Certifying Digital Rights' Expression Languages

by

Bahman Sistany

Ph.D.Thesis Proposal For the Ph.D. degree in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science Faculty of Engineering University of Ottawa

Ph.D.Thesis Advisor: Amy Felty

Abstract

This is the abstract.

${\bf Acknowledgements}$

I would like to thank all the little people who made this possible.

Dedication

This is dedicated to the one I love.

Table of Contents

Li	st of	Tables	
List of Figures			
1	Intr	roduction	1
2	Middle		3
	2.1	Logic Based Semantics for ODRL	3
	2.2	Pucella 2006	3
	2.3	what will I do?	4
		2.3.1 Coq	4
	2.4	Abstract Syntax	4
	2.5	Coq Version	8
	2.6	Semantics	9
\mathbf{A}	PPE:	NDICES	17
References			17

List of Tables

List of Figures

Chapter 1

Introduction

Digital rights management, DRM, refers to the digital management of rights associated with the access or usage of digital assets. There are various aspects of rights management however. According to the authors of the whitepaper "A digital rights management ecosystem model for the education community," digital rights management systems cover the following four areas: 1) defining rights 2) distributing/acquiring rights 3) enforcing rights and 4) tracking usage [1].

Rights Expression Languages, RELs, or more precisely when dealing with digital assets, Digital Rights Expression Languages DRELs deal with the "rights definition" aspect of the DRM ecosystem. A DREL, allows the expression and definition of digital asset usage rights such that other areas of the DRM ecosystem namely the enforcement mechanism and the usage tracking components can function correctly.

Currently the most popular RELs are the eXtensible rights Markup Language, XrML [bib], and the Open Digital Rights Language, ODRL [bib]. Both of these languages are XML based and are considered declarative languages. XrML has been selected to be the REL for MPEG-21 which is an ISO standard for multimedia applications. ODRL is also a standards based REL which has been accepted as part of the W3C community with the mandate of standardizing how rights and policies, related to the usage of digital content on the Open Web Platform, OWP, are expressed [wikipedia]. ODRL 2.0 supports expression of rights and also privacy rules for social media while ODRL 1.0 was only dealing with the mobile ecosystem – ODRL 1.0 was adopted by the Open Mobile Alliance, OMA in 2000.

As popular as both XrML and ODRL are, their adoption and usage is still somewhat limited in practice. Both Apple and Microsoft for example have defined their own lightweight RELs [problem with RELs paper] in Fair Play (Apple) and in PlayReady (Microsoft). The authors of [the problem with RELs] argue that both these RELs and other ones are simply too complex to be used effectively since they try to cover much of the DRM ecosystem.

Another issue with the current batch of RELs are due to their semantics being expressed in a natural language (e.g. English). By necessity natural languages are ambiguous and open to interpretation.

To formalize the semantics of RELS several approaches have been attempted by various authors. The main categories are logic based, operational semantics based interpreters and finally web ontology based (from the Knowledge Representation Field). In this thesis we will focus on the logic based approach to formalizing semantics and will study a specific logic based language that is a translation from a subset of ODRL.

Chapter 2

Middle

2.1 Logic Based Semantics for ODRL

Formal logic can represent the statements and facts we express in a natural language like English. Propositional logic is expressive enough to express simple facts as propositions and allows uses connectives to allow for the negation, conjunction and disjunction of the facts. However propositional logic is not expressive enough to express policies of the kind used in languages like ODRL and XrML. For example, a simple policy expressed in English like "All who pay 5 dollars can watch the movie Toy Story" cannot be expressed in propositional logic because the concept of variables doesn't exist.

The higher order logic called "Predicate Logic" or "First Order Logic" FOL is more suitable and has the expressive power to represent policies written in English. Moreover, FOL can be used to capture the meaning of policies in an unambiguous way.

Halpern and Weissman [Using First Order Logic to Reason about Policies] propose a fragment of FOL to represent and reason about policies. The fragment of FOL they arrive at is called *Lithium* which is decidable and allows for efficiently answering interesting queries. Lithium restricts policies to be written based on the concept of "bipolarity" which disallows by construction policies that both permit and deny an action on an object.

2.2 Pucella 2006

Pucella and Weissman [2] specify a predicate logic based based language that represents a subset of ODRL.

