Capturing & Inferring the Proof Process

Leo Freitas

School of Computing Science
Newcastle University, UK

with contributions from AI4FM members

Motivation

- Extract & reuse high-level proof strategy (cf. Gudmund's talk)
- Capture enough information to facilitate understanding of high-level reasoning (cf. Andrius' talk)
- meta-proof data gathering

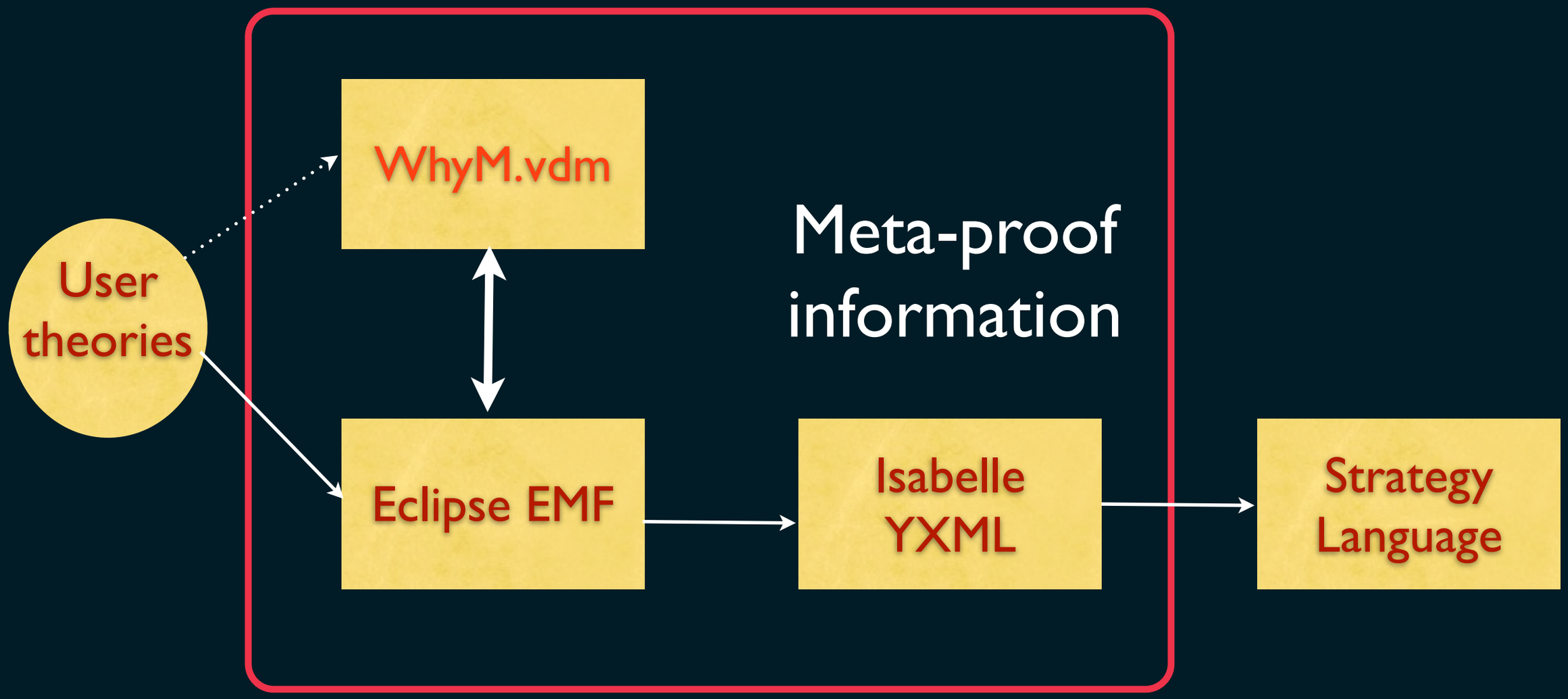
What to capture?

- categorisation of POs / hypotheses
- ranking of hypotheses per interest
- what kind of meta-data will be useful for AI techniques?
- APOLOGY: no evidence (/strategy) yet.

Models of Why

- abstract data structures to capture proof intent
- partial proof attempts as important as final proof
- described as an executable VDM model (in Overture)
- informal translations to EMF and Eclipse plugins
- reason about captured data, yet trust a prover will take care of ensuring all is okay (e.g. loosely defined)

From data to strategies



Proof engineering and abstractions

- commonly occurring categories / patterns:
  - map-op-circumscribe
  - feasibility witnessing
  - type bridges / definition morphisms
  - chase-bindings-of-interest
- proof engineering (e.g. system centric)
- proof planning (e.g. method centric)
- user's proofs (e.g. domain specific/centric)

Case studies

- R^+ lemmas used in various problems
- Union/Find characterisations
- e-science central Cloud workflows
- EMV (Europay-Master-Visa) NFC
- Tokeneer abstract models
- Part of the Xenon Security Hypervisor

R^+ lemmas

- mathematical operators properties used in various ex.
- isomorphic defs.: $R \in \mathbb{P}(X \times X) \Rightarrow R^+ = \cup \{ i : \mathbb{N}_1 \bullet R^i \}$
- lemmas useful for mechanisation
  - $i > 1 \Rightarrow R^i = R \fatsemi R^{(i-1)} = R^{(i-1)} \fatsemi R$
  - $f \in \circledcirc\, X \Rightarrow f^i \in \circledcirc\, X$
- lemmas about mathematical operators
  - $\neg\, (x, y) \in R^+ \land (x , z) \in R^+ \Rightarrow (x, y) \in (R \cup \{(z , y)\})^+$
  - $\neg\, s \in \mathrm{ran}\ R \Rightarrow (\{s\} \lhd R)^+ = \{s\} \lhd R^+$

Cliff's version of Union/find

- from partitions of sets of sets to forests
- data structure representation variations
- morphisms to aid proof / clarity
- exercise in both ZEves and Isabelle

e-science Central workflows

- cloud workflows for science as a service
- workflow deployment on federated clouds
- workflow topology representation
- representation properties of interest
  - security + cost + dependability + etc

EMV near-field communication

- model of terminal + card + transaction
- security properties / checks of protocol
- rep. of device driver algorithm
- found couple of protocol vulnerabilities

Tokeneer abstract models

- Audited secure enclave of workstations
- NSA contract for EAL5+ demonstration
- modelled in Z, coded in SPARKAda
- Tokeneer dist. has no proof, though (!)
- mechanisation of abstract Tokeneer
  - found errors/problems with audit logging

Xenon Security Hypervisor

- NRL reengineering of Xen hypervisor
- concurrency and info. flow security
- concurrency in CSP + data types in Z
- enforcement of security policies

Conclusion

- collecting meta-proof data from variety of proof exercises
- serves as input to Gudmund's strategy language
- hopefully we will find common strategies there --- they are present!