# Solution Count for Multiset Unification with Trailing Multiset Variables

Iliano Cervesato
ITT Industries, Inc. — AES Division

No Institute Given

## 1 Motivations and Background

In this extended abstract, we consider a subproblem of AC1 unification [2, 4] applied to multisets [3], where the expressions being unified mention at most one multiset variable. This extends the unification problem of lists in Prolog-like languages with commutativity. We give an upper bound to the number of solutions and outline an algorithm for their generation.

This problem emerged in an attempt to automate the verification of the Group Diffie-Hellman cryptographic protocol [1], where an attacker may combine expressions of the form $g^{x_1 \cdots x_n X}$ to mount an attack against the protocol. Each $x_i$ is a number in $\mathbb{Z}_p$ ($p$ large prime) and $X$ is a product of such numbers of unknown size. We symbolically model such expressions as the multiset of their exponents.

## 2 The Problem

A *multiset* over a set $S$ is a function $M : S \longrightarrow \mathbb{N}$. The set $S$ is called the *support* of $M$ and will usually be assumed. In analogy with the usual extensional notation for finite sets, where we write $\{a_1, \ldots, a_n\}$ for the set consisting of the elements $a_1, \ldots, a_n$, we will denote a multiset with elements $a_1, \ldots, a_n$ (possibly with replications) as $\wr a_1, \ldots, a_n \smallint$. We write $|M|$ for the number of (possibly repeated) elements of $M$.

In this paper, we consider the problem of unifying *multiset expressions* of the form $\wr t_1, \ldots, t_n | R \smallint$ where $t_i$ is either a constant $a$ from the support set $S$ or an *object variable* $X$ that can be instantiated only to elements of $S$, and $R$ is either the empty multiset $\wr\smallint$ or a *trailing multiset variable* whose values range over multisets over $S$. Multiset expressions where $R = \wr\smallint$ are called *simple*, and we will abbreviate them as $\wr t_1, \ldots, t_n \smallint$.

A *multiset equation* is then an expression of the form

$$\wr t_1, \ldots, t_n | R \smallint \overset{?}{=} \wr t'_1, \ldots, t'_{n'} | R' \smallint$$

where the two sides are multiset expressions. A *unifier* for this equation is a substitution $\theta$ that maps object variables to either constants or object variables, and multiset variables to multisets or multiset expressions such that

$$\wr t_1, \ldots, t_n | R \smallint [\theta] = \wr t'_1, \ldots, t'_{n'} | R' \smallint [\theta].$$

Given a multiset $M$, a *sequence* $s_M$ on $M$ is any total ordering of $M$. For any $M$ with $m = |M|$ elements, there are at most $m!$ distinct sequences on $M$. This is an exact bound if $M$ has no repeated elements. The number of distinct submultisets of $M$ with $n$ elements (with $0 \leq n \leq m$) is at most $\binom{m}{n} = \frac{m!}{n! \cdot (m-n)!}$. This bound is exact if all elements of $M$ are distinct. This corresponds to the number of ways of picking $n$ items out of a collection containing $m$ distinct objects.

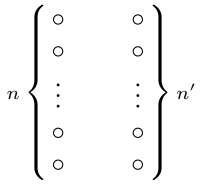
## 3 Solution Count

A multiset unification problem does not admit a unique unifier, in general, not even when limiting ourselves to most general unifiers. For example, the problem $\Lbag a, b \Rbag \overset{?}{=} \Lbag X, Y \Rbag$ has two solutions $(X \mapsto a, Y \mapsto b)$ and $(X \mapsto b, Y \mapsto a)$. There are however always finitely many most general unifiers. The purpose of this section is to prove this, and to compute exact worst-case bounds. We distinguish cases on the form of the two sides

### 3.1 Simple-Simple: $\Lbag t_1, \ldots, t_n \Rbag \overset{?}{=} \Lbag t'_1, \ldots, t'_{n'} \Rbag$

We consider first the situation in which both sides of the equation are simple terms (no multiset variables are present).

A convenient way to picture this situation, in particular when dealing with more general cases, is to rely on the graphical abstraction at right. We represent each of the $t_i$'s an hollow bullet on the left and each of the $t'_i$'s as a similar bullet on the right:



The unification problem is then reduced to drawing edges from the left half of the graph to its right-hand side so that exactly one edge enters/exits each node. The number of solutions depends on the values of $n$ and $n'$:

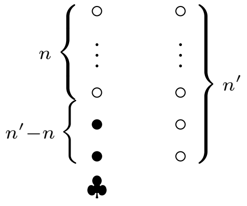
$\mathbf{n \neq n'}$**:** 0 solutions. Clearly, if $n \neq n'$, at least one node will be left out.

$\mathbf{n = n'}$**:** $n!$ solutions. In this case, we can pair up the two sets of nodes in $n!$ ways (if we keep the left string fixed, each permutation of the right sequence is a potential solution).

