

*E0:227, Program Analysis and Verification*

*3:1, August - December 2017*

*<http://www.csa.iisc.ac.in/~raghavan/pav2017/>*

K. V. Raghavan and Deepak D'Souza

## *Software defects*

- Defects are very common, and are a major bane of the software industry.
- Some defects are ongoing irritants, while some are disastrous.
- Defects are **hard to detect and fix**
  - Not enough good and usable tools for programmers
  - Often get detected only after release
  - When a program crashes or gives wrong answer, hard to identify the root cause.

## *Software defects*

- Defects are very common, and are a major bane of the software industry.
- Some defects are ongoing irritants, while some are disastrous.
- Defects are **hard to detect and fix**
  - Not enough good and usable tools for programmers
  - Often get detected only after release
  - When a program crashes or gives wrong answer, hard to identify the root cause.

Testing, finding and fixing bugs (i.e., Quality Assurance) consumes 50% of total cost and time of software development.

## *Kinds of defects*

- **Low-level** errors
  - Null pointers, uninitialized values
  - Array index out of bounds, buffer overrun
  - Memory leaks
  - Misuse of pointers and buffers (in languages like C)
  - Unreachable code

## *Kinds of defects*

- **Low-level** errors
  - Null pointers, uninitialized values
  - Array index out of bounds, buffer overrun
  - Memory leaks
  - Misuse of pointers and buffers (in languages like C)
  - Unreachable code
- High-level errors
  - Does not satisfy user's requirements
  - Algorithmic/design errors
  - Does not interact with other software or libraries correctly
  - Performs poorly



## *Ariane 5 explosion report*

On 4 June 1996, the maiden flight of the Ariane 5 launcher ended in a failure about 40 seconds after launch...

The failure of the Ariane 501 was caused by the complete loss of guidance and attitude information ... This loss of information was due to specification and design errors in the software of the inertial reference system.

The internal SRI\* software exception was caused during execution of a data conversion from 64-bit floating point to 16-bit signed integer value. The floating point number which was converted had a value greater than what could be represented by a 16-bit signed integer.

## *A common approach to software validation: Testing*

- A test suite (set of test cases) is created, and executed for each version.
- **Black box** testing: Test cases are created manually by user, or generated randomly.
- **White box** testing: Test cases are generated by an analysis of the program code to increase code coverage.
  - Typically needs tool support.
- What's good about testing? *All bugs found are real bugs.*
- What's bad about testing?



## *A common approach to software validation: Testing*

- A test suite (set of test cases) is created, and executed for each version.
- **Black box** testing: Test cases are created manually by user, or generated randomly.
- **White box** testing: Test cases are generated by an analysis of the program code to increase code coverage.
  - Typically needs tool support.
- What's good about testing? *All bugs found are real bugs.*
- What's bad about testing?

- 100% coverage of the program's behavior is impossible.
- Therefore, cannot find all bugs or prove the absence of bugs.

- Very hard to test the portion inside the “if” statement!

```
input x
if (hash(x) == 10) {
    ...
}
```

## *Program verification*

The algorithmic discovery of properties of a program by inspection of the source text.

– *Manna and Pnueli, "Algorithmic Verification"*

# *Program verification*

The algorithmic discovery of properties of a program by inspection of the source text.

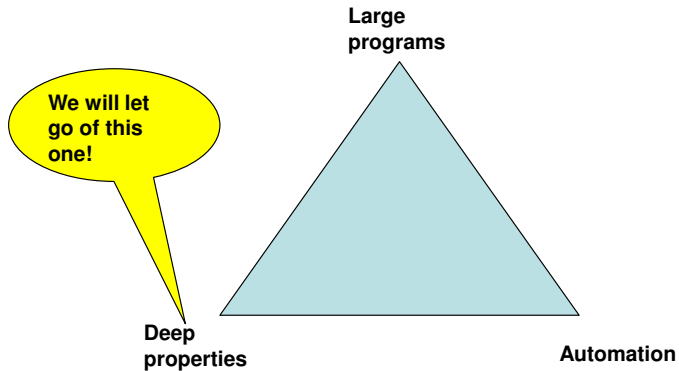
– *Manna and Pnueli, “Algorithmic Verification”*

Also known as: static analysis, static program analysis, formal methods, . . .

