# 02245 – PROGRAM VERIFICATION

#### **Christoph Matheja**

(some slides have been developed together with Peter Müller)

#### Fall 2022

### Outline

- 1. Why Program Verification?
- 2. Course Overview
- 3. Course Organization
- 4. Getting Started

#### more confidence

#### Testing is insufficient

1994 Intel® Pentium® Floating-point Division bug



- Estimate: 1 in 9 billion floating-point divisions inaccurate
- Issue: missing entries in the lookup table
- Recall losses: \$475 million (> 5 billion DKK in 2019)
- Bug was detected during experiments on number theory

#### extensive testing

#### no confidence

#### more confidence

OpenJDK's java.utils.Collection.sort() is broken: The good, the bad and the worst case<sup>\*</sup>

Stijn de Gouw<sup>1,2</sup>, Jurriaan Rot<sup>3,1</sup>, Frank S. de Boer<sup>1,3</sup>, Richard Bubel<sup>4</sup>, and Reiner Hähnle<sup>4</sup>

- TimSort: default sorting algorithm in OpenJDK and Android SDK
- Certain large arrays (>= 67M) lead to index-out-of-bounds errors
- Multiple attempts to fix related errors were ineffective

#### extensive testing

no confidence

Program testing can be very effective to show the presence of bugs, but it is hopelessly inadequate for showing their absence.



Edsger W. Dijkstra

#### more confidence



correctness arguments

extensive testing

#### no confidence

The only effective way to raise the confidence level of a program is to give a convincing proof of its correctness.



Edsger W. Dijkstra

PARTITION(A, p, r)

1 
$$x = A[r]$$
  
2  $i = p - 1$   
3 for  $j = p$  to  $r - 1$   
4 if  $A[j] \le x$   
5  $i = i + 1$   
6 exchange  $A[i]$  with  $A[j]$   
7 exchange  $A[i + 1]$  with  $A[r]$   
8 return  $i + 1$ 

At the beginning of each loop iteration:

```
1. If p \le k \le i, then A[k] \le x.
```

**2.** If  $i + 1 \le k \le j - 1$ , then A[k] > x.

3. If k = r, then A[k] = x.

credits: Cormen et al., Introduction to Algorithms, 2009

### Textbook-style correctness arguments are insufficient

- Binary search in java.util.Arrays (2006)
- Faithful implementation of algorithm from Programming Pearls, Bentley, 1986

Is this implementation correct?

```
public static int binarySearch(
    int[] a, int key) {
  int low = 0;
  int high = a.length - 1;
  while (low <= high) {</pre>
    int mid = (low + high) / 2;
    int midVal = a[mid];
    if (midVal < key)</pre>
      low = mid + 1;
    else if (midVal > key)
      high = mid -1;
    else
      return mid; // key found
  return -(low + 1); // key not found
}
```

more confidence

The only effective way to raise the confidence level of a program is to give a convincing proof of its correctness.



Edsger W. Dijkstra

#### Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications

Ion Stoica; Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan<sup>†</sup> MIT Laboratory for Computer Science chord@lcs.mit.edu http://pdos.lcs.mit.edu/chord/

Three features that distinguish Chord from many other peer-topeer lookup protocols are its simplicity, provable correctness, and provable performance. Chord is simple, routing a key through a se-

All 7 claimed invariants turned out to be incorrect!

