Provenance Analysis for Logic and Games

Erich Grädel
RWTH Aachen University, Germany
graedel@logic.rwth-aachen.de

Val Tannen
Univ. of Pennsylvania, U.S.A.
val@cis.upenn.edu

Abstract

A model checking computation checks whether a given logical sentence is true in a given finite structure. Provenance analysis abstracts from such a computation mathematical information on how the result depends on the atomic data that describe the structure. In database theory, provenance analysis by interpretations in commutative semirings has been rather successful for positive query languages (such as unions of conjunctive queries, positive relational algebra, or datalog). However, it did not really offer an adequate treatment of negation or missing information. Here we propose a new approach for the provenance analysis of logics with negation, such as first-order logic and fixed-point logics. It is closely related to a provenance analysis of the associated model-checking games, and based on new semirings of dual-indeterminate polynomials or dual-indeterminate formal power series. These are obtained by taking quotients of traditional provenance semirings by congruences that are generated by products of positive and negative provenance tokens. Beyond the use for model-checking problems in logics, provenance analysis of games is of independent interest. Provenance values in games provide detailed information about the number and properties of the strategies of the players, far beyond the question whether or not a player has a winning strategy from a given position.

2012 ACM Subject Classification Theory of Computation → Finite Model Theory

Keywords and phrases Finite Model Theory, Provenance, Games

1 Introduction

Provenance analysis aims at understanding how the result of a computational process with a complex input, consisting of multiple items, depends on the various parts of this input. In database theory, provenance analysis based on interpretations in commutative semirings has been developed for positive database query languages, to understand which combinations of the atomic facts in a database can be used for deriving the result of a given query. In this approach, atomic facts are interpreted not just by true or false, but by values in an appropriate semiring, where 0 is the value of false statements, whereas any element \( a \neq 0 \) of the semiring stands for some shade of truth. These values are then propagated from the atomic facts to arbitrary queries in the language, which permits to answer questions such as the minimal cost of a query evaluation, the confidence one can have that the result is true, the number of different ways in which the result can be computed, or the clearance level that is required for obtaining the output, under the assumption that some facts are labelled as confidential, secret, top secret, etc. We refer to [15] for a recent account and many references on the semiring framework for database provenance.

Scenarios to which the semiring provenance approach has been successfully applied include unions of conjunctive queries, positive relational algebra, nested relations, Datalog, XQuery, SQL-aggregates and several others, and it has been implemented in software systems such as Orchestra and Propolis. For details, see e.g. [2, 6, 12, 14, 17]. A main limitation of this approach is that is has been largely confined to positive query languages. Attempts to add operations that capture difference of relations have led to interesting and
Provenance Analysis for Logic and Games

algebraically challenging, but divergent approaches [1, 8, 9, 13]. In particular there has been no systematic approach in database theory for tracking negative information, and no convincing provenance analysis for languages with full negation.

Here, we would like to develop a new approach for a semiring provenance analysis for model checking problems of logics with negation, in particular first-order logic and fixed-point logic. This approach is based on several ideas:

- Provenance analysis of logics is intimately connected to provenance analysis of games. In the same way as formula evaluation or model checking can be formulated in game theoretic terms, also the propagation of provenance values from atomic facts to arbitrary formulae can be viewed as a process on the associated games. Also the typical results of a provenance analysis of database queries or logical formulae, concerning for instance confidence scores, costs, required clearance level, or number of ‘proof trees’ have natural game-theoretic interpretations. In fact, provenance analysis of games is of independent interest, and provenance values of positions in a game provide detailed information about the number and properties of the strategies of the players, far beyond the question whether or not a player has a winning strategy from a given position.

- We deal with negation by transformation to negation normal form. This is the common approach for the design of model checking games and game-based evaluation algorithms. But while this is there mainly a matter of convenience (to avoid role switches between players during a play), provenance semantics imposes even stronger reasons for transformations to negation normal form. Indeed, beyond Boolean semantics, negation is not a compositional logical operation: the provenance value of \( \neg \varphi \) is not necessarily determined by the provenance value of \( \varphi \).

- On the algebraic side, we introduce new provenance semirings of polynomials and formal power series, which take negation into account. They are obtained by taking quotients of traditional provenance semirings by congruences generated by products of positive and negative provenance tokens; they are called semirings of dual-indeterminate polynomials or dual-indeterminate power series.

Preliminary accounts of our approach, confined to first-order logic and without the connection to games, but discussing potential applications to issues such as reverse provenance analysis, model updates, and confidence maximization, have been given in [18] and [10]. Here we put also the provenance analysis of games into focus, in fact we develop our approach here from the perspectives of games. We shall first discuss the case of finite acyclic games which are sufficient for the provenance analysis of first-order logic and its fragments. Most of the central issues of our approach, in particular the view of provenance values in terms of valuations of strategies and plays, appear already in this simple scenario. We shall then discuss reachability games on graphs that admit cycles. These are the games that are relevant for the provenance analysis of logics with least (but without greatest) fixed points. For these it will be necessary to restrict from arbitrary commutative semirings to \( \omega \)-continuous ones. Such an analysis has previously been carried out for Datalog, but to deal with (atomic) negation we have to combine this with the idea of taking quotients by the duality on indeterminates, which will lead us to semirings of dual-indeterminate power series. Finally we shall outline a provenance approach for safety games and greatest fixed points. Our central algebraic tools here are absorptive semirings, in particular the semiring \( S^\infty[X] \) of generalized absorptive polynomials, admitting also infinite exponents.

This paper is intended to lay foundations for our general approach to a provenance analysis of logic and games, that should take us far beyond the specific cases studied here. The application of the acyclic case to modal and guarded logics has been analysed in [5]. In
our approach has been applied to database repairs; it has been shown how our treat-
ment of negation, or absent information, can be used to provide missing answers and repair
the failure of integrity constraints in databases. Further, the potential of the provenance
methods developed here for applications in knowledge representation and description logics
has been discussed in [4]. Work in progress includes the provenance analysis of temporal
and dynamic logics in the setting of absorptive semirings, the study of logics of dependence
and independence from the point of view of provenance, and the algorithmic analysis of
computing provenance values in various settings.

2 Commutative semirings

Definition 1. A commutative semiring is an algebraic structure \( (K, +, \cdot, 0, 1) \), with \( 0 \neq 1 \),
such that \( (K, +, 0) \) and \( (K, \cdot, 1) \) are commutative monoids, \( \cdot \) distributes over +, and \( 0 \cdot a = a \cdot 0 = 0 \). A semiring is +-positive if \( a + b = 0 \) implies \( a = 0 \) and \( b = 0 \). This excludes
rings. A semiring is root-integral if \( a \cdot a = 0 \) implies \( a = 0 \). All semirings considered in this
paper are commutative, +-positive and root-integral. Further, a commutative semiring is positive
if it is +-positive and has no divisors of \( 0 \) (i.e. \( a \cdot b = 0 \) implies that \( a = 0 \) and
\( b = 0 \)). The standard semirings considered in provenance analysis are in fact positive, but
for an appropriate treatment of negation we shall introduce later in this paper semirings (of
dual-indeterminate polynomials or power series) that have divisors of \( 0 \).

Notice that a semiring \( K \) is positive if, and only if, the unique function \( h : K \to \{0, 1\} \) with \( h^{-1}(0) = \{0\} \) is a homomorphism from \( K \) into the Boolean semiring \( \mathbb{B} = (\{0, 1\}, \lor, \land, 0, 1) \). A semiring \( K \) is (+)-idempotent if \( a + a = a \), for all \( a \in K \), and (+, ·)-
idempotent if, in addition, \( a \cdot a = a \) for all \( a \). Further, \( K \) is absorptive if \( a + ab = a \), for all
\( a, b \in K \). Obviously, every absorptive semiring is (+)-idempotent.

Elements of a commutative semiring will be used as truth values for logical statements
and as values for positions in games. The intuition is that + describes the alternative use
of information, as in disjunctions or existential quantifications, or for different possible choices
of a player in a game, whereas · stands for the joint use of information, as in conjunctions
or universal quantifications, or for choices in a game that are controlled by the opponent of
the given player. Further, 0 is the value of false statements or losing positions, whereas any
element \( a \neq 0 \) of a semiring \( K \) stands for a “nuanced” interpretation of true or as a value of
a non-losing position.

Application semirings. We briefly discuss some specific semirings that provide interesting
information but about a logical statement or a position in a game.

- The Boolean semiring \( \mathbb{B} = (\{0, 1\}, \lor, \land, 0, 1) \) is the standard habitat of logical truth.
- \( \mathbb{N} = (\mathbb{N}, +, \cdot, 0, 1) \) is used here for counting winning strategies in games. It also plays an
important role for bag semantics in databases.
- \( \mathbb{T} = (\mathbb{R}_{\geq 0}, \min, +, \infty, 0) \) is called the tropical semiring. It has many applications in several
areas of computer science. It is used here for measuring the cost of strategies.
- The Viterbi semiring \( \mathbb{V} = (\{0, 1\}, \max, \cdot, 0, 1) \) is isomorphic to \( \mathbb{T} \) via \( x \mapsto e^{-x} \) and \( y \mapsto −\ln y \). We will think of the elements of \( \mathbb{V} \) as confidence scores and use it to describe the
confidence that a player can win from a given position or the confidence assigned to a
logical statement.
- The \( \min\text{-max} \) semiring on a totally ordered set \((A, \leq)\) with least element \( a \) and greatest
element \( b \) is the semiring \((A, \max, \min, a, b)\).
Provenance semirings. Beyond the traditional application semirings, there are some important universal provenance semirings of polynomials that are used for a general provenance analysis. They admit to compute provenance values once in a general semiring and then to specialise these via homomorphisms (i.e. evaluation of the polynomials) to specific application semirings as needed.

- For any set $X$, the semiring $\mathbb{N}[X] = (\mathbb{N}[X], +, \cdot, 0, 1)$ consists of the multivariate polynomials in indeterminates from $X$ and with coefficients from $\mathbb{N}$. This is the commutative semiring freely generated by the set $X$.
- By dropping coefficients from $\mathbb{N}[X]$, we get the semiring $\mathbb{B}[X]$ whose elements are just finite sets of distinct monomials. It is the free $(+)$-idempotent semiring over $X$.
- By dropping also exponents, we get the semiring $\mathbb{W}[X]$ of finite sums of monomials that are linear in each argument. It is sometimes called the Why-semiring.
- The free absorptive semiring $\mathbb{S}[X]$ over $X$ consists of $0, 1$ and all antichains of monomials with respect to the component-wise order on their exponents. It is the quotient of $\mathbb{N}[X]$ by the congruence induced by $p \sim q$ for monomials $p, q$ with $p = qr$.
- Finally $\text{PosBool}(X) = (\text{PosBool}(X), \lor, \land, \bot, \top)$ is the semiring whose elements are classes of equivalent positive (monotone) boolean expressions with variables from $X$ (its elements are in bijection with the positive boolean expressions in irredundant disjunctive normal form). This is the distributive lattice freely generated by the set $X$.

3 Games

We consider two-player turn-based games on graphs. Such a game is defined by the game graph on which it is played, and by the objectives of the players.

► Definition 2. A game graph is a structure $G = (V, V_0, V_1, T, E)$, where $V = V_0 \cup V_1 \cup T$ is the set of positions, partitioned into the sets $V_0$, $V_1$ of the two players and the set $T$ of terminal positions, and where $E \subseteq V \times V$ is the set of moves. We denote the set of immediate successors of a position $v$ by $vE := \{w : (v, w) \in E\}$ and require that $vE = \emptyset$ if, and only if, $v \in T$. A play from an initial position $v_0$ is a finite or infinite path $v_0 v_1 v_2 \ldots$ through $G$ where the successor $v_{i+1} \in v_i E$ is chosen by Player 0 if $v_i \in V_0$ and by Player 1 if $v_i \in V_1$. A play ends when it reaches a terminal node $v_m \in T$.

