula ϕ_t, which is induced by the (possibly negated) successor constraints occurring in t. We conjoin to this formula the set constraint $X_{r_1} \cup \ldots \cup X_{r_n} = \mathcal{U}$, where $N_R = \{r_1, \ldots, r_n\}$.[3] For the resulting formula ϕ'_t, we compute the number N_t that bounds the number of Venn regions that need to be non-empty in a solution of ϕ'_t (see Lemma 1).

Definition 8. *Let $\mathcal{R}$ be an $\mathcal{ALCSCC}$ RCBox. An* augmented type *(t, V) for $\mathcal{R}$ consists of a type t for $\mathcal{R}$ together with a set of Venn region V such that $|V| \leq N_t$ and the formula ϕ'_t has a solution in which exactly the Venn regions in V are non-empty.*

The existence of a solution of ϕ'_t in which exactly the Venn regions in V are non-empty can obviously be checked (within NP) by adding to ϕ'_t conjuncts that state non-emptiness of the Venn regions in V and the fact that the union of these Venn regions is the universal set (see the description of the PSpace algorithm in the proof of Theorem 1 in [1]). Another easy to show observation is that there are only exponentially many augmented types (see the accompanying technical report of [1] for a proof of the following lemma).

Lemma 9. *Let $\mathcal{R}$ be an $\mathcal{ALCSCC}$ RCBox. The set of augmented types for $\mathcal{R}$ contains at most exponentially many elements in the size of $\mathcal{R}$ and it can be computed in exponential time.*

Basically, type elimination starts with the set of all augmented types, and then successively eliminates augmented types (i) whose Venn regions are not realized by the currently available augmented types, or (ii) whose first component is forced to be empty by the constraints in $\mathcal{R}$. To make the first reason for elimination more precise, assume that $\mathcal{A}$ is a set of augmented types and that v is a Venn region. The Venn region v yields a set of concept descriptions S_v that contains, for every set variable X_D occurring in v, the element D in case v contains X_D and the element $\neg D$ in case v contains X_D^c. It is easy to see that S_v is actually a subset of $\mathcal{M}$ (modulo removal of double negation). We say that v is *realized by* $\mathcal{A}$ if there is an augmented type $(t, V) \in \mathcal{A}$ such that $S_v \subseteq t$.

Algorithm 10. *Let $\mathcal{R}$ be an $\mathcal{ALCSCC}$ RCBox. The following steps decides consistency of $\mathcal{R}$:*

1. *Compute the set $\mathcal{M}$ consisting of all subdescriptions of $\mathcal{R}$ as well as the negations of these subdescriptions, and continue with the next step.*
2. *Based on $\mathcal{M}$, compute the set $\mathcal{A}$ of all augmented types for $\mathcal{R}$, and continue with the next step.*
3. *If the current set $\mathcal{A}$ of augmented types is empty, then the algorithm* fails. *Otherwise, check whether $\mathcal{A}$ contains an element (t, V) such that not all the Venn regions in V are realized by $\mathcal{A}$. If there is no such element (t, V) in $\mathcal{A}$, then continue with the next step. Otherwise, let (t, V) be such an element, and set $\mathcal{A} := \mathcal{A} \setminus \{(t, V)\}$. Continue with this step, but now using the new current set of augmented types.*
4. *Let $T_{\mathcal{A}} := \{t \mid$ there is V such that $(t, V) \in \mathcal{A}\}$, and let $\phi_{T_{\mathcal{A}}}$ be obtained from $\mathcal{R}$ by replacing each $|C|$ in $\mathcal{R}$ with $\sum_{t \in T_{\mathcal{A}} \text{ s.t. } C \in t} v_t$ and adding $v_t \geq 0$ for each $t \in T_{\mathcal{A}}$. Check whether $T_{\mathcal{A}}$ contains an element t such that $\phi_{T_{\mathcal{A}}} \wedge v_t \geq 1$ has no solution. If this is the case for t, then remove all augmented types of the form $(t, \cdot)$ from $\mathcal{A}$, and continue with the previous step. If no type t is removed in this step, then the algorithm* succeeds.