2.3 what will I do?

2.3.1 Coq

âĂć Program correctness âĂć Formal verification of software âĂć Certified programs âĂć Proof assistant âÁć Interactive and mechanized theorem proving âÁć Examples of machine assisted proofs: CompCert, four-color theorem proof âAć Coq is based on a higher-order functional programming language âÅć Dependent Types âÜN Subset types âÜN Easier than writing explicit proofs âAć Write formal specification and proofs that programs comply to their specification (a-short-intro-to-coq) âÁć Automatically extract code from specifications as Ocaml or Haskell (a-short-intro) âAć Properties, programs and proofs are all formalized in the same language called CIC (Calculus of inductive Constructions). (ashort-intro) â Áć Coq uses a sort called Prop for propositions â Áć Coq art: â Áć Well-formed propositions are assertions one can express about values such as mathematical objects or even programs e.g. 3 < 8 âUN Note that assertions may be true, false or simply conjectures âŬŃ An assertion is only true in general if a proof is provided âŬŃ However hand written proofs are difficult to verify âUN Coq provides an environment for developing proofs including a formal language to express proofs in, the language itself being built using proof theory making it possible to step by step verification of the proofs âUN Mechanized proof verification requires a "proof" that the verification algorithm is correct itself in applying all the formal rules correctly

2.4 Abstract Syntax

[2] uses abstract syntax instead of XML to express statements in the ODRL language. The abstract syntax used is a more compact representation than XML based language ODRL policies are written in and furthermore it simplifies specifying the semantics as we shall see. As an example here is an agreement written in ODRL and the comparable agreement expressed in the abstract syntax [2].

Listing 2.1: agreement for Mary Smith in XML

```
</print>
<requirement>
  <prepay>
        <payment> <amount currency="AUD"> 5.00</amount> </payment>
        </prepay>
        </requirement>
        </permission>
        <party> <context> <name> Mary Smith </name> </context> 

</agreement>
```

The agreement 2.1 is shown below using the syntax from [2].

Listing 2.2: agreement for Mary Smith as BNF (as used in [2])

```
agreement
for Mary Smith
about Treasure Island
with prePay[5.00] -> and[cpu[Mary's Computer] => display,
count[2] => print].
```

In the following we will cover the abstract syntax of a subset of ODRL expressed as Coq's constructs such as Inductive Types and Definitions. We will call this subset ODRL0 both because it is a variation of Pucella's ODRL language and also because it is missing some ODRL constructs such as Requirements and Conditions - we will add the missing pieces making up what we will call ODRL1 and perhaps ODRL2 (the latter only if needed). We will also describe ODRL0 in a BNF grammar that looks more like Pucella's ODRL grammar. BNF style grammars are less formal as they give some suggestions about the surface syntax of expressions [Pierce1] without getting into lexical analysis and parsing related aspects such as precedence order of operators. The Coq version in contrast is more formal and could be directly used for building compilers and interpreters. We will present both the BNF version and the Coq version for each construct of ODRL0 [Pierce1]. To get started let's see what the listing 2.2 would look like in ODRL0's Coq version.

Listing 2.3: Coq version of agreement for Mary Smith

The top level ODRL0 production is the *agreement*. An agreement expresses what actions a set of subjects may perform on an object and under what conditions. Syntactically

an agreement is composed of a set of subjects/users called a *principal* (*prin*), an *asset* and a *Policy Set* (*PolicySet*).

Listing 2.4: agreement

Principals or prins are composed of *subjects* which are specified based on the application e.g. Alice, Bob, etc for the DRM application we will be using throughout.

```
Listing 2.5: prin
```

```
{\tt \langle prin \rangle} ::= {\tt \langle subject_1 \rangle, \ldots, \langle subject_m \rangle}
```

Listing 2.6: subject

```
<subject> ::= N
```

Assets are also application specific but similar to subjects we will use specific ones for the DRM application (taken from [2]). ebook, $The\ Report$ and latestJingle are examples of specific subjects we will be using throughout. Syntactically an asset is just a positive number (N).

