Solution Count for Multiset Unification with Trailing Multiset Variables

Iliano Cervesato iliano@itd.nrl.navy.mil

ITT Industries, Inc @ NRL - Washington DC

http://www.cs.stanford.edu/~iliano/

Outline

- Background
  - Notation
  - The problem
  - Motivations
- Solution Count
  - Simple-Simple
  - Simple-General
  - General-General
- Further work

Background

Multisets

- Set with repeated elements

$$M : S \to \mathbb{N}$$

Support set

  - $M(a)$: # occurrences of a in M
  - $|M|$: total # of elements in M

- Extensional notation: $\lfloor a_1, \ldots, a_n \rfloor$

Multisets with Variables

Constant a or
Object variable x
- range over support

Trailing Mset Variable
- range over msets
- at most 1 in each mset

$$\lfloor t_1, \ldots, t_n \mid R \rfloor$$

[Cf. Prolog's lists]

- Substitutions θ
  - Map object variables to objects
  - Map trailing variables to msets
  - E.g.:
    - $\lfloor a, b, x, y \mid R \rfloor \underbrace{[a/x, c/y, \lfloor z \mid R' \rfloor / R]}_{\theta} = \lfloor a, b, a, c, z \mid R' \rfloor$

The Problem

- Equation: $\lfloor t_1, \dots, t_n \mid R \rfloor \;=^?=\; \lfloor t'_1, \dots, t'_{n'} \mid R' \rfloor$
- Solution: θ s.t.

  $$\lfloor t_1, \dots, t_n \mid R \rfloor \theta = \lfloor t'_1, \dots, t'_{n'} \mid R' \rfloor \theta$$

- Solution count

How many solutions are there, at most?

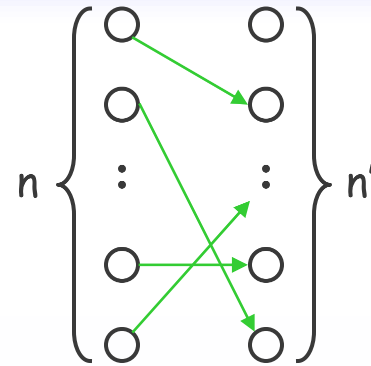
Motivations

Verification of Security Protocols [Meadows, '01]

- Group Diffie-Hellman protocol
- System of equations of the form
  - $(a^{x_1 x_2 \dots x_n \alpha}) \bmod p =^?= (a^{y_1 y_2 \dots y_{n'} \beta}) \bmod p$
  - Base a, prime p are fixed
  - x_i, y_j are either
    - known values in $\mathbb{N}_p$ or
    - known to exist (occur in other equations)
  - α, β are an unknown product
- Essential aspects
  - Commutativity / associativity of $\times$
  - x_i, y_j are "atomic"
  - Only α, β range over products of unknown size
- Represent as $\lfloor x_1, x_2, \dots, x_n \mid \alpha \rfloor =^?= \lfloor y_1, y_2, \dots, y_{n'} \mid \beta \rfloor$

Solution Count

Simple-Simple



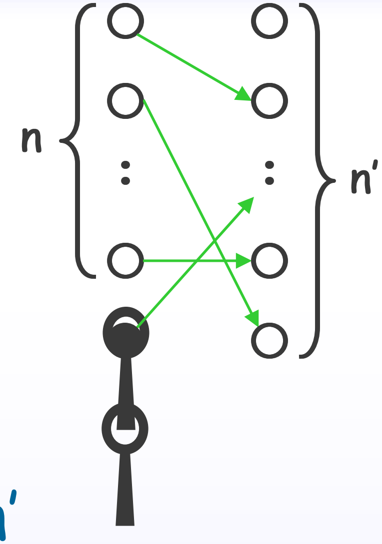
$$\lfloor t_1, \ldots, t_n \rfloor \;=?=\; \lfloor t'_1, \ldots, t'_{n'} \rfloor$$

- Map each node on the left with a node on the right

- $n = n'$
  - n! solutions

- $n \neq n'$
  - 0 solutions

Simple-General



$$\lfloor t_1, \ldots, t_n \mid R \rfloor \;=?=\; \lfloor t'_1, \ldots, t'_{n'} \rfloor$$

- Use R as an object var. factory till $n=n'$

- $n > n'$
  - 0 solutions
- $n = n'$
  - $n!$ solutions

- $n < n'$
  - $\binom{n'}{n} \times n!$ solutions

Example: $\lfloor c \mid R \rfloor \;=?=\; \lfloor a, b, c \rfloor$

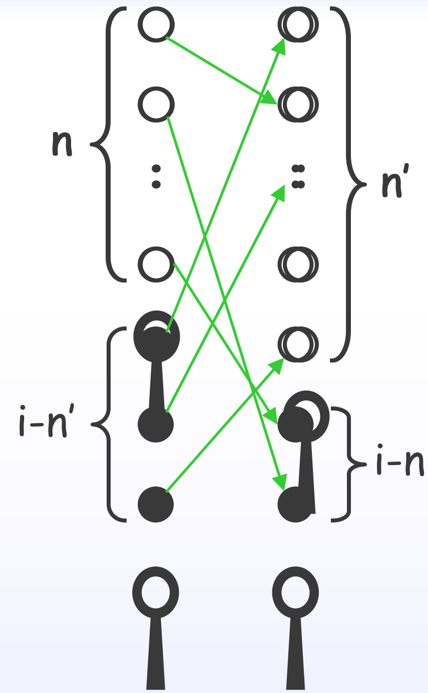
- Sol: $[\lfloor a, b \mid R' \rfloor / R]$
- Computation:
  - $\lfloor c, X, Y \mid R' \rfloor \;=?=\; \lfloor a, b, c \rfloor$
  - $[a/X, b/Y] \approx [b/X, a/Y]$

General-Simple

$$\lfloor t_1, \ldots, t_n \rfloor \;=?=\; \lfloor t'_1, \ldots, t'_{n'} \mid R' \rfloor$$

Dual

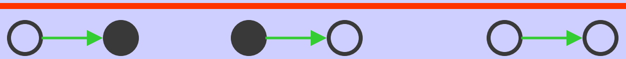
General-General



$\lfloor t_1, \ldots, t_n \mid R \rfloor$ =?= $\lfloor t'_1, \ldots, t'_{n'} \mid R' \rfloor$

- Pull new object var. out of R and R′ to get same *size* msets
- What size?
  - At least max(n,n′)
  - At most n+n′

$$\boxed{\sum_{i=\max(n,n')}^{n+n'} \binom{i-n'}{n} \times \binom{i-n}{n'} \times (n+n'+i)! \qquad \text{solutions}}$$



Further Work

Conclusions

- Solution generation cost
  - $O(2^n)$ [n = size of equation]
    - Possible solutions
    - Attempts in any case
  - Generally done at each computation step
- This is a worst case scenario
- Comparison
  - Free algebra: $O(n + \varepsilon)$
  - λ-calculus: undecidable

Can we do better?

- In many cases, probably yes
  - Indexing
  - Smart data structures
  - Lazy unification