From Model Checking to Model Learning

Two Basic Techniques in Model-Based Development of Embedded Systems

Frits Vaandrager

fvaan@cs.ru.nl

Model-Based System Development

Vision: Models as primary artifacts throughout engineering lifecycle of computer-based systems

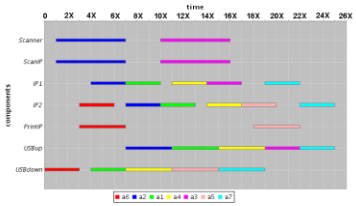
Models used for:

- Communication between stakeholders
- Simulation, verification and validation
- Design space exploration
- Code generation
- Testing
- Reuse

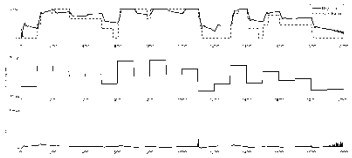
Model-Based Development at Océ

Happy flow model of paper path

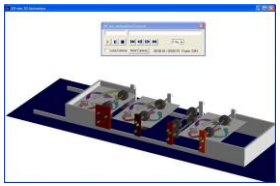
Use of Uppaal for data path scheduling

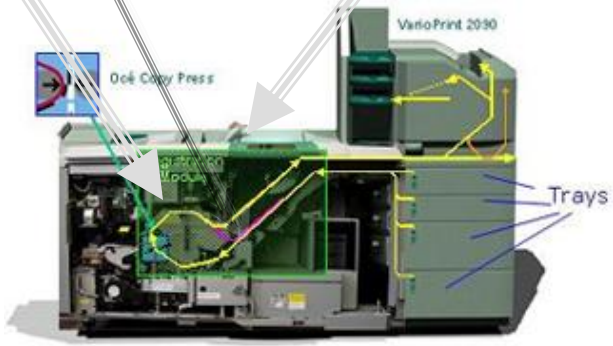


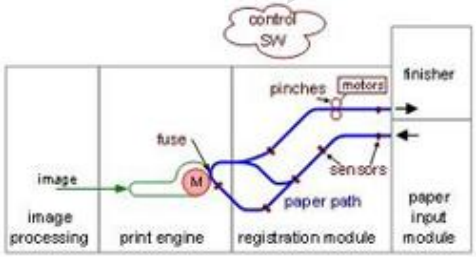
Adaptive control of printer behaviour using Bayesian networks



Co-simulation of paper path with VDM++ and bond graphs







"The modelling approach from Boderc has enabled Océ to skip a complete physical machine-building iteration cycle, saving many man-years of effort"

Two Fundamental Research Questions

- How can we prove interesting behavioral properties of complex models?

  model checking

  Embedded Systems INSTITUTE

- How can we obtain valid models of existing (legacy, black box, ..) components?

  model learning

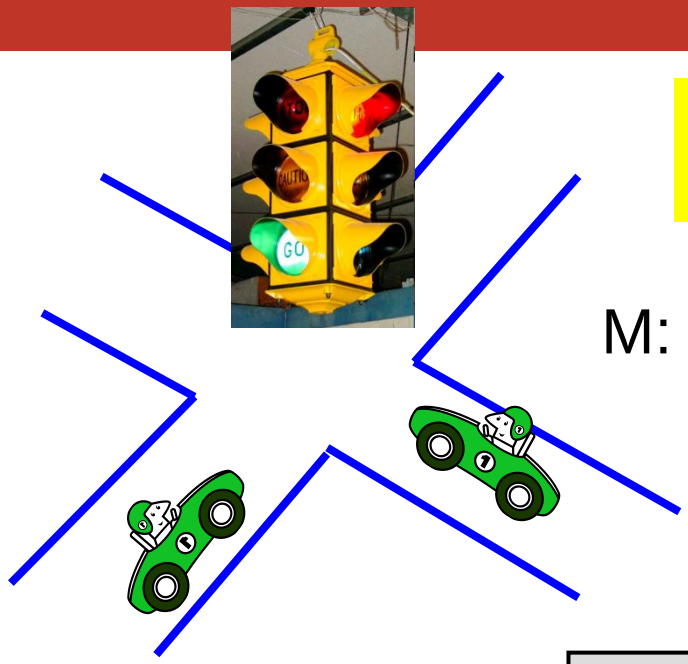
  stw

Overview Talk

- Model Checking
- Model Learning
- Research Funding

Part 1: Model Checking

Model Checking

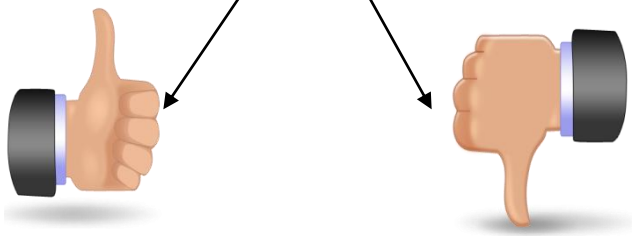


M: Traffic Light

Controller P: No Collisions

Model Checker

Does M satisfy P?



Yes!

No, and here's an example of why not.

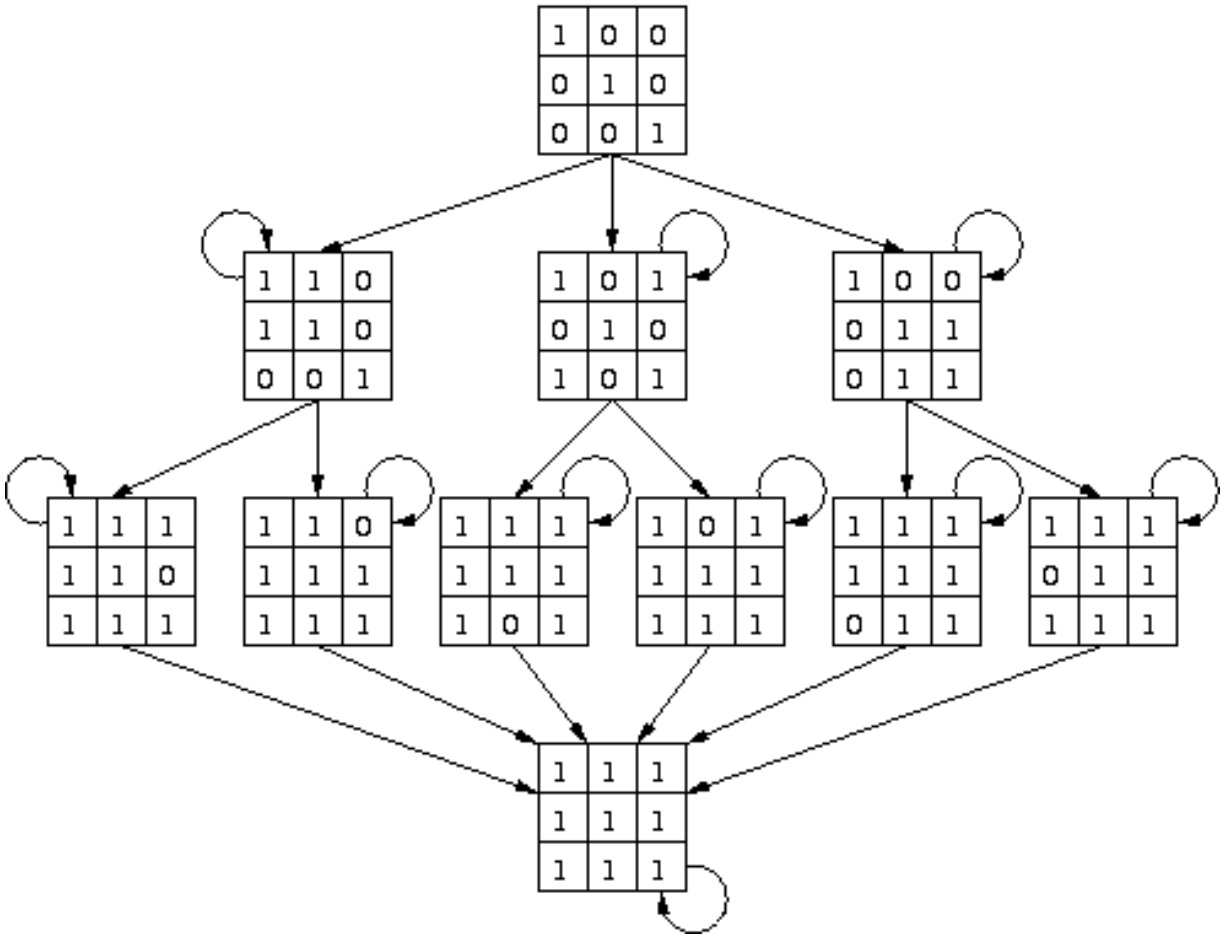
Example: Gossip Girl Problem

Five girls all have a gossip of their own.

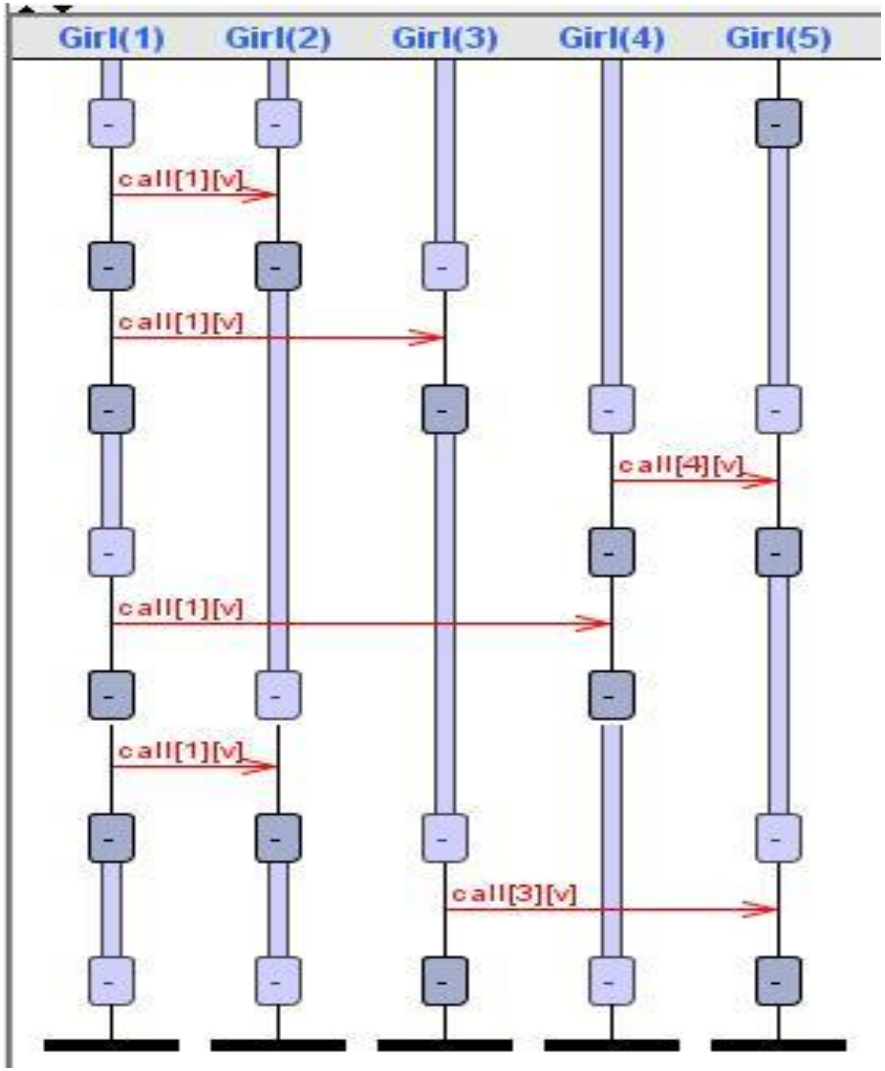
They call each other over the phone. Whenever two girls talk they exchange all gossips they know.