Lemma 11 (Soundness). *Let $\mathcal{R}$ be an $\mathcal{ALCSCC}$ RCBox. If the Algorithm 10 succeeds on input $\mathcal{R}$, then $\mathcal{R}$ is consistent.*

Proof. Assume that the algorithm succeeds on input $\mathcal{R}$, and let $\mathcal{A}$ be the final set of augmented types when the algorithm stops successfully. We show how $\mathcal{A}$ can be used to construct a model $\mathcal{I}$ of $\mathcal{R}$. For this construction, we first consider the formula $\phi_{T_{\mathcal{A}}}$, which is obtained from $\mathcal{R}$ by replacing each $|C|$ in $\mathcal{R}$ with $\sum_{t \in T_{\mathcal{A}} \text{ s.t. } C \in t} v_t$ and adding $v_t \geq 0$ for each $t \in T_{\mathcal{A}}$. Since the algorithm has terminated successfully, we know for all $t \in T_{\mathcal{A}}$ that the formula $\phi_{T_{\mathcal{A}}} \wedge v_t \geq 1$ has a solution. By Lemma 7 this implies that $\phi_{T_{\mathcal{A}}}$ has a solution in which all variables v_t for $t \in T_{\mathcal{A}}$ have a value ≥ 1 and all variables v_t with $t \notin T_{\mathcal{A}}$ have value 0. In addition, given an arbitrary number $N \geq 1$, we know that there is a solution σ_N of $\phi_{T_{\mathcal{A}}}$ such that $\sigma_N(v_t) \geq 1$ and $N | \sigma_N(v_t)$ holds for all $t \in T_{\mathcal{A}}$. To see this, note that we can just multiply with N a given solution satisfying the properties mentioned in the previous sentence.

We use the augmented types in $\mathcal{A}$ to determine the right N:

- For each augmented type (t, V), we know that the formula ϕ'_t has a solution where exactly the Venn regions in V are non-empty (see Definition 8). Assume that this solution assigns a set of cardinality $k_{(t,V)}$ to the universal set.
- For each $t \in T_{\mathcal{A}}$, let n_t be the cardinality of the set $\{V \mid (t, V) \in \mathcal{A}\}$, i.e., the number of augmented types in $\mathcal{A}$ that have t as their first component.

We now define N as $N := (\max\{k_{(t,V)} \mid (t, V) \in \mathcal{A}\}) \cdot \prod_{t \in T_{\mathcal{A}}} n_t$, and use the solution σ_N of $\phi_{T_{\mathcal{A}}}$ to construct a finite interpretation

[3] Without loss of generality we assume that N_R contains only the role names occurring in $\mathcal{R}$.

$\mathcal{I}$ as follows. The domain of $\mathcal{I}$ is defined as

$$\Delta^{\mathcal{I}} := \{(t,V)^i \mid (t,V) \in \mathcal{A} \text{ and } 1 \leq i \leq \sigma_N(v_t)/n_t\}.$$

Note that $\sigma_N(v_t)/n_t$ is a natural number since $N|\sigma_N(v_t)$ implies $n_t|\sigma_N(v_t)$. In addition, $\Delta^{\mathcal{I}} \neq \emptyset$ because $\mathcal{A} \neq \emptyset$ and $\sigma_N(v_t)/n_t \geq 1$ since $\sigma_N(v_t) \geq 1$. Moreover, for each type $t \in T_{\mathcal{A}}$, the set $\{(t,V)^i \mid (t,V)^i \in \Delta^{\mathcal{I}}\}$ has cardinality $\sigma_N(v_t)$.

The interpretation of the concept names A is based on the occurrence of these names in the first component of an augmented type, i.e., $A^{\mathcal{I}} := \{(t,V)^i \in \Delta^{\mathcal{I}} \mid A \in t\}$.