► Definition 3. For every game graph $G = (V, V_0, V_1, T, E)$, and every initial position $v_0 \in V$, the tree unraveling of $G$ from $v_0$ is the game tree $T(G, v_0)$ consisting of all finite paths from $v_0$. More precisely, $T(G, v) = (V^#, V_0^#, V_1^#, T^#, E^#)$, where $V^#$ is the set of all finite paths $\pi = v_0 v_1 \ldots v_m$ through $G$, with $V_0^# = \{\pi v \in V^# : v \in V_0\}$, $T^# = \{\pi t \in V^# : t \in T\}$, and $E^# = \{(\pi v, \pi v') : (v, v') \in E\}$. For most game-theoretic considerations, the games played on $G$ and its unravelings are equivalent, via the canonical projection from $T(G, v_0)$ to $G$ that maps every path $\pi v$ to its end point $v$.

A strategy for a player in a game is a function that selects moves at points that are controlled by that player. A strategy need not be defined at all positions of a player, but it should be closed in the sense that it defines a move from each position that is reachable by a play that is admitted by the strategy. There are several possibilities to define the notion of a strategy formally. For our purposes it is convenient to identify a strategy with the histories of plays that it admits, i.e. to view it as an appropriate subtree of $T(G, v_0)$.

► Definition 4. A strategy of Player $\sigma$ (for $\sigma \in \{0, 1\}$) from $v_0$ in a game $G$ is a subtree of $T(G, v_0)$, of the form $S = (W, F)$ with $W \subseteq V^#$ and $F \subseteq (W \times W) \cap E^#$, satisfying the following conditions:
We first study the provenance analysis of games for well-founded games, i.e., games that
are played on finite acyclic graphs and hence do not admit infinite plays. Let $K$ be a
commutative semiring, and let $G = (V, V_0, V_1, T, E)$ be a finite acyclic game graph.

A $K$-valuation of $G$ for Player $\sigma$ provides a value $f_\sigma(v) \in K$ for every position $v \in V$. Such a valuation is induced by its values on the terminal positions, i.e., by a function $f_\sigma : T \to K$, and by a valuation of the moves, i.e., by a function $h_\sigma : E \to K \setminus \{0\}$. In many cases valuations of moves are not relevant; we then just put $h_\sigma(vw) = 1$ for all edges $(v, w) \in E$.

The functions $f_\sigma : T \to K$ (for $\sigma \in \{0, 1\}$), define the value of every terminal position from the point of view of Player $\sigma$. Intuitively, $f_\sigma(t) = 0$ means that position $t$ is losing for Player $\sigma$. In the simplest case, we can specify reachability objectives $T_\sigma$ by setting $f_\sigma(t) = 1$ for $t \in T_\sigma$ and $f_\sigma(t) = 0$ otherwise. The functions $h_\sigma : E \to K \setminus \{0\}$ provide a value (or cost) for Player $\sigma$ of the moves.

The extension of the basic valuations $f_\sigma : T \to K$ and $h_\sigma : E \to K \setminus \{0\}$ to valuations $f_\varepsilon : V \to K$ for all positions relies on the idea that a move from $v$ to $w$ contributes to $f_\varepsilon(v)$ the value $h_\varepsilon(vw) \cdot f_\varepsilon(w)$. These contributions are summed up in the case that $v$ is a position

- $W$ is closed under predecessors: if $\pi v \in W$ then also $\pi \in W$.
- If $\pi v \in W \cap V_0^\#$, then $|\pi v| F = 1$.
- If $\pi v \in W \cap V_0^\# \setminus \pi v$, then $\pi v) F = (\pi v) E^\#$.

A strategy can also be viewed as a function $S : W \cap V_0^\# \to V$ such that $S(\pi v) \in v E$ defines
the node to which Player $\sigma$ moves that are admitted by the strategy. A strategy $S$ is well-founded if it does not admit any infinite plays; this is always the case on finite acyclic game graphs, but need not be the case otherwise. The set of possible outcomes of a strategy $S$ is the set of terminal nodes that are reachable by a play that is consistent with $S$.

The simplest objectives of players are reachability and safety objectives.

**Definition 5.** A reachability objective for Player $\sigma$ is given by a set $T_\sigma \subseteq T$ of winning terminal positions. With such an objective, Player $\sigma$ wins every play that reaches a position in $T_\sigma$. Dually, a safety objective for Player $\sigma$ is given by a set $L_\sigma \subseteq T$ of ‘losing’ positions that the player has to avoid, or equivalently, by its complement $S_\sigma = V \setminus L_\sigma$, the region of safe positions inside of which the Player has to keep the play. With such an objective Player $\sigma$ wins every play, finite or infinite, that never reaches a position in $L_\sigma$.

Notice that the difference between reachability and safety objectives is relevant only in cases where infinite plays are possible. Indeed, in a game that admits only finite plays, Player $\sigma$ wins a play with the reachability objective $T_\sigma$ if, and only if, she wins that play with the safety objective given by $L_\sigma = T \setminus T_\sigma$, so we can always reformulate reachability by safety and vice versa. However, in a game that admits infinite plays, Player $\sigma$ wins with a reachability objective $T_\sigma$ if, and only if, her opponent, Player $1 - \sigma$, loses with the safety condition $L_{1 - \sigma} = T_\sigma$. Hence winning with a reachability objective corresponds to defeating an opponent who plays with a safety objectives. If both players play with reachability objectives, then infinite plays are won by neither player.

## 4 Provenance for well-founded games

We first study the provenance analysis of games for well-founded games, i.e., games that are played on finite acyclic graphs and hence do not admit infinite plays. Let $K$ be a
commutative semiring, and let $G = (V, V_0, V_1, T, E)$ be a finite acyclic game graph.

A $K$-valuation of $G$ for Player $\sigma$ provides a value $f_\sigma(v) \in K$ for every position $v \in V$. Such a valuation is induced by its values on the terminal positions, i.e., by a function $f_\sigma : T \to K$, and by a valuation of the moves, i.e., by a function $h_\sigma : E \to K \setminus \{0\}$. In many cases valuations of moves are not relevant; we then just put $h_\sigma(vw) = 1$ for all edges $(v, w) \in E$.

The functions $f_\sigma : T \to K$ (for $\sigma \in \{0, 1\}$), define the value of every terminal position from the point of view of Player $\sigma$. Intuitively, $f_\sigma(t) = 0$ means that position $t$ is losing for Player $\sigma$. In the simplest case, we can specify reachability objectives $T_\sigma$ by setting $f_\sigma(t) = 1$ for $t \in T_\sigma$ and $f_\sigma(t) = 0$ otherwise. The functions $h_\sigma : E \to K \setminus \{0\}$ provide a value (or cost) for Player $\sigma$ of the moves.

The extension of the basic valuations $f_\sigma : T \to K$ and $h_\sigma : E \to K \setminus \{0\}$ to valuations $f_\varepsilon : V \to K$ for all positions relies on the idea that a move from $v$ to $w$ contributes to $f_\varepsilon(v)$ the value $h_\varepsilon(vw) \cdot f_\varepsilon(w)$. These contributions are summed up in the case that $v$ is a position
for Player $\sigma$ (i.e. when she chooses herself the successors), and multiplied in the case that $v$ is a position of the opponent (i.e. when she has to cope with any of the possible successors). Thus the valuations are defined via backwards induction, by

$$f_\sigma(v) := \begin{cases} \sum_{w \in vE} h_\sigma(vw) \cdot f_\sigma(w) & \text{if } v \in V_\sigma \\ \prod_{w \in vE} h_\sigma(vw) \cdot f_\sigma(w) & \text{if } v \in V_{1-\sigma}. \end{cases}$$

An equivalent characterization of the provenance values $f_\sigma(v)$ is obtained by defining provenance values for plays and strategies. For a play $x = v_0v_1 \ldots v_m$ from $v_0$ to a terminal node $v_m$, we define its valuation as $f_\sigma(x) := h_\sigma(v_0v_1) \cdots h_\sigma(v_{m-1}v_m) \cdot f_\sigma(v_m)$. Recall that $\text{Strat}_\sigma(v)$ is the set of all strategies of Player $\sigma$ from $v$, and Plays($\mathcal{S}$) denotes the set of all plays from $v$ that are consistent with $\mathcal{S}$.

**Theorem 6.** For any commutative semiring $K$ and any finite acyclic game $\mathcal{G}$, let $f_\sigma : V \to K$ be the valuation for Player $\sigma$, induced by the valuation $f_\sigma : T \to K$ of the terminal nodes and $h_\sigma : E \to K \setminus \{0\}$ of the moves. Then, for every position $v$

$$f_\sigma(v) = \sum_{S \in \text{Strat}_\sigma(v)} \prod_{x \in \text{Plays}(\mathcal{S})} f_\sigma(x).$$

**Proof.** For terminal positions $v$ the claim is trivially true. So suppose that $v \in V_\sigma$. Then any strategy $S \in \text{Strat}_\sigma(v)$ can be written in the form $S = v \cdot S_w$ for some successor $w \in vE$ and some strategy $S_w \in \text{Strat}_\sigma(w)$. Further any play $x \in \text{Plays}(\mathcal{S})$ has the form $vy$ for some $y \in \text{Plays}(S_w)$. By induction $f_\sigma(w) = \sum_{S_w \in \text{Strat}_\sigma(w)} \prod_{y \in \text{Plays}(S_w)} f_\sigma(y)$

$$= \sum_{v \cdot S_w \in \text{Strat}_\sigma(v)} \prod_{x \in \text{Plays}(S \cdot S_w)} f_\sigma(x) = \sum_{S \in \text{Strat}_\sigma(v)} \prod_{x \in \text{Plays}(\mathcal{S})} f_\sigma(x).$$

Finally, let $v \in V_{1-\sigma}$ with $vE = \{w_1, \ldots, w_n\}$. Every strategy $S \in \text{Strat}_\sigma(v)$ has the form $S = v(S_1 \cup \cdots \cup S_n)$ with $S_i \in \text{Strat}_\sigma(w_i)$ and every play $x \in \text{Plays}(\mathcal{S})$ has the form $vy_i$ for some $y_i \in \text{Plays}(S_i)$. It follows that

$$f_\sigma(v) = \sum_{w_i \in vE} h_\sigma(vw_i) \cdot f_\sigma(w_i) = \prod_{w_i \in vE, S_i \in \text{Strat}_\sigma(w_i)} \prod_{y_i \in \text{Plays}(S_i)} h_\sigma(vw_i) \cdot f_\sigma(y)$$

$$= \sum_{(S_1 \cup \cdots \cup S_n) \in \text{Strat}_\sigma(v)} \prod_{w_i \in vE, y_i \in \text{Plays}(S_i)} f_\sigma(vy) = \sum_{S \in \text{Strat}_\sigma(v)} \prod_{x \in \text{Plays}(\mathcal{S})} f_\sigma(x).$$

From this description, we can derive a number of applications of provenance valuations on games. We first consider the information provided by valuations in the general provenance semirings of polynomials. Let $\mathbb{N}[T]$ be the semiring of polynomials with coefficients in $\mathbb{N}$ over indeterminates $t \in T$, where $T$ is the set of terminal positions in an acyclic game graph $\mathcal{G} = (V, V_0, V_1, T, E)$. Let $f_\sigma : V \to \mathbb{N}[T]$ be the valuation induced by setting $f_\sigma(t) = t$ for $t \in T$. Further, let $h_\sigma(vw) = 1$ for all edges $(v, w)$ so that the value of a play is just its outcome, i.e. the terminal position where it ends.

