Presentation of Master's
Thesis at Nada, KTH

Stålmarck's Method versus Resolution: A Comparative Theoretical Study

Jakob Nordström

November 13, 2001

Outline of Presentation

- Formal methods
- Automated theorem proving
- Proof theory
- My Master's thesis

The Problem

Larger systems
Growing complexity
Shorter development time

Systems harder to analyze
Testing cannot provide full coverage

Product more likely to contain errors
Less time can be spent on validation

The Consequences

For example:

- Ariane 5 failure
- Pentium FDIV bug
- And others

The Solution(?)

Formal methods

Formalize the methods used for specifying and designing systems.

Use mathematical tools to *prove* correctness.

What Are Formal Methods?

According to The Free On-line Dictionary of Computing (`http://foldoc.doc.ic.ac.uk/`):

> *Mathematically based techniques for the specification, development and verification of software and hardware systems.*

Theorem Proving Approach

Express system and specification in logic.

Consider the formula

$$System \Rightarrow Specification$$

as a theorem in logic to be proved (or refuted).

Theorem proving: The process of finding proofs of logic formulas.

A Small Example

Build an XOR gate by using AND, OR and NOT gates.

And together x or y with x and y inverted.

This construction is correct if

$$
\begin{array}{rl}
 & (a \leftrightarrow (x \wedge y)) \\
\wedge & (b \leftrightarrow (x \vee y)) \\
\wedge & (c \leftrightarrow \overline{a}) \\
\wedge & (z \leftrightarrow (b \wedge c)) \\
\Rightarrow & (z \leftrightarrow (x \, XOR \, y))
\end{array}
$$

is a tautology.

(A formula is a **tautology** if it is always true regardless of the values of the variables).