Defining the interpretation of the role names is a bit more tricky. Obviously, it is sufficient to define, for each role name $r \in N_R$ and each $d \in \Delta^{\mathcal{I}}$, the set $r^{\mathcal{I}}(d)$. Thus, consider an element $(t,V)^i \in \Delta^{\mathcal{I}}$. Since (t,V) is an augmented type in $\mathcal{A}$, the formula ϕ'_t has a solution σ in which exactly the Venn regions in V are non-empty, and which assigns a set of cardinality $m := k_{(t,V)}$ to the universal set. In addition, each Venn region $w \in V$ is realized by an augmented type $(t^w, V^w) \in \mathcal{A}$. Assume that the solution σ assigns the finite set $\{d_1,\ldots,d_m\}$ to the set term $\mathcal{U}$. We consider an injective mapping π of $\{d_1,\ldots,d_m\}$ into $\Delta^{\mathcal{I}}$ such that the following holds for each element d_j of $\{d_1,\ldots,d_m\}$: if d_j belongs to the Venn region $w \in V$, then $\pi(d_j) = (t^w,V^w)^\ell$ for some $1 \leq \ell \leq \sigma_N(v_{t^w})/n_{t^w}$. Such a bijection exists since $\sigma_N(v_{t^w})/n_{t^w} \geq \max\{k_{(t',V')} \mid (t',V') \in \mathcal{A}\} \geq k_{(t,V)} = m$. We now define

$$r^{\mathcal{I}}((t,V)^i) := \{\pi(d_j) \mid d_j \in \sigma(X_r)\}.$$

Soundness of Algorithm 10 is now an easy consequence of the following claim:

Claim: *For all $C \in \mathcal{M}$, $(t,V) \in \mathcal{A}$, and i, $1 \leq i \leq \sigma_N(v_t)/n_t$ we have $C \in t$ iff $(t,V)^i \in C^{\mathcal{I}}$.*

We prove the claim by induction on the size of C:

- The cases $C = A$, $C = \neg D$, $C = D_1 \sqcap D_2$, and $C = D_1 \sqcup D_2$ can be handled as in the proof of (3) in the proof of Lemma 4.
- Now assume that $C = succ(c)$ for a set or cardinality constraint c.
  - If $C \in t$, then this constraint is part of the QFBAPA formula ϕ'_t obtained from t, and thus satisfied by the solution σ of ϕ'_t used to define the role successors of $(t,V)^i$. According to this definition, there is a 1–1 correspondence between the elements of $\sigma(\mathcal{U})$ and the role successors of $(t,V)^i$. This bijection π also respects the assignment of subsets of $\sigma(\mathcal{U})$ to set variables of the form X_r (for $r \in N_R$) and X_D (for concept descriptions D) occurring in ϕ'_t, i.e.,

    (∗) $d_j \in \sigma(X_r)$ iff $\pi(d_j) \in r^{\mathcal{I}}((t,V)^i)$ and $d_j \in \sigma(X_D)$ iff $\pi(d_j) \in D^{\mathcal{I}}$.

    Once (∗) is shown it is clear that $(t,V)^i \in succ(c)^{\mathcal{I}} = C^{\mathcal{I}}$. In fact, the translation ϕ_c of c, where r is replaced by X_r and D by X_D, is a conjunct in ϕ'_t and thus σ satisfies ϕ_c. Now (∗) shows that (modulo the application of the bijection π), when checking whether $(t,V)^i \in succ(c)^{\mathcal{I}}$, roles r and concepts D in ϕ_c are interpreted in the same way as the set variables X_r and X_D in c, respectively. Thus the fact that σ satisfies ϕ_c implies that the role successors of $(t,V)^i$ satisfy c, i.e., $(t,V)^i \in succ(c)^{\mathcal{I}}$ holds.

    For role names r, property (∗) is immediate by the definition of $r^{\mathcal{I}}((t,V)^i)$. Now consider a concept description D such that X_D occurs in ϕ'_t. Then D occurs in c, and is thus smaller than C, which means that we can apply induction to it. If $d_j \in \sigma(X_D)$, then the Venn region w to which d_j belongs contains X_D positively. Consequently, S_w contains D, and the augmented type (t^w,V^w) realizing w satisfies $D \in t_w$. By induction, we obtain $\pi(d_j) = (t^w,V^w)^\ell \in D^{\mathcal{I}}$. Conversely, assume that $\pi(d_j) = (t^w,V^w)^\ell \in D^{\mathcal{I}}$, where w is the Venn region to which d_j belongs w.r.t. σ. By induction, we obtain $D \in t_w$, and thus the Venn region w contains X_D positively. Since d_j belongs to this Venn region, we obtain $d_j \in \sigma(X_D)$.
  - The case where $C \notin t$ can be treated similarly. In fact, in this case the constraint $\neg c$ is part of the QFBAPA formula ϕ'_t obtained from t, and we can employ the same argument as above, just using $\neg c$ instead of c.

