Lecture 15

Satisfiability Modulo Theories

1. **Motivation: Path Sensitivity Analysis**
2. **A Basic SMT Solver**
3. **Optimizing the SMT Solver**

Thanks to Clark Barrett, Nikolaj Bjørner Leonardo de Moura, Bruno Dutertre, Albert Oliveras, and Cesare Tinelli for contributing material used in this lecture.
What is Satisfiability Modulo Theories (SMT)?

• Satisfiability
  – the problem of determining whether a formula has a model (an assignment that makes the formula true)

• SAT: Satisfiability of **propositional formulas**
  – A model is a truth assignment to Boolean variables
  – SAT solvers: check satisfiability of propositional formulas
    • Decidable, NP-complete

• SMT: Satisfiability modulo theories
  – Satisfiability of first-order formulas containing operations from background theories such as arithmetic, arrays, uninterpreted functions, etc.
  
  E.g. \( g(a) = c \land f(g(a)) \neq f(c) \lor g(a) = d \land c \neq d \)
  – SMT Solvers:
    • check satisfiability of SMT formulas with respect to a theory
Use of SMT for Program Correctness & Test Generation

• Precision: Path sensitivity

• Given an assertion $A$, can we generate an input that triggers an error on a given path $p$?
  - Let $F$ be the formula representing the execution of $p$
  - Is the formula $F \land \neg A$ satisfiable?
    • Not satisfiable? No error on that path
    • Satisfiable? Find 1 assignment that satisfies the formula
      (1 set of test inputs)
Each Statement is a Logical Clause

**Program**
Assume data array bound is [0, N-1]

```c
1 void ReadBlocks(int data[], int cookie) {
3   int i = 0;
4     while (true) {
6       int next;
7         next = data[i];
8         if (!(i < next && next < N)) return;
9         i = i + 1;
10        for (; i < next; i = i + 1) {
11          if (data[i] == cookie) {
12             i = i + 1;
13             else Process(data[i]);
14           }
15       }
16   }
17 }
```

**One execution path**

**Static Single Assignment (SSA)**

```c
3 i₁ = 0;
7 next₁ = data₀ [i₁];
8 i₁ < next₁ && next₁ < N₀
9 i₂ = i₁ + 1;
10 i₂ < next₁;
11 data₀ [i₂] = cookie₀;
12 i₃ = i₂ + 1;
10 i₄ = i₃ + 1;
10 !(i₄ < next₁);
7 next₂ = data₀ [i₄];
```
An Execution Path as a Logic Formula

Program
Assume data array bound is [0, N-1]

```c
1 void ReadBlocks(int data[], int cookie)
2 {
3    int i = 0;
4    while (true)
5    {
6        int next;
7        next = data[i];
8        if (!(i < next && next < N)) return;
9        i = i + 1;
10       for (; i < next; i = i + 1){
11          if (data[i] == cookie)
12             i = i + 1;
13           else
14               Process(data[i]);
15       }
16    }
17 }
```

One execution path (SSA)

```
\begin{align*}
F &= \begin{cases}
3 & i_1 = 0; \\
7 & next_1 = data_0 [i_1]; \\
8 & i_1 < next_1 \&\& next_1 < N_0 \\
9 & i_2 = i_1 + 1; \\
10 & i_2 < next_1; \\
11 & data_0 [i_2] = cookie_0; \\
12 & i_3 = i_2 + 1; \\
10 & i_4 = i_3 + 1; \\
10 & !(i_4 < next_1); \\
7 & next_2 = data_0 [i_4];
\end{cases}
\end{align*}
```

Line 7: Array bound assertion $A$:

\[(0 \leq i_1 \land i_1 < N_0)\]
Checking for Out-of-Bound Array Access (Line 7, iteration 1)

Program
Assume data array bound is $[0, N-1]$

```c
1 void ReadBlocks(int data[], int cookie)
2 {
3   int i = 0;
4   while (true)
5     {
6       int next;
7       next = data[i];
8       if (!((i < next && next < N)) return;
9       i = i + 1;
10      for (; i < next; i = i + 1){
11        if (data[i] == cookie)
12         i = i + 1;
13        else
14         Process(data[i]);
15      }   
16   }
17 }
```

One execution path (SSA)

```c
F = 
3 i_1 = 0;
```

Line 7: Array bound assertion $A$:

$$(0 \leq i_1 \land i_1 < N_0)$$

Check: Is $F \land \neg A$ satisfiable?