A Slightly More Realistic Example

$(x_{1,1} \vee x_{1,2} \vee x_{1,3} \vee x_{1,4} \vee x_{1,5} \vee x_{1,6} \vee x_{1,7}) \wedge (x_{2,1} \vee x_{2,2} \vee x_{2,3} \vee x_{2,4} \vee x_{2,5} \vee x_{2,6} \vee x_{2,7}) \wedge (x_{3,1} \vee x_{3,2} \vee x_{3,3} \vee x_{3,4} \vee x_{3,5} \vee x_{3,6} \vee x_{3,7}) \wedge (x_{4,1} \vee x_{4,2} \vee x_{4,3} \vee x_{4,4} \vee x_{4,5} \vee x_{4,6} \vee x_{4,7}) \wedge (x_{5,1} \vee x_{5,2} \vee x_{5,3} \vee x_{5,4} \vee x_{5,5} \vee x_{5,6} \vee x_{5,7}) \wedge (x_{6,1} \vee x_{6,2} \vee x_{6,3} \vee x_{6,4} \vee x_{6,5} \vee x_{6,6} \vee x_{6,7}) \wedge (x_{7,1} \vee x_{7,2} \vee x_{7,3} \vee x_{7,4} \vee x_{7,5} \vee x_{7,6} \vee x_{7,7}) \wedge (x_{8,1} \vee x_{8,2} \vee x_{8,3} \vee x_{8,4} \vee x_{8,5} \vee x_{8,6} \vee x_{8,7}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{2,1}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{3,1}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{4,1}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{5,1}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{6,1}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{7,1}) \wedge (\overline{x}_{1,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{2,1} \vee \overline{x}_{3,1}) \wedge (\overline{x}_{2,1} \vee \overline{x}_{4,1}) \wedge (\overline{x}_{2,1} \vee \overline{x}_{5,1}) \wedge (\overline{x}_{2,1} \vee \overline{x}_{6,1}) \wedge (\overline{x}_{2,1} \vee \overline{x}_{7,1}) \wedge (\overline{x}_{2,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{3,1} \vee \overline{x}_{4,1}) \wedge (\overline{x}_{3,1} \vee \overline{x}_{5,1}) \wedge (\overline{x}_{3,1} \vee \overline{x}_{6,1}) \wedge (\overline{x}_{3,1} \vee \overline{x}_{7,1}) \wedge (\overline{x}_{3,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{4,1} \vee \overline{x}_{5,1}) \wedge (\overline{x}_{4,1} \vee \overline{x}_{6,1}) \wedge (\overline{x}_{4,1} \vee \overline{x}_{7,1}) \wedge (\overline{x}_{4,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{5,1} \vee \overline{x}_{6,1}) \wedge (\overline{x}_{5,1} \vee \overline{x}_{7,1}) \wedge (\overline{x}_{5,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{6,1} \vee \overline{x}_{7,1}) \wedge (\overline{x}_{6,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{7,1} \vee \overline{x}_{8,1}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{2,2}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{3,2}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{4,2}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{5,2}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{6,2}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{7,2}) \wedge (\overline{x}_{1,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{2,2} \vee \overline{x}_{3,2}) \wedge (\overline{x}_{2,2} \vee \overline{x}_{4,2}) \wedge (\overline{x}_{2,2} \vee \overline{x}_{5,2}) \wedge (\overline{x}_{2,2} \vee \overline{x}_{6,2}) \wedge (\overline{x}_{2,2} \vee \overline{x}_{7,2}) \wedge (\overline{x}_{2,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{3,2} \vee \overline{x}_{4,2}) \wedge (\overline{x}_{3,2} \vee \overline{x}_{5,2}) \wedge (\overline{x}_{3,2} \vee \overline{x}_{6,2}) \wedge (\overline{x}_{3,2} \vee \overline{x}_{7,2}) \wedge (\overline{x}_{3,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{4,2} \vee \overline{x}_{5,2}) \wedge (\overline{x}_{4,2} \vee \overline{x}_{6,2}) \wedge (\overline{x}_{4,2} \vee \overline{x}_{7,2}) \wedge (\overline{x}_{4,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{5,2} \vee \overline{x}_{6,2}) \wedge (\overline{x}_{5,2} \vee \overline{x}_{7,2}) \wedge (\overline{x}_{5,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{6,2} \vee \overline{x}_{7,2}) \wedge (\overline{x}_{6,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{7,2} \vee \overline{x}_{8,2}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{2,3}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{3,3}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{4,3}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{5,3}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{6,3}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{7,3}) \wedge (\overline{x}_{1,3} \vee \overline{x}_{8,3}) \wedge (\overline{x}_{2,3} \vee \overline{x}_{3,3}) \wedge (\overline{x}_{2,3} \vee \overline{x}_{4,3}) \wedge (\overline{x}_{2,3} \vee \overline{x}_{5,3}) \wedge (\overline{x}_{2,3} \vee \overline{x}_{6,3}) \wedge (\overline{x}_{2,3} \vee \overline{x}_{7,3}) \wedge (\overline{x}_{2,3} \vee \overline{x}_{8,3}) \wedge (\overline{x}_{3,3} \vee \overline{x}_{4,3}) \wedge (\overline{x}_{3,3} \vee \overline{x}_{5,3}) \wedge (\overline{x}_{3,3} \vee \overline{x}_{6,3}) \wedge (\overline{x}_{3,3} \vee \overline{x}_{7,3}) \wedge (\overline{x}_{3,3} \vee \overline{x}_{8,3}) \wedge (\overline{x}_{4,3} \vee \overline{x}_{5,3}) \wedge (\overline{x}_{4,3} \vee \overline{x}_{6,3}) \wedge (\overline{x}_{4,3} \vee \overline{x}_{7,3}) \wedge (\overline{x}_{4,3} \vee \overline{x}_{8,3}) \wedge (\overline{x}_{5,3} \vee \overline{x}_{6,3}) \wedge (\overline{x}_{5,3} \vee \overline{x}_{7,3}) \wedge (\overline{x}_{5,3} \vee \overline{x}_{8,3}) \wedge$