Clearly, we can write $f_\sigma(v)$ as a sum of monomials $m \cdot t_1^{j_1} \cdots t_k^{j_k}$. This provides a detailed description of the number and properties of the strategies that Player $\sigma$ has from position $v$. 


Theorem 7. The valuation $f_\sigma(v) \in \mathbb{N}[T]$ is the sum of those monomials $m \cdot t_1^{j_1} \ldots t_k^{j_k}$ (with $m, j_1, \ldots, j_k > 0$) such that Player $\sigma$ has precisely $m$ strategies $S \in \text{Strat}_\sigma(v)$ with the property that the set of possible outcomes for $S$ is exactly $\{t_1, \ldots, t_k\}$, and precisely $j_i$ plays that are consistent with $S$ have the outcome $t_i$.

This is an immediate consequence of Theorem 6. In many cases, somewhat less detailed information is sufficient, which can be obtained by valuations in less informative provenance semirings than $\mathbb{N}[T]$:

- Evaluating $f_\sigma(v)$ in the idempotent semiring $\mathbb{B}[T]$ gives us the sum of monomials $t_1^{j_1} \ldots t_k^{j_k}$ for which Player $\sigma$ has at least one strategy whose multiset of admitted outcomes consists of $t_1, \ldots, t_k$ with multiplicities $j_1, \ldots, j_k$, respectively.
- If we evaluate $f_\sigma(v)$ in $\mathbb{W}[T]$ we get the sum of monomials $t_1 \ldots t_m$ such that Player $\sigma$ has a strategy whose set of outcomes is $\{t_1, \ldots, t_m\}$. The information on multiplicities of strategies and outcomes is dropped.
- An interesting case is the evaluation in the absorptive semiring $\mathbb{S}[X]$. For two strategies $S, S' \in \text{Strat}_\sigma(v)$, we say that $S$ absorbs $S'$ if for every terminal position $t \in T$, $S$ admits less plays with outcome $t$ than $S'$. We call $S$ absorption-dominant if it is not absorbed by any other strategy. Now, $f_\sigma(v) \in \mathbb{S}[X]$ is the sum of monomials $t_1^{j_1} \ldots t_k^{j_k}$ that describe precisely the (multiset of outcomes of the) absorption-dominant strategies of Player $\sigma$ from $v$. See Sect. 11 below for a more detailed analysis of absorption among strategies.
- Finally, the evaluation of $f_\sigma(v)$ in $\mathbb{PosBool}[T]$ consists of those monomials $t_1 \ldots t_k$ such that $\{t_1, \ldots, t_k\}$ a minimal set among the sets of outcomes of strategies $S \in \text{Strat}_\sigma(v)$.

Fix any reachability objective $W \subseteq T$. In any of these provenance semirings, we can write the polynomial $f_\sigma(v)$ as a sum $f_\sigma(v) = f^W_\sigma(v) + g^W_\sigma(v)$ where $f^W_\sigma(v)$ is the sum of those monomials that only contain indeterminates in $W$ and $g^W_\sigma(v)$ contains the rest.

Theorem 8. For every subset $W \subseteq T$ and every $v \in V$, Player $\sigma$ has a strategy to reach $W$ from $v$ if, and only if, $f^W_\sigma(v) \neq 0$ (in any of the provenance semirings given above). Moreover, if we set $f(t) = 1$ for $t \in W$ and $f(t) = 0$ for $t \in T \setminus W$, and evaluate $f_\sigma$ in the semiring $\mathbb{N}$ of natural numbers, then $f_\sigma(v)$ is the number of distinct winning strategies for Player $\sigma$ to reach $W$ from $v$.

Evaluation in other application semirings gives further interesting information about strategies:

Cost of strategies. Given a game $G$, we associate with Player 0 cost functions $f_0 : T \to \mathbb{R}_+$ and $h : E \to \mathbb{R}_+$ for the terminal positions and the moves. Further, we define the cost for Player 0 of a play $x = v_0v_1 \ldots v_m$ from an initial position $v_0$ to a terminal position $v_m$ as $c(x) := \sum_{i=0}^{m-1} h(v_i,v_{i+1}) + f_0(v_m)$, and the cost of a strategy $S \in \text{Strat}_0(v)$ is the sum of the costs of all plays that it admits.

Proposition 9. The cost of an optimal strategy from $v$ in $G$ is given by the valuation $f_0(v)$ in the tropical semiring $\mathbb{T} = (\mathbb{R}_+^\infty, \min, +, \infty, 0)$.

Proof. Since the product in $\mathbb{T}$ is addition in $\mathbb{R}_+^\infty$, the cost of a play $x$ for Player 0, as defined above, coincides with the valuation $f_0(x)$ in $\mathbb{T}$. The summation in $\mathbb{T}$ is minimization in $\mathbb{R}_+^\infty$, so from Theorem 8 we get that

$$f_0(v) = \min_{S \in \text{Strat}_0(v)} \sum_{x \in \text{Plays}(S)} f_0(x)$$

describes indeed the minimal cost of a strategy for Player 0 from position $v$. ▶
Clearance levels. The access control semiring is $\mathbb{A} = \{\{P < C < S < T < 0\}, \min, \max, 0, P\}$ where $P$ is “public”, $C$ is “confidential”, $S$ is “secret”, $T$ is “top secret”, and $0$ is “so secret that nobody can access it!” Let $f_\sigma : T \rightarrow \mathbb{A}$ and $h_\sigma : E \rightarrow \mathbb{A} \setminus \{0\}$ define access levels for the terminal positions and the moves for Player $\sigma$, in the sense that Player $\sigma$ can make a move $e$ if, and only if, his personal clearance level is at least $h(e)$ and similarly, he can access a terminal position $t$ if, and only if, his clearance level is at least $f_\sigma(t)$.

**Proposition 10.** The valuation $f_\sigma(v) \in \mathbb{A}$ describes the minimal clearance level that Player $0$ needs to win from position $v$, i.e. to have a strategy that guarantees to reach a terminal position that is accessible for him.

The proof is a straightforward induction.

Confidence in games. Suppose that $f_\sigma : T \rightarrow [0, 1]$ describes the confidence that Player $\sigma$ puts into $t$ being a winning position for her. We want to compute confidence scores $f_\sigma(v)$ to describe the confidence of Player $\sigma$ that she can win from $v$. It is natural to define the confidence score $f_\sigma(v)$ as the maximum of the confidence scores of the successors $w \in vE$ in the case that $v \in V_\sigma$. For confidence scores of combinations of events whose choice is taken by an opponent, such as for the possible moves from a position $v \in V_{1-\sigma}$, there are different approaches in the literature. A popular one, with which we work here, takes the product of the confidence scores of the events from which the opponent chooses. Adopting this definition, the following proposition is immediate.

**Proposition 11.** Confidence scores are computed as semiring valuations $f_\sigma : V \rightarrow \mathbb{V}$ in the Viterbi semiring $\mathbb{V} = ([0, 1], \max, \cdot, 0, 1)$.

Min-Max Games. Finally note that valuations in a min-max semiring $(\mathbb{A}, \max, \min, a, b)$ describe the value of positions in games where Player 0 tries to maximize and Player 1 tries to minimize the outcome of the play.

**Definition 12.** Let $\mathcal{G}$ be a game graph, with valuations $f_0, f_1$ for the two players in a semiring $\mathbb{K}$, and let $U \subseteq V$ be a set of positions. We say that

1. $f_0, f_1$ for the two players are separating on $U$ if for all $u \in U$, either $f_0(u) = 0$ or $f_1(u) = 0$.
2. $f_0, f_1$ are weakly separating on $U$ if $f_0(u)f_1(u) = 0$ for all $u \in U$. Notice that in the case where $K$ has no divisors of 0, weakly separating valuations are in fact separating. 
3. $f_0$ and $f_1$ are strongly separating on $U$, if they are separating, and in addition, $f_0(u) + f_1(u) \neq 0$ for all $u \in U$.

**Proposition 13.** If two valuations $f_0$ and $f_1$ are (weakly) separating on the the terminal positions of $\mathcal{G}$, then they are (weakly) separating on all positions of $\mathcal{G}$.

**Proof.** Recall that all our semirings are assumed to be +-positive. For $v \in V_\sigma$, we have that

$$f_\sigma(v) = \sum_{w \in vE} h(vw) f_\sigma(w) \text{ and } f_{1-\sigma}(v) = \prod_{w \in vE} h(vw) f_{1-\sigma}(w).$$

It follows that $f_0$ and $f_1$ are separating on $v$ if they are separating on all $w \in vE$. Further,

$$f_\sigma(v)f_{1-\sigma}(v) = \left(\sum_{w \in vE} h_\sigma(vw) f_\sigma(w)\right) \left(\prod_{w \in vE} h_{1-\sigma}(vw)f_{1-\sigma}(w)\right) =$$

$$\sum_{w \in vE} \left(h_\sigma(vw)f_\sigma(w)\prod_{w' \in vE} h_{1-\sigma}(vw')f_{1-\sigma}(w')\right) =$$

$$\sum_{w \in vE} \left(h_\sigma(vw)h_{1-\sigma}(vw)f_\sigma(w)f_{1-\sigma}(w)\prod_{w' \in vE \setminus \{w\}} h_{1-\sigma}(vw')f_{1-\sigma}(w').\right).$$
This proves that \(f_0\) and \(f_1\) are weakly separating on \(v\) if they are so on all \(w \in vE\).

The corresponding implication for strongly separating valuations does not hold for all \(^+\)-positive semirings, but it holds for positive ones.

\textbf{Proposition 14.} If two valuations \(f_0\) and \(f_1\) into a positive semiring are strongly separating on the \textit{the terminal positions of \(G\),} then they are so on all positions of \(G\).

\textbf{Proof.} By induction. Assume that \(f_0\) and \(f_1\) are strongly separating on all \(w \in vE\). Then \(f_\sigma(v) + f_{1-\sigma}(v) = 0\) only if \(f_\sigma(w) = 0\) for all \(w \in vE\) and \(f_{1-\sigma}(w) = 0\) for at least one \(w \in vE\). But this implies that \(f_0(w) + f_1(w) = 0\) for some \(w \in vE\) which contradicts our assumption.

Note that for the Boolean semiring \(K = \mathcal{B}\), this is just Zermelo’s Theorem on the determinacy of reachability games on well-founded graph games: from every position, one of the two players has a winning strategy.

\textbf{Counting positional winning strategies?} A strategy is \textit{positional} if it only depends on the current position, and not on the history of the play, i.e. if \(S(\pi v) = S(\pi' v)\) for all \(v\) and all paths \(\pi v, \pi' v\) that lead to \(v\). A positional strategy can be described by a function \(s : V_v \rightarrow V\) or by a subgraph \(S\) of \(G\) (rather than of \(T(G, v_0)\)).

Given that in the study of games there is (for instance for algorithmic reasons) a strong interest in positional strategies, it is reasonable to ask whether there exist valuations in different semirings that would allow us to count just the positional strategies. However, invariance under counting bisimulation shows that this is not possible.

\textbf{Definition 15.} Let \(G = (V, V_0, V', T, E)\) and \(G' = (V', V_0', V'_1, T', E')\) be two game graphs. A \textit{counting bisimulation} between \(G\) and \(G'\) is a relation \(Z \subseteq V \times V\) such that for every pair \((v, v') \in Z\) we have that

\begin{enumerate}
\item \(v \in V\) if, and only if, \(v' \in V_0'\) and \(v \in T\) if, and only if, \(v' \in T'\), and
\item there is a local bijection \(z_{vv'} : vE \rightarrow v'E'\) between the immediate successors of \(v\) and \(v'\) such that \((w, z_{vv'}(w)) \in Z\), for every \(w \in vE\).
\end{enumerate}

We write \(G, v \sim G', v'\) if there is a counting bisimulation \(Z\) between \(G\) and \(G'\) such that \((v, v') \in Z\). Notice that for any game graph \(G\), the relation \(Z = \{(v, \pi v) : v \in V, \pi v \in V^\#\}\) is a counting bisimulation between \(G\) and its unraveling \(T(G, v_0)\).