$$i_1 = 0 \land \neg(0 \leq i_1 \land i_1 < N_0)$$
Answer for Out-of-Bound Array Access (Line 7, iteration 1)

Program Assume data array bound is [0, N-1]

1 void ReadBlocks(int data[], int cookie) {
2     int i = 0;
3     while (true) {
4         int next;
5         next = data[i];
6         if (!$((i < next && next < N))$) return;
7         i = i + 1;
8         for (; i < next; i = i + 1){
9             if (data[i] == cookie)
10                 i = i + 1;
11             else
12                 Process(data[i]);
13         }
14     }
15 }

One execution path (SSA)

F = \begin{align*}
3 & i_1 = 0; \\
7 & \text{next}_1 = \text{data}_0 [i_1]; \\
8 & i_1 < \text{next}_1 \&\& \text{next}_1 < N_0 \\
9 & i_2 = i_1 + 1; \\
10 & i_2 < \text{next}_1; \\
11 & \text{data}_0 [i_2] = \text{cookie}_0; \\
12 & i_3 = i_2 + 1; \\
10 & i_4 = i_3 + 1; \\
10 & !(i_4 < \text{next}_1); \\
7 & \text{next}_2 = \text{data}_0 [i_4];
\end{align*}

Line 7: Array bound assertion $A$:

$$(0 \leq i_1 \land i_1 < N_0)$$

$\Rightarrow$ maps to

Check: Is $F \land \neg A$ satisfiable?

$$i_1 = 0 \land \neg(0 \leq i_1 \land i_1 < N_0)$$

Yes! \{i_1 \mapsto 0, N_0 \mapsto 0\} BUG!!
Checking for Out-of-Bound Array Access (Line 7, iteration 2)

Program Assume data array bound is [0, N-1]
1 void ReadBlocks(int data[], int cookie)
2 {
3    int i = 0;
4    while (true)
5    {
6        int next;
7        next = data[i];
8        if (!((i < next && next < N)) return;
9        i = i + 1;
10       for (; i < next; i = i + 1){
11          if (data[i] == cookie)
12              i = i + 1;
13          else
14              Process(data[i]);
15       }
16    }
17 }

One execution path (SSA)

3 i₁ = 0;
7 next₁ = data₀ [i₁];
8 i₁ < next₁ && next₁ < N₀
9 i₂ = i₁ + 1;
10 i₂ < next₁;
11 data₀ [i₂] = cookie₀;
12 i₃ = i₂ + 1;
10 i₄ = i₃ + 1;
10 !(i₄ < next₁);
7 next₂ = data₀ [i₄];

Line 7: Array bound assertion A:
(0 ≤ i₄ ∧ i₄ < N₀)

Check: Is F ∧ ¬A satisfiable?
F ∧ ¬(0 ≤ i₄ ∧ i₄ < N₀)
Answer for Out-of-Bound Array Access (Line 7, iteration 2)

Program: Assume data array bound is [0, N-1]

```c
1 void ReadBlocks(int data[], int cookie)
2 {
3     int i = 0;
4     while (true)
5         {
6             int next;
7             next = data[i];
8             if (!((i < next && next < N)) return;
9             i = i + 1;
10            for (; i < next; i = i + 1){
11                if (data[i] == cookie)
12                    i = i + 1;
13                else
14                    Process(data[i]);
15            }
16         }
17 }
```

One execution path (SSA)

- \( i_1 = 0; \)
- \( next_1 = data_0 [i_1]; \)
- \( i_1 < next_1 \) \&\& \( next_1 < N_0 \)
- \( i_2 = i_1 + 1; \)
- \( i_2 < next_1; \)
- \( data_0 [i_2] = cookie_0; \)
- \( i_3 = i_2 + 1; \)
- \( i_4 = i_3 + 1; \)
- \( !(i_4 < next_1); \)
- \( next_2 = data_0 [i_4]; \)

Line 7: Array bound assertion \( A: \)

- \( 0 \leq i_4 \land i_4 < N_0 \)

<table>
<thead>
<tr>
<th>Var</th>
<th>( \mapsto )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_0 )</td>
<td>3</td>
</tr>
<tr>
<td>( i_1 )</td>
<td>0</td>
</tr>
<tr>
<td>( i_2 )</td>
<td>1</td>
</tr>
<tr>
<td>( i_3 )</td>
<td>2</td>
</tr>
<tr>
<td>( i_4 )</td>
<td>3</td>
</tr>
<tr>
<td>( next_1 )</td>
<td>2</td>
</tr>
<tr>
<td>( data_0 )</td>
<td>&lt;2,0,0&gt;</td>
</tr>
<tr>
<td>( cookie_0 )</td>
<td>0</td>
</tr>
</tbody>
</table>

