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.

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

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.

1.1 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.

1.2 Pucella 2006

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

1.3 what will I do?

1.3.1 Coq

âĂć Program correctness âĂć Formal verification of software âĂć Certified programs âĂć Proof assistant âĂć Interactive and mechanized theorem proving âĂć Examples of machine assisted proofs: CompCert, four-color theorem proof âĂć Coq is based on a higher-order functional programming language âĂć Dependent Types âŮŃ Subset types âŮŃ Easier than writing explicit proofs âĂć Write formal specification and proofs that programs comply to their specification (a-short-intro-to-coq) âĂć Automatically extract code from specifications as Ocaml or Haskell (a-short-intro) âĂć Properties, programs and proofs are

all formalized in the same language called CIC (Calculus of inductive Constructions). (a-short-intro) âĂć Coq uses a sort called Prop for propositions âĂć Coq art: âĂć Well-formed propositions are assertions one can express about values such as mathematical objects or even programs e.g. $3 < 8$ âŮŃ Note that assertions may be true, false or simply conjectures âŮŃ An assertion is only true in general if a proof is provided âŮŃ However hand written proofs are difficult to verify âŮŃ 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 âŮŃ Mechanized proof verification requires a "proof" that the verification algorithm is correct itself in applying all the formal rules correctly

1.4 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 1.1: agreement for Mary Smith in XML

```
<agreement>
 <asset> <context> <uid> Treasure Island </uid> </context> </
    asset>
 <permission>
   <display>
    <constraint>
     <cpu> <context> <uid> Mary's computer </uid> </context> <
        /cpu>
    </constraint>
   </display>
   <print>
    <constraint> <count> 2 </count> </constraint>
   </print>
  <requirement>
   <prepay>
    <payment> <amount currency="AUD"> 5.00</amount> </payment>
   </prepay>
  </requirement>
 </permission>
 <party> <context> <name> Mary Smith </name> </context> </
    party>
</agreement>
```

The agreement 1.1 is shown below using the syntax from [2].

Listing 1.2: 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 1.2 would look like in ODRL0's Coq version.

Listing 1.3: Coq version of agreement for Mary Smith

```
Agreement (Single MarySmith) Treasure Island
 (PrimitivePolicySet (Constraint (PrePay 5.00))
  (AndPolicy
   (NewList (PrimitivePolicy (Constraint
                     (Principal
                      (Single MarysComputer))) id1 Display)
   (Single (PrimitivePolicy (Constraint (Count 2)) id2 Print))))).
```

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 1.4: agreement

```
<agreement> ::= 'agreement' 'for' <prin> 'about' <asset> 'with' <policySet
    >
```

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 1.5: prin

```
<prin> ::=  { <subject_1>, ..., <subject_m> }
```

Listing 1.6: 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 1.7: 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 1.8: policySet

```
<policySet> ::=
   | <PrimitivePolicySet> : <preRequisite> → <policy>
   | <PrimitiveExclusivePolicySet> : <preRequisite> ↦ <policy>
   | <AndPolicySet> : 'and'[ <policySet_1>, ..., <policySet_m> ]
```

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 1.9: policy

```
<policy> ::=
   | <PrimitivePolicy> : <preRequisite> ⇒_<policyId> <act>
   | <AndPolicy> : 'and'[ <policy_1>, ..., <policy_m> ]
```

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 1.10: act

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

A *Policy Id* (*policyId*) is a unique identifier specified as (increasing) positive integers.