## *Difficulty of program verification*

- What will we prove?
  - “Deep” specifications of complex software are as complex as the software itself
  - Are difficult to prove
  - State of the art tools and automation are not good enough
- We will focus on “shallow” properties
  - That is, we will prove “partial correctness”, or absence of certain classes of low-level errors (e.g., null pointer dereferences)

# Elusive triangle



*Example: Determining whether variables are odd (o) or even (e)*

<code>p = oddNatInput()</code>	<code>(p,o)</code>	
<code>q = evenNatInput()</code>	<code>(p,o)</code>	<code>(q,e)</code>
<code>if (p &gt; q)</code>	<code>(p,o)</code>	<code>(q,e)</code>
<code>p = p*2 + q</code>	<code>(p,e)</code>	<code>(q,e)</code>
<code>write(p)</code>	<code>(p,oe)</code>	<code>(q,e)</code>
<code>if (p &lt;= q)</code>	<code>(p,o)</code>	<code>(q,e)</code>
<code>p = p+1</code>	<code>(p,e)</code>	<code>(q,e)</code>
<code>write(p)</code>	<code>(p,e)</code>	<code>(q,e)</code>
<code>q = q+p</code>	<code>(p,e)</code>	<code>(q,e)</code>

## *A verification approach: abstract interpretation*

- A kind of program execution in which variables store *abstract* values from bounded domains, not concrete values
- Input values are also from the abstract domains
- Program statement semantics are modified to work on abstract variable values
- We execute the program on *all* (abstract) inputs and observe the program properties from these runs

## Example: An abstraction

- Abstract value domain  $V_1$  for a single variable:  $\{o, e, oe\}$ .
- Abstract domain:

$$L_1 = Var \rightarrow V_1$$

where  $Var$  is the set of variables in the program.

- Modified operator semantics:

+	<i>o</i>	<i>e</i>	<i>oe</i>
<i>o</i>	<i>e</i>	<i>o</i>	<i>oe</i>
<i>e</i>	<i>o</i>	<i>e</i>	<i>oe</i>
<i>oe</i>	<i>oe</i>	<i>oe</i>	<i>oe</i>

*	<i>o</i>	<i>e</i>	<i>oe</i>
<i>o</i>	<i>o</i>	<i>e</i>	<i>oe</i>
<i>e</i>	<i>e</i>	<i>e</i>	<i>e</i>
<i>oe</i>	<i>oe</i>	<i>e</i>	<i>oe</i>

- From the operator semantics, we can construct an abstract transfer function, with signature  $L_1 \rightarrow L_1$ , for each possible statement in the language.



## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

<pre>p=oddNatInput() q=evenNatInput() if (p &gt; q)   p = p*2 + q write(p) if (p &lt;= q)   p = p+1 write(p) q = q+p</pre>	$\langle (p, oe), (q, oe) \rangle$
--	------------------------------------

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

<pre>p=oddNatInput() q=evenNatInput() if (p &gt; q)   p = p*2 + q write(p) if (p &lt;= q)   p = p+1 write(p) q = q+p</pre>	<pre>&lt;(p,oe), (q,oe)&gt; &lt;(p,o), (q,oe)&gt;</pre>
--	---

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

<pre>p=oddNatInput() q=evenNatInput() if (p &gt; q)   p = p*2 + q write(p) if (p &lt;= q)   p = p+1 write(p) q = q+p</pre>	<pre>&lt;(p,oe), (q,oe)&gt; &lt;(p,o), (q,oe)&gt; &lt;(p,o), (q,e)&gt;</pre>
--	--

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

	$\langle (p, oe), (q, oe) \rangle$
<code>p=oddNatInput()</code>	$\langle (p, o), (q, oe) \rangle$
<code>q=evenNatInput()</code>	$\langle (p, o), (q, e) \rangle$
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$
<code>p = p*2 + q</code>	
<code>write(p)</code>	
<code>if (p &lt;= q)</code>	
<code>p = p+1</code>	
<code>write(p)</code>	
<code>q = q+p</code>	