How many calls are needed before every girl knows every gossip?

State Machine

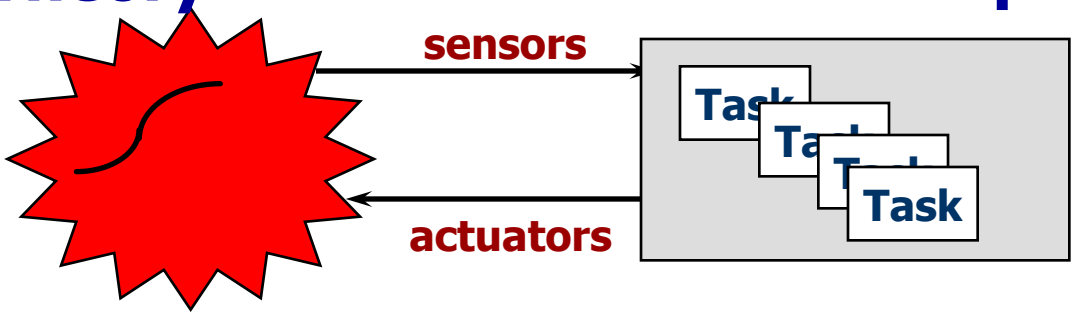


Solution Model Checker



Real Time Systems

Control Theory **Computer Science**



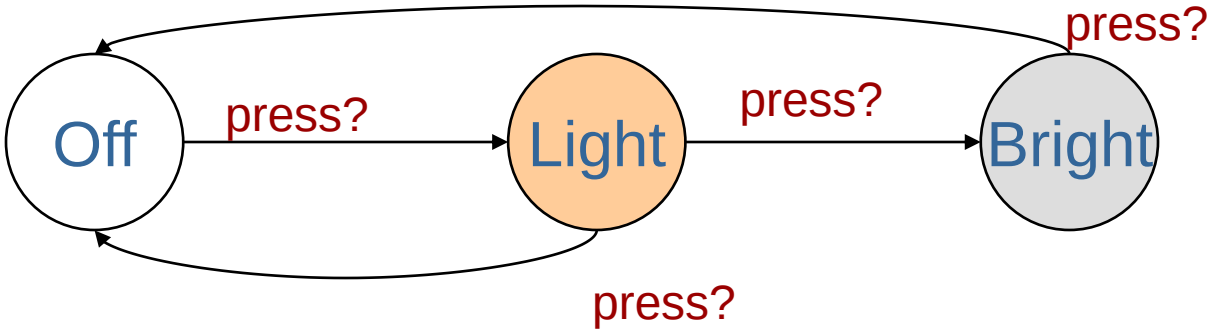
Plant
Continuous

Controller Program
Discrete

Eg.: Pump Control
Air Bags
Robots
Cruise Control
ABS
CD Players
Production Lines

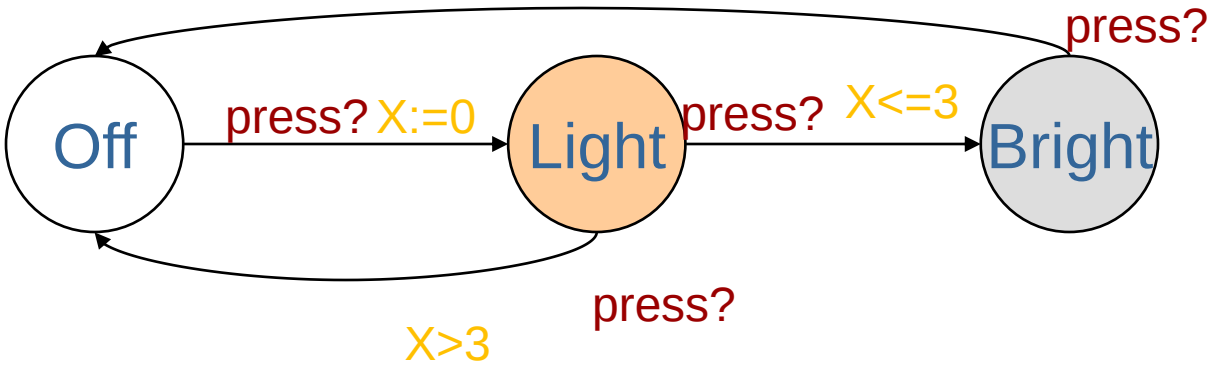
Real Time System
A system where correctness not only depends on the logical order of events but also on their timing

Timed Automata



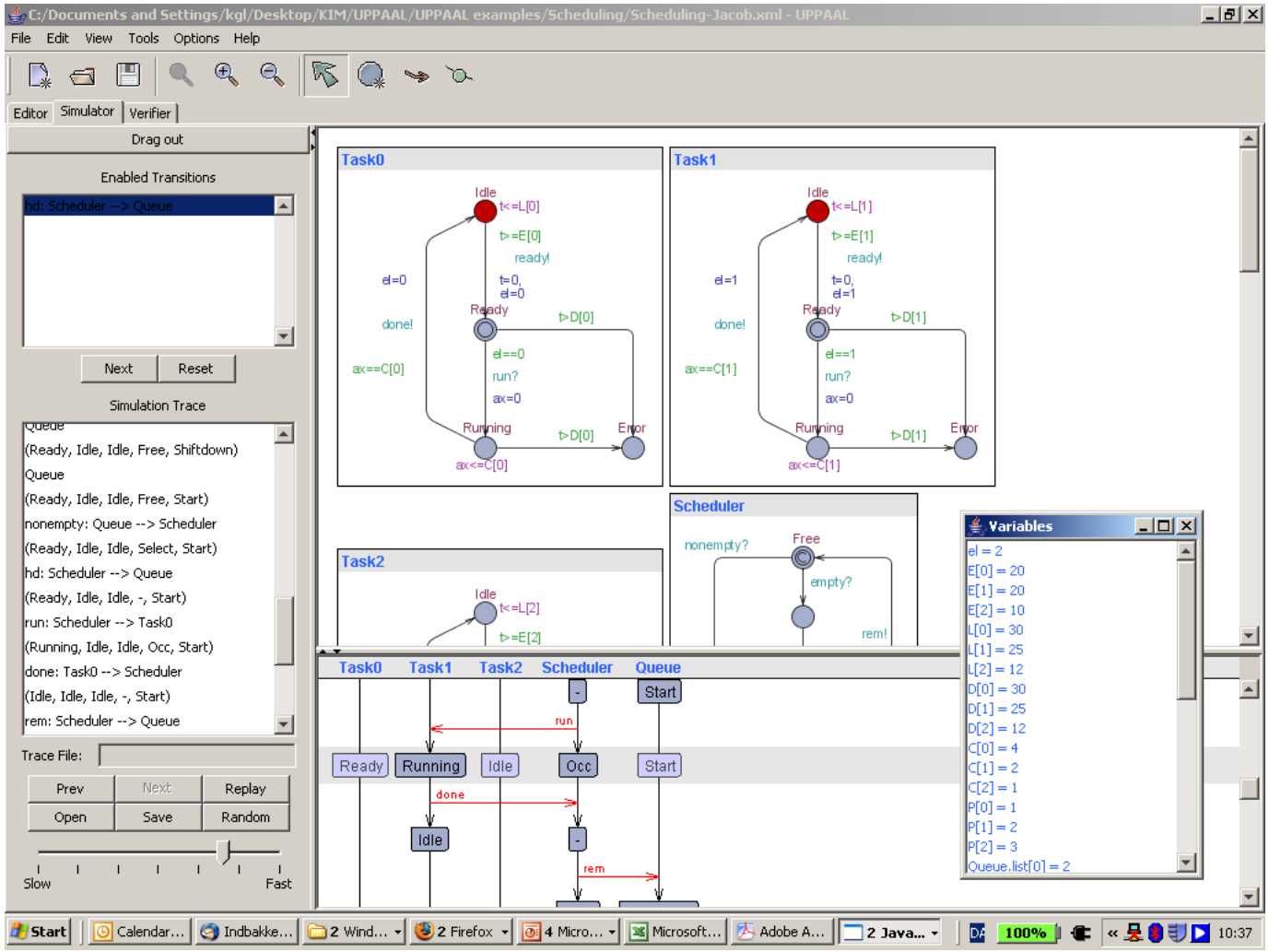
WANT: if press is issued twice quickly then the light will get brighter; otherwise the light is turned off.