#### correctness proofs

correctness arguments

extensive testing

no confidence

more confidence

machine-checked proofs

correctness proofs

correctness arguments

extensive testing

no confidence

#### our focus: deductive verification tools



## Interactive verification

- Success stories:
  - CompCert: formally verified C compiler (2008)
  - seL4: formally verified high-performance operating system microkernel (2009)
  - EveryCrypt: formally verified crypto library (2020)
- Strengths:
  - Can handle complex systems and properties
  - Well-established trusted code base
- Weaknesses
  - Requires expert knowledge
  - Very labor-intensive (CompCert: > 6 person years)
  - Possible detachment from production code or vendor lock-in









# Automated (or auto-active) Verification

- Idea: "use verification like compilation"
  - Specifications take the form of source code annotations
  - Analogies: TypeScript, Rust ownership & traits, Python type hints
- Strengths:
  - Substantially less effort than interactive verification
  - Integrates into existing development processes
  - More annotations → more correctness guarantees
- Weaknesses:
  - Less expressive than interactive verification
  - May produce false positives (due to undecidability)
  - Still requires effort and expertise



P\*rust-\*i





### Prusti – a Rust Verifier



#### (live demo)

(more examples in teaser video)

## Outline

- 1. Why Program Verification?
- 2. Course Overview
- 3. Course Organization
- 4. Getting Started

## **Course objectives**



# Architecture of automated program verifiers



- Automated verifiers are often implemented as a tool stack
- Stepwise compilation of programs into logical formulas (and back for error reporting)
- Each transformation deals with one verification problem
- Requirements:
  - reasoning principles
  - verification methodologies
  - engineering practices

### Roadmap



- 1. We learn how to build and use a verification tool for a small programming language
  - Core reasoning principles
  - Generation of proof obligations
  - Working with SMT solvers
  - Error reporting
- 2. We extend the language by advanced features
  - Verification challenges
  - Advanced reasoning and specification principles
  - Automation via encoding to lower levels

### Tentative course outline



### Outline

- 1. Why Program Verification?
- 2. Course Overview
- 3. Course Organization
- 4. Getting Started

### Infrastructure

- Website: <u>http://courses.compute.dtu.dk/02245</u>
  - Course material (slides + webpage) is self-contained; reading references is optional
  - Material will be available at least one day before each lecture

■ 7.5 ETCS course → involves homework

- Classes
  - Lectures: Thursday 13:00 17:00, room B321-H033
  - Question time (for help with material, homework, etc.)
    - Physical: Monday 13:00 14:00, room B321-017
    - Online: Tuesday 18:00 19:00, MS Teams

#### Lectures are meant to be interactive (red slides and boxes)

- Many in-class exercises involve verification tools
  - Make sure to have them at hand when coming to class
  - Typically 5 30 min for each exercise
  - Teamwork is encouraged
- Discuss exercise solutions
- Feel free to ask questions at any time
- Feedback is highly appreciated
  - This is new material, your feedback will improve it  $\ensuremath{\textcircled{\odot}}$

Think about questions in these boxes before the lecture



### Examination

- Completeness and quality of group projects (size: 2-3)
  - 15% Homework: preparation for projects
    - Weekly deadline until project release
    - Solutions will be marked and discussed in class
  - 40% Project A: build a verification tool from scratch
  - 60% Project B: design a new verification methodology
  - Yes, the total is 115% ©
  - Project deadline: November 27, 23:59
  - No reports but submissions must be well-documented and justified
- Individual oral exam
  - Project presentation (ca. 7min, no slides needed)
  - Discussion of projects and course content (ca. 20 min)

### Outline

- 1. Why Program Verification?
- 2. Course Overview
- 3. Course Organization
- 4. Getting Started

### Tentative course outline

#### But first: using a verifier



# The Viper Verification Framework

- Viper language
  - Models verification problems
  - Some statements are not executable
- Two verification backends
  - Carbon (close to what you will build)
  - Silicon
- For now: Programming language with a built-in verifier
- Later: Automate new methodologies



# **Installing Viper**

- Install <u>Java 11+</u> (64-bit)
  - set Java\_HOME and PATH
- Install <u>Visual Studio Code</u> (64-bit)
- In Visual Studio Code:
  - Open the extensions browser (û+Ctrl+X or û+ ℜ+X)
  - Search for Viper
  - Install the extension and restart
- Create and verify the file test.vpr (right)
- Switch to carbon and verify test.vpr again
  - click on silicon (bottom left) to switch



#### Viper methods



### Viper methods



#### Assertions

```
method triple(x: Int, flag: Bool)
  returns (r: Int)
{
    if (flag) {
        r := 3 * x
        assert r > 0
    } else {
        var y: Int
        y := x + x
        r := x + y
        assert r == 3 * x
```

- assert expr tests if expr evaluates to true
  - Yes: no effect
  - No: runtime error
- Testing: no assertion error for *chosen* inputs
- Verification: no assertion error for *all* inputs

Which assertions hold?