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

<code>p=oddNatInput()</code>	$\langle (p, oe), (q, oe) \rangle$
<code>q=evenNatInput()</code>	$\langle (p, o), (q, oe) \rangle$
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$
<code>p = p*2 + q</code>	$\langle (p, o), (q, e) \rangle$
<code>write(p)</code>	$\langle (p, e), (q, e) \rangle$
<code>if (p &lt;= q)</code>	
<code>p = p+1</code>	
<code>write(p)</code>	
<code>q = q+p</code>	

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

<code>p=oddNatInput()</code>	$\langle (p, oe), (q, oe) \rangle$	
<code>q=evenNatInput()</code>	$\langle (p, o), (q, oe) \rangle$	
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$	
<code>p = p*2 + q</code>		$\langle (p, e), (q, e) \rangle$
<code>write(p)</code>	$\langle (p, oe), (q, e) \rangle$	
<code>if (p &lt;= q)</code>		
<code>p = p+1</code>		
<code>write(p)</code>		
<code>q = q+p</code>		

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_1$

<code>p=oddNatInput()</code>	$\langle (p, oe), (q, oe) \rangle$	
<code>q=evenNatInput()</code>	$\langle (p, o), (q, oe) \rangle$	
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$	
<code>p = p*2 + q</code>		$\langle (p, e), (q, e) \rangle$
<code>write(p)</code>	$\langle (p, oe), (q, e) \rangle$	
<code>if (p &lt;= q)</code>	$\langle (p, oe), (q, e) \rangle$	
<code>p = p+1</code>		
<code>write(p)</code>		
<code>q = q+p</code>		

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_1$

	$\langle (p, oe), (q, oe) \rangle$	
<code>p=oddNatInput()</code>	$\langle (p, o), (q, oe) \rangle$	
<code>q=evenNatInput()</code>	$\langle (p, o), (q, e) \rangle$	
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$	
<code>p = p*2 + q</code>		$\langle (p, e), (q, e) \rangle$
<code>write(p)</code>	$\langle (p, oe), (q, e) \rangle$	
<code>if (p &lt;= q)</code>	$\langle (p, oe), (q, e) \rangle$	
<code>p = p+1</code>		$\langle (p, oe), (q, e) \rangle$
<code>write(p)</code>		
<code>q = q+p</code>		



## Example: The abstract interpretation

Abstract interpretation, using domain  $L_1$

	$\langle (p, oe), (q, oe) \rangle$	
$p = \text{oddNatInput}()$	$\langle (p, o), (q, oe) \rangle$	
$q = \text{evenNatInput}()$	$\langle (p, o), (q, e) \rangle$	
if ( $p > q$ )	$\langle (p, o), (q, e) \rangle$	
$p = p * 2 + q$		$\langle (p, e), (q, e) \rangle$
write( $p$ )	$\langle (p, oe), (q, e) \rangle$	
if ( $p \leq q$ )	$\langle (p, oe), (q, e) \rangle$	
$p = p + 1$		$\langle (p, oe), (q, e) \rangle$
write( $p$ )	$\langle (p, oe), (q, e) \rangle$	
$q = q + p$		

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_1$

	$\langle (p, oe), (q, oe) \rangle$	
$p = \text{oddNatInput}()$	$\langle (p, o), (q, oe) \rangle$	
$q = \text{evenNatInput}()$	$\langle (p, o), (q, e) \rangle$	
if ( $p > q$ )	$\langle (p, o), (q, e) \rangle$	
$p = p * 2 + q$		$\langle (p, e), (q, e) \rangle$
write( $p$ )	$\langle (p, oe), (q, e) \rangle$	
if ( $p \leq q$ )	$\langle (p, oe), (q, e) \rangle$	
$p = p + 1$		$\langle (p, oe), (q, e) \rangle$
write( $p$ )	$\langle (p, oe), (q, e) \rangle$	
$q = q + p$	$\langle (p, oe), (q, oe) \rangle$	