Listing 2.7: asset

```
<asset> ::= N
```

Agreements include policy sets. Each policy set specifies a prerequisite and a policy. In general if the prerequisite holds the policy is taken into consideration. Otherwise the policy will not be looked at. Some policy sets are specified as exclusive. The Primitive Exclusive Policy Sets are exclusive to agreement's users in that only those users may perform the actions specified in the policy set. The implication is that all other users who are not specified in the agreement's principal (prin) are forbidden from performing the specified actions. Finally policy sets could be grouped together in a conjunction allowing a single agreement to be associated with many policy sets.

Listing 2.8: policySet

```
<policySet> ::=
    | <PrimitivePolicySet> : <preRequisite> \rightarrow <policy>
    | <PrimitiveExclusivePolicySet> : <preRequisite> \mapsto <policy>
    | <AndPolicySet> : 'and'[ <policySet<sub>1</sub>>, ..., <policySet<sub>m</sub>> ]
```

A policy specifies an action to be performed on an asset, depending of whether the policy's prerequisite holds or not. If the prerequisite holds the agreement's user is permitted to perform the action on the agreement's asset; otherwise permission is denied. Similar to policy sets, policies could also be grouped together in a conjunction. The policy also includes a unique identifier. The policy identifier is added to help the translation (from agreements to formulas) but is optional in ODRL proper.

Listing 2.9: policy

```
<policy> ::=
  | <PrimitivePolicy> : <preRequisite> \Rightarrow_{<policyId>} <act>
  | <AndPolicy> : 'and'[ <policy<sub>1</sub>>, ..., <policy<sub>m</sub>> ]
```

An Action (act) is simply a positive number. Similar to assets and subjects, actions are application specific. Some example actions taken from [2] are Display and Print.

```
Listing 2.10: act
```

```
<act> ::= N
```

A $Policy\ Id\ (policy\ Id)$ is a unique identifier specified as (increasing) positive integers.

```
Listing 2.11: policyId
```

```
<policyId> ::= N
```

In ODRL0 a prerequisite is either true or it is a constraint. The true prerequisite always holds. A constraint is an intrinsic part of a policy and cannot be influenced by agreement's user. Minimum height requirements for popular attractions and rides are examples of we would consider a constraint. The constraint ForEachMember is interesting in its expressive power but has complicated semantics as we shall see in the 2.6 section. Roughly speaking, ForEachMember takes a prin (a list of subjects) and a list L of constraints. The ForEachConstraint holds if each subject in prin satisfies each constraint in L.NotCons is a negation of a constraint. The set of prerequisites are closed under conjunction (AndPrqs), disjunction (OrPrqs) and exclusive disjunction (XorPrqs).

Listing 2.12: preRequisite

Constraints are either *Principal*, *Count* or *CountByPrin*. Principal constraints basically require matching to specified prins. For example, the user being Alice is a Principal constraint. A count constraint refers to a set of policies P and specifies the number of times the user of an agreement has invoked the policies in P to justify her actions. If the count constraint is part of a policy then the set P is composed of the single policy. In the case that the count constraint is part of a policy set, the set P is the set of policies specified in the policy set.

Listing 2.13: constraint

```
<constraint> ::=
    | <Principal> : <prin>
    | <Count> : 'Count' [N]
    | <CountByPrin> : <prin> ('Count' [N])
```

2.5 Coq Version

Listing 2.14: Coq version of agreement

Listing 2.15: prin

```
Definition prin := nonemptylist subject.
```

Listing 2.16: asset

```
Definition asset := nat.
```

Listing 2.17: subject

```
Definition subject := nat.
```

Listing 2.18: policySet

```
Inductive policySet : Set :=
    | PrimitivePolicySet : preRequisite → policy → policySet
    | PrimitiveExclusivePolicySet : preRequisite → policy → policySet
    | AndPolicySet : nonemptylist policySet → policySet.
```

Listing 2.19: policy

Listing 2.20: act

```
Definition act := nat.
```

Listing 2.21: policyId

```
{\tt Definition\ policyId} := {\tt nat}.
```

Listing 2.22: preRequisite

```
Inductive preRequisite : Set :=
    | TruePrq : preRequisite
    | Constraint : constraint → preRequisite
    | ForEachMember : prin → nonemptylist constraint → preRequisite
    | NotCons : constraint → preRequisite
    | AndPrqs : nonemptylist preRequisite → preRequisite
    | OrPrqs : nonemptylist preRequisite → preRequisite
    | XorPrqs : nonemptylist preRequisite → preRequisite.
```

Listing 2.23: constraint

```
Inductive constraint : Set :=
    | Principal : prin → constraint
    | Count : nat → constraint
    | CountByPrin : prin → nat → constraint.
```

2.6 Semantics

In this section, we describe the semantics of ODRL0 language by a translation from each language object (e.g. agreement) to a proposition in Coq. The semantics will help answer queries of the form "may subject s perform action act to asset a?". If the answer is yes, we say permission is granted. Otherwise permission is denied.