#### Postconditions

```
method triple(x: Int) returns (r: Int)
   ensures r == 3 * x
{
 var y: Int
 y := x + x
  r := x + y
}
method client() {
  var z: Int
  z := triple(7)
  assert z == 21
```

 Postconditions specify how returned outputs are related to inputs

- Default: true

#### Postconditions



- Postconditions specify how returned outputs are related to inputs
  - Default: true
- Checked against implementation for all possible parameters
- Guaranteed to hold after method calls for supplied parameters

### **Alternative Implementation**

```
method triple(x: Int) returns (r: Int)
   ensures r == 3 * x
{
                   x = 7
    r := x / 2
                 x = 3
    r := 6 * r
                  x = 18
}
method client() {
  var z: Int
  z := triple(7)
  assert z == 21
}
```

- Some implementations do not work for arbitrary inputs
- A precondition filters out undesirable inputs

#### Preconditions

```
method triple(x: Int) returns (r: Int)
   requires x % 2 == 0
   ensures r == 3 * x
{
 r := x / 2
 r := 6 * r
}
method client() {
  var z: Int
  z := triple(7)
  assert z == 21
}
```

 Preconditions specify on what inputs a method can be called

- Default: true

#### Preconditions



- Preconditions specify on what inputs a method can be called
  - Default: true
- Guaranteed at the beginning of method implementation
- Checked before method calls for supplied parameters

#### Exercise

Write at least two Viper implementations for the method below that verify. Try to find one that does *not* compute the maximum.





#### Contracts

A method contract consist of the method's

- name,
- input and output parameters, and
- pre- and postconditions.

Contracts must be upheld by method calls and implementations.

```
method triple(x: Int) returns (r: Int)
  requires x % 2 == 0
  ensures r == 3 * x
```

```
// implementation
r := x / 2
r := 6 * r
```



{

### Underspecification



- Implementation details are often irrelevant
- Contracts may
  - require more than an implementation needs
  - ensure less than an implementation gives

Give another contract implementation.

### Verifying Method Calls


## **Abstract Methods**

```
method triple(x: Int) returns (r: Int)
  ensures r == 3 * x
```

```
method isqrt(x: Int) returns (r: Int)
    requires x >= 0
    ensures x >= r * r
    ensures x < (r+1) * (r+1)</pre>
```

| <pre>method foo(a: Int)</pre> | returns | (b: | Int) |
|-------------------------------|---------|-----|------|
| <pre>requires a &gt; 0</pre>  |         |     |      |
| ensures b > a                 |         |     |      |
| {                             |         |     |      |
| b := isqrt(a)                 |         |     |      |
| <pre>b := triple(a)</pre>     |         |     |      |
| }                             |         |     |      |

- Contracts without Implementations
  - abstract from hard-to-verify code
  - abstract from unknown implementation
- Verification and good software engineering facilitate each other
  - Incremental development by refinement
  - Contracts become simpler if every method has a *single responsibility*
  - Avoid premature optimizations

#### Exercise

#### Consider the method maxSum with the following signature:

method maxSum(x: Int, y: Int) returns (sum: Int, max: Int)

maxSum is supposed to store the sum of x and y in variable sum and the maximum of x and y in variable max, respectively.

- a) Define a reasonable contract for maxSum.
- b) Implement a method that calls maxSum on 1723 and 42. Test your contract by adding assertions after the call. Improve your contract if any assertion fails.
- c) Implement maxSum.