**BUG!!**
Checking the Whole Program All at Once

• A program has many execution paths
• Conditional statements
  – Represent alternative paths symbolically with one formula using SSA
• Loops
  – Optimistically: Unroll a few times
  – Catches many errors, but not all errors
Conditional Statements

- **Conditional statements**: $\varphi$ functions in SSA
  
  ```
  if (i > 0) {
    a = 2;
    b = 3;
  } else {
    a = 3;
    b = 2;
  }
  c = a+b;
  ```

  - **Assert $A$**: $c_3 = 5$
  - **Is $F \land \neg A$ satisfiable?**
    
    $$
    \varphi_1 = (i_0 > 0) \land (\varphi_1 \rightarrow c_3 = 5) \land (\neg \varphi_1 \rightarrow c_3 = 5) \land (c_3 \neq 5)
    $$
Applying the Resolution Rule to Example

• A resolution rule in propositional logic:

\[
\text{Given } p \lor A \text{ and } \neg p \lor B, \text{ add the resolvent } A \lor B
\]

\[
\begin{array}{c}
\text{Resolve} \\
p \lor A & \neg p \lor B \\
\hline
A \lor B
\end{array}
\]

• Is \( F \land \neg A \) satisfiable?

\[
\varphi_1 = (i_0 > 0) \land (\varphi_1 \rightarrow c_3 = 5) \land (\neg \varphi_1 \rightarrow c_3 = 5) \land (c_3 \neq 5)
\]

• Recall: \( p \rightarrow q \equiv \neg p \lor q \)

\[
\varphi_1 = (i_0 > 0) \land (\neg \varphi_1 \lor c_3 = 5) \land (\varphi_1 \lor c_3 = 5) \land (c_3 \neq 5)
\]

\[
\varphi_1 = (i_0 > 0) \land (c_3 = 5) \land (c_3 \neq 5)
\]

• \( F \land \neg A \) is not satisfiable
• The assertion \( A \) is true.
**Loops**

- **Optimistically: Unroll two times**

```plaintext
for (; i < next; i = i + 1) {
  if (data[i] == cookie) {
    i = i + 1;
  } else {
    Process(data[i]);
  }
}
```

```plaintext
if (i < next) {
  if (data[i] == cookie) {
    i = i + 1;
  } else {
    Process(data[i]);
    i = i + 1;
  }
}
```

```plaintext
if (i < next) {
  if (data[i] == cookie) {
    i = i + 1;
  } else {
    Process(data[i]);
    i = i + 1;
  }
}
```
Loops: Apply SSA

```
1 if (i < next) {
2   if (data[i] == cookie)
3     i = i + 1;
4   else
5     Process(data[i]);
6   i = i + 1;
7   if (i < next) {
8     if (data[i] == cookie)
9       i = i + 1;
10      else
11         Process(data[i]);
12      i = i + 1;
13     }
14   }
15 i = i + 1;
16 }
```

```
1 φ₁ = (i₀ < next₀);
2 φ₂ = (data₀[i₀] == cookie₀);
3 i₁ = i₀ + 1;
4 i₂ = φ₂ ? i₁ : i₀;
5 i₃ = i₂ + 1;
6 φ₃ = (i₃ < next₀);
7 φ₄ = (data₀[i₃] == cookie₀);
8 i₄ = i₃ + 1;
9 i₅ = φ₄ ? i₄ : i₃;
10 i₆ = i₅ + 1;
11 i₇ = φ₃ ? i₆ : i₃;
12 i₈ = φ₁ ? i₇ : i₀;
```
# Major Categories of Program Analysis Tools