Whether a permission is granted or denied depends on the agreements in question but also on the facts recorded in the environment. For ODRL0 those facts revolve around the number of times a policy has been used to justify an action. We encode this information in an *environment* which is a conjunction of equalities of the form count(s, policyId) = n.

The Coq version of the count equality is a new inductive type called *count_equality*. An environment is defined to be a non-empty list of count_equality objects.

Listing 2.24: Environments and Counts

```
Inductive count_equality : Set :=
    | CountEquality : subject → policyId → nat → count_equality.

Inductive environment : Set :=
    | SingleEnv : count_equality → environment
    | ConsEnv : count_equality → environment → environment.
```

At high-level, an agreement is translated into a conjunction of formulas of the form $\forall x (prerequisites(x) \to P(x))$ where P(x) itself is a conjunction of formulas of the form $prerequisites(x) \to (\neg) Permitted(x, act, a)$, where x is a variable of type subject that occurs free in P(x) and Permitted is a predicate. A predicate in Coq is any function that returns a term of type Prop. Here is how we have specified the Permitted predicate in

Coq: Parameter Permitted : subject -> act -> asset -> Prop. So Permitted predicate takes a subject, an act (action) and an asset and returns a Prop. Note that the notation (\neg) indicates that Permitted may be negated.

More specifically translation of the top level agreement element proceeds by case analysis on the structure of the agreement. Note that each translation function takes an environment parameter.

Listing 2.25: Translation of agreement

```
Definition trans_agreement (e:environment)(ag:agreement) : Prop := match ag with | Agreement prin_u a ps ⇒ trans_ps e ps prin_u a end.
```

Translation of a *policySet* proceeds with case analysis of different policySet constructors. We then recurse into translation functions for the composing elements. A policySet is either a *PrimitivePolicySet*, *PrimitiveExclusivePolicySet* or a *AndPolicySet*.

Translation of a PrimitivePolicySet yields a formula

Listing 2.26: Policy Set Translation: PrimitivePolicySet

Listing 2.27: Policy Set Translation: PrimitiveExclusivePolicySet

Listing 2.28: Policy Set Translation: AndPolicySet

```
 \llbracket and[policySet_1,...,policySet_m] \rrbracket^{e,prin_u,a} \triangleq \llbracket policySet_1 \rrbracket^{e,prin_u,a} \wedge ... \wedge \llbracket policySet_m \rrbracket^{e,prin_u,a}
```

Listing 2.29: Translation of Policy Set

```
Fixpoint trans_ps
  (e:environment)(ps:policySet)(prin_u:prin)(a:asset){struct ps} : Prop :=

let trans_ps_list := (fix trans_ps_list (ps_list:nonemptylist policySet)(prin_u:prin)
        (a:asset){struct ps_list}:=

match ps_list with
    | Single ps1 \Rightarrow trans_ps e ps1 prin_u a
    | NewList ps ps_list' \Rightarrow ((trans_ps e ps prin_u a) /\ (trans_ps_list ps_list' prin_u a
    ))
```

Listing 2.30: Prin Translation: Single subject

```
[[subject]]_x \triangleq x = subject
```

Listing 2.31: Prin Translation: List of subjects

```
[\![\{subject_1,...,subject_m\}]\!]_x \triangleq [\![subject_1]\!]_x \vee ... \vee [\![subject_m]\!]_x
```

Listing 2.32: Translation of a Prin

```
\begin{tabular}{ll} Fixpoint trans_prin & $(x:subject)(p:prin): Prop: = \\ \hline match p with & |Single s \Rightarrow (x=s) \\ |NewList s rest \Rightarrow ((x=s) \setminus / trans_prin x rest) \\ end. & \\ \end{tabular}
```