\(K\)-valuations of games are invariant under counting bisimilarity in the following sense. Let \(G\) and \(G'\) be two acyclic game graphs with \(K\)-valuations \(f_\sigma : T \rightarrow K\) and \(f'_\sigma : T' \rightarrow K\) of the terminal positions and \(h : E \rightarrow K\) and \(h' : E' \rightarrow K\) of the moves. We say that a counting bisimulation \(Z \subseteq V \times V'\) respects these valuations if \(f_\sigma(t) = f'_\sigma(t')\) for all \((t, t') \in Z \cap T \times T'\), and \(h_\sigma(vw) = h'_\sigma(v'w')\) whenever \((v, v') \in Z\) and \((w, w') \in Z\).

\textbf{Proposition 16.} Let \(Z\) be a counting bisimulation between \(G\) and \(G'\) that respects the basic valuations of the terminal positions and the moves. Then \(Z\) respects the valuations of all positions, i.e. \(f_\sigma(v) = f'_\sigma(v')\) for all \((v, v') \in Z\).

\textbf{Proof.} Let \((v, v') \in Z\). If \(v\) and \(v'\) are terminal positions, then \(f_\sigma(v) = f'_\sigma(v')\) by assumption. Otherwise, \(v\) and \(v'\) are both positions of the same player. If they belong to Player \(\sigma\), then \(f_\sigma(v) = \sum_{w \in vE} f_\sigma(w)\). The local bijection \(z_{vv'}\) maps every \(w \in vE\) to some \(w' \in v'E'\) such that, by induction hypothesis, \(f_\sigma(w) = f'_\sigma(w')\). Hence \(f_\sigma(v) = \sum_{w \in vE} f'_\sigma(w)\). If \(v\) and \(v'\) belong to Player \((1 - \sigma)\) the reasoning is completely analogous, taking a product rather than a sum.
In particular $K$-valuations of acyclic games do not change if we replace a game graph $G$ by one of its unravelings $T(G, v)$. Indeed, every valuation $f_\tau : T \rightarrow K$ on the terminal positions of a game graph $G$ extends to the same valuation for $v$ on $G$ as on the tree unraveling $T(G, v)$. On the other side, every strategy on a tree-shaped game graph is positional. Thus the number of positional winning strategies is certainly not invariant under unraveling and hence not definable by valuations in a semiring.

5 Provenance for first-order logic via model checking games and dual-indeterminate polynomials

Given a finite relational vocabulary $\tau$ and a finite non-empty universe $A$, we denote by $\text{Atoms}_A(\tau)$ the set of all atoms $R\pi$ with $R \in \tau$ and $\pi \in A^\ell$. Further, let $\text{NegAtoms}_A(\tau)$ be the set of all negated atoms $\neg R\pi$ where $R\pi \in \text{Atoms}_A(\tau)$, and consider the set of all $\tau$-literals on $A$,

$$\text{Lit}_A(\tau) := \text{Atoms}_A(\tau) \cup \text{NegAtoms}_A(\tau) \cup \{a \text{ op } b : a, b \in A\},$$

where op stands for $=$ or $\neq$.

Definition 17. Given any commutative semiring $K$, a $K$-interpretation (for $\tau$ and $A$) is a function $\pi : \text{Lit}_A(\tau) \rightarrow K$ that maps equalities and inequalities to their truth values 0 or 1.

We have defined in [10] how a semiring interpretation extends to a full valuation $\pi : \text{FO}(\tau) \rightarrow K$ mapping any fully instantiated formula $\psi(\pi)$ (or equivalently, any first-order sentence of vocabulary $\tau \cup A$), to a value $\pi[\psi]$, by setting

$$\pi[\psi \lor \varphi] := \pi[\psi] + \pi[\varphi] \quad \pi[\psi \land \varphi] := \pi[\psi] \cdot \pi[\varphi] \quad \pi[\exists x \varphi(x)] := \sum_{a \in A} \pi[\varphi(a)] \quad \pi[\forall x \varphi(x)] := \prod_{a \in A} \pi[\varphi(a)].$$

Negation is handled via negation normal forms: we set $\pi[\neg \varphi] := \pi[\text{nnf}(\neg \varphi)]$ where $\text{nnf}(\varphi)$ is the negation normal form of $\varphi$.

This is equivalent to the game provenance, as defined above, for the model checking game associated with the formula $\psi$ and the $K$-interpretation $\pi : \text{Lit}_A(\tau) \rightarrow K$. Notice that classically, model checking games are defined for a formula (assumed to be given in negation normal form) and a fixed structure $\mathfrak{A}$ (see e.g. [3, Chap. 4]). However, the game graph of such a model checking game depends only on the formula $\psi$ and the universe $A$ of the given structure $\mathfrak{A}$. It is only the labelling of the terminal positions of the game, as winning for either the Verifier (Player 0) or the Falsifier (Player 1), that depends on which of the literals in $\text{Lit}_A(\tau)$ are true in $\mathfrak{A}$. Hence the definition of a model checking game readily generalizes to our more abstract provenance scenario.

Definition 18. Let $\psi(\pi) \in \text{FO}(\tau)$ be a first-order formula in negation normal form with a relational vocabulary $\tau$, and let $A$ be a (finite) universe. The model checking game $G(A, \psi)$ has positions $\varphi(\pi)$, obtained from a subformula $\varphi(\pi)$ of $\psi$, by instantiating the free variables $\pi$ by a tuple $\pi$ of elements of $A$. At a disjunction $(\psi \lor \varphi)$, Player 0 (Verifier) moves to either $\psi$ or $\varphi$, and at a conjunction, Player 1 (Falsifier) makes an analogous move. At a position $\exists x \varphi(\pi, x)$, Verifier selects an element $b$ and moves to $\varphi(\pi, b)$, whereas at positions $\forall x \varphi(\pi, x)$ the move to to the next position $\varphi(\pi, b)$ is done by Falsifier. The terminal positions of $G(A, \psi)$ are the literals in $\text{Lit}_A(\tau)$. 

A $K$-interpretation $\pi : \text{Lit}_A(\tau) \to K$ thus provides a valuation of the set $T \subseteq \text{Lit}_A(\tau)$ of terminal positions of the model checking game $G(A, \psi)$, for any sentence $\psi \in \text{FO}(\tau \cup A)$. We view it as a valuation $f_0$ for Player 0. The associated valuation $f_1$ for Player 1 is obtained by setting $f_1(\varphi) = \pi[\neg \varphi]$ for any literal $\varphi \in \text{Lit}_A(\tau)$. Both valuations extend to full valuations $f_0$ and $f_1$ of all positions of $G(A, \psi)$, including position $\psi$ itself. The following result is proved by a straightforward induction on formulae.

\textbf{Theorem 19.} For all positions $\varphi$ of $G(A, \psi)$ we have that $f_0(\varphi) = \pi[\varphi]$ and $f_1(\varphi) = \pi[\neg \varphi]$.

Although this theorem holds without any restrictions on the semiring $K$ and the $K$-interpretation $\pi$, not all such $K$-interpretations are really meaningful for logic. Indeed the provenance value of complementary literals $R(\pi)$ and $\neg R(\pi)$ have to be related in a reasonable way, and as a consequence also the general provenance semirings of polynomials need to be modified. In the simplest case a $K$-interpretation defines a unique $\tau$-structure.

\textbf{Definition 20.} A semiring interpretation $\pi : \text{Lit}_A(\tau) \to K$ is \emph{model-defining} if for every atom $\varphi \in \text{Atoms}_A(\tau)$ one of $\pi(\varphi)$ and $\pi(\neg \varphi)$ is 0, and the other is $\neq 0$. It uniquely defines the $\tau$-structure $A_\pi$ that has universe $A$, and in which precisely those literals $\varphi$ are true for which $\pi(\varphi) \neq 0$.

Notice that, if $K$ is not the Boolean semiring, then several different $K$-interpretations may define the same structure. Further, $K$-interpretations are interesting, and have a number of applications, also in cases where they do not specify a single model, see [10] and the references given there.

\textbf{Dual-Indeterminate Polynomials.} Let $X, \overline{X}$ be two disjoint sets together with a one-to-one correspondence $X \leftrightarrow \overline{X}$. We denote by $p \in X$ and $\overline{p} \in \overline{X}$ two elements that are in this correspondence. We refer to the elements of $X \cup \overline{X}$ as \emph{provenance tokens} and we shall use “positive” and “negative” tokens $p$ and $\overline{p}$ to annotate atoms $R(\varphi) \in \text{Atoms}_A(\tau)$ and negated atoms $\neg R(\varphi) \in \text{NegAtoms}_A(\tau)$, respectively. By convention, if we annotate $R(\varphi)$ with $p$ then the “negative” token $\overline{p}$ can only be used to annotate $\neg R(\varphi)$, and vice versa. We refer to $p$ and $\overline{p}$ as \emph{complementary} tokens.

\textbf{Definition 21.} The semiring $\mathbb{N}[X, \overline{X}]$ is the quotient of the semiring of polynomials $\mathbb{N}[X \cup \overline{X}]$ by the congruence generated by the equalities $p \cdot \overline{p} = 0$ for all $p \in X$. This is the same as quotienting by the ideal generated by the polynomials $p\overline{p}$ for all $p \in X$. Observe that two polynomials $g, g' \in \mathbb{N}[X \cup \overline{X}]$ are congruent if, and only if, they become identical after deleting from each of them the monomials that contain complementary tokens. Hence, the congruence classes in $\mathbb{N}[X, \overline{X}]$ are in one-to-one correspondence with the polynomials in $\mathbb{N}[X \cup \overline{X}]$ such that none of their monomials contain complementary tokens. We shall call these \emph{dual-indeterminate polynomials}.

Note that $\mathbb{N}[X, \overline{X}]$ is $+$-positive and root-integral, but not positive, since it has divisors of 0. Further we have the following \emph{universality property}:

\textbf{Proposition 22.} Every function $f : X \cup \overline{X} \to K$ into any commutative semiring $K$ with the property that $f(p) \cdot f(\overline{p}) = 0$ for all $p \in X$ extends uniquely to a semiring homomorphism $h : \mathbb{N}[X, \overline{X}] \to K$ that coincides with $f$ on $X \cup \overline{X}$.

\textbf{Definition 23.} A \emph{provenance-tracking} interpretation is a mapping $\pi : \text{Lit}_A(\tau) \to X \cup \overline{X} \cup \{0, 1\}$ such that $\pi(\text{Atoms}_A(\tau)) \subseteq X \cup \{0, 1\}$ and $\pi(\text{NegAtoms}_A(\tau)) \subseteq \overline{X} \cup \{0, 1\}$. Further, $\pi$ maps equalities and inequalities to their truth values 0 or 1.
The idea is that if \( \pi \) annotates a positive or negative atom with a token, then we wish to track that literal through the model-checking computation. On the other hand annotating with 0 or 1 is done when we do not track the literal, yet we need to recall whether it holds or not in the model. See [10] for more details and potential applications of provenance-tracking interpretations.

