# On the Limitations of Provenance for Queries With Difference


Yael Amsterdamer Tel Aviv University and INRIA

Daniel Deutch Ben Gurion University and INRIA

Val Tannen University of Pennsylvania


TaPP 2011

# Starting Point: Provenance Semirings

- Provenance semirings [(K,+,·,0,1)] were originally defined for the positive relational algebra
- Two important features of semirings
  - Algebraic uniformity
  - A correspondence between the semiring axioms and query (bag) equivalence identities: the semiring axioms are dictated by the identities!

# Correspondence of identities

| | Query Identities | Algebraic Identities |
|---|---|---|
| 1 | $R \cup (S \cup T) = (R \cup S) \cup T$ | $a+(b+c) = (a+b)+c$ |
| 2 | $R \cup \phi = R$ | $a+0 = a$ |
| 3 | $R \cup S = S \cup R$ | $a+b = b+a$ |
| 4 | $R \bowtie (S \bowtie T) = (R \bowtie S) \bowtie T$ | $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ |
| 5 | $R \bowtie 1 = R$ | $a \cdot 1 = a$ |
| 6 | $R \bowtie S = S \bowtie R$ | $a \cdot b = b \cdot a$ |
| 7 | $R \bowtie (S \cup T) = (R \bowtie S) \cup (R \bowtie T)$ | $a \cdot (b+c) = a \cdot b + a \cdot c$ |
| 8 | $R \bowtie \phi = \phi$ | $a \cdot 0 = 0$ |

**Semiring axioms!**

Security = (S, MIN, MAX, 0,1)

S = {1,C,S,T,0}

$1 < C < S < T < 0$

Emps

| Dep. | Emp | Prov. |
|---|---|---|
| Eng. | Alice | S |
| Eng. | Bob | T |
| Sales | Carol | S |

GoodEmps

| Emp | Prov. |
|---|---|
| Alice | C |
| Bob | S |
| Carol | T |

$\pi_{Dep}(Emps \bowtie GoodEmps)$

| Dep. | Prov. |
|---|---|
| Eng. | S·C+T·S<br>= S + T<br>= S |
| Sales | S·T = T |

# Suggested semantics for difference

- m-semirings [Geerts Poggi '10]

  a−b is the smallest c such that $a \leq b+c$

  (works for naturally ordered cases:

  $a \leq b \Leftrightarrow \exists c\ a + c = b$ is an order relation)
- By encoding as a nested aggregate query [Amsterdamer D. Tannen PODS '11]

  a-b=a if b=0, otherwise 0 (for positive semirings)

  – Also suggested for SPARQL

    [Theoharis, Fundulaki, Karvounarakis, Christophides '10]
- Z-semantics [Green Ives Tannen '09]

# Abstracting away

- Can we extend the framework to support difference?
- Work with a structure $(K,+,\cdot,0,1,-)$
- We still want $(K,+,\cdot,0,1)$ to be a semiring
- How do we define the additional operator?
- Let us try to throw in more axioms
  - A subset of those that hold for bag and set semantics

# Additional Identities

| | Query Identities | Algebraic Identities |
|---|---|---|
| 9 | $R - R = \phi$ | $a - a = 0$ |
| 10 | $\phi - R = \phi$ | $0 - a = 0$ |
| 11 | $R \cup (S - R) =$<br>$S \cup (R - S)$ | $a+(b - a) =$<br>$b+(a - b)$ |
| 12 | $R - (S \cup T) =$<br>$(R - S) - T$ | $a - (b+c) =$<br>$(a - b) - c$ |
| 13 | $R \bowtie (S - T) =$<br>$(R \bowtie S) - (R \bowtie T)$ | $a \cdot (b - c) =$<br>$(a \cdot b) - (a \cdot c)$ |

# Impossibility of satisfying the axioms

- Distributive lattices are particular semirings with an order relation such that
  - a+b is the least upper bound of a and b
  - a·b is the greatest lower bound of a and b
  - The security semiring, Three Value Logic are concrete examples
- Theorem If (K,+, ·, 0, 1,−) is an (extension of a) distributive lattice such that axioms 1-12 hold, and there exists in K two distinct elements a, b s.t. $a > b$ and $(a - b) \cdot b = 0$ then axiom 13 fails in K.

# Key observation

- Let (K,+,0) be a naturally ordered commutative monoid
  - Commutative monoid means axioms 1-3 hold
  - Naturally ordered means

    $a \leq b \Leftrightarrow \exists c\ a + c = b$ is an order relation

Theorem [Bosbach '65]: Axioms 9-12 hold if and only if

$a - b$ is the smallest c such that $a \leq b + c$

# Key Observation (cont.)

- For the security semiring, with
  $a = S, b = T$ we get
  $a - b = S$ and $(a - b) \cdot b = T = 0$

And indeed: $(S - T) \cdot T = S \cdot T = T$ but

$S \cdot T - T \cdot T = T - T = 0$

(S, MIN, MAX, 0,1)

S = {1,C,S,T,0}

$1 < C < S < T < 0$

**Emps**

| Emp | Prov. |
|---|---|
| Alice | S |
| Bob | T |
| Carol | S |

**GoodEmps**

| Emp | Prov. |
|---|---|
| Alice | C |
| Bob | S |
| Carol | T |

**FiredEmps**

| Emp | Prov. |
|---|---|
| Alice | C |
| Bob | S |
| Carol | T |

(Emps– FiredEmps) ⋈ GoodEmps

| Emp | Prov. |
|---|---|
| .. | .. |
| Carol | T |

Emps ⋈ GoodEmps – FiredEmps ⋈ GoodEmps

| Emp | Prov. |
|---|---|
| ... | ... |
| Carol | 0 |

# Where do solutions fail?

| Query Identities | Algebraic Identities | Fail for: |
|---|---|---|
| $R - R = \phi$ | $a - a = 0$ | |
| $\phi - R = \phi$ | $0 - a = 0$ | Z-Semantics |
| $R \cup (S - R) =$<br>$S \cup (R - S)$ | $a + (b - a) =$<br>$b + (a - b)$ | Agg, SPARQL |
| $R - (S \cup T) =$<br>$(R - S) - T$ | $a - (b+c) =$<br>$(a - b) - c$ | |
| $R \bowtie (S - T) =$<br>$(R \bowtie S) - (R \bowtie T)$ | $a \cdot (b - c) =$<br>$(a \cdot b) - (a \cdot c)$ | m-semirings |

# So what can we do?

- Work with a restricted class of semirings
  - We show in the paper another security semiring that is not a lattice; we use sets of security levels
  - Can we characterize the class for which bag equivalences hold?
- Give up on some of the equivalence axioms
- Give up on a uniform definition of difference