DE LA RECHERCHE À L'INDUSTRIE

Real Behavior of Floating Point Numbers

SMT 2017 | Bruno Marre, Bobot François, Zakaria Chihani

23 July 2017

www.cea.fr

COLIBRI (Bruno Marre)

- Started in 2000 for test case generation
- Used only as a library in PathCrawler and Gatel
- CP solver uses Eclipse Prolog
- Proprietary with the help of IRSN

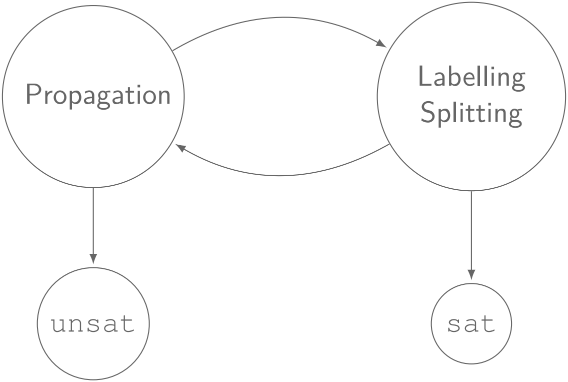
- No test case that use `NaN` or infinities
- Only `fp.eq`, no $=$, only RNE, $+0 = -0$, only 32/64 bit
- integer modulo, real

COLIBRI (Bruno Marre)

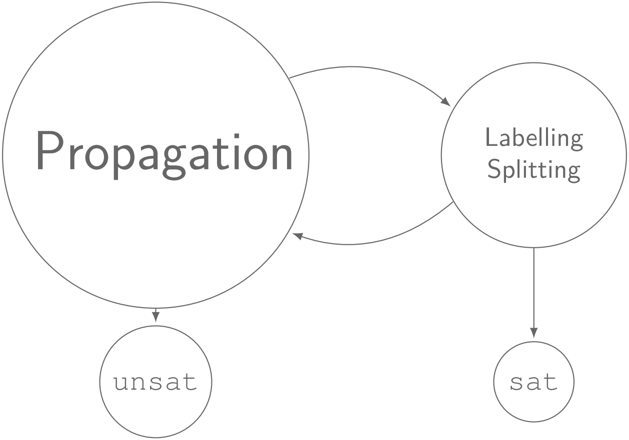
- Started in 2000 for test case generation
- Used ~~only~~ as a library in PathCrawler and Gatel
- CP solver uses Eclipse Prolog
- ~~Proprietary~~ freeware for academic with the help of IRSN

- ~~No test case that use `NaN` or infinities~~
- ~~Only `fp.eq`, no $=$, only RNE, $+0 = -0$~~, only 32/64 bit
- integer modulo, real

Architecture



Architecture



Floating Points

✔ Clear Semantic: $x \oplus y = \mathrm{o}(x + y)$

Floating Points

- ✔ Clear Semantic: $x \oplus y = \circ(x + y)$
- ✘ Few algebraic properties: not associative, $x \oplus y = x \not\Rightarrow y = 0$

Floating Points

- ✔ Clear Semantic: $x \oplus y = \circ(x + y)$
- ✘ Few algebraic properties: not associative, $x \oplus y = x \not\Rightarrow y = 0$
- ✘ Counter-intuitive: $\overbrace{0.1 \oplus \cdots \oplus 0.1}^{10} \neq 0.1 \otimes 10. = 1.$

Floating Points

- ✔ Clear Semantic: $x \oplus y = \circ(x + y)$
- ✘ Few algebraic properties: not associative, $x \oplus y = x \not\Rightarrow y = 0$
- ✘ Counter-intuitive: $\overbrace{0.1 \oplus \cdots \oplus 0.1}^{10} \neq 0.1 \otimes 10. = 1.$
- ✘ State of the art: current bit-blasting doesn't scale