$(\overline{x}_{6,3} \vee \overline{x}_{7,3}) \wedge (\overline{x}_{6,3} \vee \overline{x}_{8,3}) \wedge (\overline{x}_{7,3} \vee \overline{x}_{8,3}) \wedge (\overline{x}_{1,4} \vee \overline{x}_{2,4}) \wedge (\overline{x}_{1,4} \vee \overline{x}_{3,4}) \wedge$
$(\overline{x}_{1,4} \vee \overline{x}_{4,4}) \wedge (\overline{x}_{1,4} \vee \overline{x}_{5,4}) \wedge (\overline{x}_{1,4} \vee \overline{x}_{6,4}) \wedge (\overline{x}_{1,4} \vee \overline{x}_{7,4}) \wedge (\overline{x}_{1,4} \vee \overline{x}_{8,4}) \wedge$
$(\overline{x}_{2,4} \vee \overline{x}_{3,4}) \wedge (\overline{x}_{2,4} \vee \overline{x}_{4,4}) \wedge (\overline{x}_{2,4} \vee \overline{x}_{5,4}) \wedge (\overline{x}_{2,4} \vee \overline{x}_{6,4}) \wedge (\overline{x}_{2,4} \vee \overline{x}_{7,4}) \wedge$
$(\overline{x}_{2,4} \vee \overline{x}_{8,4}) \wedge (\overline{x}_{3,4} \vee \overline{x}_{4,4}) \wedge (\overline{x}_{3,4} \vee \overline{x}_{5,4}) \wedge (\overline{x}_{3,4} \vee \overline{x}_{6,4}) \wedge (\overline{x}_{3,4} \vee \overline{x}_{7,4}) \wedge$
$(\overline{x}_{3,4} \vee \overline{x}_{8,4}) \wedge (\overline{x}_{4,4} \vee \overline{x}_{5,4}) \wedge (\overline{x}_{4,4} \vee \overline{x}_{6,4}) \wedge (\overline{x}_{4,4} \vee \overline{x}_{7,4}) \wedge (\overline{x}_{4,4} \vee \overline{x}_{8,4}) \wedge$
$(\overline{x}_{5,4} \vee \overline{x}_{6,4}) \wedge (\overline{x}_{5,4} \vee \overline{x}_{7,4}) \wedge (\overline{x}_{5,4} \vee \overline{x}_{8,4}) \wedge (\overline{x}_{6,4} \vee \overline{x}_{7,4}) \wedge (\overline{x}_{6,4} \vee \overline{x}_{8,4}) \wedge$
$(\overline{x}_{7,4} \vee \overline{x}_{8,4}) \wedge (\overline{x}_{1,5} \vee \overline{x}_{2,5}) \wedge (\overline{x}_{1,5} \vee \overline{x}_{3,5}) \wedge (\overline{x}_{1,5} \vee \overline{x}_{4,5}) \wedge (\overline{x}_{1,5} \vee \overline{x}_{5,5}) \wedge$
$(\overline{x}_{1,5} \vee \overline{x}_{6,5}) \wedge (\overline{x}_{1,5} \vee \overline{x}_{7,5}) \wedge (\overline{x}_{1,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{2,5} \vee \overline{x}_{3,5}) \wedge (\overline{x}_{2,5} \vee \overline{x}_{4,5}) \wedge$
$(\overline{x}_{2,5} \vee \overline{x}_{5,5}) \wedge (\overline{x}_{2,5} \vee \overline{x}_{6,5}) \wedge (\overline{x}_{2,5} \vee \overline{x}_{7,5}) \wedge (\overline{x}_{2,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{3,5} \vee \overline{x}_{4,5}) \wedge$
$(\overline{x}_{3,5} \vee \overline{x}_{5,5}) \wedge (\overline{x}_{3,5} \vee \overline{x}_{6,5}) \wedge (\overline{x}_{3,5} \vee \overline{x}_{7,5}) \wedge (\overline{x}_{3,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{4,5} \vee \overline{x}_{5,5}) \wedge$
$(\overline{x}_{4,5} \vee \overline{x}_{6,5}) \wedge (\overline{x}_{4,5} \vee \overline{x}_{7,5}) \wedge (\overline{x}_{4,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{5,5} \vee \overline{x}_{6,5}) \wedge (\overline{x}_{5,5} \vee \overline{x}_{7,5}) \wedge$
$(\overline{x}_{5,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{6,5} \vee \overline{x}_{7,5}) \wedge (\overline{x}_{6,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{7,5} \vee \overline{x}_{8,5}) \wedge (\overline{x}_{1,6} \vee \overline{x}_{2,6}) \wedge$