Now, consider a method reconstructMaxSum that tries to determine the values of maxSum's input parameters from the output parameters, i.e. it reconstructs x and y from sum and max.

- d) Write an abstract method with a postcondition specifying the behaviour of reconstructMaxSum.
- e) Can you give an implementation of reconstructMaxSum? If not, can you implement it after adding a precondition?
- f) Write a client to test your implementation of reconstructMaxSum.

## More abstract methods

```
method unsound(x: Int)
  returns (r: Int)
  ensures r != r
method test() {
  var a: Int
  a := unsound(17)
  assert 2 != 2
}
```

# Wrap-up: Informal Overview

## Tentative course outline



# Outline

1. Why do we need formal foundations?

- 2. Formalizing contracts
- 3. Reasoning about contracts
- 4. Epilogue

# The Program Verification Task



and a **specification spec**,

give a proof

that all program executions

comply with spec



spec: abs(x) returns |x|

Does every execution comply with **spec**?

# Verification must be rooted in rigorous mathematics



# Outline

- 1. Why do we need formal foundations?
- 2. Formalizing contracts
- 3. Reasoning about contracts
- 4. Epilogue

Program states assign values to variables in Var

States = {  $\sigma: V \rightarrow Int | V \subseteq Var and V finite$  }

#### Program semantics describes how states evolve during program execution



**Predicates** capture properties of program states

$$\mathsf{Pred} = \{ P \mid P : \mathsf{States} \rightarrow \mathsf{Bool} \}$$



Set characterization  $P = \{ \sigma \in \text{States} \mid \sigma(\mathbf{x}) \neq 0 \}$ 

Floyd-Hoare triples capture properties of (possibly infinitely many) executions



The triple { *Pre* } S { *Post* } is **valid** if and only if when program S is started in any state in *Pre*, then S terminates in a state in *Post*.



# Which pictures correspond to valid Floyd-Hoare triples?

{ *Pre* } S { *Post* } is **valid** iff when program S is started in any state in *Pre*, then S terminates in a state in *Post*.





Christoph Matheja – 02245 – Program Verification

## Which triples are valid?





# Outline

- 1. Why do we need formal foundations?
- 2. Formalizing contracts
- 3. Reasoning about contracts
- 4. Epilogue

## Reasoning about triples

- Argue as rigorously as possible that the Floyd-Hoare triple described by the following Viper method is valid.
- Hint: annotate the file 03-quintuple.vpr

```
method quintuple(x: Int) returns (r: Int)
    requires x > 0
    ensures r > 4 * x
{
    var y: Int
    y := 2 * x
    var z: Int
    z := 3 * x
    r := y + z
}
```



# How do we systematically prove a triple valid?

- Determine a verification condition VC
  - VC is a predicate
  - VC is **valid** iff it is true for *all* states
- Soundness: VC is valid → triple is valid
- **Completeness:** triple is valid  $\rightarrow$  VC is valid
- Predicate transformers describe how predicates evolve during program execution



## Forward Reasoning



Forward VC: is the strongest postcondition SP(Pre, S) (all final states that we can reach from Pre) of Pre and program S contained in Post?

# Informal Forward Reasoning



## **Backward Reasoning**



Backward VC: is *Pre* included in the weakest precondition *WP*(S, *Post*) (all initial states from which we must terminate in *Post*) of program S and *Post*?

# Informal Backward Reasoning



# PLO: a first programming language

| x is a variable in <b>Var</b>                   | z is a constant in <b>Int</b> |
|---|-------------------------------|
| Arithmetic expressions                          |                               |
| a ::= x   z   a + a   a - a   a / a             | a % a                         |
|   |                               |
| Boolean expressions                             |                               |
| b ::= true   false   a < a   a = a   b && b   b | b   !b                        |
|   |                               |
| Predicates (incomplete)                         |                               |
| P, Q, R ::= b   P && P   P ==> P   exists x ::  | P   <b>forall</b> x :: P      |
|   |                               |
| Statements in PL0                               |                               |
| S ::= var x   x := a   S;S   S [] S   as        | sert P   assume P             |
|   |                               |

# Local variable declarations: var x



 $\{ x == 5 \& \& y > x \}$ 

var x;

 $\{ y > 5 \}$ 



#### Assertions: assert R



*Crashes* if R does not hold in the current state; otherwise, *no effect*.