This finishes the proof of the claim. As an easy consequence of this claim we have for all C occurring in $\mathcal{R}$ that

$$C^{\mathcal{I}} = \{(t,V)^i \mid C \in t, (t,V) \in \mathcal{A}, \text{ and } 1 \leq i \leq \sigma_N(v_t)/n_t\}.$$

Consequently, $|C^{\mathcal{I}}| = \sum_{t \in T_{\mathcal{A}} \text{ s.t. } C \in t} \sigma_N(v_t)$, which shows that $\mathcal{I}$ satisfies $\mathcal{R}$ since σ_N solves $\phi_{T_{\mathcal{A}}}$. □

Lemma 12 (Completeness). *Let $\mathcal{R}$ be an $\mathcal{ALCSCC}$ RCBox. If $\mathcal{R}$ is consistent, then Algorithm 10 succeeds on input $\mathcal{R}$.*

Proof. Assume that $\mathcal{I}$ is a model of $\mathcal{R}$. Consider the set of all types of elements of $\mathcal{I}$, i.e.,

$$T_{\mathcal{I}} := \{t_{\mathcal{I}}(d) \mid d \in \Delta^{\mathcal{I}}\}.$$

As mentioned below Definition 3, the elements of $T_{\mathcal{I}}$ are indeed a set of types according to Definition 3. In addition, since $\Delta^{\mathcal{I}} \neq \emptyset$, we also know that $T_{\mathcal{I}} \neq \emptyset$.

First, note that no element of $T_{\mathcal{I}}$ can be removed in Step 4 of our algorithm. This is an easy consequence of the following observation. Let T be a set of types such that $T_{\mathcal{I}} \subseteq T$, and let ϕ_T be obtained from $\mathcal{R}$ by replacing each $|C|$ in $\mathcal{R}$ with $\sum_{t \in T \text{ s.t. } C \in t} v_t$ and adding $v_t \geq 0$ for each $t \in T$. Since $\mathcal{I}$ is a model of $\mathcal{R}$, it is easy to see that ϕ_T has a solution that also satisfies $v_t \geq 1$ for all $t \in T_{\mathcal{I}}$.

Regarding Step 3 of our algorithm, we want to show that for every type $t \in T_{\mathcal{I}}$ there is at least one set of Venn regions V such that the augmented type (t,V) is not removed in this step. To this purpose, let us now extend the types in $T_{\mathcal{I}}$ by adding appropriate Venn regions as second components. Consider $t := t_{\mathcal{I}}(d)$ for an element $d \in \Delta^{\mathcal{I}}$. Then the QFBAPA formula ϕ'_t corresponding to t has a solution σ in which the universal set $\mathcal{U}$ consists of all the role successors of d, and the other set variables are assigned sets according to the interpretations of roles and concept descriptions in the model $\mathcal{I}$. Let $\{d_1,\ldots,d_m\} = \sigma(\mathcal{U})$ be the set of all role successors of d, and w_i the Venn region to which d_i belongs w.r.t. σ. By Lemma 1, there is a solution σ' of ϕ'_t such that the set V of non-empty Venn regions w.r.t. σ' has cardinality $\leq N_t$ and each of these non-empty Venn regions in V is one of the Venn regions w_i, i.e., $V \subseteq \{w_1 \ldots, w_m\}$. By construction, (t,V) is an augmented type. Let $\mathcal{A}_{\mathcal{I}}$ denote the set of augmented types obtained by extending the types in $T_{\mathcal{I}}$ in this way for every $d \in \Delta^{\mathcal{I}}$. By construction, for every $t \in T_{\mathcal{I}}$ there is a set of Venn regions V such that $(t, V) \in \mathcal{A}_{\mathcal{I}}$. In particular, this yields $\mathcal{A}_{\mathcal{I}} \neq \emptyset$ and $T_{\mathcal{A}_{\mathcal{I}}} = T_{\mathcal{I}}$.