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_1$

Ideal results

	$\langle (p, oe), (q, oe) \rangle$	
<code>p=oddNatInput()</code>	$\langle (p, o), (q, oe) \rangle$	$(p, o)$
<code>q=evenNatInput()</code>	$\langle (p, o), (q, e) \rangle$	$(p, o) \quad (q, e)$
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$	$(p, o) \quad (q, e)$
<code>p = p*2 + q</code>		$(p, e) \quad (q, e)$
<code>write(p)</code>	$\langle (p, oe), (q, e) \rangle$	$(p, oe) \quad (q, e)$
<code>if (p &lt;= q)</code>	$\langle (p, oe), (q, e) \rangle$	$(p, o) \quad (q, e)$
<code>p = p+1</code>		$(p, e) \quad (q, e)$
<code>write(p)</code>	$\langle (p, oe), (q, e) \rangle$	$(p, e) \quad (q, e)$
<code>q = q+p</code>	$\langle (p, oe), (q, oe) \rangle$	$(p, e) \quad (q, e)$

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_1$

	$\langle (p, oe), (q, oe) \rangle$	
<code>p=oddNatInput()</code>	$\langle (p, o), (q, oe) \rangle$	
<code>q=evenNatInput()</code>	$\langle (p, o), (q, e) \rangle$	
<code>if (p &gt; q)</code>	$\langle (p, o), (q, e) \rangle$	
<code>p = p*2 + q</code>		$\langle (p, e), (q, e) \rangle$
<code>write(p)</code>	$\langle (p, oe), (q, e) \rangle$	
<code>if (p &lt;= q)</code>	<b>X</b> $\langle (p, oe), (q, e) \rangle$	
<code>p = p+1</code>		<b>X</b> $\langle (p, oe), (q, e) \rangle$
<code>write(p)</code>	<b>X</b> $\langle (p, oe), (q, e) \rangle$	
<code>q = q+p</code>	<b>X</b> $\langle (p, oe), (q, oe) \rangle$	

Ideal results

$(p, o)$	
$(p, o)$	$(q, e)$
$(p, o)$	$(q, e)$
$(p, e)$	$(q, e)$
$(p, oe)$	$(q, e)$
$(p, o)$	$(q, e)$
$(p, e)$	$(q, e)$
$(p, e)$	$(q, e)$
$(p, e)$	$(q, e)$

## *Example: Another abstraction*

- Abstract value domain  $V_2$  for a single variable:  $\{o, e\}$ .
- The alternative domain:

$$L_2 = 2^{Var \rightarrow V_2}$$

where  $Var$  is the set of variables in the program.

- Same operator tables as before.
- From the operator semantics, we can construct an abstract transfer function,  $L_2 \rightarrow L_2$ , for each possible statement in the language.

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

<pre>p=oddNatInput() q=evenNatInput() if (p &gt; q)   p = p*2 + q write(p) if (p &lt;= q)   p = p+1 write(p) q = q+p</pre>	<pre>{&lt;(p,o), (q,o)&gt;, &lt;(p,o), (q,e)&gt; &lt;(p,e), (q,o)&gt;, &lt;(p,e), (q,e)&gt;} ,</pre>
--	--

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

<pre>p=oddNatInput() q=evenNatInput() if (p &gt; q)   p = p*2 + q write(p) if (p &lt;= q)   p = p+1 write(p) q = q+p</pre>	<pre>{&lt;(p,o), (q,o)&gt;, &lt;(p,o), (q,e)&gt; &lt;(p,e), (q,o)&gt;, &lt;(p,e), (q,e)&gt;} {&lt;(p,o), (q,o)&gt;, &lt;(p,o), (q,e)&gt;}</pre>
--	---

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	
$p = p*2 + q$	
write( $p$ )	
if ( $p \leq q$ )	
$p = p+1$	
write( $p$ )	
$q = q+p$	



## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	
write( $p$ )	
if ( $p \leq q$ )	
$p = p+1$	
write( $p$ )	
$q = q+p$	

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$
write( $p$ )	
if ( $p \leq q$ )	
$p = p+1$	
write( $p$ )	
$q = q+p$	

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \}$
if ( $p \leq q$ )	
$p = p+1$	
write( $p$ )	
$q = q+p$	

## *Example: The abstract interpretation*

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
if ( $p \leq q$ )	
$p = p+1$	
write( $p$ )	
$q = q+p$	

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
if ( $p \leq q$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = p+1$	
write( $p$ )	
$q = q+p$	

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
if ( $p \leq q$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = p+1$	$\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,e) \rangle \}$
write( $p$ )	
$q = q+p$	