<table>
<thead>
<tr>
<th>Complete (Small programs)</th>
<th>Static Property Based</th>
<th>Dynamic Execution Based</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verification</strong></td>
<td>Prove a property in a program</td>
<td><strong>(Symbolic) Model Checking (SMT/BDD)</strong></td>
</tr>
<tr>
<td>Floyd-Hoare logic:</td>
<td>{pre-condition} s {post-condition}</td>
<td>Given a system model (sw/hw), simulate the execution to check if a property is true for all possible inputs. Symbolic: many states all at once</td>
</tr>
<tr>
<td>Applicable to small programs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incomplete (Large programs)</th>
<th>Static Analysis (Data flow)</th>
<th>Test case generation (SMT/BDD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract the program conservatively</td>
<td>Check a property</td>
<td>Check a property opportunistically (e.g. unroll loops twice)</td>
</tr>
<tr>
<td>Check a property</td>
<td>Sound: no false-negatives--find all bugs</td>
<td>Use analysis to generate test inputs</td>
</tr>
<tr>
<td>False-positives: false warnings</td>
<td>Too imprecise is useless</td>
<td>No false-positives: generate a test</td>
</tr>
<tr>
<td>Too imprecise is useless</td>
<td></td>
<td>False-negatives: cannot find all bugs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No correctness/security guarantees</td>
</tr>
</tbody>
</table>
2. A Basic SMT Solver

• **SMT**: Satisfiability modulo theories
  – Satisfiability of first-order formulas containing operations from background theories such as arithmetic, arrays, uninterpreted functions, etc.

• **SMT Solvers**:
  – check satisfiability of SMT formulas with respect to a theory
Uninterpreted function theory:
  Functions assumed to be pure:
    A function always returns the same value for a given input

Example: \[ x \geq 0 \land f(x) \geq 0 \land f(y) \geq 0 \land x \neq y \]

Is this formula satisfiable?
Uninterpreted function theory:
Functions assumed to be pure:
A function always returns the same value for a given input

Example: $x \geq 0 \land f(x) \geq 0 \land f(y) \geq 0 \land x \neq y$
This formula is satisfiable
An example model satisfying the formula
$x \mapsto 0$
$y \mapsto 1$
$f(0) \mapsto 0$
$f(1) \mapsto 0$
Adding a Theory of Arrays

Notation: write(v, i, x) means \( v[i] := x; \)
read(v, i) means returns \( v[i] \)

Array theory axioms:
read(write(v, i, x), i) = x
\( i \neq j \rightarrow \text{read(write(v, i, x), j)} = \text{read(v, j)} \)
\( v \neq w \rightarrow \text{read(v, k)} \neq \text{read(w, k)} \) for some \( k \)

Example: \( b + 2 = c \land f(\text{read(write(a, b, 3), c - 2)}) \neq f(c - b + 1) \)

Is this formula satisfiable?
Adding a Theory of Arrays

Notation: write(v, i, x) means v[i] := x;
          read(v, i) means returns v[i]

Array theory axioms:
read(write(v, i, x), i) = x
i ≠ j → read(write(v, i, x), j) = read(v, j)
v ≠ w → read(v, k) ≠ read(w, k) for some k

Example: b + 2 = c ∧ f(read(write(a, b, 3), c − 2)) ≠ f(c − b + 1)

By arithmetic reasoning, this is equivalent to
b + 2 = c ∧ f(read(write(a, b, 3), b)) ≠ f(3)

By first array axiom, b + 2 = c ∧ f(3) ≠ f(3)

But in the theory of uninterpreted functions, f(3) ≠ f(3) is not true

Therefore, this formula is not satisfiable
SMT Solvers

• Input: a first-order formula F
• Output
  – F is satisfiable, optionally: a model M
  – F is unsatisfiable, optionally: a proof of unsatisfiability
• Which is easier?
• Main issues
  – formula size (e.g. thousands of atoms or more)
  – formulas with complex Boolean structure
  – combination of theories
Overview of a SMT Solver

- SMT Solver = SAT Solver + Theory Solver
  - Given a formula F,
    the SAT solver enumerates possible truth assignments (M)
  - The theory solver checks whether the truth assignments are satisfiable in the theories
Example of a Basic Algorithm

\[ F : g(a) = c \land f(g(a)) \neq f(c) \lor g(a) = d \land c \neq d \]

1. Choose a model \( M \)
2. \( \text{unsat} \)
3. Choose a model \( M \)
4. \( \text{unsat} \)

SAT SOLVER

- Choose a model \( M \)
- \( 1, \overline{2} \lor 3, \overline{4} \)
- \( 1, \overline{2}, \overline{4} \)
- \( 1, \overline{2} \lor 3, \overline{4}, \overline{1} \lor 2 \lor 4 \)
- \( 1, 3, \overline{4} \)
- \( 1, \overline{2} \lor 3, \overline{4}, \overline{1} \lor 2 \lor 4, \overline{1} \lor 3 \lor 4 \)
- \( \text{unsat} \)
- \( \text{send F} \)

THEORY SOLVER

(Uninterpreted Functions)

- \( \text{send F} \)
- \( \text{unsat} \)
- \( \text{send F} \)
- \( \text{unsat} \)
- \( \text{send F} \)
Basic Algorithm

- **DEFINITION**
  T-conflict: check for conflicts with respect to theory T

- **DESIGN**
  Two independent solvers:
  - SAT solver that is independent of theory
  - Theory solver that checks for T-conflicts clause by clause

- **ALGORITHM**
  Repeat
  SAT Solver: propose a full propositional model $M$ for formula $F$
  if no $M$ is found, $F$ is unsatisfiable.