Next, we show that the Venn regions occurring in some augmented type in $\mathcal{A}_{\mathcal{I}}$ are realized by $\mathcal{A}_{\mathcal{I}}$. Thus, let (t, V) be an augmented type constructed from a type $t = t_{\mathcal{I}}(d)$ as described above, and let $w \in V$ be a Venn region occurring in this augmented type. Then there is a role successor d_i of d such that d_i belongs to the Venn region $w = w_i$ w.r.t. the solution σ of ϕ'_t induced by $\mathcal{I}$. We know that $d_i \in D^{\mathcal{I}}$ for all $D \in S_w$, and thus $S_w \subseteq t_{\mathcal{I}}(d_i)$. Since $\mathcal{A}_{\mathcal{I}}$ contains an augmented type with first component $t_{\mathcal{I}}(d_i)$, this shows that w is realized by $\mathcal{A}_{\mathcal{I}}$.

We claim that, during the run of Algorithm 10, we always have $\mathcal{A}_{\mathcal{I}} \subseteq \mathcal{A}$ and $T_{\mathcal{I}} \subseteq T_{\mathcal{A}}$. Obviously, this is true after Step 2 of our algorithm. In addition, in Step 3 of our algorithm, no element of $\mathcal{A}_{\mathcal{I}}$ can be removed since we have seen that the Venn regions occurring in some augmented type in $\mathcal{A}_{\mathcal{I}}$ are realized by $\mathcal{A}_{\mathcal{I}}$. Finally, we have also seen above that in Step 4 of our algorithm, no element of $T_{\mathcal{I}} = T_{\mathcal{A}_{\mathcal{I}}}$ can be removed.

Since $\mathcal{A}_{\mathcal{I}}$ is non-empty, this shows that the algorithm cannot fail on input $\mathcal{R}$, and thus must succeed. This completes the proof of completeness. □

Theorem 13. *Consistency of $\mathcal{ALCSCC}$ RCBoxes is ExpTime-complete.*

Proof. It remains to prove that Algorithm 10 indeed runs in exponential time. To see this, first note that, according to Lemma 9, there are only exponentially many augmented types, and they can be computed in exponential time. Thus, the first two steps of the algorithm take only exponential time. In addition, the iteration between Steps 3 and 4 can happen only exponentially often since in each iteration at least one augmented type is removed. A single Step 3 takes only exponential time since for each of the exponentially many augmented types (t, V), only exponentially many other augmented types need to be considered. Finally, a single Step 4 takes only exponential time. In fact, we need to consider exponentially many systems of linear inequalities $\phi_{T_{\mathcal{A}}} \wedge v_t \geq 1$. Each of these systems may be of exponential size, but its solvability can be tested in time that is polynomial in this size (according to Lemma 7), and thus exponential in the size of the input. □

5 CONCLUSION

In two previous papers we have shown how to use QFBAPA constraints to extend the expressive power of number restrictions, on the hand, and of cardinality restrictions on concepts, on the other hand. In the present paper, we have shown that these two extensions can be combined without increasing the complexity of reasoning. In fact, reasoning w.r.t. ECBoxes (RCBoxes) in $\mathcal{ALCSCC}$, the extension of $\mathcal{ALC}$ with more expressive number restrictions, has the same complexity as in $\mathcal{ALC}$: NExpTime-complete for ECBoxes and ExpTime-complete for RCBoxes. This combination is non-trivial since role successors required by local constraints might actually be prevented to exist by global constraints. For ECBoxes, we have shown that this can be taken care of by inequalities that basically relate cardinalities of global types with cardinalities of local Venn regions (see the definition of the formulae γ_t in (1)). For RCBoxes, stating such inequalities explicitly is not necessary since in this restricted case the cardinalities of global types can be made as large as is needed to satisfy the local constraints (see the definition of the solution σ_N in the proof of Lemma 11). Technically, in both cases the fact that the interaction between global and local constraints is handled appropriately shows up in the definition of the interpretation of roles in the soundness proofs, where we see that sufficiently many copies of the (augmented) types belong to the domain of the interpretation. The combined logic can, for instance, be used to check the correctness of statistical statements. For example, if a German car company claims that they have produced more than N cars in a certain year, and P% of the tires used for their cars were produced by Betteryear, this may be contradictory to a statement of Betteryear that they have sold less than M tires in Germany. This information can be expressed using $\mathcal{ALCSCC}$ ECBoxes, and thus the contradiction can be found using our consistency algorithm. It would not have been possible to express this using $\mathcal{ALC}$ ECBoxes since stating how many tires a car has requires local number restrictions.