### 3.2 Simple-General: $\Lbag t_1, \ldots, t_n | R \Rbag \overset{?}{=} \Lbag t'_1, \ldots, t'_{n'} \Rbag$
### General-Simple: $\Lbag t_1, \ldots, t_n \Rbag \overset{?}{=} \Lbag t'_1, \ldots, t'_{n'} | R' \Rbag$

We now consider the case where one side of the equation is a simple multiset expression, while the other side is unrestricted. By symmetry, we can concentrate on the first possibility.

We extend our graphical reasoning by thinking about $R$ as a magic hat from which we can extract new variables as the need arises. Indeed, if we know that $R$ has at least one element, it is certainly the case that $R = \Lbag X | R^* \Rbag$, where $X$ and $R^*$ are new variables.



This possibility somewhat relaxes the cardinality constraints of the *simple-simple* case. Since the two multisets must eventually have the same number of elements, we shall be able to pull $n' - n$ new variables out of $R$. Let us denote them with filled bullets in our graph, and write the residual multiset variable as $\clubsuit$ (clearly, it shall be instantiated with the empty multiset).

We can now count the solutions:

$\mathbf{n} > \mathbf{n}'$**:** 0 solutions. If there are originally more nodes on the left than on the right, there is no way we can obtain a one-to-one matching.

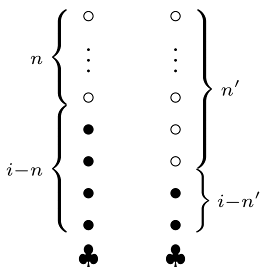
$\mathbf{n} = \mathbf{n}'$**:** $n!$ solutions. If the two side have originally the same number of hollow nodes, then we are essentially in the *simple-simple* situation.

$\mathbf{n} < \mathbf{n}'$**:** $\binom{n'}{n} \cdot n!$ solutions. If there are originally more nodes on the right than on the left, then we can create the missing filled bullets so that the two sides have the same number of nodes. Observe that the number of solutions is not $n'!$: while the precise way in which we relate hollow nodes counts, this does not apply to filled nodes. Indeed, if $a$ and $b$ correspond to hollow nodes, and $X$ and $Y$ to filled nodes, the mappings $(X \mapsto a, Y \mapsto b)$, and $(X \mapsto b, Y \mapsto a)$ are the same solution since $X$ and $Y$ are ultimately part of $R$.
Therefore, we shall pick $n' - n$ nodes out of $n'$ from the left-hand side to be associated with the filled bullets (or equivalently pick $n$ out $n'$ nodes to be mapped to the hollow bullets). For each possibility, we shall consider all the permutations of the remaining $n$ hollow bullets. Therefore, we obtain the reported total number of solutions. Observe that this situation implies the case where $n = n'$ as $\binom{n}{n} = 1$ for any $n$.

### 3.3 General-General: $\wr \boldsymbol{t_1, \ldots, t_n | R} \int \overset{?}{=} \wr \boldsymbol{t'_1, \ldots, t'_{n'} | R'} \int$

We will now consider the fully general case where both sides of the equation mention multiset variables. Observe that each $t_i$ can be mapped to a $t'_j$ but can also be related to a yet undefined member of $R'$. Dually, each $t'_i$ can be associated with either a $t_j$ or with some indistinct element of $R$. We will proceed by examining the cardinality of the multiset, $M$ say, resulting from the instantiation of the two sides of the equation. For this purpose, we will again align hollow bullets for each $t_i$ and $t'_j$, and we will extend each alignment with as many filled bullets as needed to achieve the hypothetical cardinality of $M$. Again, trailing multiset variables are rendered as $\clubsuit$.

First notice that $|M|$ must be at least $\max(n, n')$ for any solution to exist. On the other hand, values of $|M|$ greater than $n + n'$ do not contribute to the solution since this would imply that some of the filled bullets on the left are mapped to filled bullets on the right: they would be subsumed by solutions of cardinality $|M| - 1$ which push back these nodes into the respective $\clubsuit$'s.



Let $\max(n, n') \leq i \leq n + n'$ be the hypothesized cardinality of $M$. We count solutions as follows: first we shall pick $i - n$ out of $n'$ hollow bullets from the right-hand side and map them to the filled bullets on the left; then we dually pick $i - n'$ out of $n$ hollow bullets on the left-hand side and map them to the filled bullets on the right; finally, we associate the remaining $n + n' - i$ hollow bullets to each other in all possible ways.

The total number of most general solutions is then given by:

$$\sum_{i=\max(n,n')}^{n+n'} \binom{i-n'}{n} \cdot \binom{i-n}{n'} \cdot (n+n'-i)! \quad \text{solutions.}$$

Observe that this case always has a solution. Observe also that this situation subsumes the previous two as soon as we impose a strict bound on the cardinality of the solution multiset.

## 4 Algorithm

The previous solution count suggests an immediate algorithm for producing these solutions. The cost is clearly very high (exponential), and there are likely to be heuristics and data structures to lower this cost (at least when some constant elements are present).