  Theory Solver:
  - Check for T-conflict on model $M$
  - If $M$ is satisfiable: $F$ is satisfiable
  - If $M$ has a T-conflict, add constraint to $F$
3. Improvements (Example, Algorithm, Rules)

A. Incremental model decision:

Don’t just guess the entire model (all the assignments)
Check each assignment incrementally, not all at once.
Two kinds of assignments
  - Propagated: deduced from the Boolean expression
  - Arbitrary decisions (marked with •)
If unsatisfiable, backtrack on decisions (Propagate, Decide, T-Conflict, Learn, Restart)

B. Use the theory to propagate and learn (T-Propagate)

C. Backtrack to conflicting decision (Conflict, Explain, Backjump)
### A. Incremental: Example

\( F: \quad g(a) = c \land f(g(a)) \neq f(c) \lor g(a) = d \land c \neq d \)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>C</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 2 v 3, 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 4</td>
<td>1, 2 v 3, 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 4 • 2</td>
<td>1, 2 v 3, 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 4 • 2</td>
<td>1, 2 v 3, 4, \top v 2 v 4</td>
<td>\top v 2 v 4</td>
<td>T-Conflict</td>
</tr>
<tr>
<td>1 4 • 2</td>
<td>1, 2 v 3, 4, \top v 2 v 4</td>
<td>\top v 2 v 4</td>
<td>Learn</td>
</tr>
<tr>
<td>1 4</td>
<td>1, 2 v 3, 4, \top v 2 v 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 4 2 3</td>
<td>1, 2 v 3, 4, \top v 2 v 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 4 2 3</td>
<td>1, 2 v 3, 4, \top v 2 v 4, \top \top v 3 v 4 v 2</td>
<td>\top \top v 3 v 4 v 2</td>
<td>T-Conflict, Learn</td>
</tr>
<tr>
<td>fail</td>
<td></td>
<td></td>
<td>Fail</td>
</tr>
</tbody>
</table>
A. Incremental: Algorithm

- Build incrementally a satisfying truth assignment $M$ for a CNF formula $F$
  - CNF: conjunction of disjunctions of literals

- Algorithm
  Apply rules until there is a satisfying model or Fail, in decreasing priority
  
  T-conflict: if all the literals $l_1, \ldots, l_n$ in $M$ cannot be satisfied by $T$, set the conflict clause $C := \overline{T_1} \lor \ldots \lor \overline{T_n}$
  
  Learn: add the new conflict constraint to $F$
  Restart: Restart the SAT solver after learning a new constraint
  Propagate: deduce the truth value of a literal from $M$ and $F$
  Decide: guess a truth value
  
  Fail: if there is no decision to roll back
# A. Incremental: Rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
</table>
| Propagate | Deduce the truth value of a literal from M and F  
\[ l_1 \lor \ldots \lor l_n \lor l \in F \quad \mathcal{T}_1, \ldots, \mathcal{T}_n \in M \quad l, \mathcal{T} \notin M \]
\[ M := M \cup \{ l \} \] |
| Decide | Guess a truth value  
\[ l \in \text{Lit}(F) \quad l, \mathcal{T} \notin M \]
\[ M := M \cup \{ l \} \] |
| T-Conflict | If all the literals \( l_1, \ldots, l_n \) in M cannot be satisfied by T, set the conflict clause  
\[ C := \mathcal{T}_1 \lor \ldots \lor \mathcal{T}_n \]
\[ C = \text{no} \quad l_1, \ldots, l_n \in M \quad l_1, \ldots, l_n \models T \models \bot \]
\[ C := \mathcal{T}_1 \lor \ldots \lor \mathcal{T}_n \] |
| Learn | Add the new learned constraint to formula F  
\[ F \models P C \quad C \notin F \]
\[ F := F \cup \{ C \} \] |
| Restart | Restart the SAT solver  
\[ M := M^{[0]} \quad C := \text{no} \] | Each Decide defines a new level  
\[ M^{[i]} \] means Model M up to level \( i \) |
A. Incremental: Rules