### 6 Semirings of dual-indeterminate power series and least fixed point solutions

It is known that the general properties of commutative semirings are not sufficient to deal with unbounded iterations as they occur in fixed-point logic. Even for Datalog, one of the simplest fixed-point formalisms that omit the complications arising with universal quantification and negation, appropriate semirings have the additional property of being \( \omega \)-continuous. The general \( \omega \)-continuous provenance semirings are no longer semirings of polynomials, but semirings of formal power series, such as \( \mathbb{N}^\mathbf{\infty}[X] \). We combine this here with our approach for dealing with negation by taking quotients with respect to the congruence generated by products \( \mathcal{P} \) of positive and negative provenance tokens. What we obtain are \( \omega \)-continuous provenance semirings of dual-indeterminate power series, such as \( \mathbb{N}^\mathbf{\infty}[X, \overline{X}] \), as well as idempotent, absorptive, and other variants thereof.

A semiring \( K \) is **naturally ordered** if the relation \( a \leq b :\Leftrightarrow \exists x(a + x = b) \) is a partial order. Note that this relation is reflexive and transitive in every semiring, but it is not always antisymmetric. An \( \omega \)-chain is a sequence \( (a_i)_{i \in \omega} \) with \( a_i \leq a_{i+1} \) for all \( i \in \omega \).

**Definition 24.** A commutative semiring \( K \) is **\( \omega \)-continuous** if it is naturally ordered and satisfies the following additional conditions:

- Every \( \omega \)-chain \( (a_i)_{i \in \omega} \) in \( K \) has a supremum \( \sup_{i \in \omega} a_i \) in \( K \). As a consequence, we have a well-defined infinite summation operator \( \sum \), such that for every sequence \( (b_i)_{i \in \omega} \),

\[
\sum_{i \in \omega} b_i := \sup\{a_0 + \cdots + a_n : n \in \omega\}
\]

- For every sequence \( (a_i)_{i \in \omega} \) in \( K \), every \( c \in K \), and every partition \( (I_j)_{j \in J} \) of \( \omega \), we have that
\[
c \cdot \sum_{i \in \omega} a_i = \sum_{i \in \omega} c \cdot a_i \quad \text{and} \quad \sum_{j \in J} \sum_{i \in I_j} a_i = \sum_{i \in \omega} a_i.
\]
In an \( \omega \)-continuous semiring we further have the Kleene star operation, \( a^* := \sum_{i \in \omega} a^i = \sup_{i \in \omega} (1 + a + a^2 + \cdots + a^i) \). A function \( f : K \rightarrow K \) is \( \omega \)-continuous if, \( \sup_{i \in \omega} f(a_i) = f(\sup_{i \in \omega} a_i) \) for every \( \omega \)-chain \( (a_i)_{i \in \omega} \). A consequence of the definition is that any function defined by a polynomial or a power series is \( \omega \)-continuous in each argument.

**Definition 25.** Given a semiring \( K \) and a finite set \( X \) of indeterminates, we denote by \( K[X] \) the semiring of formal power series (i.e. possibly infinite sums of monomials) with coefficients in \( K \) and indeterminates in \( X \), with addition and multiplication defined in the obvious way. If \( K \) is \( \omega \)-continuous and \( |X| = n \), then every formal power series \( f \in K[X] \) induces a well-defined function \( f : K^n \rightarrow K \) which is \( \omega \)-continuous in each argument. Further, if \( K \) is \( \omega \)-continuous, then so is \( K[X] \) [10].

A **system of power series** with indeterminates \( X_1, \ldots, X_n \) is a sequence \( F = (f_1 \ldots f_n) \) with \( f_i \in K[X] \) for each \( i \). It induces a function \( F : K^n \rightarrow K^n \) that is monotone in each argument. By Kleene’s Fixed-Point Theorem \( F \) has a least fixed point \( lfp F \) which coincides with the supremum of the Kleene approximants \( F^k \), defined by \( F^0 = 0, F^{k+1} = F(F^k) \), i.e.
\( \text{lp} F = \sup_{k \in \omega} F^k \). We also refer to \( \text{lp} F \) as the *least fixed-point solution* of the equation system
\[
X_1 = f_1(X_1, \ldots, x_n), \ldots, X_n = f_n(X_1, \ldots, X_n),
\]
in short, \( X = F(X) \).

**Dual-indeterminate power series.** Semirings \( K[[X]] \) of power series turn out to be appropriate as general provenance semirings for (not necessarily acyclic) reachability games, without any further structure on the terminal nodes, as well as for purely positive fixed-point formalisms, without negation even on the atomic level. However, as soon as we want to deal with fixed-point logics with (atomic) negation we again need to take quotients with respect to the congruence generated by an appropriate correspondence \( X \leftrightarrow \overline{X} \) between positive and negative tokens (with the same conventions as in Definition 21).

**Definition 26.** The semiring \( K[[X, \overline{X}]] \) is the quotient of the semiring of power series \( K[[X \cup \overline{X}]] \) by the congruence generated by the equalities \( p \cdot \overline{p} = 0 \) for all \( p \in X \). The congruence classes in \( K[[X, \overline{X}]] \) are in one-to-one correspondence with the power series in \( K[[X \cup \overline{X}]] \) such that none of their monomials contain complementary tokens. We call these *dual-indeterminate power series*.

Again we have a universality property.

**Proposition 27.** Every function \( f : X \cup \overline{X} \to K \) into an \( \omega \)-continuous semiring \( K \) with the property that \( f(p) \cdot f(\overline{p}) = 0 \) for all \( p \in X \) extends uniquely to an \( \omega \)-continuous semiring homomorphism \( h : \mathbb{N}[X, \overline{X}] \to K \) that coincides with \( f \) on \( X \cup \overline{X} \).

### 7 Provenance for reachability games with cycles

We now extend our provenance approach to games that admit infinite plays. We assume that the game graphs are finite, but no longer acyclic. Given a valuation \( f_\sigma : T \to K \) for the terminal nodes of a game graph \( G \), the rules defining valuations for the other nodes have now to be read as an equation system \( F_\sigma \) in indeterminates \( X_v \) (for \( v \in V \)):

\[
\begin{align*}
X_v &= f_\sigma(v) \quad \text{for } v \in T \\
(F_\sigma) \quad X_v &= \sum_{w \in vE} h_\sigma(vw) \cdot X_w \quad \text{if } v \in V_\sigma \\
X_v &= \prod_{w \in vE} h_\sigma(vw) \cdot X_w \quad \text{if } v \in V_{1-\sigma}
\end{align*}
\]

If we assume that the underlying semiring \( K \) is \( \omega \)-continuous, then such a system \( F_\sigma \) always has a least fixed-point solution \( \text{lp} F_\sigma \).

**Valuations of plays and strategies.** As in section 4, a finite play \( x = v_0v_1 \ldots v_m \) from \( v_0 \) to a terminal node \( v_m \) gets the valuation \( f_\sigma(x) = h_\sigma(v_0v_1) \cdots h_\sigma(v_{m-1}v_m) \cdot f_\sigma(v_m) \). The provenance value of an infinite play is defined to be 0. For a strategy \( S \in \text{Strat}_\sigma(v) \), we put

\[
f_\sigma(S) := \prod \{ f_\sigma(x) : x \in \text{Plays}(S) \}.
\]

As a consequence, a strategy \( S \) can have a non-zero provenance value only if it admits only finite plays. By König’s Lemma, it then admits only a finite number of plays. Although the number of different strategies \( S \in \text{Strat}_\sigma(v) \) may well be infinite, Theorem 6 generalizes to reachability games with cycles, with a proof based on Kleene’s fixed-point theorem.
Theorem 28. For every game graph $G$ with basic valuations $f_\sigma$ and $h_\sigma$ of the terminal positions and moves in an $\omega$-continuous semiring $K$, we have that, for every position $v$

$$f_\sigma(v) := (\text{lfp } F_\sigma)(v) = \sum_{S \in \text{Strat}_0(v)} f_\sigma(S) = \sum_{S \in \text{Strat}_o(v)} \prod_{x \in \text{Plays}(S)} f_\sigma(x).$$

Proof. For any $n \geq 1$, let $\text{Strat}_o^n(v)$ be the restriction of $\text{Strat}_o(v)$ to strategies for at most $n - 1$ moves. Formally, for any strategy $S = (W, F) \in \text{Strat}_o(v)$, let $S_n = (W \cap V \leq n, F \cap (V \leq n \times V \leq n))$ and put

$$\text{Strat}_o^n(v) := \{ S_m : S \in \text{Strat}_o(v) \}.$$ 

The set $\text{Plays}(S)$ for a strategy $S \in \text{Strat}_o^n(v_0)$ only contains plays $v_0 \ldots v_m$ such that either $m < n$ and $v_m \in T$, or $m = n$. Note that these plays have at most $n - 1$ moves and need not be complete, i.e. have not necessarily reached a terminal position.

Let $(F^n)_n < \omega$ be the sequence of Kleene approximants for the least fixed-point solution $(\text{lfp } F_\sigma)$. We extend these approximants $F^n : V \rightarrow K$ to plays: for $n \geq 1$ and any play $x = (v_0 \ldots v_m) \in \text{Plays}(S)$ for an $(n - 1)$-move strategy $S \in \text{Strat}_o^n(v_0)$, we set $F^n(x) = h_\sigma(v_0 v_1) \ldots h_\sigma((v_{m-1} - v_m) \cdot F^n(v_m))$. It then suffices to prove that

$$F^n(v) = \sum_{S \in \text{Strat}_o^n(v)} \prod_{x \in \text{Plays}(S)} F^n(x)$$

for all $v$ and all $n$ with $1 \leq n < \omega$.

For $n = 1$ we observe that, since $F^0(v) = 0$ for all $v$, we have that $F^1(v) = f_\sigma(v)$ for $v \in T$ and $F^1(v) = 0$ otherwise. On the other side $\text{Strat}_o^1(v)$ consists of the single consistent play that consists just of the node $v$, so we obviously have that $F^1(v) = f_\sigma(S_0(v)) = F^1(x)$ for the unique play $x \in \text{Plays}(S_0(v))$.

Let now $n > 1$. For $v \in V_\sigma$, any strategy $S \in \text{Strat}_o^n(v)$ can be written in the form $S = v \cdot S_w$ for some successor $w \in vE$ and some strategy $S_w \in \text{Strat}_o^{n-1}(w)$, and any play $x \in \text{Plays}(S)$ has the form $vy$ for some $y \in \text{Plays}(S_w)$. By induction $F^{n-1}(w) = \sum_{S_w \in \text{Strat}_o^{n-1}(w)} \prod_{y \in \text{Plays}(S_w)} F^{n-1}(y)$. Hence

$$F^n(v) = \sum_{w \in vE} h_\sigma(vw) \cdot F^{n-1}(w) = \sum_{w \in vE} \sum_{S_w \in \text{Strat}_o^{n-1}(w)} h_\sigma(vw) \cdot \prod_{y \in \text{Plays}(S_w)} F^{n-1}(y) = \sum_{v \cdot S_w \in \text{Strat}_o^n(v)} \prod_{y \in \text{Plays}(v \cdot S_w)} F^n(vy) = \sum_{S \in \text{Strat}_o^n(v)} \prod_{x \in \text{Plays}(S)} F^n(x).$$

For $v \in V_{1-\sigma}$ with $vE = \{ w_1, \ldots, w_k \}$, the strategies $S \in \text{Strat}_o^n(v)$ have the form $S = v(S_1 \cup \ldots \cup S_k)$ with $S_i \in \text{Strat}_o^{n-1}(w_i)$ and every play $x \in \text{Plays}(S)$ has the form $vy$, for some $y_i \in \text{Plays}(S_i)$. It follows that

$$F^n(v) = \prod_{w_i \in vE} h_\sigma(vw_i) \cdot F^{n-1}(w_i) = \prod_{w_i \in vE} \sum_{S_i \in \text{Strat}_o^{n-1}(w_i)} \prod_{y_i \in \text{Plays}(S_i)} h_\sigma(vw_i) \cdot F^{n-1}(y) = \sum_{v(S_1 \cup \ldots \cup S_k) \in \text{Strat}_o^n(v)} \prod_{w_i \in vE} \prod_{y_i \in \text{Plays}(v \cdot S_i)} F^n(vy) = \sum_{S \in \text{Strat}_o^n(v)} \prod_{x \in \text{Plays}(S)} F^n(x).$$

For the case of game valuations $f_\sigma : V \rightarrow \mathbb{N}[T]$, given by the basic valuations $f_\sigma(t) = t$ for terminal positions $t \in T$ and $h_\sigma(vw) = 1$ for all moves $(v, w) \in E$, we again get precise information about the number of strategies that a player has for a specific outcome. Indeed, $f_\sigma(v)$ is a (possibly) infinite sum of monomials $m \cdot t^n \ldots t^n$.
Corollary 29. Let \( f_\sigma : V \rightarrow \mathbb{N}[T] \) be the valuation of Player \( \sigma \) for the game \( \mathcal{G} \) in \( \mathbb{N}[T] \). For every monomial \( m = t_1^{j_1} \ldots t_k^{j_k} \) in \( f_\sigma(v) \) (with \( m \in \mathbb{N} \) and \( j_i > 0 \)) Player \( \sigma \) has precisely \( m \) strategies \( S \) from \( v \) with the property that the set of possible outcomes for \( S \) is precisely \( \{t_1, \ldots, t_k\} \), and precisely \( j_i \) plays that are consistent with \( S \) have the outcome \( t_i \).