Floating Points

- ✔ Clear Semantic: $x \oplus y = \circ(x + y)$
- ✘ Few algebraic properties: not associative, $x \oplus y = x \not\Rightarrow y = 0$
- ✘ Counter-intuitive: $\overbrace{0.1 \oplus \cdots \oplus 0.1}^{10} \neq 0.1 \otimes 10. = 1.$
- ✘ State of the art: current bit-blasting doesn't scale
- ✘ Pervasives in programs

Domain Specific Approach of CP

$$X_i \in [1;10] \implies X_0 \oplus X_1 \oplus X_2 \oplus X_3 \oplus X_4 \oplus X_5 \oplus X_6 \oplus X_7 \in [8;80]$$

Z3 : 3s
COLIBRI: < 0.1s $(+0.25$s$)$

Domain Specific Approach of CP

$$X_i \in [1;10] \implies X_0 \oplus X_1 \oplus X_2 \oplus X_3 \oplus X_4 \oplus X_5 \oplus X_6 \oplus X_7 \in [8;80]$$

Z3 : 3s
COLIBRI: < 0.1s ($+0.25$s)

$$X_i \in [1;10] \implies X_0 \otimes X_1 \otimes X_2 \otimes X_3 \otimes X_4 \otimes X_5 \otimes X_6 \otimes X_7 \in [1;10^8]$$

Z3 : 31min
COLIBRI: < 0.1s ($+0.25$s)

COLIBRI: Floating Point

- Precise domain propagation:
  $x \oplus y = 0.05 \implies x, y \in [-0.1259..; 0.175....]$

COLIBRI: Floating Point

- Precise domain propagation:
  $x \oplus y = 0.05 \implies x, y \in [-0.1259..; 0.175....]$
  0.05: $0x3fa999999999999a$

COLIBRI: Floating Point

- Precise domain propagation:
  $x \oplus y = 0.05 \implies x, y \in [-0.1259..; 0.175....]$
  0.05: $0x3fa999999999999a$
- Distance graph on floating-point numbers

Distance graph on floating-point numbers

x	IEEE-format, num(x)
0.	0

$$\mathrm{num}(x) - \mathrm{num}(\texttt{fp.mul\_2}\, x) = 2^{52}$$

Distance graph on floating-point numbers

x	IEEE-format, num(x)
0.	0
$+1p - 1074$	1
$+1p - 1073$	2
1.0	`0x3ff0000000000000`
2.0	`0x4000000000000000`

$$\mathrm{num}(x) - \mathrm{num}(\texttt{fp.mul\_2}\,x) = 2^{52}$$

Distance graph on floating-point numbers

x	IEEE-format, num(x)
-2.0	`-0x4000000000000000`
-1.0	`-0x3ff0000000000000`
$-1p-1073$	`-2`
$-1p-1074$	`-1`
$0.$	`0`
$+1p-1074$	`1`
$+1p-1073$	`2`
1.0	`0x3ff0000000000000`
2.0	`0x4000000000000000`

$$\mathrm{num}(x) - \mathrm{num}(\texttt{fp.mul\_2}\,x) = 2^{52}$$

Distance graph on floating-point numbers

x	IEEE-format, num(x)
-2.0	`-0x4000000000000000`
-1.0	`-0x3ff0000000000000`
$-1p-1073$	`-2`
$-1p-1074$	`-1`
$-0.$	`-0`
$0.$	`0`
$+1p-1074$	`1`
$+1p-1073$	`2`
1.0	`0x3ff0000000000000`
2.0	`0x4000000000000000`

$$\mathrm{num}(x) - \mathrm{num}(\texttt{fp.mul\_2}\,x) = 2^{52}$$

Distance graph on floating-point numbers

$$x \in [1;10],\ \texttt{fp.mul}\ \mathrm{RNE}\ x\ 2 = y$$

$$\begin{tikzcd} x \arrow[rr, bend left, "\{2^{52}\}"] & & y \end{tikzcd}$$

$$w \in [1;10],\ \texttt{fp.add}\ \mathrm{RNE}\ w\ 3 = z$$

$$\begin{tikzcd} w \arrow[rr, bend left, "{[\mathrm{num}(13) - \mathrm{num}(10); \mathrm{num}(4) - \mathrm{num}(1)]}"] & & z \end{tikzcd}$$