Timed Automata



Solution: Add real-valued clock xx

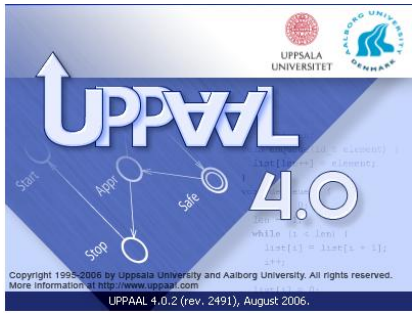
UPPAAL 4.0

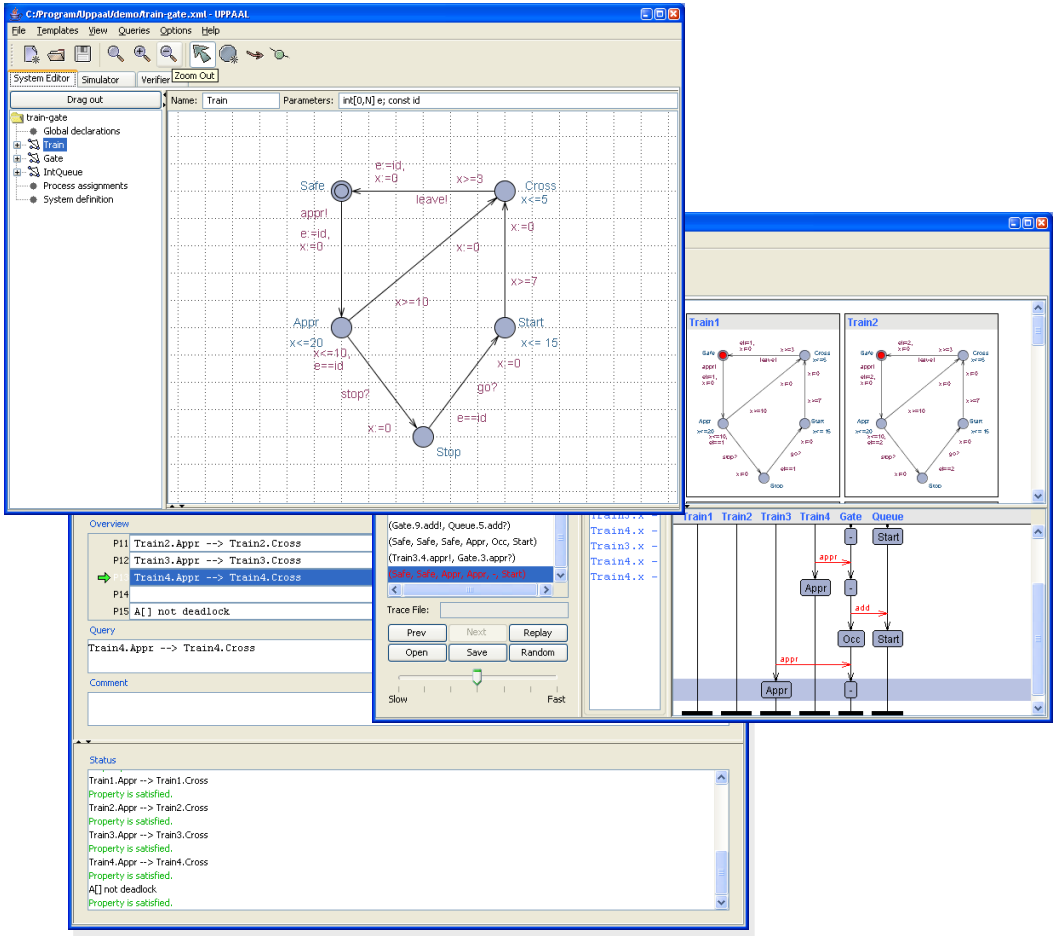

C:/Documents and Settings/kgl/Desktop/KIM/UPPAAL/UPPAAL examples/Scheduling/Scheduling-Jacob.xml - UPPAAL
File Edit View Tools Options Help
Editor Simulator Verifier
Drag out
Enabled Transitions
hd: Scheduler --> Queue
Next Reset
Simulation Trace
(Ready, Idle, Idle, Free, Shiftdown)
Queue
(Ready, Idle, Idle, Free, Start)
nonempty: Queue --> Scheduler
(Ready, Idle, Idle, Select, Start)
hd: Scheduler --> Queue
(Ready, Idle, Idle, -, Start)
run: Scheduler --> Task0
(Running, Idle, Idle, Occ, Start)
done: Task0 --> Scheduler
(Idle, Idle, Idle, -, Start)
rem: Scheduler --> Queue
Trace File:
Prev Next Replay
Open Save Random
Slow Fast
Task0
Idle
t<=L[0]
t>=E[0]
ready!
el=0
t=0, el=0
Ready
done!
t>D[0]
el==0
ax==C[0]
run?
ax=0
Running
Error
ax<=C[0]
Task1
Idle
t<=L[1]
t>=E[1]
ready!
el=1
t=0, el=1
Ready
done!
t>D[1]
el==1
ax==C[1]
run?
ax=0
Running
Error
ax<=C[1]
Task2
Idle
t<=L[2]
t>=E[2]
Scheduler
nonempty?
Free
empty?
rem!
Variables
el = 2
E[0] = 20
E[1] = 20
E[2] = 10
L[0] = 30
L[1] = 25
L[2] = 12
D[0] = 30
D[1] = 25
D[2] = 12
C[0] = 4
C[1] = 2
C[2] = 1
P[0] = 1
P[1] = 2
P[2] = 3
Queue.list[0] = 2
Task0 Task1 Task2 Scheduler Queue
-
Start
run
Ready Running Idle Occ Start
done
Idle
-
rem
Start
Calendar... Indbakke... 2 Wind... 2 Firefox 4 Micro... Microsoft... Adobe A... 2 Java...
100%
10:37

Impact

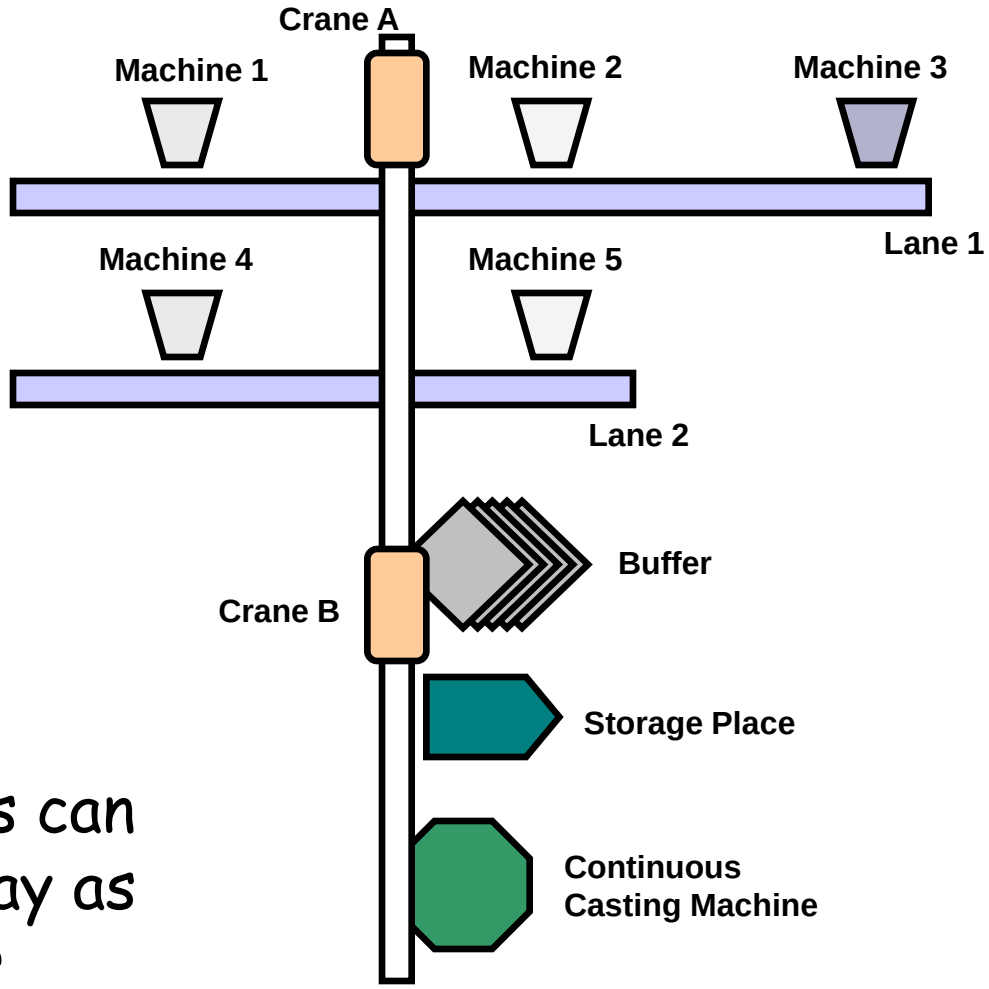
Company Downloads

Mecel
Jet
Symantec
SRI
Relogic
Realwork
NASA
Verified Systems
Microsoft
ABB
Airbus
PSA
Saab
Siemens
Volvo
Lucent Technologies
Etc etc





SIDMAR Steel Production Plant



Observation

(EU VHS project)

Many scheduling problems can be phrased in a natural way as reachability problems for timed automata.

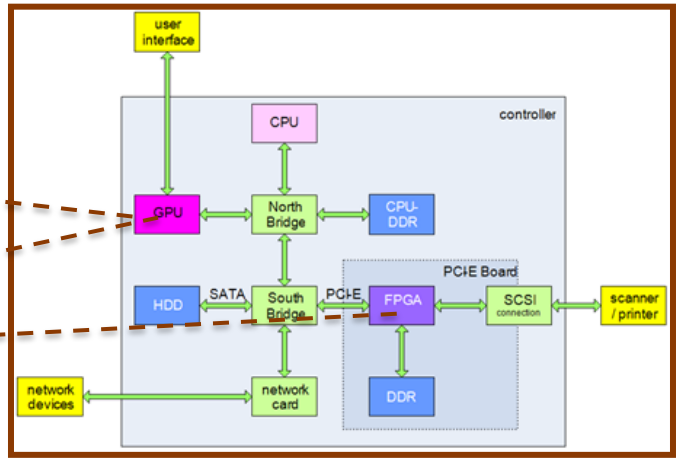
Challenge Océ: Design of printer data path

Use cases:

Printing

Scanning

Hardware platform:
board template



Mapping constraints

Typical requirements:

1. Throughput must be X pages per minute
2. Scanning should not disrupt printing

Dimensioning constraints:

1. FPGA buffer memory size
2. CPU/GPU count & speed
3. Bus bandwidth

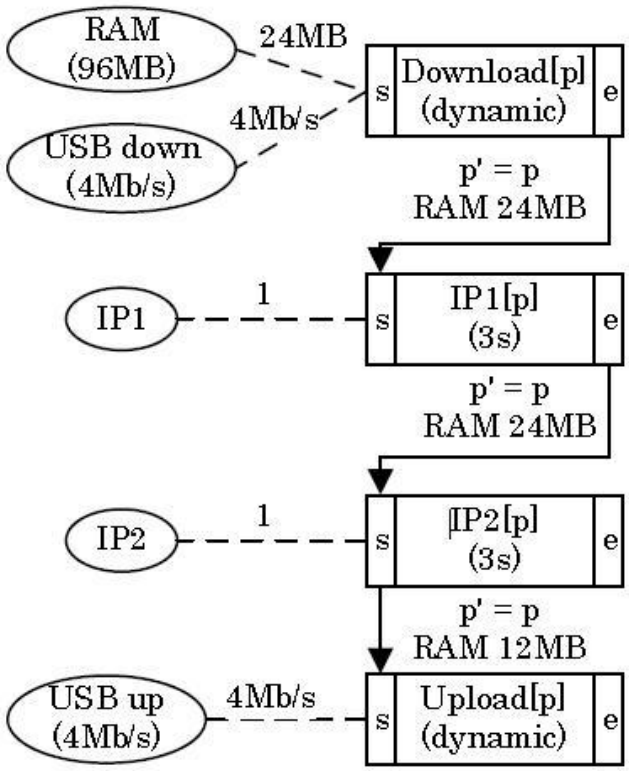
good mapping and scheduling

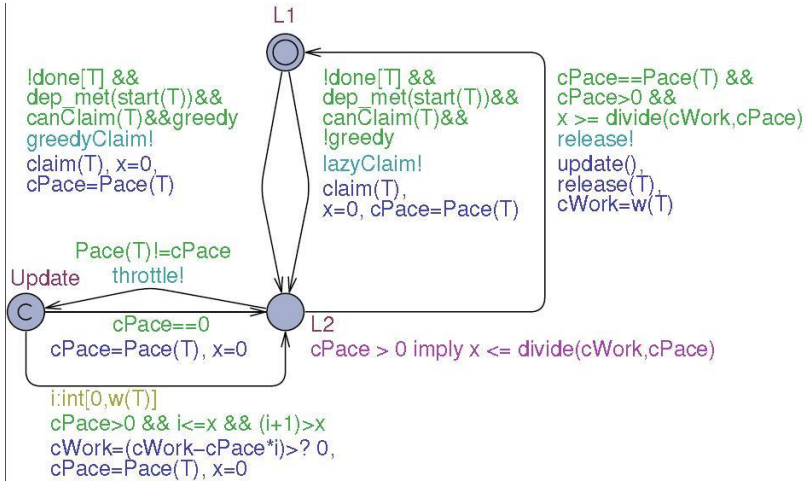
satisfy requirements

right sizes

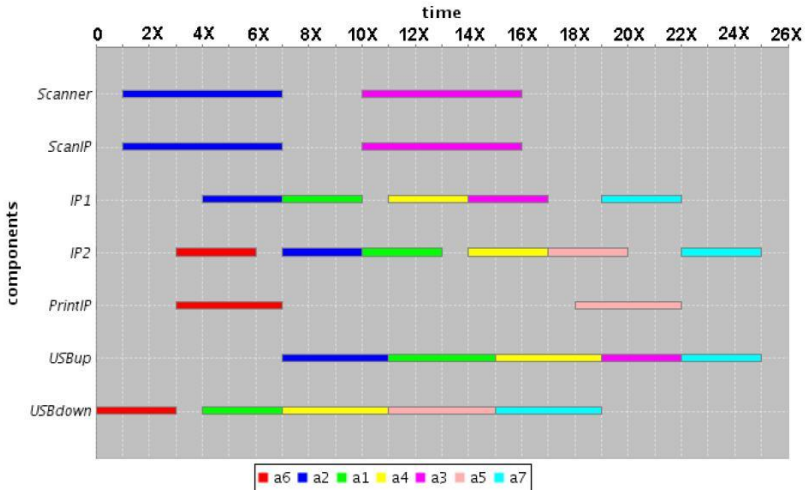
DESIGN OBJECTIVE

From Printer Models to Uppaal





From Uppaal to Gantt Charts

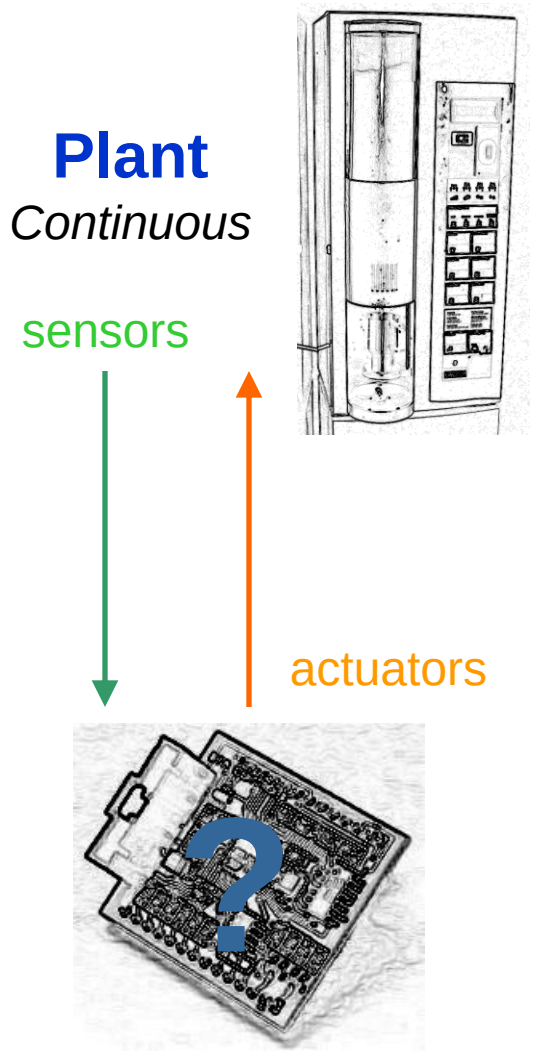


Challenges

- Dynamic resources
- Timing uncertainty
- Repetitive behavior

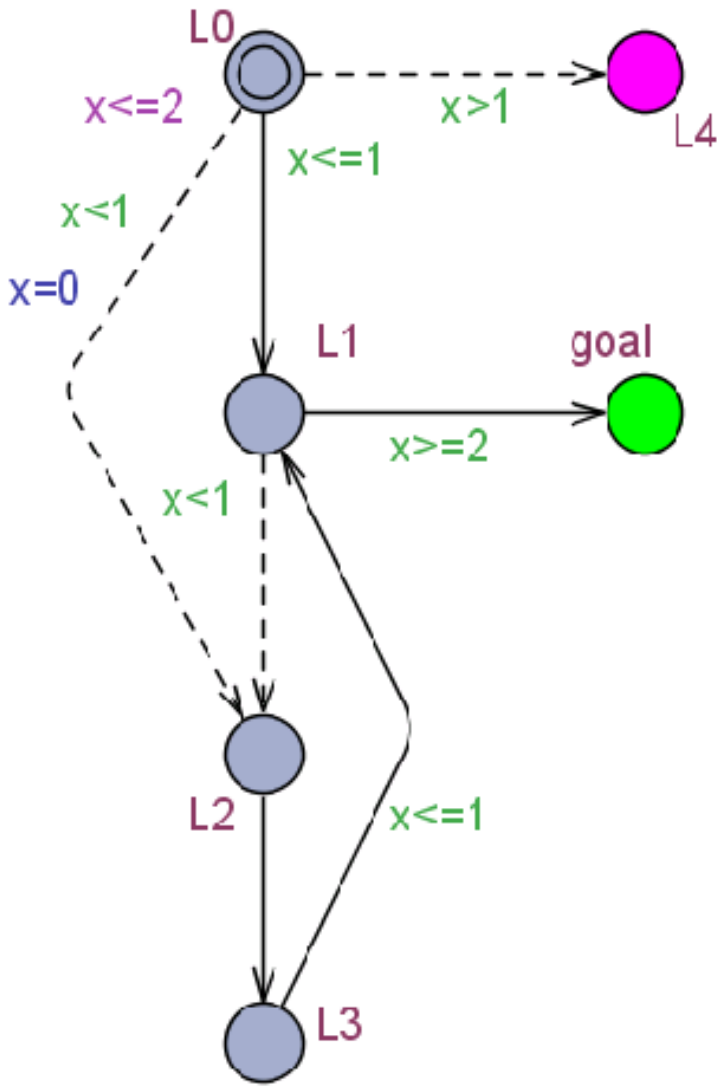
Why Timed Games?

- Controller synthesis:
  - Model the environment + what a controller *can* do.
  - Generate controller that meets control objective
    Generate the *right* code automatically.
  - 2-player timed game: environment moves vs controller moves
    ⇒ Timed Game Automata

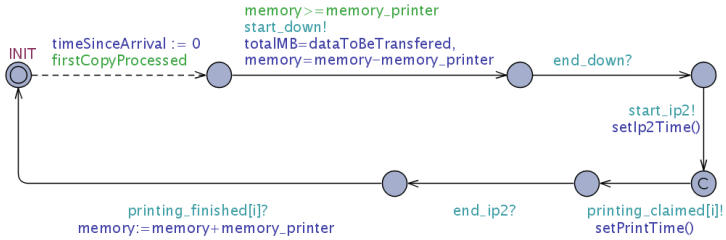


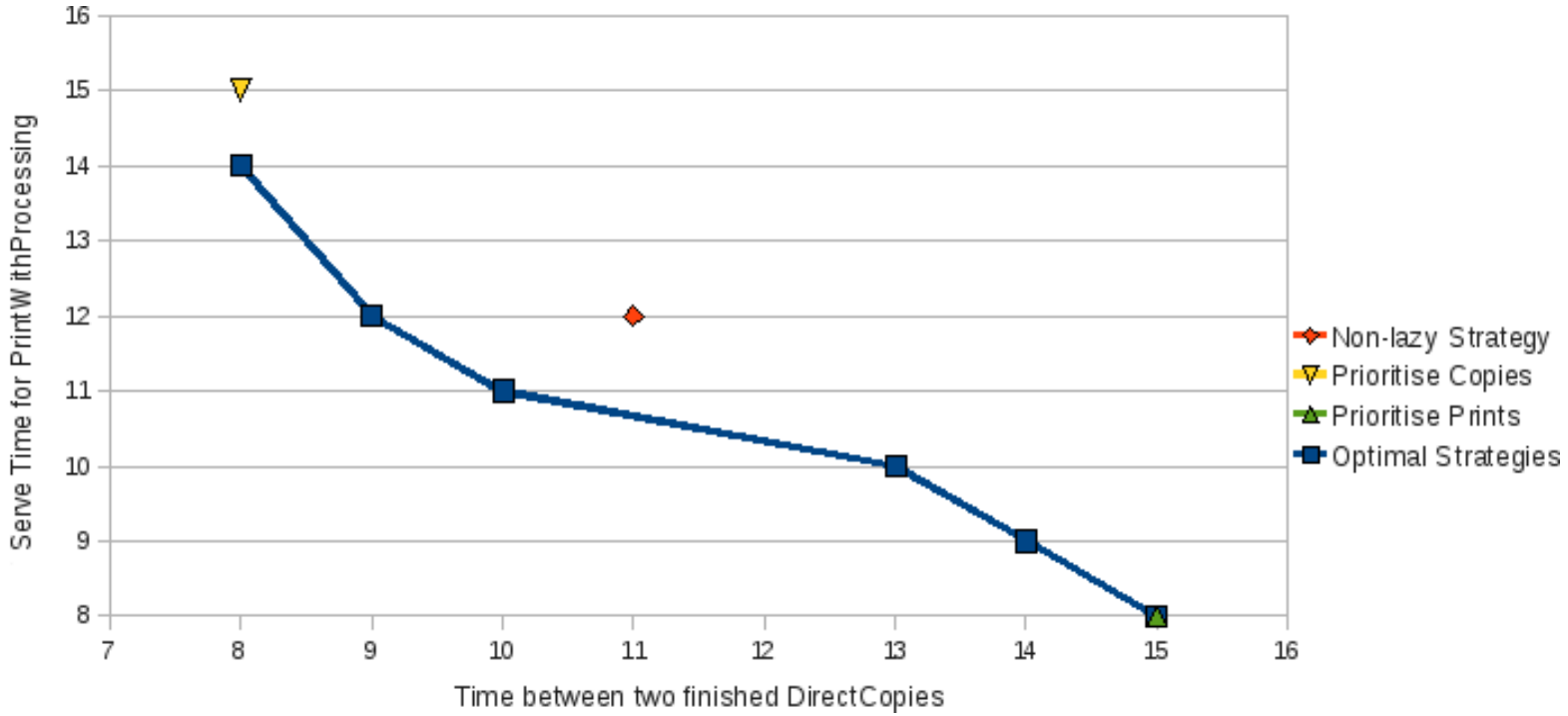
Timed Game Automata

- Timed automata with controllable and uncontrollable transitions
- Reachability & safety games
  - control: A<> TGA.goal
  - control: A[] not TGA.L4
- Memoryless strategy:
  - state → action