Fail if there is no decision to roll back

\[
\begin{align*}
\text{Fail} & \quad l_1 \lor \ldots \lor l_n \in F \quad T_1, \ldots, T_n \in M \quad \bullet \notin M \\
\text{fail} & 
\end{align*}
\]
Improvements (Example, Algorithm, Rules)

A. Incremental model decision
   (Propagate, Decide, T-Conflict, Learn, Restart)

B. Use the theory to propagate and learn (T-Propagate)
   In A, propagation is based only on the Boolean expression;
   Here, we add propagation due to the Theories

C. Backtrack to conflicting decision (Conflict, Explain, Backjump)
B: T-Propagate: Example

\[
g(a) = c \land f(g(a)) \neq f(c) \lor g(a) = d \land c \neq d
\]

<table>
<thead>
<tr>
<th>M</th>
<th>F</th>
<th>C</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1, 2 ∨ 3, 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1, 2 ∨ 3, 4</td>
<td></td>
<td>Propagate+</td>
</tr>
<tr>
<td>1</td>
<td>1, 2 ∨ 3, 4</td>
<td></td>
<td>T-Propagate (1 ⊨_T 2)</td>
</tr>
<tr>
<td>1</td>
<td>1, 2 ∨ 3, 4</td>
<td></td>
<td>T-Propagate (1, 4 ⊨_T 3)</td>
</tr>
<tr>
<td>Fail</td>
<td>1, 2 ∨ 3, 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notation:
1 ⊨_T 2: predicate 1 entails predicate 2 under theory T
If predicate 1 is true, predicate 2 is true under theory T
B. T-Propagate: Algorithm

- Add T-Propagate to increase deduced values using theory T

- Algorithm
  
  Apply rules until there is a satisfying model or Fail, in decreasing priority
  
  T-conflict: if all the literals $l_1, \ldots, l_n$ in $M$ cannot be satisfied by $T$, set the conflict clause $C := \overline{T_1} \lor \ldots \lor \overline{T_n}$

  Learn: add the new conflict constraint to $F$

  Restart: Restart the SAT server after learning a new constraint

  Propagate: deduce the truth value of a literal from $M$ and $F$

  T-Propagate: deduce the truth value of a literal using theory $T$

  Decide: guess a truth value

  Fail: if there is no decision to roll back
B. T-Propagate: Rules

Deduce the truth value of a literal using theory T

T-Propagate \( l \in \operatorname{Lit}(F) \quad M \models_T l, \overline{T} \notin M \)

\[ M := M \uplus l \]
Improvements (Example, Algorithm, Rules)

A. Incremental model decision
   (Propagate, Decide, T-Conflict, Learn, Restart)

B. Use the theory to propagate and learn (T-Propagate)

C. Backtrack to conflicting decision (Conflict, Explain, Backjump)
   Find the root cause that causes the conflict
   Backtrack by skipping decisions immaterial to the conflict
### C. Backjumping: Example

\[ F := \{1, 1 \lor 2, 3 \lor 4, 5 \lor 6, 1 \lor 5 \lor 7, 2 \lor 3 \lor 6 \lor 7\} \]

<table>
<thead>
<tr>
<th>M</th>
<th>F</th>
<th>C</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12\cdot3</td>
<td>( F )</td>
<td></td>
<td>Decide</td>
</tr>
<tr>
<td>12\cdot34</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12\cdot34\cdot5</td>
<td>( F )</td>
<td></td>
<td>Decide</td>
</tr>
<tr>
<td>12\cdot34\cdot5\overline{6}</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12\cdot34\cdot5\overline{67}</td>
<td>( F )</td>
<td>( 2 \lor 3 \lor 6 \lor 7 )</td>
<td>Conflict</td>
</tr>
<tr>
<td>12\cdot34\cdot5\overline{67}</td>
<td>( F )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C. Backjumping: Example Details

\[ F := \{1, \overline{1} \vee 2, \overline{3} \vee 4, \overline{5} \vee 6, \overline{1} \vee \overline{5} \vee 7, \overline{2} \vee \overline{5} \vee 6 \vee 7\} \]
\[ M := 12 \cdot 34 \cdot 567 \]
\[ C := \overline{2} \vee \overline{5} \vee 6 \vee 7 \]