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_2$

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
if ( $p \leq q$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$
$p = p+1$	$\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,e) \rangle \}$
write( $p$ )	$\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,e) \rangle \}$
$q = q+p$	

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_2$

Ideal results

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$	
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$	
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$	$(p,o)$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$	$(p,o) \quad (q,e)$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$	$(p,o) \quad (q,e)$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$	$(p,e) \quad (q,e)$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$	$(p,oe) \quad (q,e)$
if ( $p \leq q$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$	$(p,o) \quad (q,e)$
$p = p+1$	$\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,e) \rangle \}$	$(p,e) \quad (q,e)$
write( $p$ )	$\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,e) \rangle \}$	$(p,e) \quad (q,e)$
$q = q+p$	$\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,o) \rangle \}$	$(p,e) \quad (q,e)$



## Example: The abstract interpretation

Abstract interpretation, using domain  $L_2$

Ideal results

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$	
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$	
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$	$(p,o)$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$	$(p,o) (q,e)$
if ( $p > q$ )	$\{ \langle (p,o), (q,e) \rangle \}$	$(p,o) (q,e)$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$	$(p,e) (q,e)$
write( $p$ )	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$	$(p,oe) (q,e)$
if ( $p \leq q$ )	$\{ \langle (p,o), (q,e) \rangle, \mathbf{X} \langle (p,e), (q,e) \rangle \}$	$(p,o) (q,e)$
$p = p+1$	$\{ \langle (p,e), (q,e) \rangle, \mathbf{X} \langle (p,o), (q,e) \rangle \}$	$(p,e) (q,e)$
write( $p$ )	$\{ \langle (p,e), (q,e) \rangle, \mathbf{X} \langle (p,o), (q,e) \rangle \}$	$(p,e) (q,e)$
$q = q+p$	$\{ \langle (p,e), (q,e) \rangle, \mathbf{X} \langle (p,o), (q,o) \rangle \}$	$(p,e) (q,e)$

## Example: The abstract interpretation

Abstract interpretation, using domain  $L_2$

Ideal results

	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle$	
	$\langle (p,e), (q,o) \rangle, \langle (p,e), (q,e) \rangle \}$	
$p = \text{oddNatInput}()$	$\{ \langle (p,o), (q,o) \rangle, \langle (p,o), (q,e) \rangle \}$	$(p,o)$
$q = \text{evenNatInput}()$	$\{ \langle (p,o), (q,e) \rangle \}$	$(p,o) (q,e)$
if $(p > q)$	$\{ \langle (p,o), (q,e) \rangle \}$	$(p,o) (q,e)$
$p = p*2 + q$	$\{ \langle (p,e), (q,e) \rangle \}$	$(p,e) (q,e)$
write(p)	$\{ \langle (p,o), (q,e) \rangle, \langle (p,e), (q,e) \rangle \}$	$(p,oe) (q,e)$
if $(p \leq q)$	$\{ \langle (p,o), (q,e) \rangle, \mathbf{X} \langle (p,e), (q,e) \rangle \}$	$(p,o) (q,e)$
$p = p+1$	$\{ \langle (p,e), (q,e) \rangle, \mathbf{X} \langle (p,o), (q,e) \rangle \}$	$(p,e) (q,e)$
write(p)	$\{ \langle (p,e), (q,e) \rangle, \mathbf{X} \langle (p,o), (q,e) \rangle \}$	$(p,e) (q,e)$
$q = q+p$	$\{ \langle (p,e), (q,e) \rangle, \mathbf{X} \langle (p,o), (q,o) \rangle \}$	$(p,e) (q,e)$

In comparison to the  $L_1$  domain

- $L_2$  is a *more precise* domain. Result at the end of the program was  $\langle (p,oe), (q,oe) \rangle$  with  $L_1$ , which *over-approximates*  $\{ \langle (p,e), (q,e) \rangle, \langle (p,o), (q,o) \rangle \}$ .
- However,  $L_1$  is *more efficient*.
- Both are less precise than ideal!