Sequential composition: S1;S2



# Nondeterministic choice: S1 [] S2



Executes *either* S1 *or* S2.

# Assumptions: assume R

- Verification-specific statement
- Not executable
- Part of trusted code base



Nothing happens if R holds in the

#### Assignment: x := a

{ y > 0 }
x := 17 + y
{ y > 0 && x == 17 + y }

| { y < 23 } | { x + 1 > 42 } |
|------------|----------------|
| x := 23    | x := x + 1     |
| { y < x }  | { x > 42 }     |

{ 
$$x > 42$$
 }  
x := x + 1  
{  $x > 42$  && x == x + 1 }

Assigns the value of a (evaluated in the initial state) to x in the final state.

$$WP(x := a, Q) ::= Q[x / a]$$

E[x / F]: E where every x is replaced by F

#### Assignment: x := a

{ y > 0 }
x := 17 + y
{ y > 0 && x == 17 + y }

| { y < 23 } | $\{ x + 1 > 42 \}$ |
|------------|--------------------|
| x := 23    | x := x + 1         |
| { y < x }  | { x > 42 }         |

{ 
$$x > 42$$
 }  
x := x + 1  
{  $x > 42$  && x == x + 1 }

Assigns the value of a (evaluated in the initial state) to x in the final state.

$$WP(x := a, Q) ::= Q[x / a]$$

E[x / F]: E where every x is replaced by F

#### Assignment: x := a

{ y > 0 } x := 17 + y { y > 0 && x == 17 + y }

| { y < 23 } | $\{ x + 1 > 42 \}$ |
|------------|--------------------|
| x := 23    | x := x + 1         |
| { y < x }  | { x > 42 }         |

{ x > 42 } x := x + 1 { x > 42 && x == x + 1 } Assigns the value of a (evaluated in the initial state) to x in the final state.

$$WP(x := a, Q) ::= Q[x / a]$$

E[x / F]: E where every x is replaced by F

# Proof annotations via overlapping Floyd-Hoare triples



#### Exercise

What is wrong with the following proof?

Exercise

*Left* half of room: use *WP* to check which triples are valid

*Right* half of room: use *SP* to check which triples are valid

$$\left\{ \begin{array}{c} 0 <= x \end{array} \right\} \\ x := x + 1 \\ \left\{ \begin{array}{c} -2 <= x \end{array} \right\} \\ y := 0 \\ \left\{ \begin{array}{c} -10 <= x \end{array} \right\} \end{array} \right\} \left\{ \begin{array}{c} 0 <= x \end{array} \right\} \\ x := x + 1 \\ \left\{ \begin{array}{c} true \end{array} \right\} \\ y := 0 \\ \left\{ \begin{array}{c} -10 <= x \end{array} \right\} \end{array} \right\} \left\{ \begin{array}{c} x == X \& \& y == Y \end{array} \right\} \\ x := x - X; \\ y := y - X; \\ x := x + y \\ \left\{ \begin{array}{c} x == Y \& \& y == X \end{array} \right\} \end{array} \right\}$$

Christoph Matheja – 02245 – Program Verification

# Outline

- 1. Why do we need formal foundations?
- 2. Formalizing contracts
- 3. Reasoning about contracts
- 4. Epilogue

#### Strongest Post vs. Weakest Pre – Does it matter?
## Wrap-up

## Where are we?



- Viper language
- **WP** (our preference)
- SP (used later)
- next lecture