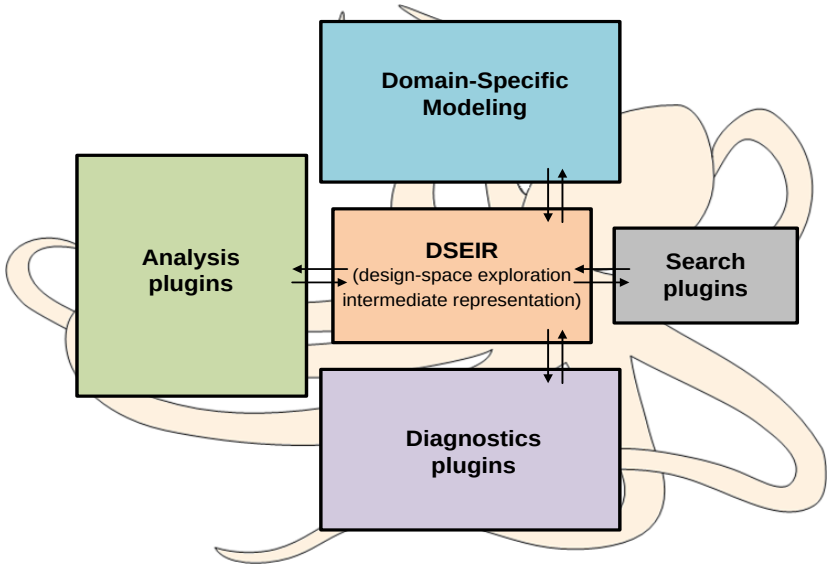
Using Timed Games for Printers

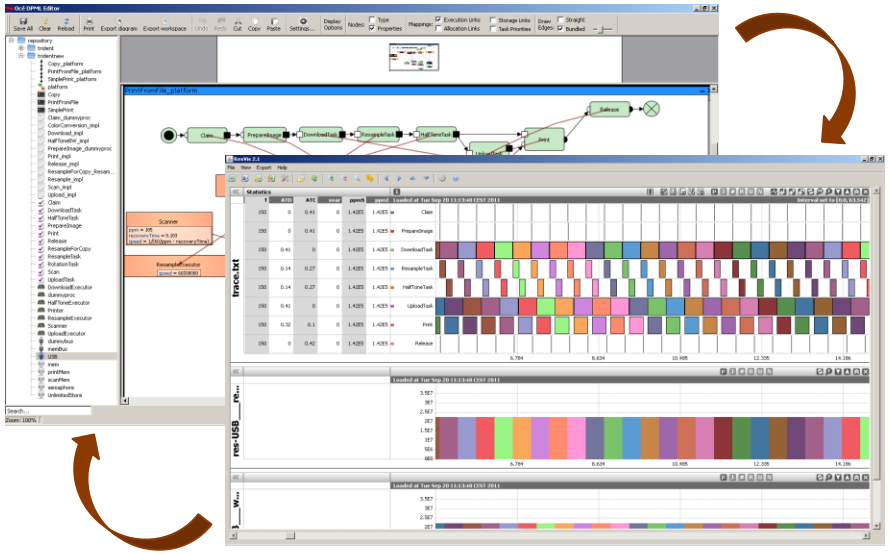




Outcomes ESI project Octopus

- Toolset for Model-Driven Design-Space Exploration
- Addresses Océ challenges





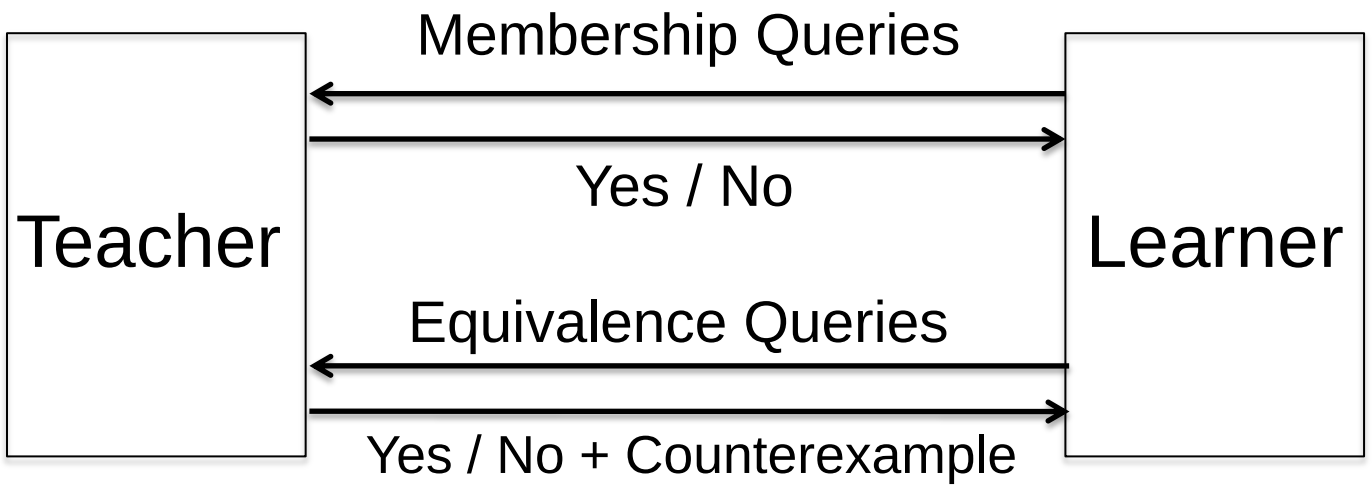
Part 2: Model Learning

Children Learn How to Use Computers

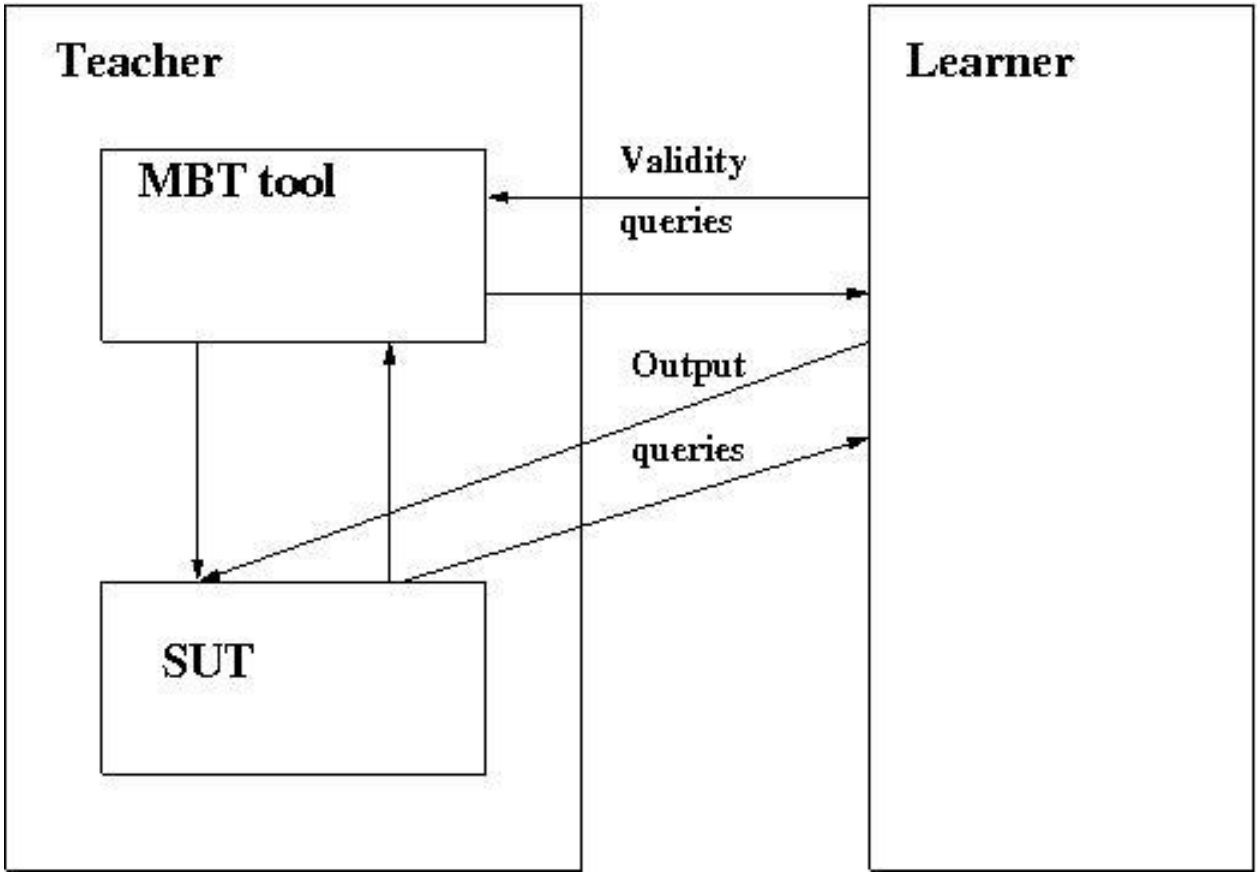
Can computers learn state diagrams as well?

Angluin's L* Algorithm (1987)

Learning Finite Automata:



Learning Models of Reactive Systems

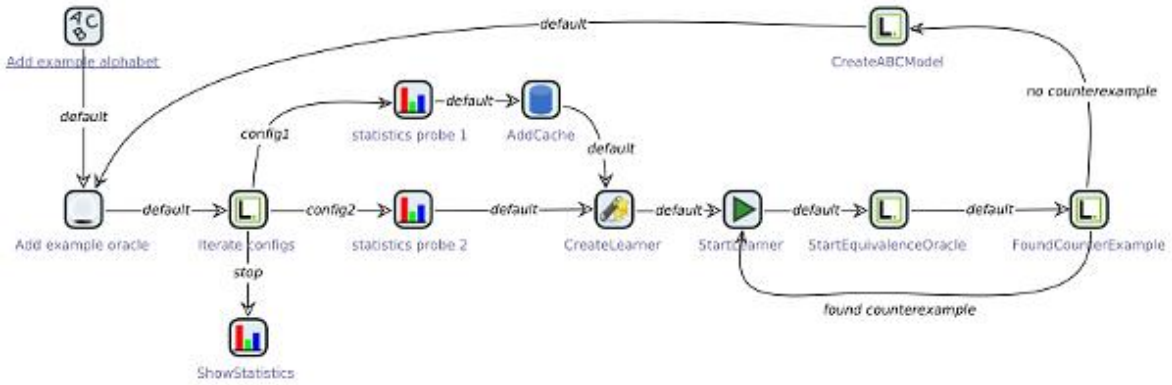


Learner: Formulate hypothesis

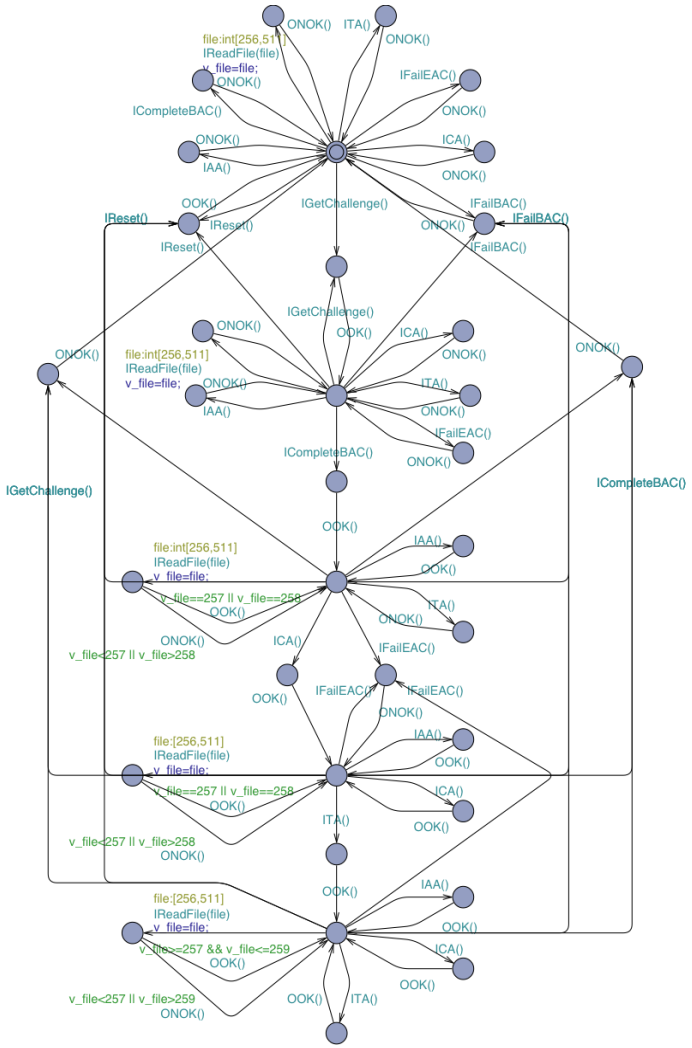
Model-Based Testing: Test hypothesis

LearnLib Tool

- Tool for active learning of Mealy machine models
- Developed by group Bernhard Steffen (University Dortmund)
- Able to learn models with up to 10.000 states



Case Study: Biometric Passport



Case Study: Banking Cards

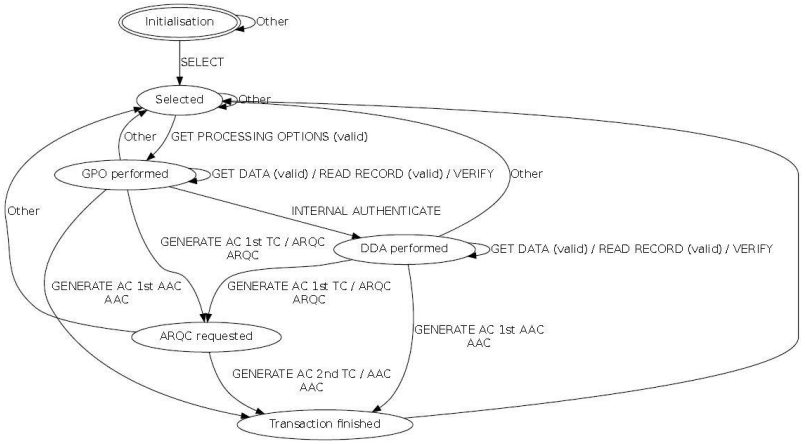
• Reverse engineering of EMV protocol applications for collection of banking and credit cards

•Simple models with up to 7 states
(NB EMV standard has over 700 pages)

• At most 1500 membership queries, less than 30 minutes

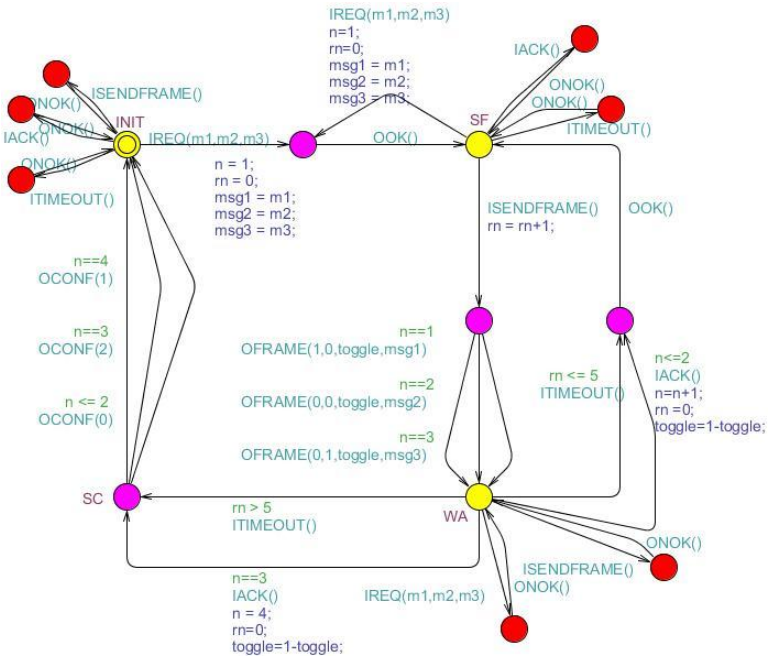
• Models provide unique fingerprints of different cards

•Useful as part of security evaluation

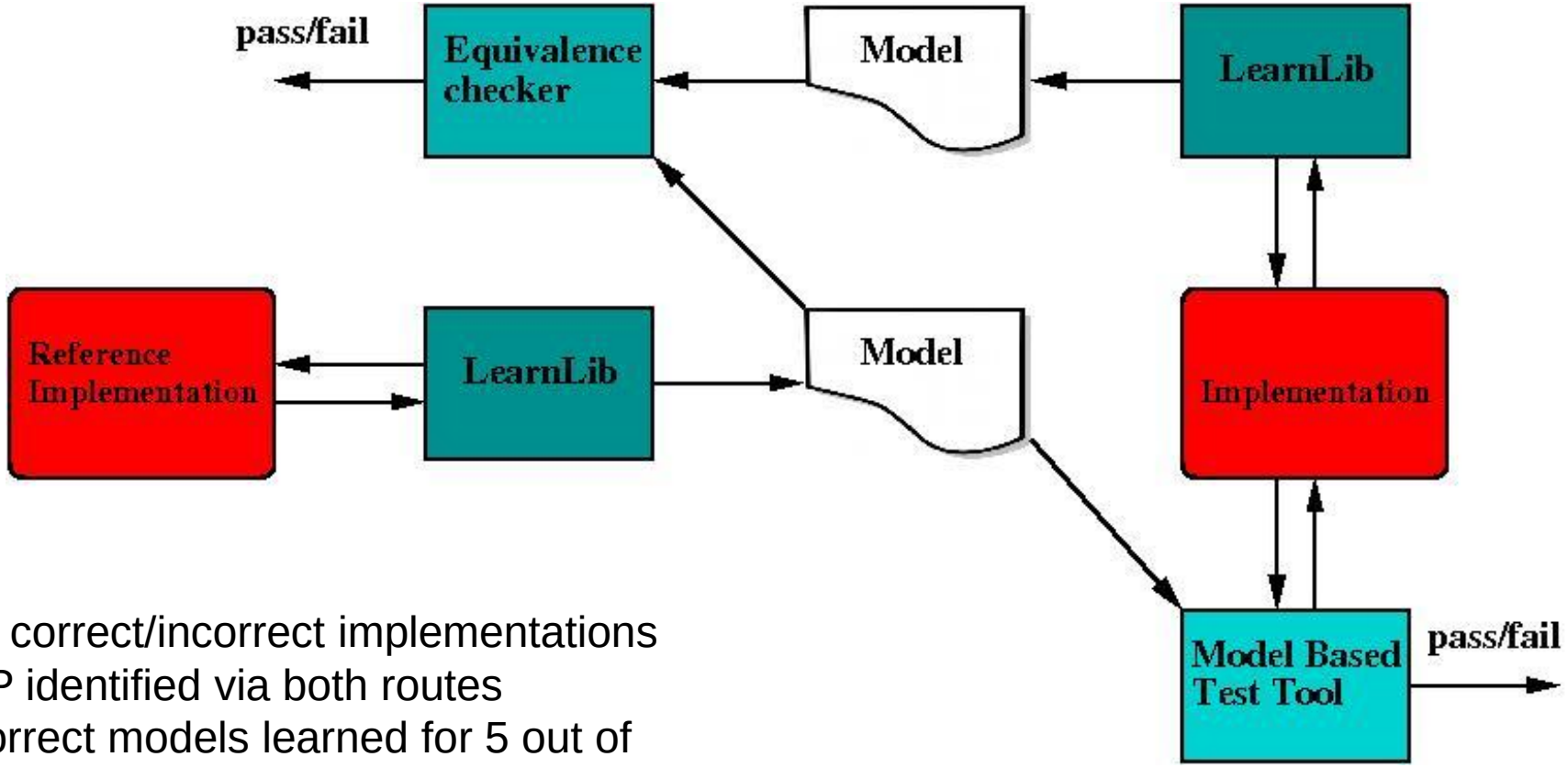


Case Study: Philips Bounded Retransmission Protocol

Use learning to test whether BRP protocol implementations are correct relative to reference implementation



Experimental Setting



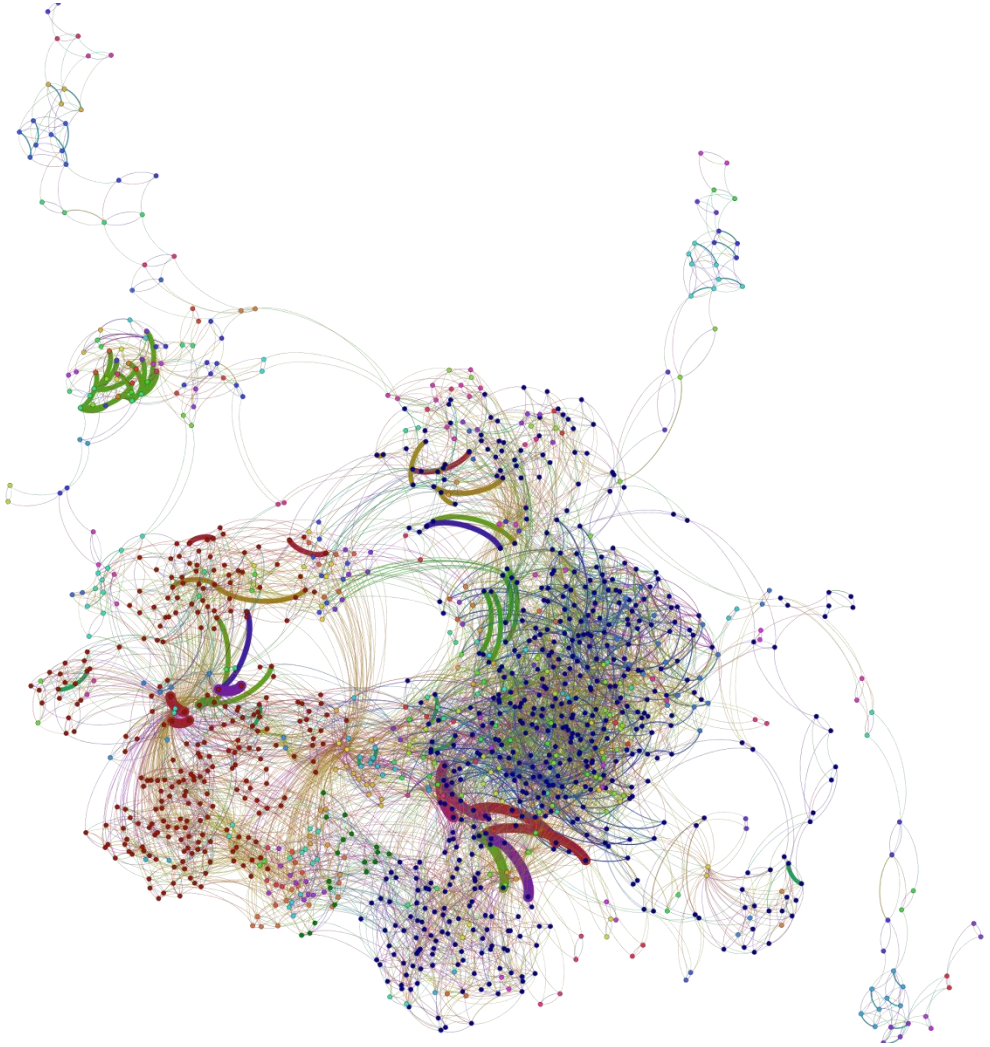
- All correct/incorrect implementations BRP identified via both routes
- Correct models learned for 5 out of 6 mutant implementations of BRP