REFERENCES

[1] Franz Baader. 2017. A New Description Logic with Set Constraints and Cardinality Constraints on Role Successors. In *Proceedings of the 11th International Symposium on Frontiers of Combining Systems (FroCoS'17) (Lecture Notes in Computer Science)*, Clare Dixon and Marcelo Finger (Eds.), Vol. 10483. Springer-Verlag, Brasília, Brazil, 43–59.

[2] Franz Baader, Martin Buchheit, and Bernhard Hollunder. 1996. Cardinality Restrictions on Concepts. *Artificial Intelligence* 88, 1–2 (1996), 195–213.

[3] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (Eds.). 2003. *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge University Press.

[4] Franz Baader and Andreas Ecke. 2017. Extending the Description Logic ALC with More Expressive Cardinality Constraints on Concepts. In *GCAI 2017. 3rd Global Conference on Artificial Intelligence (EPiC Series in Computing)*, Vol. 50. EasyChair, 6–19.

[5] Alexander Borgida, Ronald J. Brachman, Deborah L. McGuinness, and Lori Alperin Resnick. 1989. CLASSIC: A Structural Data Model for Objects. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data.* 59–67.

[6] Stéphane Demri and Denis Lugiez. 2010. Complexity of modal logics with Presburger constraints. *J. Applied Logic* 8, 3 (2010), 233–252.

[7] Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. 2015. The Role of Ontologies in Biological and Biomedical Research: A Functional Perspective. *Brief. Bioinform.* 16, 6 (2015), 1069–1080.

[8] Bernhard Hollunder and Franz Baader. 1991. Qualifying Number Restrictions in Concept Languages. In *Proc. of the 2nd Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'91).* 335–346.

[9] Bernhard Hollunder, Werner Nutt, and Manfred Schmidt-Schauß. 1990. Subsumption Algorithms for Concept Description Languages. In *Proc. of the 9th Eur. Conf. on Artificial Intelligence (ECAI'90).* Pitman, London (United Kingdom), 348–353.

[10] Viktor Kuncak and Martin C. Rinard. 2007. Towards Efficient Satisfiability Checking for Boolean Algebra with Presburger Arithmetic. In *Proc. of the 21st Int. Conf. on Automated Deduction (CADE-07) (Lecture Notes in Computer Science)*, Frank Pfenning (Ed.), Vol. 4603. Springer, 215–230.

[11] Klaus Schild. 1991. A Correspondence Theory for Terminological Logics: Preliminary Report. In *Proc. of the 12th Int. Joint Conf. on Artificial Intelligence (IJCAI'91).* 466–471.

[12] Stephan Tobies. 1999. A PSPACE Algorithm for Graded Modal Logic. In *Proc. of the 16th Int. Conf. on Automated Deduction (CADE'99) (Lecture Notes in Artificial Intelligence)*, Harald Ganzinger (Ed.), Vol. 1632. Springer-Verlag, 52–66.

[13] Stephan Tobies. 2000. The Complexity of Reasoning with Cardinality Restrictions and Nominals in Expressive Description Logics. *J. of Artificial Intelligence Research* 12 (2000), 199–217.

[14] Stephan Tobies. 2001. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation.* Ph.D. Dissertation. LuFG Theoretical Computer Science, RWTH-Aachen, Germany. http://lat.inf.tu-dresden.de/research/phd/Tobies-PhD-2001.pdf

[15] Kuat Yessenov, Ruzica Piskac, and Viktor Kuncak. 2010. Collections, Cardinalities, and Relations. In *Proc. of the 11th International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI'10) (Lecture Notes in Computer Science)*, Gilles Barthe and Manuel V. Hermenegildo (Eds.), Vol. 5944. Springer-Verlag, 380–395.