Let \( \mathcal{G} = (V, V_0, V_1, T, E) \) be a game with reachability objectives \( T_0, T_1 \) for the two players, such that \( T_0 \cap T_1 = \emptyset \). Let \( W_0, W_1 \subseteq V \) be the winning regions for the two players, i.e., \( W_\sigma \) is the set of those positions \( v \in V \) such that Player \( \sigma \) has a strategy from \( v \) to force the play to \( T_\sigma \). Note that \( V \) is the disjoint union of the \( W_0, W_1 \) and \( U \), the set of those positions from which none of the two players has a winning strategy. By Zermelo’s Theorem both players have strategies to guarantee that each play from \( U \) will be at least a draw.

Corollary 30. Let \( f_\sigma : T \rightarrow K \) be a valuation of the terminal positions of \( \mathcal{G} \) in an \( \omega \)-continuous semiring, with \( f_\sigma(t) \neq 0 \) if, and only if, \( t \in T_\sigma \). The least fixed point solution of the equation system \( F_\sigma \) extends this to a valuation \( f_\sigma : V \rightarrow K \), with \( f_\sigma(v) \neq 0 \) if, and only if, \( v \in W_\sigma \).

Notice that weakly contradictory valuations \( f_0 \) an \( f_1 \) on the terminal positions extend to weakly contradictory valuations on all positions. However, even valuations into \( \omega \)-continuous semirings that are strongly contradictory on the terminal positions, are in general only weakly contradictory on the set of all positions, unless \( W_0 \cup W_1 = V \), since \( f_0(U) = f_1(U) = 0 \).

Example 31. We illustrate our findings by the following very simple example of a game where Player 0 moves from \( v \), Player 1 moves from \( w \), and \( s \) and \( t \) are terminal nodes.

The corresponding equation system for Player 0 has the equations \( X_v = s + X_w \) and \( X_w = t \cdot X_v \).

For \( f_0(s) = 0 \) and \( f_1(w) = 0 \), the least fixed-point solution is \( f(v) = s \cdot (1 + t + t^2 + \ldots) \) and \( f(w) = s \cdot (t + t^2 + \ldots) \).

If we evaluate it for the reachability objectives \( \{s\} \) and \( \{t\} \), respectively, we obtain \( f(v)[0,t] = f(w)[0] = 0 \) which illustrates that neither from \( v \) nor from \( w \), Player 0 has a strategy to reach \( t \). On the other side, \( f(v)[s,0] = s \cdot t \) and \( f(w)[s,0] = 0 \) which is consistent with the fact that Player 0 has a strategy to reach \( s \) from \( v \) but not from \( w \).

But the formal power series \( f(v) \) and \( f(w) \) reveal more information than that. For instance, the fact that \( f(v) \) contains, for every \( n \), the monomial \( s \cdot t^n \) implies that Player 0 has precisely one strategy \( S \) from \( v \) that admits precisely \( n + 1 \) consistent plays, one of which has outcome \( s \) and the other \( n \) have outcome \( t \); this is the strategy where Player 0 moves from \( v \) to \( w \) the first \( n \) times, and then to \( s \). Notice that Player 0 also has one further strategy, namely the (positional) strategy to move always to \( w \). However, this strategy does not guarantee that the play terminates and therefore has value 0, so it is not visible in the provenance values \( f(v) \) and \( f(w) \).

8 Least fixed-point logic for positive LFP

Least fixed-point logic, denoted LFP, extends first order logic by least and greatest fixed points of definable monotone operators on relations: If \( \psi(R, \overline{x}) \) is a formula of vocabulary \( \tau \cup \{R\} \), in which the relational variable \( R \) occurs only positively, and if \( \overline{x} \) is a tuple of variables such that the length of \( \overline{x} \) matches the arity of \( R \), then \( \text{lfp}(\overline{x} \cdot \psi(R, \overline{x})) \) and \( \text{gfp}(\overline{x} \cdot \psi(R, \overline{x})) \) are also formulae (of vocabulary \( \tau \)). The semantics of these formulae is that \( \overline{x} \) is contained in the least (respectively the greatest) fixed point of the update operator \( F_\psi : R \rightarrow \{\overline{x} : \psi(R, \overline{x})\} \). Due to the positivity of \( R \) in \( \psi \), any such operator \( F_\psi \) is monotone and therefore has, by
the Knaster-Tarski-Theorem, a least fixed point \( \text{lfp}(F_\psi) \) and a greatest fixed point \( \text{gfp}(F_\psi) \). See e.g. [11] for background on LFP.

Note that in formulae \([\text{lfp} \varphi(x).\psi](\overline{x})\) one may allow \( \psi \) to have other free variables besides \( \overline{x} \); these are called parameters of the fixed-point formula. However, at the expense of increasing the arity of the fixed-point predicates and the number of variables one can always eliminate parameters. For the construction of model-checking games and also for provenance analysis it is convenient to assume that formulae are parameter-free. The duality between least and greatest fixed point implies that for any \( \psi \),

\[
[\text{gfp} \varphi(x).\psi](\overline{x}) \equiv \neg[\text{lfp} \varphi(x).\neg \psi(R/R)](\overline{x}).
\]

Using this duality together with de Morgan’s laws, every LFP-formula can be brought into negation normal form, where negation applies to atoms only.

**The fragment of positive least fixed points.** We denote by \( \text{posLFP} \) the fragment of LFP consisting of formulae in negation normal form such that all its fixed-point operators are least fixed-points. It is known that, on finite structures (but not in general), \( \text{posLFP} \) has the same expressive power as full LFP, and thus captures all polynomial-time computable properties of ordered finite structures [11].

An advantage of dealing with \( \text{posLFP} \), rather than full LFP, is that it admits much simpler model checking games. Indeed the appropriate games for LFP are simpler model checking games. Indeed the appropriate games for LFP are parity games.

A \( K \)-interpretation \( \pi : \text{Lit}_A(\tau) \rightarrow K \) into an \( \omega \)-continuous semiring thus provides a valuation of the terminal positions of the game graph \( G(A, \psi) \) for any \( \psi \in \text{posLFP}(\tau) \). By Theorem 30 this extends to a valuation \( f_0 : V \rightarrow K \) on the set \( V \) of all positions \( \psi(\overline{x}) \) of \( G(A, \psi) \), including position \( \psi \) itself.

**Definition 32.** For any instantiated subformula \( \varphi \) of a sentence \( \psi \in \text{posLFP} \), we define the provenance value \( \pi[\varphi] \) by its game valuation: \( \pi[\varphi] := f_0(\varphi) \).

In particular, if \( \pi \) is model-defining, then \( f_0 \) provides truth values for all fully instantiated subformula \( \varphi \) of \( \psi \) on the structure \( \mathfrak{A}_\pi \) that \( \pi \) describes. Indeed \( \mathfrak{A}_\pi \models \varphi \) if, and only if, \( \pi[\varphi] \neq 0 \), and in that case the value \( \pi[\varphi] \) gives us additional information, how and why \( \varphi \) holds in \( \mathfrak{A} \), for instance by information on the winning strategies that Verifier has available for establishing the truth of \( \varphi \) in \( \mathfrak{A}_\pi \). However, contrary to the case of first-order logic, in the case where \( \mathfrak{A}_\pi \models \varphi \), and hence \( \pi[\varphi] = 0 \) we do not get additional information on the reasons why \( \varphi \) is false. The possibility to move to \( \neg \varphi \) (or more precisely, its negation normal form) and to do the provenance analysis for that formula, does not exist here since \( \neg \varphi \) is not a formula of \( \text{posLFP} \). In fact, the model checking-game for \( \neg \varphi \) is not a reachability game, but a safety game. To deal with safety games and greatest fixed points we shall have to impose additional restrictions on the underlying semirings. We shall discuss this below.

One can define provenance values for \( \text{posLFP} \)-sentences also directly by a fixed-point interpretation in \( \omega \)-commutative semirings. The goal is to extend, by induction over the
syntax, a $K$-interpretation $\pi : \text{Lit}_A(\tau) \to K$ to valuations $\pi[\psi] \in K$ for all sentences $\psi \in \text{posLFP}(\tau \cup A)$. The rules for first-order operations are defined already, so we just have to consider sentences of form $\psi([\pi]) = [\text{lfp} R\pi.\varphi(R, [\pi])]([\pi])$, with $\varphi \in \text{posLFP}(\tau \cup \{R\})$. If $R$ has arity $m$, then its $K$-interpretations of $A$ are functions $g : A^m \to K$. These functions are ordered, by $g \leq g'$ if, and only if, $g(\overline{a}) \leq g'(\overline{a})$ for all $\overline{a} \in A^m$. Given a $K$-interpretation $\pi : \text{Lit}_A(\tau) \to K$, we denote by $[\pi[R \mapsto g]]$ the $K$-interpretation of $\text{Lit}_A(\tau) \cup \text{Atoms}_A(\{R\})$ obtained from $\pi$ by adding values $g(\overline{a})$ for the atoms $R\overline{a}$. (Notice that $R$ appears only positively in $\varphi$, so negated atoms are not needed).

The formula $\varphi(R, [\pi])$ now defines, together with $\pi$, a monotone update operator $F_\varphi^\pi$ on functions $g : A^m \to K$. More precisely, it maps $g$ to

$$F_\varphi^\pi (g) : [\pi \mapsto \pi[R \mapsto g]][\varphi(R, [\pi])].$$

By Kleene’s Fixed-Point Theorem, the operator $F_\varphi^\pi$ has a least fixed point $\text{lfp} (F_\varphi^\pi)$ which coincides with the limit of the sequence $(g^n)_{n<\omega}$ with $g^0 := 0$ and $g^{n+1} := F_\varphi^\pi (g^n)$, and which we may define as the provenance value of $[\text{lfp} R\pi.\varphi(R, [\pi])]([\pi])$. The two definitions coincide.

**Proposition 33.** For every formula $[\text{lfp} R\pi.\varphi(R, [\pi])] \in \text{posLFP}$ and every $K$-interpretation $\pi : \text{Lit}_A(\tau) \to K$ into an $\omega$-continuous semiring, $\pi[[\text{lfp} R\pi.\varphi(R, [\pi])]([\pi])] = \text{lfp} (F_\varphi^\pi)([\pi])$.