COLIBRI: Floating Point

- Precise domain propagation:
  $x \oplus y = 0.05 \implies x, y \in [-0.1259..; 0.175....]$
  0.05: $0x3fa999999999999a$
- Distance graph on floating-point numbers
- Monotonic functions:
  $\mathrm{o}(f(x)) < \mathrm{o}(y) \implies \mathrm{o}(x) \leq \mathrm{o}(f^{-1}(\mathrm{o}(y)))$

COLIBRI: Floating Point

- Precise domain propagation:
  $x \oplus y = 0.05 \implies x, y \in [-0.1259..; 0.175....]$
  0.05: $0x3fa999999999999a$
- Distance graph on floating-point numbers
- Monotonic functions:
  $o(f(x)) < o(y) \implies o(x) \leq o(f^{-1}(o(y)))$
- Instantiated for many functions

COLIBRI: Floating Point

- Precise domain propagation:
  $x \oplus y = 0.05 \implies x, y \in [-0.1259..; 0.175....]$
  0.05: $0x3fa999999999999a$
- Distance graph on floating-point numbers
- Monotonic functions:
  $o(f(x)) < o(y) \implies o(x) \leq o(f^{-1}(o(y)))$
- Instantiated for many functions
- Linearization of constraints for simplex

Interesting and Simple Real Examples

```
/*@ requires 0 ≤ x ≤ 1000;
    requires 0 ≤ y ≤ 1000;
    ensures  0 ≤ \result ≤ 1;   @*/
double x_normalisation(double x,double y){

  return x/sqrt(x*x + y*y);

}
```

COLIBRI: Example of Reasoning

$$0 \leq x, y \leq 1000 \implies \sqrt{x^2 \oplus y^2} \geq x \ ?$$

COLIBRI: Example of Reasoning

$$0 \leq x, y \leq 1000 \implies \sqrt{x^2 \oplus y^2} \geq x \ ?$$

$$\mathrm{o}\left(\sqrt{\mathrm{o}(x^2) + \mathrm{o}(y^2)}\right) < x$$

$$\mathrm{o}(x^2) + \mathrm{o}(y^2) \leq \mathrm{o}(x^2)$$

$$\mathrm{o}(x^2) + \mathrm{o}(y^2) = \mathrm{o}(x^2)$$

$$\mathrm{o}\left(\sqrt{\mathrm{o}(x^2)}\right) < x$$

$x < x$ if $\mathrm{o}(x^2)$ is normalized

$\mathrm{o}(x^2)$ is denormalized

x the minimum of the remaining values is a solution

COLIBRI: Example of Reasoning

$$0 \leq x, y \leq 1000 \implies \sqrt{x^2 \oplus y^2} \geq x \ ?$$

$$\mathrm{o}\left(\sqrt{\mathrm{o}(x^2) + \mathrm{o}(y^2)}\right) < x$$

$$\mathrm{o}(x^2) + \mathrm{o}(y^2) \leq \mathrm{o}(x^2)$$

$$\mathrm{o}(x^2) + \mathrm{o}(y^2) = \mathrm{o}(x^2)$$

$$\mathrm{o}\left(\sqrt{\mathrm{o}(x^2)}\right) < x$$

$x < x$ if $\mathrm{o}(x^2)$ is normalized

$\mathrm{o}(x^2)$ is denormalized

x the minimum of the remaining values is a solution

There is a counter-example!