Listing 1.11: 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 1.6 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 1.12: preRequisite

```
<preRequisite> ::=
   | <TruePrq> : 'True'
   | <Constraint> : <constraint>
   | <ForEachMember> : 'ForEachMember' [<prin> ; <constraint_1>, ..., <
   constraint_m> ]
   | <NotCons> : 'not' [ <constraint> ]
   | <AndPrqs> : 'and'[ <preRequisite_1>, ..., <preRequisite_m> ]
   | <OrPrqs> : 'or'[ <preRequisite_1>, ..., <preRequisite_m> ]
   | <XorPrqs> : 'xor'[ <preRequisite_1>, ..., <preRequisite_m> ]
```

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 1.13: constraint

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

1.5 Coq Version

Listing 1.14: Coq version of agreement

```
Inductive agreement : Set :=
  | Agreement : prin → asset → policySet → agreement.
```

Listing 1.15: prin

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

Listing 1.16: asset

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

Listing 1.17: subject

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

Listing 1.18: policySet

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

Listing 1.19: policy

```
Inductive policy : Set :=
  | PrimitivePolicy : preRequisite → policyId → act → policy
  | AndPolicy : nonemptylist policy → policy.
```

Listing 1.20: act

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

Listing 1.21: policyId

```
Definition policyId := nat.
```

Listing 1.22: 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 1.23: constraint

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

1.6 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 1.24: 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.
```

The translation starts with the top level agreement element and proceeds by case analysis on the structure of the agreement. Note that each translation function takes an environment parameter.

Listing 1.25: 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 policy set proceeds with case analysis of different Policy Set constructors. We then recurse into translation functions for the composing elements. The specific Coq propositions for each constructor is taken from the formula translation for each constructor defined in [2].

Listing 1.26: Policy Translation As Formulas

$$[\![prq \rightarrow p]\!]^{prin_u,a} \triangleq \forall\ ((([\![prin_u]\!]_x \wedge [\![prq]\!]_x))$$

Listing 1.27: 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 ⇒ trans_ps e ps1 prin_u a
  | NewList ps ps_list' ⇒ ((trans_ps e ps prin_u a) /\ (trans_ps_list ps_list' prin_u a
   ))
 end) in
  match ps with
  | PrimitivePolicySet prq p ⇒ ∀ x, (((trans_prin x prin_u) /\
                        (trans_preRequisite e x prq (getId p) prin_u)) →
                        (trans_policy_positive e x p prin_u a))

  | PrimitiveExclusivePolicySet prq p ⇒ ∀ x, ((((trans_prin x prin_u) /\
                                (trans_preRequisite e x prq (getId p) prin_u)) →
                               (trans_policy_positive e x p prin_u a)) /\
                              ((not (trans_prin x prin_u)) → (trans_policy_negative e
    x p a)))

  | AndPolicySet ps_list ⇒ trans_ps_list ps_list prin_u a
  end.
```

Chapter 2

Observations

This would be a good place for some figures and tables.

Some notes on figures and photographs...

- A well-prepared PDF should be
  1. Of reasonable size, *i.e.* photos cropped and compressed.
  2. Scalable, to allow enlargment of text and drawings.
- Photos must be bit maps, and so are not scaleable by definition. TIFF and BMP are uncompressed formats, while JPEG is compressed. Most photos can be compressed without losing their illustrative value.
- Drawings that you make should be scalable vector graphics, *not* bit maps. Some scalable vector file formats are: EPS, SVG, PNG, WMF. These can all be converted into PNG or PDF, that pdflatex recognizes. Your drawing package probably can export to one of these formats directly. Otherwise, a common procedure is to print-to-file through a Postscript printer driver to create a PS file, then convert that to EPS (encapsulated PS, which has a bounding box to describe its exact size rather than a whole page). Programs such as GSView (a Ghostscript GUI) can create both EPS and PDF from PS files. Appendix B shows how to generate properly sized Matlab plots and save them as PDF.
- It's important to crop your photos and draw your figures to the size that you want to appear in your thesis. Scaling photos with the includegraphics command will cause loss of resolution. And scaling down drawings may cause any text annotations to become too small.