$(\overline{x}_{1,6} \vee \overline{x}_{3,6}) \wedge (\overline{x}_{1,6} \vee \overline{x}_{4,6}) \wedge (\overline{x}_{1,6} \vee \overline{x}_{5,6}) \wedge (\overline{x}_{1,6} \vee \overline{x}_{6,6}) \wedge (\overline{x}_{1,6} \vee \overline{x}_{7,6}) \wedge$
$(\overline{x}_{1,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{2,6} \vee \overline{x}_{3,6}) \wedge (\overline{x}_{2,6} \vee \overline{x}_{4,6}) \wedge (\overline{x}_{2,6} \vee \overline{x}_{5,6}) \wedge (\overline{x}_{2,6} \vee \overline{x}_{6,6}) \wedge$
$(\overline{x}_{2,6} \vee \overline{x}_{7,6}) \wedge (\overline{x}_{2,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{3,6} \vee \overline{x}_{4,6}) \wedge (\overline{x}_{3,6} \vee \overline{x}_{5,6}) \wedge (\overline{x}_{3,6} \vee \overline{x}_{6,6}) \wedge$
$(\overline{x}_{3,6} \vee \overline{x}_{7,6}) \wedge (\overline{x}_{3,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{4,6} \vee \overline{x}_{5,6}) \wedge (\overline{x}_{4,6} \vee \overline{x}_{6,6}) \wedge (\overline{x}_{4,6} \vee \overline{x}_{7,6}) \wedge$
$(\overline{x}_{4,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{5,6} \vee \overline{x}_{6,6}) \wedge (\overline{x}_{5,6} \vee \overline{x}_{7,6}) \wedge (\overline{x}_{5,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{6,6} \vee \overline{x}_{7,6}) \wedge$
$(\overline{x}_{6,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{7,6} \vee \overline{x}_{8,6}) \wedge (\overline{x}_{1,7} \vee \overline{x}_{2,7}) \wedge (\overline{x}_{1,7} \vee \overline{x}_{3,7}) \wedge (\overline{x}_{1,7} \vee \overline{x}_{4,7}) \wedge$
$(\overline{x}_{1,7} \vee \overline{x}_{5,7}) \wedge (\overline{x}_{1,7} \vee \overline{x}_{6,7}) \wedge (\overline{x}_{1,7} \vee \overline{x}_{7,7}) \wedge (\overline{x}_{1,7} \vee \overline{x}_{8,7}) \wedge (\overline{x}_{2,7} \vee \overline{x}_{3,7}) \wedge$
$(\overline{x}_{2,7} \vee \overline{x}_{4,7}) \wedge (\overline{x}_{2,7} \vee \overline{x}_{5,7}) \wedge (\overline{x}_{2,7} \vee \overline{x}_{6,7}) \wedge (\overline{x}_{2,7} \vee \overline{x}_{7,7}) \wedge (\overline{x}_{2,7} \vee \overline{x}_{8,7}) \wedge$
$(\overline{x}_{3,7} \vee \overline{x}_{4,7}) \wedge (\overline{x}_{3,7} \vee \overline{x}_{5,7}) \wedge (\overline{x}_{3,7} \vee \overline{x}_{6,7}) \wedge (\overline{x}_{3,7} \vee \overline{x}_{7,7}) \wedge (\overline{x}_{3,7} \vee \overline{x}_{8,7}) \wedge$
$(\overline{x}_{4,7} \vee \overline{x}_{5,7}) \wedge (\overline{x}_{4,7} \vee \overline{x}_{6,7}) \wedge (\overline{x}_{4,7} \vee \overline{x}_{7,7}) \wedge (\overline{x}_{4,7} \vee \overline{x}_{8,7}) \wedge (\overline{x}_{5,7} \vee \overline{x}_{6,7}) \wedge$
$(\overline{x}_{5,7} \vee \overline{x}_{7,7}) \wedge (\overline{x}_{5,7} \vee \overline{x}_{8,7}) \wedge (\overline{x}_{6,7} \vee \overline{x}_{7,7}) \wedge (\overline{x}_{6,7} \vee \overline{x}_{8,7}) \wedge (\overline{x}_{7,7} \vee \overline{x}_{8,7})$