- **Conflict:** \( \overline{2} \vee \overline{5} \vee 6 \vee 7 \) last literal choice is 7
- **Explain:** Choice of 7 is due to \( \overline{1} \vee \overline{5} \vee 7 \)
- **Learn:** \( \overline{1} \vee \overline{2} \vee \overline{5} \vee 6 \) = resolvent of \( \overline{2} \vee \overline{5} \vee 6 \vee 7 \) and \( \overline{1} \vee \overline{5} \vee 7 \)
- **Conflict:** \( \overline{1} \vee \overline{2} \vee \overline{5} \vee 6 \) last literal choice is 6
- **Explain:** Choice of 6 is due to \( \overline{5} \vee 6 \)
- **Learn:** \( \overline{1} \vee \overline{2} \vee 5 \) = resolvent of \( \overline{1} \vee \overline{2} \vee 5 \vee 6 \) and \( \overline{5} \vee 6 \)
- **Conflict:** \( \overline{1} \vee \overline{2} \vee 5 \)
- **Backjump:** Choice of 5 was a decision
  - Conflict involves literals 1, 2, 5, the decision of 5 is at level 2
  - 1, 2 are both level 0
  - Back jump to level 0, propagate 1,2 and choose \( \overline{5} \)
**C. Backjumping: Example**

\[ F := \{1, \top \lor 2, \overline{3} \lor 4, \overline{5} \lor 6, \top \lor \overline{5} \lor 7, \overline{2} \lor \overline{5} \lor 6 \lor 7\} \]

<table>
<thead>
<tr>
<th>M</th>
<th>F</th>
<th>C</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12 \cdot 3</td>
<td>( F )</td>
<td></td>
<td>Decide</td>
</tr>
<tr>
<td>12 \cdot 34</td>
<td>( F )</td>
<td></td>
<td>Propagate</td>
</tr>
<tr>
<td>12 \cdot 34 \cdot 56</td>
<td>( F )</td>
<td></td>
<td>Decide</td>
</tr>
<tr>
<td>12 \cdot 34 \cdot 567</td>
<td>( F )</td>
<td>( \overline{2} \lor \overline{3} \lor 6 \lor 7 )</td>
<td>Conflict</td>
</tr>
<tr>
<td>12 \cdot 34 \cdot 567</td>
<td>( F )</td>
<td>( \top \lor \overline{2} \lor 5 \lor 6 )</td>
<td>Explain with ( \top \lor \overline{5} \lor 7 )</td>
</tr>
<tr>
<td>12 \cdot 34 \cdot 567</td>
<td>( F )</td>
<td>( \top \lor \overline{2} \lor 5 )</td>
<td>Explain with ( \overline{3} \lor 6 )</td>
</tr>
<tr>
<td>125</td>
<td>( F )</td>
<td></td>
<td>Backjump</td>
</tr>
<tr>
<td>125 \cdot 3</td>
<td>( F )</td>
<td></td>
<td>Decide</td>
</tr>
<tr>
<td>125 \cdot 34</td>
<td>( F )</td>
<td></td>
<td>Propagate (SAT)</td>
</tr>
</tbody>
</table>
C. Backjumping: Algorithm

• If $M$ is $T$-unsatisfiable, backtrack to some point where the assignment was still $T$-satisfiable.

• Trace back to the decision that causes the conflict $C$.
  • Let $l$ be the last literal choice that causes conflict $C$, if $l$ is a decision, proceed to the next step.
  • **Explain:**
    if $l$ is chosen due to clause $C_1$ in $F$ (explanation), new conflict $C = \text{resolvent of } C$ and $C_1$ (eliminating $l$).
  • Repeat the above.

• Backtrack by skipping decisions immaterial to conflict $C$.
  • **Backjump:** Keep model up to level $i$, (highest level of satisfiable decisions involved in $C$);
    add the latest literal $l$ in $C$. 