Case Study: Engine Status Manager of Océ Printer

Goal: learn models of realistic printer controllers

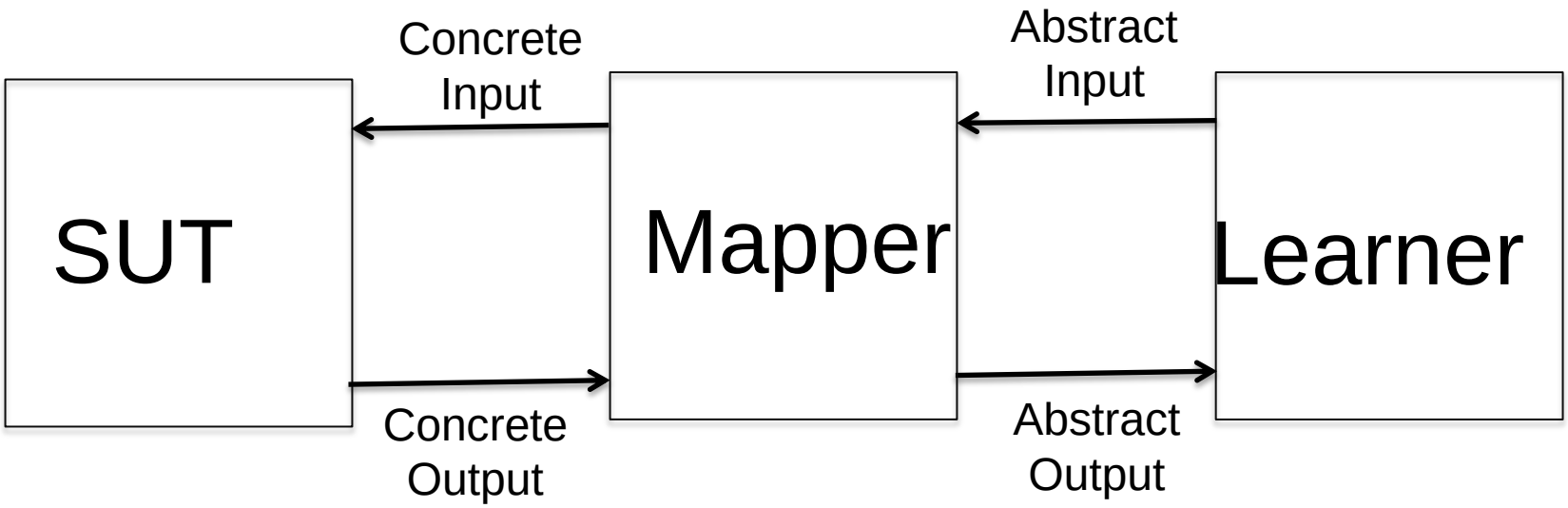
Possible use: regression testing, generation of new implementations,..

Visualisation of Learned Model



1314 states, 64.136.029 member queries

Use of Mappers to Handle Data



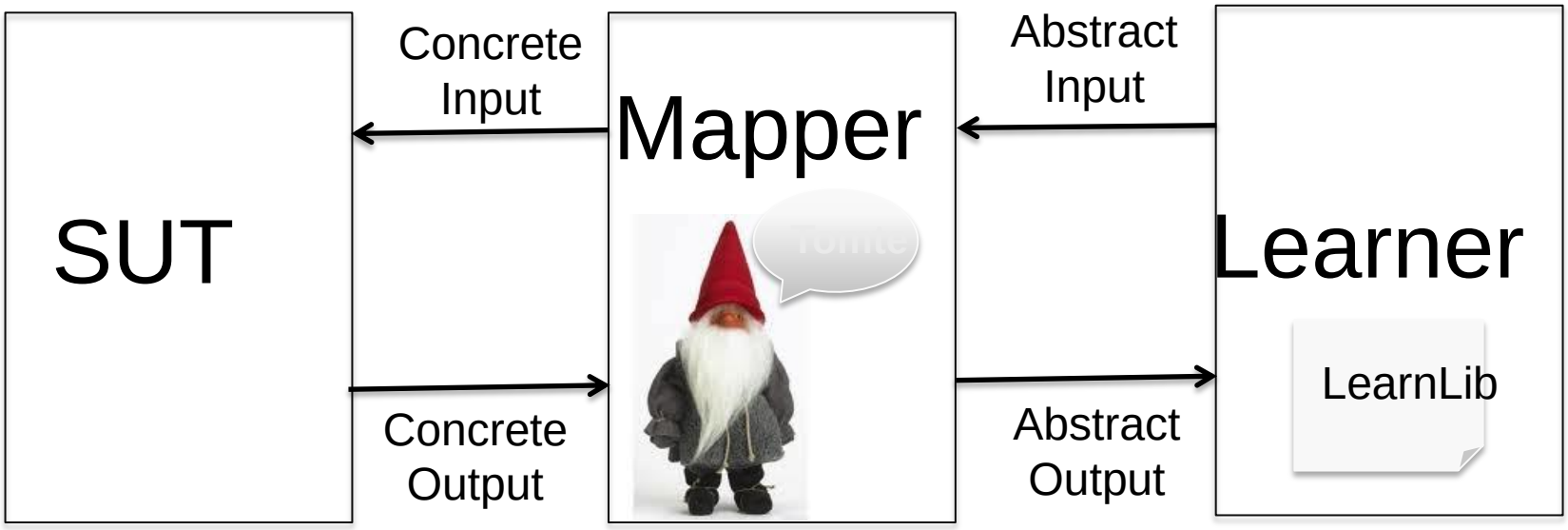
Theory

- Mapper induces two transformations:
  - An abstraction operator α_A that transforms concrete models into abstract models
  - A concretization operator γ_A that transforms abstract models into concrete models

- **Theorem 1.** *Suppose* $\alpha_{\mathcal{A}}(\mathcal{M}) \leq \mathcal{H}$. *Then* $\mathcal{M} \leq \gamma_{\mathcal{A}}(\mathcal{H})$.

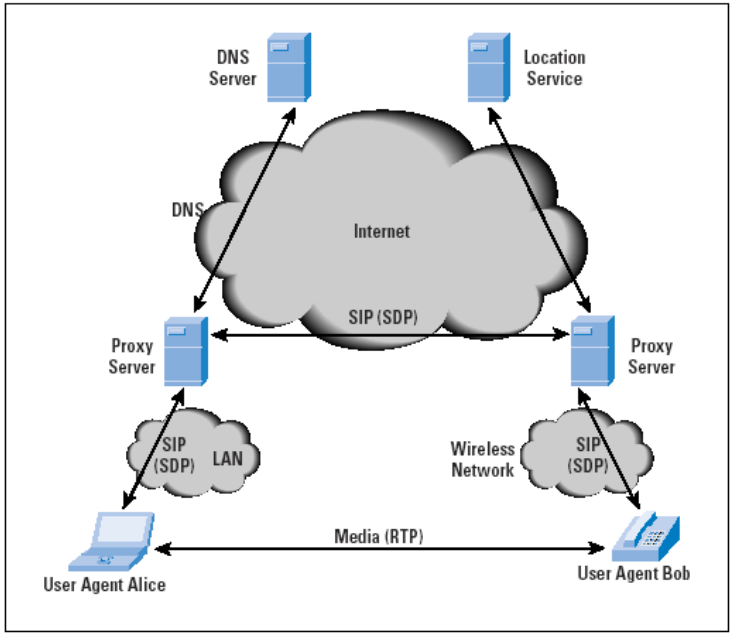
Tomte Tool

Compute abstractions automatically



Case Study: Session Initiation Protocol (SIP)

Figure 1: SIP Components and Protocols



Learned model is extended finite state machine with 29 states, 3741 transitions, and 17 state variables with various types

Research Challenges

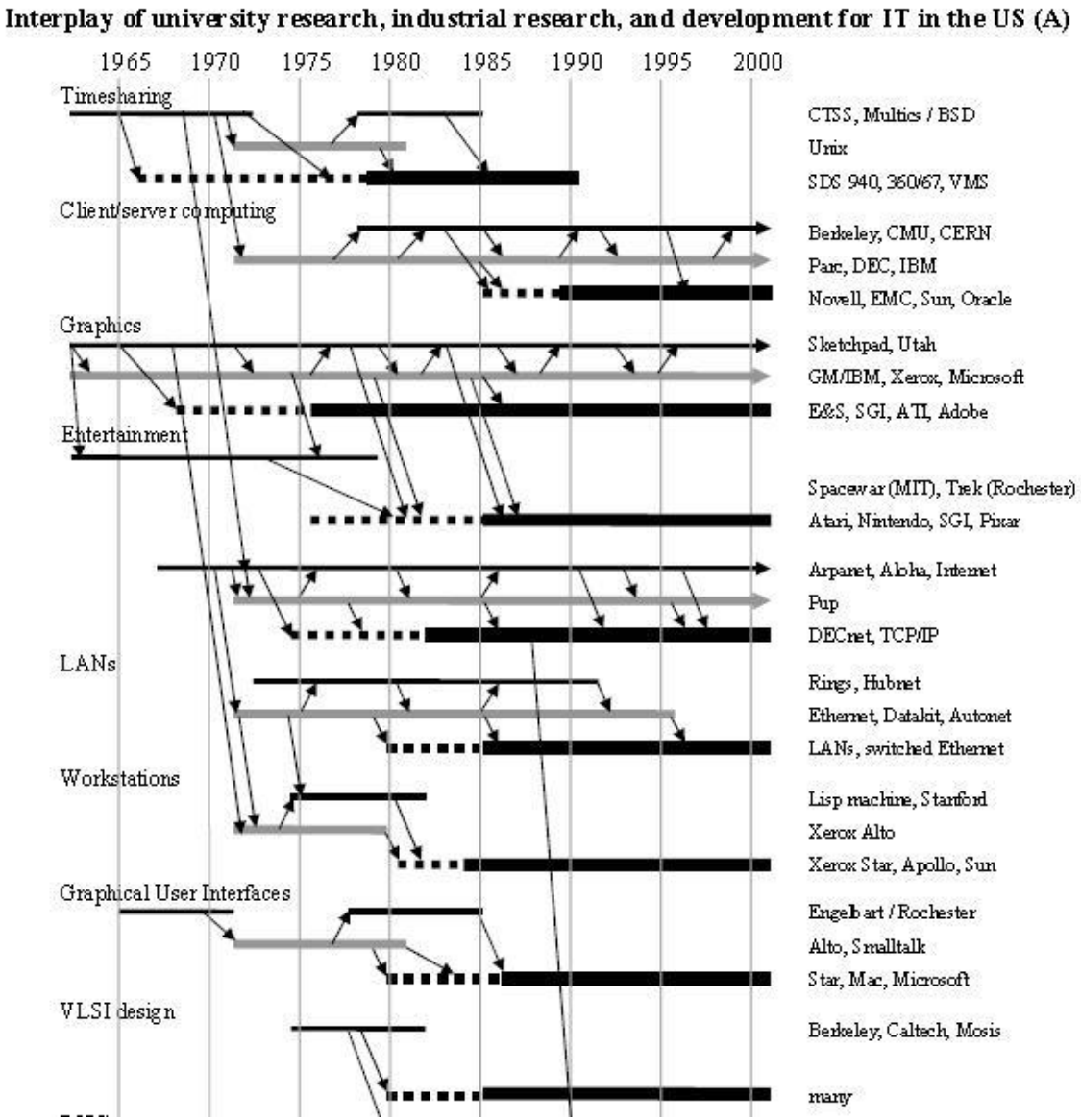
- Learn models with structure (data,..)
- Cope with state space explosion
- Visualisation of learned models
- Cope with nondeterminism SUTs
- …