Automated Theorem Proving

The second example formula had 56 variables. Real world problems have millions of variables. We need computer assistance.

Automated theorem provers are computer programs which perform automated logical deduction.

Used in for instance:

- Formal methods
- Artificial intelligence
- Theoretical mathematics

Proof System

An automated theorem prover is an algorithm for logical reasoning.

But which reasoning rules does it use?

And what do the proofs produced look like?

This is described by the *proof system* used by the algorithm.

Proof system:

Format for writing down proofs

$$+$$

Algorithm for checking correctness

Conjunctive Normal Form

A **literal** is a variable x or its negation $\overline{x}$.

A **clause** is a disjunction of literals.

$$x \vee y \vee \overline{z}$$

A clause is true if at least one literal in it is true.

A **CNF formula** is a conjunction of clauses.

$$\begin{aligned}
& (\overline{x} \vee \overline{s}) \\
\wedge\ & (s \vee t) \\
\wedge\ & (x \vee y \vee \overline{z}) \\
\wedge\ & (s \vee \overline{t}) \\
\wedge\ & (x \vee \overline{y}) \\
\wedge\ & (x \vee s) \\
\wedge\ & (x \vee z)
\end{aligned}$$

A CNF formula is true if all its clauses are true.

Resolution

Transform the formula

$$System \Rightarrow Specification$$

to a CNF formula F.

F says that $System \Rightarrow Specification$ does *not* hold (i.e. that the design is incorrect).

We want to prove that F is false (i.e. that the design is correct).

Resolution (continued)

Start with clauses in CNF formula F.

Derive new clauses from the clauses in F by the **resolution rule:**

$$\frac{B \vee x \quad C \vee \overline{x}}{B \vee C}$$

A **resolution refutation** of F is a derivation of a contradiction (x and $\overline{x}$) from F.

Example of Resolution Refutation

$\overline{x} \vee \overline{s}$

$s \vee t$

$x \vee y \vee \overline{z}$

$s \vee \overline{t}$

$x \vee \overline{y}$

$x \vee s$

$x \vee z$

s

$\overline{x}$

$x \vee \overline{z}$

x

Proof Method

Proof system: Non-constructive definition of what a proof is.

Proof method: Constructive algorithm to *find* a proof (used by automated theorem prover).

Proof method $A_{\mathcal{P}}$ for proof system $\mathcal{P}$:

- Deterministic algorithm
- Input: Propositional logic formula F
- Output: Proof of F in $\mathcal{P}$ if F is true or counter-example otherwise.

Proof Systems and Proof Methods

A proof method is an algorithm that searches for proofs in a proof system.

Lower bounds in proof system $\mathcal{P}$

$$\Downarrow$$

Lower bounds for proof method $A_{\mathcal{P}}$

(The algorithm cannot be faster than the smallest proof it can possibly find.)

So What about My Master's Thesis?

Written at Prover Technology (`www.prover.com`)

Study of proof system underlying Stålmarck's proof method

Comparison to resolution

Also (theoretical) comparison of Stålmarck's method to resolution-based algorithms

References

Master's thesis at URL

`www.student.nada.kth.se/~md93-jno/`

in file

`docs/mastersthesis.pdf`

Chapter 1 in the thesis and the references for further reading given there