Listing 2.33: Positive Policy Translation: Single policy

```
    \begin{bmatrix} preRequisite \Rightarrow_{policyId} act \end{bmatrix}_{x}^{positive,e,prin_{u},a} \triangleq \\ ([preRequisite]_{x}^{e,policyId,prin_{u}}) \Rightarrow Permitted(x, [act], a)
```

Listing 2.34: Positive Policy Translation: List of policies

```
 \llbracket and[policy_1,...,policy_m] \rrbracket^{positive,e,prin_u,a} \triangleq \llbracket policy_1 \rrbracket^{positive,e,prin_u,a} \wedge ... \wedge \llbracket policy_m \rrbracket^{positive,e,prin_u,a}
```

Listing 2.35: Translation of a positive policy

```
Fixpoint trans_policy_positive
 (e:environment)(x:subject)(p:policy)(prin_u:prin)(a:asset){struct p} : Prop :=
let trans_p_list := (fix trans_p_list (p_list:nonemptylist policy)(prin_u:prin)(a:
   asset){struct p_list}:=
            match p_list with
             | Single p1 ⇒ trans_policy_positive e x p1 prin_u a
             | NewList p p_list' ⇒
                ((trans_policy_positive e x p prin_u a) /\
                 (trans_p_list p_list' prin_u a))
            end) in
 match p with
  PrimitivePolicy prq policyId action ⇒ ((trans_preRequisite e x prq (Single
   policyId) prin_u) →
                               (Permitted x action a))
  | AndPolicy p_list ⇒ trans_p_list p_list prin_u a
 end.
```

Listing 2.36: Negative Policy Translation: Single policy

```
[preRequisite \Rightarrow_{policyId} act]_x^{negative,e,prin_u,a} \triangleq ([preRequisite]_x^{e,policyId,prin_u}) \Rightarrow \neg (Permitted(x, [act], a))
```

Listing 2.37: Negative Policy Translation: List of policies

```
[\![and[policy_1,...,policy_m]\!]]^{negative,e,a} \triangleq [\![policy_1]\!]^{negative,e,a} \wedge ... \wedge [\![policy_m]\!]^{negative,e,a}
```

Listing 2.38: Translation of a negative policy

```
|\  \, \text{PrimitivePolicy prq policyId action} \Rightarrow \text{not (Permitted x action a)} \\ |\  \, \text{AndPolicy p\_list} \Rightarrow \text{trans\_p\_list p\_list a} \\ |\  \, \text{end}.
```

Listing 2.39: Prerequisite Translation: Always True Prerequisite

 $[prerequisite :: true] \triangleq True$

Listing 2.40: Prerequisite Translation: Constraint

 $[\![prerequisite :: constraint]\!]_x^{[id_1, \dots, id_m], prin_u} \triangleq [\![constraint]\!]_x^{[id_1, \dots, id_m], prin_u}$

Listing 2.41: Prerequisite Translation: ForEachMember

Listing 2.42: Prerequisite Translation: Not Constraint

 $\llbracket not \ prerequisite :: constraint \rrbracket_{x}^{[id_{1},...,id_{m}],prin_{u}} \ \triangleq \ \neg \llbracket constraint \rrbracket_{x}^{[id_{1},...,id_{m}],prin_{u}}$

Listing 2.43: Prerequisite Translation: Conjunction

 $\llbracket and \ [preRequisite_1,...,preRequisite_k] \rrbracket^{[id_1,...,id_m],prin_u} \triangleq \llbracket preRequisite_1 \rrbracket^{[id_1,...,id_m],prin_u} \land ... \land \llbracket preRequisite_k \rrbracket^{[id_1,...,id_m],prin_u}$

Listing 2.44: Prerequisite Translation: Inclusive Disjunction

Listing 2.45: Prerequisite Translation : Exclusive Disjunction

Listing 2.46: Translation of a PreRequisite

```
match prq with
    | TruePrq ⇒ True
    | Constraint const ⇒ trans_constraint e x const IDs prin_u
    | ForEachMember prn const_list ⇒ trans_forEachMember e x prn const_list IDs
    | NotCons const ⇒ trans_notCons e x const IDs prin_u
    | AndPrqs prqs ⇒ True
    | OrPrqs prqs ⇒ True
    | XorPrqs prqs ⇒ True
end.
```