Interesting and Simple Real Examples: Corrected

```
/*@ requires 0.0001 ≤ x ≤ 1000;
    requires 0.0001 ≤ y ≤ 1000;
    ensures  0 ≤ \result ≤ 1;   @*/
double x_normalisation(double x,double y){

  return x/sqrt(x*x + y*y);

}
```

Other Examples: From SPARK User Rule

```
procedure User_Rule_7 (X, Y, Z, A : Float;
                       Res        : out Boolean)
 is
 begin
    pragma Assume (Z ≥ 0.0);
    pragma Assume (X ≥ Y);
    pragma Assume (Y ≥ Z);
    pragma Assume (X > Z);
    pragma Assume (A ≥ 1.0);
    Res := (X − Y) / (X − Z) ≤ A;
    pragma Assert (Res);      -- valid
 end User_Rule_7;
```

Other Examples: From SPARK User Rule

$$A \leq \frac{X \ominus Y}{X \ominus Z} \leq B \quad \text{with } \ldots$$

$$\sqrt{X^2 \ominus Y^2} \leq X \quad \text{with } \ldots$$

$$\frac{X}{\sqrt{X^2 \oplus Y^2}} \leq 1 \quad \text{with } \ldots$$

Linearization [Belaid2012]

For t a normal positive number with double precision:

$$o(t)$$

Linearization [Belaid2012]

For t a normal positive number with double precision:

$$(1 - \frac{1}{2^{52}-1}) \cdot t \leq \mathrm{o}(t) \leq (1 + \frac{1}{2^{52}+1}) \cdot t.$$

Linearization [Belaid2012]

For t a normal positive number with double precision:

$$(1 - \frac{1}{2^{52} - 1}) \cdot t \leq \mathrm{o}(t) \leq (1 + \frac{1}{2^{52} + 1}) \cdot t.$$

$$(0. \leq_f x \leq_f 10.0) \wedge (0. \leq_f y \leq_f 10.0) \Rightarrow$$
$$((x \oplus y) \ominus x) \ominus y \leq_f 0.0001$$

Linearization [Belaid2012]

For t a normal positive number with double precision:

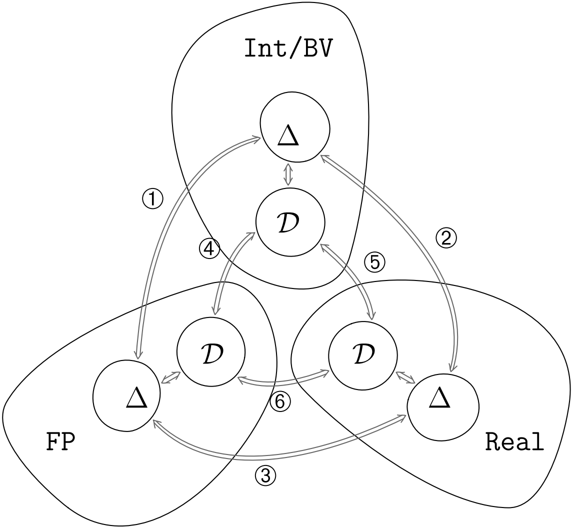
$$(1 - \frac{1}{2^{52} - 1}) \cdot t \leq \mathsf{o}(t) \leq (1 + \frac{1}{2^{52} + 1}) \cdot t.$$

$$(0. \leq_f x \leq_f 10.0) \wedge (0. \leq_f y \leq_f 10.0) \Rightarrow$$
$$\mathsf{o}(\mathsf{o}(\mathsf{o}(x + y) - x) - y) \leq_f 0.0001$$

Bitvector and Integer Arithmetic (CPAIOR17)

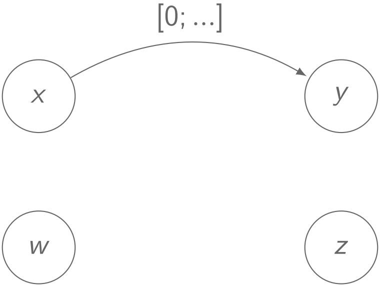
- High-level view of bitvectors
- New propagations for integers $\leftrightarrow$ bitvectors