### 4.1 Simple-Simple: $\boldsymbol{\Lbag t_1, \ldots, t_n \Rbag \overset{?}{=} \Lbag t'_1, \ldots, t'_{n'} \Rbag}$

1. If $n \neq n'$, fail, otherwise proceed as follows.
2. Pick a permutation $p_j$ of the indices $1..n$.
3. Construct the system of object equalities

$$S_j = \{t_1 \overset{?}{=} t'_{p_j(1)}, \quad \ldots \quad, t_n \overset{?}{=} t'_{p_j(n)}\}$$

4. If $S_j$ is solvable add it to the solution set, otherwise discard it.
5. Repeat steps 2 to 4 until all $n!$ permutations of $1..n$ have been examined.

### 4.2 Simple-General: $\boldsymbol{\Lbag t_1, \ldots, t_n | R \Rbag \overset{?}{=} \Lbag t'_1, \ldots, t'_{n'} \Rbag}$ General-Simple: $\boldsymbol{\Lbag t_1, \ldots, t_n \Rbag \overset{?}{=} \Lbag t'_1, \ldots, t'_{n'} | R' \Rbag}$

We will concentrate on the Simple-General case: the General-Simple case is dual.

1. If $n > n'$, fail, otherwise proceed as follows.
2. Pick $n$ elements out of $t'_1, \ldots, t'_{n'}$. Call these elements $\bar{t}'_1, \ldots, \bar{t}'_n$ and call the unselected elements $\hat{t}'_1, \ldots, \hat{t}'_{n'-n}$.
3. Pick a permutation $p_j$ of the indices $1..n$.

4. Construct the system of object equalities

$$S_j = \{t_1 \overset{?}{=} \bar{t}'_{p_j(1)}, \quad \ldots \quad, t_n \overset{?}{=} \bar{t}'_{p_j(n)}\}$$

5. If $S_j$ is solvable add it together with the multiset variable substitution

$$R \mapsto \lbag \hat{t}'_1, \ldots, \hat{t}'_{n'-n} \rbag$$

to the solution set, otherwise discard it.
6. Repeat steps 3 to 5 until all $n!$ permutations of $1..n$ have been examined.
7. Repeat steps 2 to 6 until all $\binom{n}{n'}$ ways of picking $n$ out of $n'$ elements of the right-hand side have been examined.

### 4.3 General-General: $\boldsymbol{\lbag t_1, \ldots, t_n | R \rbag \overset{?}{=} \lbag t'_1, \ldots, t'_{n'} | R' \rbag}$

1. Set $i = \max(n, n')$.
2. Pick $n + n' - i$ elements out of $t'_1, \ldots, t'_{n'}$. Call these elements $\bar{t}'_1, \ldots, \bar{t}'_{n+n'-i}$ and call the unselected elements $\hat{t}'_1, \ldots, \hat{t}'_{i-n}$.
Similarly, pick $n + n' - i$ elements out of $t_1, \ldots, t_n$. Call these elements $\bar{t}_1, \ldots, \bar{t}_{n+n'-i}$ and call the unselected elements $\hat{t}_1, \ldots, \hat{t}_{i-n'}$.
3. Pick a permutation $p_j$ of the indices $1..(n + n' - i)$.
4. Construct the system of object equalities

$$S^i_j = \{t_1 \overset{?}{=} \bar{t}'_{p_j(1)}, \quad \ldots \quad, t_{n+n'-i} \overset{?}{=} \bar{t}'_{p_j(n+n'-i)}\}$$

5. If $S^i_j$ is solvable add it together with the multiset variable substitutions

$$R \mapsto \lbag \hat{t}'_1, \ldots, \hat{t}'_{i-n} | R^* \rbag \quad \text{and} \quad R' \mapsto \lbag \hat{t}_1, \ldots, \hat{t}_{i-n'} | R^* \rbag$$

to the solution set, otherwise discard it. $R^*$ is a new multiset variable.
6. Repeat steps 3 to 5 until all $(n + n' - i)!$ permutations of $1..(n + n' - i)$ have been examined.
7. Repeat steps 2 to 6 until all $\binom{i-n}{n'} \cdot \binom{i-n'}{n}$ ways of picking $i - n$ out of $n'$ elements on the right-hand side and $i - n'$ out of $n$ elements on the left-hand side have been examined.
8. If $i \leq n + n'$, increment $i$ by 1 and go back to step 1.

## References

1. G. Ateniese, M. Steiner, and G. Tsudik. Authenticated group key management and friends. In *Proc. Conf. on Computer and Communication Security*. ACM Press, 1998.
2. F. Baader and T. Nipkow. *Term Rewriting and All That*. Cambridge University Press, 1998.
3. A. Dovier, A. Policriti, and G. Rossi. A uniform axiomatic view of lists, multisets, and sets, and the relevant unification algorithms. *Fundamenta Informaticae*, 36(2–3):201–235, 1998.
4. D. Kapur and P. Narendran. Double-exponential complexity of computing a complete set of AC-unifiers. In *Proc. Logic in Computer Science — LICS'92*, pages 11–21, Santa Cruz, CA, 1992.