M. Lam
# C. Backjumping Rules

## Conflict

If one of the literals $T_1, \ldots, T_n$ in $M$ must be inverted in $F$, set the conflict clause $C := l_1 \lor \ldots \lor l_n$

<table>
<thead>
<tr>
<th>$C = \text{no}$</th>
<th>$l_1 \lor \ldots \lor l_n \in F$</th>
<th>$T_1, \ldots, T_n \in M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C := l_1 \lor \ldots \lor l_n$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Explain

Given conflict $C$ involving latest $l$, chosen due to a clause in $F$, their resolvent is the new conflict

<table>
<thead>
<tr>
<th>$C = l \lor D$</th>
<th>$l_1 \lor \ldots \lor l_n \lor T \in F$</th>
<th>$T_1, \ldots, T_n &lt;_M T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C := l_1 \lor \ldots \lor l_n \lor D$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Backjump

Keep model up to level $i$ (highest level of sat. decisions involved in $C$); add latest $l$ in $C$

<table>
<thead>
<tr>
<th>$C = l_1 \lor \ldots \lor l_n \lor l$</th>
<th>$\text{lev } T_1, \ldots, \text{lev } T_n \leq i &lt; \text{lev } \overline{l}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C := \text{no}$</td>
<td>$M := M^{[i]} l$</td>
</tr>
</tbody>
</table>

$l <_M l'$ if $l$ occurs before $l'$ in $M$

$M^{[i]}$ means Model $M$ up to level $i$

$\text{lev } l = i$ iff $l$ occurs in decision level $i$ of $l$
C. Backjumping Rules (cont.)

Replace

Fail if there is no decision to roll back

\[
\begin{align*}
\text{Fail} & \quad l_1 \lor \ldots \lor l_n \in F \quad T_1, \ldots, T_n \in M \quad \bullet \notin M \\
& \quad \text{fail}
\end{align*}
\]

with

Fail if there is a conflict and there is no decision to roll back

\[
\begin{align*}
\text{Fail} & \quad C \neq \text{no} \quad \bullet \notin M \\
& \quad \text{fail}
\end{align*}
\]
Putting it All Together

Apply rules until there is a satisfying model or Fail, in decreasing priority

**T-conflict:** if all the literals $l_1, \ldots, l_n$ in $M$ cannot be satisfied by $T$, set the conflict clause $C := \overline{l_1} \lor \ldots \lor \overline{l_n}$

**Explain:** If the last literal $l$ in conflict $C$ is not a decision,
If $T$ chosen due to clause $C_1$ in $F$ (explanation),
new conflict = resolvent of $C$ and $C_1$

**Backjump:** Keep model up to level $i$,
(highest level of satisfiable decisions involved in $C$);
add the latest literal $l$ in $C$

**Learn:** add the new conflict constraint to $F$

**Propagate:** deduce the truth value of a literal from $M$ and $F$

**T-Propagate:** deduce the truth value of a literal using theory $T$

**Decide:** guess a truth value

**Fail:** if there is no decision to roll back

**Restart:** Restart on the learned $F$ if too many conflicts have been found
Summary

• Use of SMT to handle path sensitivity in test generation & static analysis

• Basic optimizations in SMT Solver
  – Incremental model decision (Propagate, Decide, T-Conflict, Learn, Restart)
  – Use the theory to propagate and learn (T-Propagate)
  – Smart backtracking (Conflict, Explain, Backjump)

• More work is needed to handle combinations of theories etc

• Practical tools:
  – Z3 SMT solver
    • A widely used, open-source project from Microsoft
  – CVC4 SMT solver (developed at Stanford)
Applications of SMT

• Program Analysis and Verification!

• Security (e.g. checking access policies at AWS)

• Hardware configuration and verification

• Verification of smart contracts for blockchains

• Checking database integrity constraints

• Network configuration checking

• Program synthesis

• And many more...
Research in SMT

• New theories and theory solvers, often adapted to an application domain

• Independently checkable proofs for unsatisfiable formulas

• Extensions to reasoning about quantifiers

• Extensions to synthesis

• Better decision heuristics

• Better algorithms and performance

• Theory combination mechanisms

• Parallel SMT solving
Further Readings

• "Satisfiability Modulo Theories"
  Clark Barrett and Cesare Tinelli.
  In Handbook of Model Checking,
  (Ed Clarke, Thomas Henzinger, and Helmut Veith, eds.), 2016.
  In preparation.

• "Satisfiability Modulo Theories"
  Clark Barrett, Roberto Sebastiani, Sanjit Seshia, and Cesare Tinelli.
  In Handbook of Satisfiability,
  vol. 185 of Frontiers in Artificial Intelligence and Applications,
  (Armin Biere, Marijn J. H. Heule, Hans van Maaren, and Toby Walsh, eds.),

• Satisfiability Modulo Theories: Introduction and Applications
  Leonardo De Moura, Nikolaj Bjørner
  Communications of the ACM, Vol. 54 No. 9, Pages 69-77
  Sept 2011