Interreductions



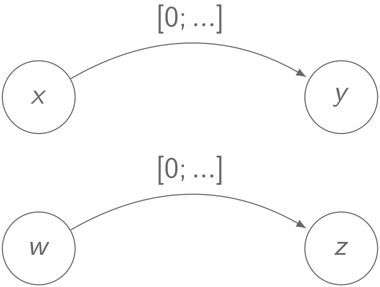
Casts

$$x, y \in [1; 1000],\ \texttt{fp.to\_sbv}\ \_\ x = w,\ \texttt{fp.to\_sbv}\ \_\ y = z$$

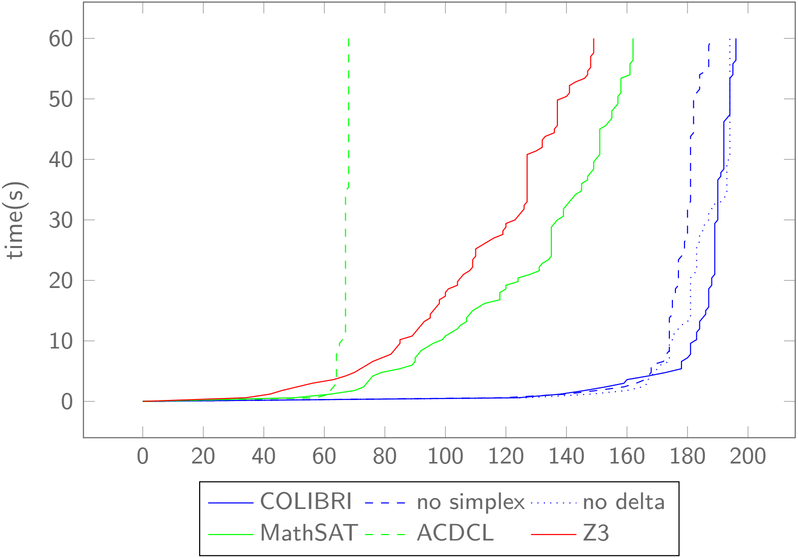


Casts

$$x, y \in [1; 1000],\ \mathtt{fp.to\_sbv}\_x = w,\ \mathtt{fp.to\_sbv}\_y = z$$



Griggio and Schanda



Future Work

- Look at the unsolved benchmarks

Future Work

- Look at the unsolved benchmarks
- More confidence in the propagation and rewrite rules

Future Work

- Look at the unsolved benchmarks
- More confidence in the propagation and rewrite rules
- Uninterpreted functions and quantifiers

Future Work

- Look at the unsolved benchmarks
- More confidence in the propagation and rewrite rules
- Uninterpreted functions and quantifiers
- MCsat

Future Work

- Look at the unsolved benchmarks
- More confidence in the propagation and rewrite rules
- Uninterpreted functions and quantifiers
- MCsat
- Reduce the loading time...

Theorem

Let $D, E \subset \mathcal{R}$, $f : D \mapsto E$ and $f^{-1} : E \mapsto D$ such that

- $\forall x : D,\ f^{-1}(f(x)) = x$
- *f increasing*

We have

- $\forall x \in D, \mathrm{o}(y) \in E, \mathrm{o}(f(x)) < \mathrm{o}(y) \implies \mathrm{o}(x) \leq \mathrm{o}(f^{-1}(\mathrm{o}(y)))$
- $\forall x \in D, y \in E, \mathrm{o}(f(x)) < \mathrm{o}(f(y)) \implies x < y$

Instantiated for many functions in COLIBRI's DBM

Interesting and Simple Real Examples

```
/*@ ensures \result ≤ (double) 1; @*/
double test2(){
  double x = read_sensor();
  /*@ assert (double) 0 ≤ x ≤ (double) 1000; @*/
  double y = read_sensor();
  double z = read_sensor();

  x = x * x + z * z + y * y + 1;

  if (z ≤ y){
    return (x−y)/(x−z);
  } else {
    return (x−z)/(x−y);
  }
}
```

The Problems