The proof is a rather straightforward adaptation of the correctness proof for model checking games for LFP, see e.g. [11] Chapter 3.3.

9 Beyond reachability: safety games and greatest fixed points

While the restriction of LFP to its positive fragment comes with no loss of expressive power (on finite structures) and while posLFP is sufficiently powerful to capture a number of interesting and relevant other fixed-point formalisms in computer science, it is nevertheless not really satisfactory. One reason is that the transformation from a fixed-point formula with non-atomic negation into one in posLFP is (contrary to transformations into negation normal form) not a simple syntactic translation. It goes through the Stage Comparison Theorem and can make a formula much longer and more complicated. Further, such transformations are not available for important fixed-point formalisms such as the modal $\mu$-calculus, stratified datalog, transitive closure logics, and even simple temporal languages such as CTL. On the game-theoretic side, reachability games are just the simplest kind of games on graphs, and in many applications players have different and more ambitious goals such as safety, Büchi, parity or Muller objectives. It is thus an important and interesting challenge to lay the foundations of a provenance analysis for full LFP and infinite games with more general objectives, and to apply this approach to the numerous other fixed-point formalisms, in particular in databases and verification.

We defer a detailed treatment of this to forthcoming work. Here we discuss some of the mathematical concepts and challenges that arise in this project, and apply them to the provenance of safety games. Recall that the computation of winning positions for safety objectives is a simple, but also in some sense universal, application of greatest fixed points.

The first observation is that we need to impose additional requirements on the semirings that we consider. While $\omega$-continuous semirings are appropriate for a provenance analysis of least fixed points and reachability objectives, they are not always adequate for greatest fixed points. The property of $\omega$-continuity is not sufficient to guarantee the existence of greatest
fixed points, and in cases where they exist they do not necessarily provide the information that we are interested in.

Example 34. We consider the game graph

with associated equation system for Player 0 consisting of \( X_w = X_w + X_z, X_w = f(s) \cdot X_w, \) and \( X_z = f(t) \cdot X_v. \) The least fixed-point solution (in whatever semiring) has values \( f(v) = f(w) = f(z) = 0 \) which reflects the fact that Player 0 has no strategy to guarantee a finite play. It is not difficult to next.

We consider the game graph

with associated equation system for Player 0 consisting of

\[ X_w = X_w + X_z, \quad X_w = f(s) \cdot X_w, \quad \text{and} \quad X_z = f(t) \cdot X_v. \]

The least fixed-point solution (in whatever semiring) has values \( f(v) = f(w) = f(z) = 0 \) which reflects the fact that Player 0 has no strategy to guarantee a finite play. It is not difficult to see that in \( \mathbb{N}^\infty [s,t] \) this in fact the unique fixed point, hence in particular the greatest one, which however gives us no information about safety strategies. In \( \mathbb{N}^\infty \) instead, under a valuation of the terminal nodes with \( f(s) = a \neq 0 \) and \( f(t) = 0, \) we get the greatest fixed point \( f(v) = f(w) = \infty \) and \( f(z) = 0. \) In particular, greatest fixed-points do not specialise correctly from \( \mathbb{N}^\infty [s,t] \) to \( \mathbb{N}^\infty. \)

We shall see below that get interesting information on safety strategies by provenance values in the absorptive semiring \( \mathbb{S}^\infty [s,t]. \)

To make sure that also greatest fixed points of polynomial equation systems exist, we shall require that our semirings are not just \( \omega \)-continuous, but also also \( \omega \)-co-continuous, i.e. that every descending \( \omega \)-chain \((a_i)_{i \in \omega}, \) with \( a_{i+1} \leq a_i \) for all \( i \in \omega, \) has an infimum \( \inf_{i \in \omega} a_i \) in \( K, \) which is compatible with the semiring operations in the sense that, for every \( c \in K, \)

\[ c + \inf_{i \in \omega} a_i = \inf_{i \in \omega} (c + a_i) \quad \text{and} \quad c \cdot \inf_{i \in \omega} a_i = \inf_{i \in \omega} (c \cdot a_i). \]

We call such semirings fully \( \omega \)-continuous. Our most important example of such a semiring is \( \mathbb{S}^\infty [X], \) the semiring of generalized absorptive polynomials, that we are going to discuss next.

10 Absorptive semirings and generalized absorptive polynomials

Recall that a semiring \( K \) is absorptive if \( a + ab = a \) for all \( a,b \in K \) which is equivalent to \( 1 + a = 1 \) for all \( a \in K. \) Examples include the Viterbi semiring, the tropical semiring, min-max semirings, further the semiring \( \mathbb{S}[X] \) of absorptive polynomials over \( X. \) Absorptive semirings are \(+\)-idempotent and naturally ordered, 1 is the top element, and multiplication decreases elements: \( ab \leq b. \) In particular, the powers of an element form a descending \( \omega \)-chain \( 1 \geq a \geq a^2 \geq \cdots. \) If this chain has an infimum then we denote it by \( a^\infty. \)

In the semiring \( \mathbb{S}[X], \) the infima of descending \( \omega \)-chains \((x^n)_{n < \omega} \) are always 0 and thus not very informative. We therefore complete \( \mathbb{S}[X] \) to the semiring \( \mathbb{S}^\infty [X] \) by admitting exponents in \( \mathbb{N}^\infty. \)

Definition 35. Let \( X \) be a finite set of provenance tokens. A monomial over \( X \) with exponents from \( \mathbb{N}^\infty \) is a function \( m : X \to \mathbb{N}^\infty. \) Informally, we write \( m \) as \( x_1^{m_1(x_1)} \cdots x_n^{m_n(x_n)}. \) Monomial multiplication adds the exponents. Observe also that \( x^\infty \cdot x^n = x^\infty. \) For any two monomials, \( m_1, m_2 \) we say that that \( m_2 \) absorbs \( m_1 \) if \( m_2 \) has smaller exponents than \( m_1. \) Formally, \( m_1 \leq m_2 \) if, and only if, \( m_1(x) \geq m_2(x) \) for all \( x \in X. \) Since monomials are functions, this is the pointwise partial order given by the order on \( \mathbb{N}^\infty. \)

Because \( \mathbb{N}^\infty \) is a lattice (with top and bottom) the monomials also inherit a lattice structure. The set of all monomials is, of course, infinite. However, it has some crucial finiteness properties.
Proposition 36. Every ascending chain and every antichain of monomials is finite.

Proof. Clearly \((\mathbb{N}^\infty, \leq)\) is a well-order. For any finite set \(X\), the set of monomials \(m : X \to \mathbb{N}^\infty\) with the reverse order than the absorption order is isomorphic to \((\mathbb{N}^\infty)^k\) with \(k = |X|\) and with the component-wise order inherited from \((\mathbb{N}^\infty, \leq)\). This is a well-quasi-order and therefore has no infinite descending chains and no infinite antichains. This implies that in the set of monomials over \(X\) with the absorption order, all ascending chains and all antichains are finite.

Definition 37. We define \(S^\infty[X]\) as the set of antichains of monomials with indeterminates from \(X\) and exponents in \(\mathbb{N}^\infty\). Writing an antichain as a (formal) sum of its monomials we identify it with a polynomial with coefficients 0 or 1, and call these generalized absorptive polynomials. We define polynomial addition and multiplication as usual, except that for coefficients \(1+1=1\), and that we keep only the maximal monomials in the result. The empty antichain corresponds to the 0 polynomial. The 1 polynomial consists of just the monomial in which every indeterminate has exponent 0.

Proposition 38. \((S^\infty[X], +, \cdot, 0, 1)\) is an absorptive commutative semiring. Further it is a complete lattice wrt. to the natural order, which is fully \(\omega\)-continuous and moreover completely distributive.

As a consequence, we can compute not only least fixed point solutions for systems of polynomial equations but also greatest fixed points. In contrast to other semirings with such properties, such as for instance the Viterbi semiring, \(S^\infty[X]\) has one further crucial property. It is chain-positive which means that the infimum of every chain of non-zero elements is also non-zero.

As in other semirings of polynomials and power series we can also here take pairs of positive and negative indeterminates, with a correspondence \(X \leftrightarrow \overline{X}\) and build the quotient with respect to the congruence generated by the equation \(x \cdot \overline{x} = 0\). We thus obtain a new semiring \(S^\infty[X, \overline{X}]\) which provides a natural framework for a provenance analysis for full LFP and other fixed point calculi. We shall develop this in forthcoming work.

Here we use the semiring \(S^\infty[T]\) to describe a provenance analysis for safety games where \(T\) is the set of terminal positions of the given game graph.

11 Absorption among strategies

Definition 39. Let \(G = (V, V_0, V_1, T, E)\) be a finite game graph, and \(v \in V\). For two strategies \(S, S' \in \text{Strat}_\sigma(v)\), we say that \(S\) absorbs \(S'\) (in symbols \(S \succeq_a S'\)) if

- for all \(t \in T\), \(S\) admits at most as many plays with outcome \(t\) as \(S'\) does, and
- if \(S\) admits an infinite play, then so does \(S'\).

We call \(S\) absorption-dominant if it is maximal with respect to \(\succeq_a\).

Absorption-dominant strategies are interesting both for games in general and for logic because they can win “with minimal effort”. As a simple example, consider a model checking game for a formula \(\varphi \lor (\varphi \land \psi)\). The Verifier can either establish \(\varphi\) or \(\varphi \land \psi\), but any strategy that establishes the truth of \(\varphi \land \psi\) will have more plays and more outcomes than one that proves just \(\varphi\), and will thus be absorbed by it. The absorption-dominant strategies for \(\varphi \lor (\varphi \land \psi)\) are thus precisely the absorption-dominant strategies for \(\varphi\).
Notice however that, despite this minimality, absorption dominant strategies need not be positional, not even in acyclic games.

\[ \text{Example 40. Consider the game} \]

\[ \begin{array}{c}
\text{u} \\
\text{v} \\
\text{w} \\
\text{s} \\
\text{f} \\
\end{array} \]

There are four strategies in Strat\(_v\) with provenance values \(s^2\), \(st\), \(st\), and \(t^2\). The positional ones are those with values \(s^2\) and \(t^2\), but all four strategies are absorption-dominant.

However, absorption-dominant strategies are weakly positional in the sense that if a node is reached several times during the same play, then, without loss of strategic power, the player can always make the same choice at that node. Absorption among strategies makes sense for both acyclic and cyclic games. In acyclic games, absorption-dominant strategies are described by provenance polynomials in \(\mathbb{S}[T]\) (with only finite exponents). But they are even more interesting for the analysis of reachability and safety games that admit infinite plays. The fundamental difference between valuations for reachability and safety strategies concerns the valuations of infinite plays. If, as we assume here, reachability and safety goals are defined for terminal nodes, then an infinite play is losing for every reachability objective but winning for every safety objective. As a consequence, the strategies \(S \in \text{Strat}_u\) that enforce all plays to be non-terminating absorb all other strategies in Strat\(_v\) that admit at least one infinite play.

We thus extend the valuations of plays to two different valuation function \(f^\mu_S\) and \(f^\nu_S\). For simplicity, we assume trivial valuations on the edges, so for a finite play \(x\) ending in \(t\), we just put \(f^\mu_S(x) = f^\nu_S(x) = f_\sigma(t)\) but if \(x\) is an infinite play, we put \(f^\mu_S(x) = 0\) and \(f^\nu_S(x) = 1\).