## *Abstract interpretation framework*

User provides

- Domain of *abstract values*
  - In example, each abstract value is a set of tuples, where each tuple maps variables to  $\{o, e\}$
- *Transfer functions* for each kind of statement
- *Join operator* ( $\cup$ , in the example)
- A *concretization* function  $\gamma$ , such that  $\gamma(a)$  is a set of *concrete stores* (memory configurations), for each abstract value  $a$ .

In the example:

$$S(e) = \{0, 2, 4, \dots\}, \quad S(o) = \{1, 3, 5, \dots\}$$
$$\gamma(a) = \{ \langle (p, i), (q, j) \rangle \mid \langle (p, m), (q, n) \rangle \in a \wedge i \in S(m) \wedge j \in S(n) \}$$

## *Result of Abstract interpretation*

Dataflow analysis algorithm computes an abstract value  $a_p$  at each program point  $p$ , by

1. computing an abstract value at  $p$  via *each* path to  $p$ , and
2. *joining* these abstract values.

Correctness condition

- For any program point  $p$ ,  $\gamma(a_p)$  is a *superset* of all possible concrete stores that may arise at  $p$  during actual execution
  - In general, computing precise results is undecidable
- Superset property is guaranteed if the user-provided framework satisfies certain *sufficient conditions*

## *Examples of verification problems*

<i>Analysis</i>	<i>Abstract domain</i>
Null-pointer deref.	$Var \rightarrow \{not\text{-}pointer, null, non\text{-}null\} \times 2^{Var}$
Array overruns	$Var \rightarrow IntegerRanges$
File IO	$File\text{-}handles \rightarrow Files, Files \rightarrow \{open, closed\},$
Reachability	Reachability condition
Mutual exclusion	set of locks taken

## *Other applications of program analysis*

- Compilers
  - Live variables analysis
    - Useful, e.g., for register allocation
  - Side-effect analysis of functions
    - Useful, e.g., for code motion
  - Interaction between statements
    - Useful, e.g., for separating sequential code into independent threads
- Code development tools
  - Refactoring; e.g., rename method, extract method
  - Generating code automatically from specifications
  - Automated generation of test cases

## *Overview of PAV course*

- Introduction (1 lecture)
- Lattice theory (2)
- Theory of abstract interpretation (3)
- Implementation of abstract interpretation:
  - Intra-procedural (2)
  - Inter-procedural (9)
- Pointer analysis (2)
- Program slicing (4)
- Type systems (4)
- Assertional reasoning — a first-order predicate logic for proving facts about programs (2)

## *Prerequisites*

- Discrete structures such as sets, relations, partially ordered sets, functions
- (Undergraduate level) algorithms
- Mathematical logic (propositional, first-order)
- General mathematical maturity: comfort with notation, understanding and writing proofs
- Familiarity with imperative languages like C or Java
- Programming project will involve Java (only basic features of Java)



## *What we will not cover*

- Software engineering
  - How to collect requirements from customers and prioritize them
  - Planning and management of software development
  - Design, architecture, coding
  - *in* Formal Methods in Software Engineering, Jan. 2015
- Programming languages
- Analysis of parallel/concurrent programs, distributed systems

## *Assignments and exams (tentative)*

- Four written assignments (40%)
- Programming assignment: 20%
- Exams: mid-sem (15%), final (25%)

## Misconduct policy

- Academic misconduct (e.g., copying) will not be tolerated
- Discussion in exams  $\Rightarrow$  automatic fail grade for both students
- Assignments
  - Work individually.
  - If necessary, you can seek clarifications on basic concepts from other students.
  - However, you must develop the solutions to the given problems on your own (without discussions), and write the answers on your own.
  - **No looking at others solutions, no showing your solution to others!**
  - If you refer to materials other than class lecture notes and text books, mention them.
- Penalties
  - For *each* instance of a violation of above policy  $\Rightarrow$  Zero for the entire assignment, *plus* one grade-point reduction in final grade (for the one who copied).
    - Grade-point reductions over multiple violations will accumulate.
- Grading: Your marks for each assignment will be based on your written answers *and* on a subsequent viva.

## *Late policy for assignments*

- 12 “free” late days for use over all assignments.
- For each late day after free days have been exhausted  $\Rightarrow$  25% penalty on the assignment marks. (Weekends and weekdays treated the same.)
- No late days allowed on final assignment.