For more information on LaTeX see the uWaterloo Skills for the Academic Workplace course notes at saw.uwaterloo.ca/latex. [1]

[1] Note that while it is possible to include hyperlinks to external documents, it is not wise to do so,

Here is an example of how to include figures in LaTeX. Figure 2.1 shows a cantilever beam of circular cross-section subjected to a point load and a uniformly distributed load, both of which are uncertain. Note that it is better not to include the extension of the figure's source file.

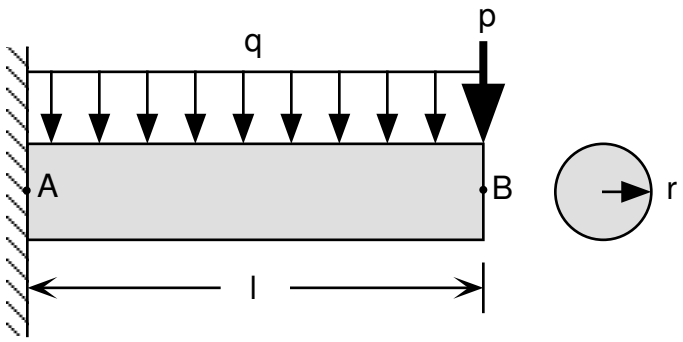

Figure 2.1: Cantilever Beam

2.1 Adding Nomenclature

The following example is part of the "nomentbl" package. Refer to the package's documentation for more details.

Let's start with equations to show how to use greek and mathematical symbols within Nomenclature.

Here is an equation

$$\dot{Q} = k \cdot A \cdot \Delta T \tag{2.1}$$

Here is another one

$$\frac{1}{k} = \left[\frac{1}{\alpha_\mathrm{i}\, r_\mathrm{i}} + \sum_{j=1}^{n} \frac{1}{\lambda_j} \ln \frac{r_{\mathrm{a},j}}{r_{\mathrm{i},j}} + \frac{1}{\alpha_\mathrm{a}\, r_\mathrm{a}} \right] \cdot r_\mathrm{reference} \tag{2.2}$$

The following example is to show how to use abbreviations within the Nomenclature. EECS is a school at the UO.

Don't forget to run:

since anything you can't control may change over time. It *would* be appropriate and necessary to provide external links to additional resources for a multimedia "enhanced" thesis. But also note that if the **hyperref** package is not included, as for the print-optimized option in this thesis template, any `\href` commands in your logical document are no longer defined. A work-around employed by this thesis template is to define a dummy `\href` command (which does nothing) in the preamble of the document, before the **hyperref** package is included. The dummy definition is then redifined by the **hyperref** package when it is included.

```
makeindex -s nomentbl.ist -o uottawa-thesis.nls uottawa-thesis.nlo
```

APPENDICES

Appendix A

Sources of Information and Help

Appendix B

Matlab Code for Making a PDF Plot

B.1 Using the GUI

Properties of Matab plots can be adjusted from the plot window via a graphical interface. Under the Desktop menu in the Figure window, select the Property Editor. You may also want to check the Plot Browser and Figure Palette for more tools. To adjust properties of the axes, look under the Edit menu and select Axes Properties.

To set the figure size and to save as PDF or other file formats, click the Export Setup button in the figure Property Editor.

B.2 From the Command Line

All figure properties can also be manipulated from the command line. Here's an example:

```
x=[0:0.1:pi];
hold on % Plot multiple traces on one figure
plot(x,sin(x))
plot(x,cos(x),'--r')
plot(x,tan(x),'.-g')
title('Some Trig Functions Over 0 to \pi') % Note LaTeX markup!
legend('{\it sin}(x)','{\it cos}(x)','{\it tan}(x)')
hold off
set(gca,'Ylim',[-3 3]) % Adjust Y limits of "current axes"
set(gcf,'Units','inches') % Set figure size units of "current figure"
set(gcf,'Position',[0,0,6,4]) % Set figure width (6 in.) and height (4 in.)
cd n:\thesis\plots % Select where to save
print -dpdf plot.pdf % Save as PDF
```

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.