A strategy \(S \in \text{Strat}_v\) may well admit an infinite set of plays. Nevertheless this set is described in \(\mathbb{S}^\infty[T]\) by monomials (or 0), namely

\[ f^\mu_S(S) := \prod_{x \in \text{Plays}(S)} f^\mu_x(x) \quad \text{and} \quad f^\nu_S(S) := \prod_{x \in \text{Plays}(S)} f^\nu_x(x). \]

We extend the absorption order \(\succeq\) on monomials by \(m \geq 0\) for all \(m\).

\[ \text{Lemma 41. For all strategies} \ S, S' \in \text{Strat}_v \text{ we have,} \]

- \(0 \neq f^\mu_S(S) \succeq f^\mu_S(S') \neq 1\).
- \(f^\mu_S(S) = 0\) if, and only if, \(S\) admits an infinite play. Otherwise \(f^\nu(S) = f^\nu(S')\).
- \(f^\nu(S) = 1\) if, and only if, \(S\) admits only infinite plays.
- \(S\) absorbs \(S'\) if, and only if, both \(f^\mu_S(S) \succeq f^\mu_S(S')\) and \(f^\nu_S(S) \succeq f^\nu_S(S')\).

**Proof.** Only the last item requires proof. Suppose that \(S\) absorbs \(S'\). If \(S\) admits only finite plays, then \(f^\mu_S(S) = f^\mu_S(S') \succeq f^\mu_S(S')\). If \(S\) admits an infinite play, then so does \(S'\) and \(f^\mu_S(S) \succeq f^\mu_S(S') \succeq f^\mu_S(S') = f^\mu_S(S) = 0\). In both cases, \(f^\nu_S(S) \succeq f^\nu_S(S')\) and \(f^\nu_S(S) \succeq f^\nu_S(S')\). Conversely, assume that \(S\) does not absorb \(S'\). Then either there
Theorem 43. Let $G = (V, V_0, V_1, T, E)$ be a game graph and let $F_\sigma$ be the associated equation system for Player $\sigma$. In the semiring $S^\infty[T]$ this system has least and greatest fixed point solutions $\lfp F_\sigma$ and $\gfp F_\sigma$ with

$$ (\lfp F_\sigma)(v) := \sum_{S \in \Strat_\sigma(v)} f^\mu_\sigma(S) \quad \text{and} \quad (\gfp F_\sigma)(v) := \sum_{S \in \Strat_\sigma(v)} f^\nu_\sigma(S). $$

The values of these sums do not change if we restrict them to the absorption-dominant strategies.

Proof. Since $S^\infty[T]$ is $\omega$-continuous, the claim for $(\lfp F_\sigma)$ follows from Theorem 28. For the greatest fixed-point solution we use that $S^\infty[T]$ is also $\omega$-co-continuous and has the structure of a complete lattice. Thus, $(\gfp F_\sigma)$ is the limit of the descending chain $(G^n)_{n<\omega}$ of approximants starting with $G^0 = 1$, and $G^{n+1}$ is defined by applying the equation system $F_\sigma$ to $G^n : V \to S^\infty[T]$.

For $n = 1$ we observe that, since $G^0(v) = 1$ for all $v$, we have that $G^1(v) = f^\nu_\sigma(v)$ for $v \in T$ and $G^1(v) = 1$ otherwise. On the other side $\Strat_1(v)$ consists of the the single strategy $S_0(v) := \{(v, \emptyset)\}$ with a single consistent play that consists just of the node $v$, so we obviously have that $G^1(v) = f^\nu_\sigma(S_0(v))$.

Consistent with the terminology from the proof of Theorem 28 set $G^n(x) := G^n(v_m)$ for any play $x = (v_0 \ldots v_m) \in \Plays(S)$ where $S \in \Strat_\sigma(v_0)$. It suffices to prove that

$$ G^n(x) = \sum_{S \in \Strat_\sigma(v_0)} \prod_{x \in \Plays(S)} G^n(x) $$

for all $v$ and all $n$ with $1 \leq n < \omega$.

Let $n > 1$. For $v \in V_\sigma$, any strategy $S \in \Strat_\sigma^n(v)$ can be written in the form $S = v \cdot S_w$ for some successor $w \in vE$ and some strategy $S_w \in \Strat_\sigma^{n-1}(w)$, and any play $x \in \Plays(S)$ has the form $vy$ for some $y \in \Plays(S_w)$. By induction $G^{n-1}(w) = \sum_{S_w \in \Strat_\sigma^{n-1}(w)} \prod_{y \in \Plays(S_w)} G^{n-1}(y)$. Hence

$$ G^n(x) = \sum_{w \in vE} G^{n-1}(w) = \sum_{w \in vE} \sum_{S_w \in \Strat_\sigma^{n-1}(w)} \prod_{y \notin \Plays(S_w)} G^{n-1}(y) $$

$$ = \sum_{w \in vE} \prod_{y \in \Plays(v)} G^n(vy) = \sum_{S \in \Strat_\sigma^n(v)} \prod_{x \in \Plays(S)} G^n(x). $$
For $v \in V\_\sigma$ with $vE = \{w_1, \ldots, w_k\}$, the strategies $S \in \text{Strat}\_\pi^\nu(v)$ have the form $S = v(S_1 \cup \cdots \cup S_k)$ with $S_i \in \text{Strat}\_\pi^{\nu-1}(w_i)$ and every play $x \in \text{Plays}(S)$ has the form $vy_1$ for some $y_1 \in \text{Plays}(S_i)$. It follows that

$$G^\nu(v) = \prod_{w_i \in vE} G^{\nu-1}(w_i) = \prod_{w_i \in vE} \sum_{S_i \in \text{Strat}\_\pi^{\nu-1}(w_i)} \prod_{y_i \in \text{Plays}(S_i)} G^{\nu-1}(y)$$

$$= \sum_{v \cdot (S_1 \cup \cdots \cup S_k) \in \text{Strat}\_\pi(v)} \prod_{i \in E(v)} \prod_{y \in \text{Plays}(v \cdot S_i)} G^{\nu}(vy) = \sum_{S \in \text{Strat}\_\pi(v)} \prod_{x \in \text{Plays}(S)} G^{\nu}(x).$$

These least and greatest fixed points give precise descriptions of the absorption-dominant reachability and safety strategies of the players for each position of the game.

**Example 44.** We return to the Example 44.

Recall that the associated equation system for Player 0 has the equations $X_v = X_w + X_z$, $X_w = f(s) \cdot X_v$, and $X_z = f(t) \cdot X_v$.

The greatest fixed-point solution in $S^\infty[s, t]$, computed by iterating from the top element $f = 1$ results in $f^\nu(v) = s^\infty + t^\infty$, $f^\nu(w) = s^\infty + st^\infty$, and $f^\nu(z) = s^\infty t + t^\infty$. Notice that indeed, $f^\nu(v) = f^\nu(w) + f^\nu(z)$ because $st^\infty$ is absorbed by $t^\infty$, and $s^\infty t$ by $s^\infty$. The greatest fixed point solution indicates that Player 0 has two absorptive strategies (move always to $w$ or move always to $z$), and gives, for each of the terminal nodes $s$ and $t$ the number of plays ending in that node that the strategy admits. For instance, if the safety objective requires to avoid $t$, then $v$ and $w$ the strategy moving to $w$ has infinitely many winning plays ending in $s$ (and one nonterminating play with value 1), but since $f(z)[s, 0] = 0$, Player 0 has no safety strategy from $z$ that avoids $t$.

**12 Outlook**

In this paper we have extended the semiring framework for provenance analysis by new elements, so that it can be applied to logics with negation, in particular first-order logic and fixed-point logics, and to an analysis of games that provides detailed information about the number and properties of the strategies of the players.

Our treatment of negation is based on transformations to negation normal form and the use of newly introduced semirings of dual-indeterminate polynomials and dual-indeterminate power series. In particular, $\omega$-continuous semirings $\mathbb{N}^\infty[X, X]$ of dual-indeterminate power series provide an adequate general framework for logics with least fixed points, such as posLFP (and Datalog) and the semiring of absorptive generalized dual-indeterminate polynomials $\mathbb{S}^\infty[X, X]$ permits an adequate treatment of greatest fixed points. We have thus laid foundations for a provenance analysis of general fixed-point logics, and we are currently applying this also to modal, temporal, and dynamic logics.

On the level of games, we have seen that provenance valuations in $\omega$-continuous and absorptive semirings give us very detailed information about strategies for possibly infinite games with reachability and safety objectives. We are currently expanding this to games with more complicated objectives, such as Büchi, Co-Büchi or parity games. Since these objectives do no longer depend on terminal nodes but on the data occurring in infinite plays, a somewhat different framework has to be used, depending on instance for basic valuations of the edges of the game graph.
References

1. Y. Amsterdamer, D. Deutch, and V. Tannen. On the limitations of provenance for queries with difference. In 3rd Workshop on the Theory and Practice of Provenance, TaPP’11, 2011. See also CoRR abs/1105.2255.

2. Y. Amsterdamer, D. Deutch, and V. Tannen. Provenance for aggregate queries. In Principles of Database Systems, PODS, pages 153–164, 2011. See also CoRR abs/1101.1110.

3. K. Apt and E. Grädel, editors. Lectures in Game Theory for Computer Scientists. Cambridge University Press, 2011.

4. K. Dannert and E. Grädel. Provenance analysis: A perspective for description logics? In C. Lutz et al., editor, Description Logic, Theory Combination, and All That, Lecture Notes in Computer Science Nr. 11560. Springer, 2019.

5. K. Dannert and E. Grädel. Semiring provenance for guarded logics. In Hajnal Andréka and István Németi on Unity of Science: From Computing to Relativity Theory through Algebraic Logic, Outstanding Contribution to Logic. Springer, 2019.

6. D. Deutch, T. Milo, S. Roy, and V. Tannen. Circuits for datalog provenance. In Proc. 17th International Conference on Database Theory ICDT, pages 201–212, 2014.

7. J. Foster, T. Green, and V. Tannen. Annotated XML: queries and provenance. In Principles of Database Systems, PODS, pages 271–280, 2008.

8. F. Geerts and A. Poggi. On database query languages for K-relations. J. Applied Logic, 8(2):173–185, 2010.

9. F. Geerts, T. Unger, G. Karvounarakis, I. Fundulaki, and V. Christophides. Algebraic structures for capturing the provenance of SPARQL queries. J. ACM, 63(1):7:1–7:63, 2016.

10. E. Grädel and V. Tannen. Semiring provenance for first-order model checking. arXiv:1712.01980 [cs.LO], 2017. URL: http://www.logic.rwth-aachen.de/pub/graedel/ET17.pdf

11. E. Grädel et al. Finite Model Theory and Its Applications. Springer-Verlag, 2007.

12. T. Green. Containment of conjunctive queries on annotated relations. Theory Comput. Syst., 49(2):429–459, 2011.

13. T. Green, Z. Ives, and V. Tannen. Reconcilable differences. In Database Theory - ICDT 2009, pages 212–224, 2009.

14. T. Green, G. Karvounarakis, and V. Tannen. Provenance semirings. In Principles of Database Systems PODS, pages 31–40, 2007.

15. T. Green and V. Tannen. The semiring framework for database provenance. In Proceedings of PODS, pages 93–99, 2017.

16. W. Kuich. Semirings and formal power series. In Handbook of Formal Languages, volume 1. Springer, 1997.

17. V. Tannen. Provenance propagation in complex queries. In In Search of Elegance in the Theory and Practice of Computation - Essays Dedicated to Peter Buneman, pages 483–493, 2013.

18. V. Tannen. Provenance analysis for FOL model checking. SIGLOG News, 4(1):24–36, 2017.

19. J. Xu, W. Zhang, A. Alawini, and V. Tannen. Provenance analysis for missing answers and integrity repairs. IEEE Data Eng. Bull., 41(1):39–50, 2018.