Listing 2.47: Constraint Translation: Principal

```
\llbracket constraint :: prin \rrbracket_{x}^{[subject_{1}, \dots, subject_{m}]} \triangleq \llbracket prin \rrbracket_{x}^{[subject_{1}, \dots, subject_{m}]}
```

Listing 2.48: Constraint Translation: Count

```
\llbracket constraint :: count[N] \rrbracket_x^{[id_1, \dots, id_m], prin_u} \triangleq \llbracket count[N] \rrbracket_x^{[id_1, \dots, id_m], prin_u}
```

Listing 2.49: Constraint Translation: Count by Principal

Listing 2.50: Translation of a Constraint

```
Fixpoint trans_constraint
  (e:environment)(x:subject)(const:constraint)(IDs:nonemptylist policyId)
  (prin_u:prin){struct const} : Prop :=
  match const with
   | Principal prn ⇒ trans_prin x prn
   | Count n ⇒ trans_count e n IDs prin_u
   | CountByPrin prn n ⇒ trans_count e n IDs prn
  end.
```

Listing 2.51: For Each Member Translation: Count by Principal

```
 [\![constraint]\!]_x^{(subject_2, constraint_m), [id_1, \dots, id_n]} \wedge \dots \wedge [\![constraint]\!]_x^{(subject_k, constraint_1), [id_1, \dots, id_n]} \\ \wedge \dots \wedge [\![constraint]\!]_x^{(subject_k, constraint_m), [id_1, \dots, id_n]}
```

Listing 2.52: Translation of forEachMember

```
Fixpoint trans_forEachMember
      (e:environment)(x:subject)(principals: nonemptylist subject)(const_list:
   nonemptylist constraint)
      (IDs:nonemptylist policyId){struct const_list} : Prop :=
let trans_forEachMember_Aux
 := (fix trans_forEachMember_Aux
      (prins_and_constraints: nonemptylist (Twos subject constraint))
      (IDs:nonemptylist policyId){struct prins_and_constraints} : Prop :=
    match prins_and_constraints with
     | Single pair1 ⇒ trans_constraint e x (right pair1) IDs (Single (left pair1))
     \mid NewList pair1 rest_pairs \Rightarrow
        (trans_constraint e x (right pair1) IDs (Single (left pair1))) /\
        (trans_forEachMember_Aux rest_pairs IDs)
    end) in
    let prins_and_constraints := process_two_lists principals const_list in
    trans_forEachMember_Aux prins_and_constraints IDs.
```

The translation for "Not Constraint" was listed in listing 2.42 earlier but we repeat it here to go along the Coq version.

Listing 2.53: Not Constraint Translation

```
[\![not\ constraint]\!]_x^{[id_1,\ldots,id_m],prin_u} \triangleq \neg [\![constraint]\!]_x^{[id_1,\ldots,id_m],prin_u}
```

Listing 2.54: Translation of not cons

```
\label{eq:decomposition} \begin{split} & \texttt{Definition trans\_notCons} \\ & (\texttt{e:environment})(\texttt{x:subject})(\texttt{const:constraint})(\texttt{IDs:nonemptylist policyId})(\texttt{prin\_u:prin}) : \\ & \texttt{Prop} := \\ & \tilde{} & (\texttt{trans\_constraint e x const IDs prin\_u}). \end{split}
```

Listing 2.55: Count Translation: subject and policyId pair

```
[[count[N]]]_x^{subject_1,id_1} \triangleq getCount(subject_1,id_1) < N
```

Listing 2.56: Count Translation: subject and policyId pairs

```
 \begin{aligned} & [ count[N] ]_x^{[id_1,...,id_m],prin_u} \triangleq \\ & ( getCount(getSubject(prin_u)_1,id_1) + ... + getCount(getSubject(prin_u)_1,id_m) + ... + \\ & getCount(getSubject(prin_u)_k,id_1) + ... + getCount(getSubject(prin_u)_k,id_m)) < N \end{aligned}
```

Listing 2.57: Translation of count

References

- [1] Robby Robson Geoff Collier, Harry Piccariello. A digital rights management ecosystem model for the education community. *DRM Whitepapers: Content Guard*, 2004.
- [2] Riccardo Pucella and Vicky Weissman. A formal foundation for ODRL. CoRR, abs/cs/0601085, 2006.