Conclusions/Perspective

- Promissing technique, many potential applications
- Useful for control intensive systems, such as network protocols and control software
- More research needed before routine use in commercial setting is feasible
- Wanted: challenging case studies!

Part 3: Research Funding

Lampson's Tire Tracks



Final Report for Period: 09/1994 - 08/1999
Principal Investigator: Garcia-Molina, Hector .
Organization: Stanford University
Title:
The Stanford Integrated Digital Library Project

Submitted on: 05/15/2000
Award ID: 9411306

pants

Senior Personnel

Name: Garcia-Molina, Hector
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Paepcke, Andreas
Worked for more than 160 Hours: Yes
Contribution to Project:
Project Director

Post-doc

Graduate Student

Name: Page, Larry
Worked for more than 160 Hours: Yes
Contribution to Project:

Larry Page

Name: Chang, Ed
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Chang, Kevin
Worked for more than 160 Hours: Yes
Contribution to Project:

Journal Publications

Please see http://www-diglib.stanford.edu for a list of publications., "Please see http://www-diglib.stanford.edu for a list of publications.",
Please see http://www-diglib.stanford.edu for a list of

Please see
(1998). Bo
Bibliograp

The Google search engine was developed as part of the project. It is now a company (www.google.com)

URL(s):

http://www-diglib.stanford.edu

Description:

Other Specific Products

Product Type:

Data or databases

Product Description:

The Google search engine was developed as part of the project.
It is now a company (www.google.com)

Sharing Information:

The engine is publicly available

How About Us (NL)?

- Read Martin Rem's book "Tegen de Stroom" : ASML, TomTom, WiFi, Bluetooth, Python, AMS-IX,..
- 21 Spin Offs CWI: Data Distilleries, MonetDB, Spinque, SIG, VectorWise,..
- ..

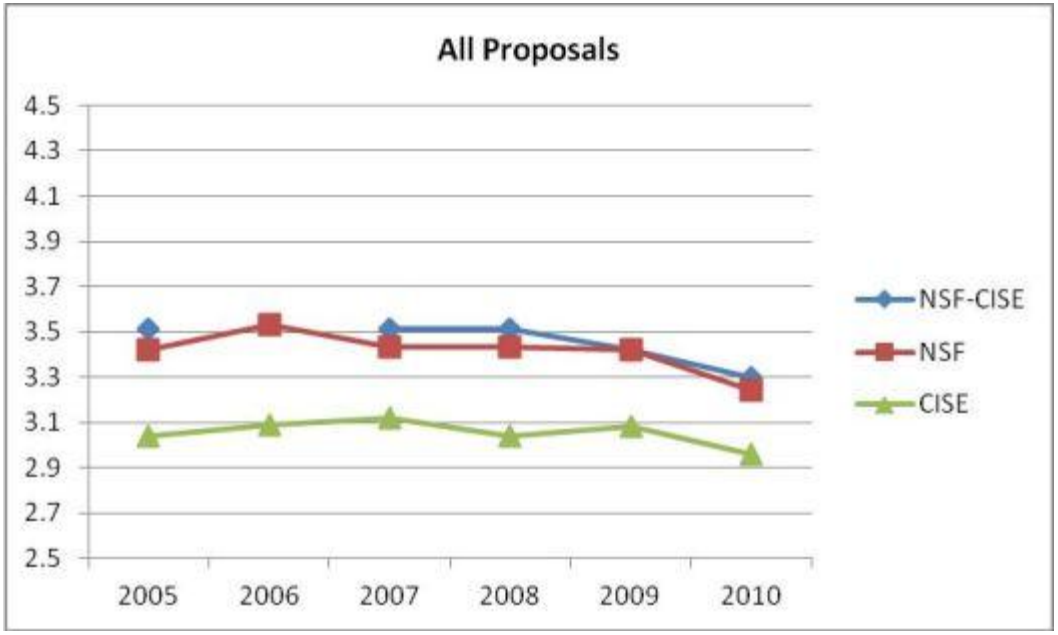
But we can do better!

Research Funding within NWO

- Other disciplines tend to view CS as less fundamental
- Applied/multidisciplinary research less favored
- CS researchers hypercritical

Need separate budget for CS research

Grades for CS Proposals in US



Representation different areas in Dutch CS unbalanced

- Intelligent systems and Formal methods overrepresented
- Minimal presence Computer systems, Database systems, Algorithms, Graphics
- Need more experimental systems research

ICT Research Funding

- Dutch government / NWO does not recognize importance ICT research
- Dutch government should invest more in education/research anyway
- Help/stimulate spin off companies

- Societal relevance should play larger role in allocation research money

  (Idea to let NWO support “Top sectors” good -up to a point; idea to stop FES bad)

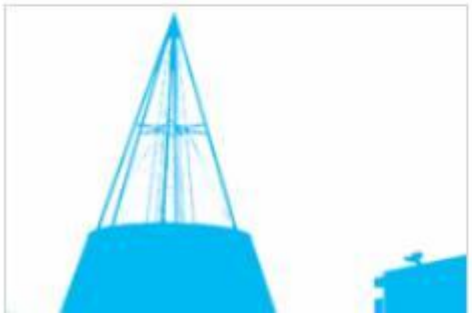
US industry support voor Dutch CS research/education

2010 Google Europe Doctoral Fellowship Recipients

- Roland Angst, Google Europe Fellowship in Computer Vision **(Swiss Federal Institute of Technology Zurich, Switzerland)**
- Arnar Birgisson, Google Europe Fellowship in Computer Security **(Chalmers University of Technology, Sweden)**
- Omar Choudary, Google Europe Fellowship in Mobile Security **(University of Cambridge, U.K.)**
- Michele Coscia, G
- Moran Feldman, G
- Neil Houlsby, Goog
- Kasper Dalgaard L
- Florian Laws, Goo
- Cynthia Liem, Goo

The project Dome is the unique research collaboration between IBM Netherlands and ASTRON. In this project fundamental research is performed in the field of Radio Astronomy and will concentrate on three domains:

1. Green supercomputing (extreme performance at minor energy costs)
2. Extreme Streaming (real time processing on a permanent basis of massive data amounts)
3. Nano Photonics.



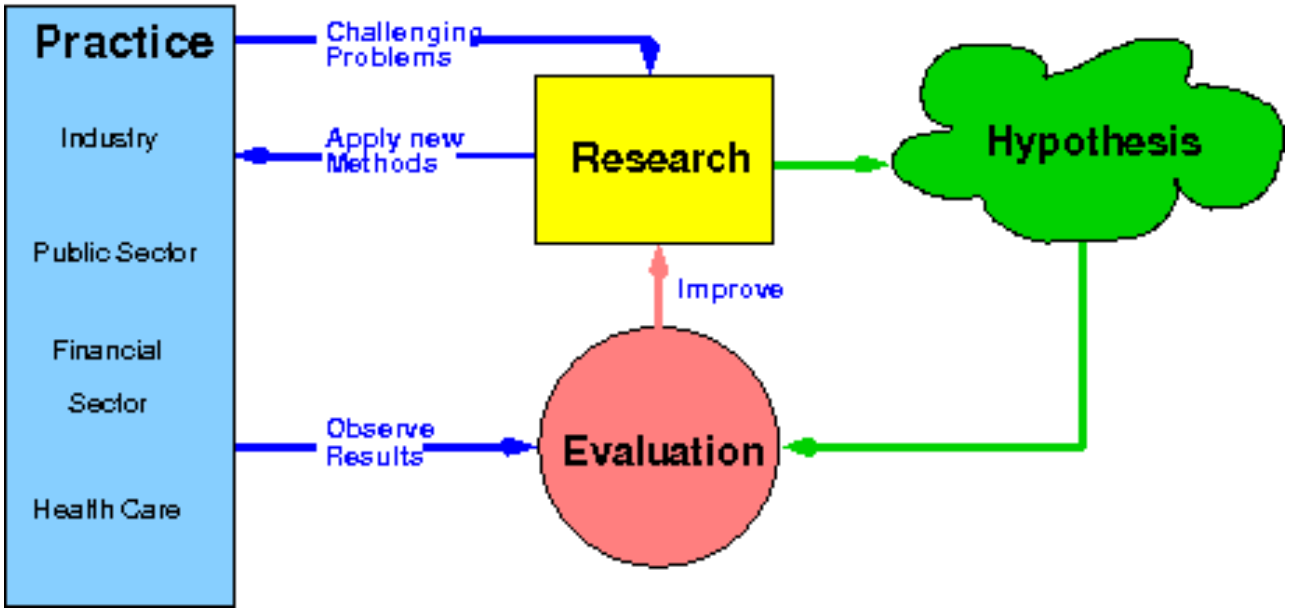
De Amerikaanse softwaregigant Microsoft is bereid een miljoen euro te steken in een onderzoek van de TU Delft naar zogenoemde quantumcomputers. De laatste handtekening van de deal is op de post gegaan, zo stelde Spinozaprijswinnaar Leo Kouwenhoven van de universiteit en onderzoeksleider vrijdag naar aanleiding van een bericht daarover in de Volkskrant.

What Should ICT Industry Do?

- Help improve image ICT profession and ICT studies
- Better lobby for more/better investments in ICT studies/research

What Should Academic Researchers Do?

- Work on fundamental problems
- Be inspired by practice
- Theory – tools – applications cycle
- Practice as laboratory



Thank You!

Questions?