

Secure Information Flow in Java bytecode

FMSS 2019-2020

Java bytecode: a simple instruction set

op	pop two operands off the stack, perform the operation, and push the result onto the stack
gog	discard the top value from the stack
push k	push the constant k onto the stack
load x	push the value of variable x onto the stack
store x	pop off the stack and store the value into x
if j	pop off the stack and jump to j if non-zero
goto j	jump to j
jsr j	at address p, jump to address j and push p+1
	onto the operand stack
ret x	jump to the address stored in x
halt	stop

Standard operational semantics

Constants	V	k, k',
Addresses	A	i, j,
Memories	$Mem = var \rightarrow V \cup A$	m, m',
Stacks	Stack = $(V \cup A)^*$	s, s',

Transition system:

A state consists of: <program counter, memory, operand stack>

State = (A x Mem x Stack) Transition: rule of the standard operational semantics

Standard Operational Semantics

state: <program counter, memory, operand stack>

 $\langle PC, [MEM(x) \ MEM(y)], STACK \rangle$

- 0 load y
- 1 if 4

x: 5

y: 1

- 2 push 1
- 3 goto 5
- 4 push 0
- 5 store x
- 6 halt

$$\begin{array}{l} \langle \mathbf{0}, [5, 1], \lambda \rangle \\ \downarrow \mathbf{load} \\ \langle \mathbf{1}, [5, 1], 1 \rangle \\ \downarrow \mathbf{if_{true}} \\ \langle \mathbf{4}, [5, 1], \lambda \rangle \\ \downarrow \mathbf{push} \\ \langle \mathbf{5}, [5, 1], 0 \rangle \\ \downarrow \mathbf{store} \\ \langle \mathbf{6}, [0, 1], \lambda \rangle \end{array}$$

1 load x
2 store y explicit flow
3 halt

x is loaded onto the stack, then it is stored into y, that is, y depends explicitly on x

1 load x
2 if 5
3 push 1
4 goto 6 implicit flow
5 push 0
6 store y
7 halt

variable x is loaded onto the stack. Depending on the value of x, either the constant 1 or the constant 0 is pushed onto the stack, and successively stored onto y

In both cases observing the final value of y reveals information on the value of x

- 1 load x
- 2 if5
- 3 push1
- 4 goto 6
- 5 push 0
- 6 storey
- 7 halt

Implicit flow starts at [2]

When implicit flow terminates?

- [6] is the first instruction that is common to both branches
- The implicit flow terminates at [6]

[6] is the first instruction that is not under the implicit flow



We use the concept of immediate postdominator on the CFG of the program to handle implicit flows



immediate postdominator of i: the first node belonging to all paths from i

ipd(i) = j

The implicit flow of an if instruction at address i terminates at the instruction with address ipd(i)

What about nested control instructions?



Nested implicit flows

The innest implicit flow if the implicit flow that terminate first

immediate postdominator of i1: the first node belonging to all paths from i

ipd(i1) = j

IPD stack:

when executing an instruction, the ipd stack mantains information on the open implict flows IPD stack is updated any time a control instruction is enetered and any time a

control instruction terminates



Execution of instructions

when an instruction j is executed: if the instruction j is the top of the ipd stack, the stack is updated by executing pop (j is removed from the stack)

before i	Stack of ipd:	λ
i: control instruction	Stack of ipd:	ipd(i)
i1: control instruction	Stack of ipd:	ipd(i1) ipd(i)
j: top of the ipd stack	Stack of ipd:	ipd(i)
j: top of the ipd stack	Stack of ipd:	λ

CONTROL REGION of a branching instruction



Execution of instructions

when an instruction j is executed: if the instruction j is the top of the ipd stack, the stack is updated by executing pop (j is removed from the stack)

before i	Stack of ipd:	λ
i: control instruction	Stack of ipd:	ipd(i)
i1: control instruction	Stack of ipd:	ipd(i1) ipd(i)
j: top of the ipd stack	Stack of ipd:	ipd(i)
j: top of the ipd stack	Stack of ipd:	λ

CONTROL REGION of a branching instruction



Influence of the implicit flow onto the operand stack

the stack may be manipulated in different ways by the branches of a branching instruction: they can perform a different number of pop and push operations, and with a different order.

The length and the content of the operand stack may be a means by which security leakages can occur

- 1 push1
- 2 load x
- 3 if5
- 4 pop
- 5 halt

The stack is empty or not, depending on the value of x

Secure Information flow

- 1 load x
- 2 if5
- 3 push 1
- 4 goto 6
- 5 push 0
- 6 storey
- 7 halt

A program $P = \langle c, H, L \rangle$ satisfies secure information flow if the final value of each low variable does not depend on the initial value of the high variables.

$$H=\{x\} \quad L=\{y\}$$

Termination Agreement H={x} L={}>

- 1 loadx 2 if1
- 2 if1 3 halt

it is not possible to leak high information by observing the termination of the program

Timing Agreement

- 1 loadx
- 2 if5
- 3 push 1
- 4 goto 6
- 5 push 0
- 6 storey
- 7 halt

$H=\{x\} L=\{y\}>$

the number of instructions executed in a computation may reveal information on the value of the high variables

Domains of the concrete semantics

Security levels Constants Addresses Concrete Values Concrete Addresses Concrete Memories Concrete Stacks Environments $\mathcal{L} = \{ L < H \}$ σ, τ, ... k, k', ... V Α i, j, .. $\boldsymbol{\mathcal{V}}^{=} \vee \times \boldsymbol{\mathcal{L}}$ (k, σ) $a = A \times L$ (i, **σ**) $\mathcal{M}^{=}$ var $\rightarrow (\mathcal{V} \cup \mathcal{a})$ M, M', ... $S = (\mathcal{V} \cup \mathcal{A})^*$ S, S', .. $\mathcal{L} = \mathcal{L}$ σ, τ, ..

Concrete Semantics

 $\begin{array}{ll} \text{STATES} \qquad \quad \mathcal{L} \times \mathsf{A} \times \mathcal{M} \times \mathcal{S} \times \ \mathcal{A}^* \\ < \sigma, \ \mathsf{PC}, \ \mathsf{M}, \ \mathsf{S}, \ \rho > \end{array}$

 σ environmentPCprogram counterMmemorySoperand stack (σ 1 σ n) ρ ipd stack (j, σ).....(j', σ ')

IPD Stack ρ if $\rho = (j1, \sigma 1) \dots (jn, \sigma n)$ there are n open implicit flows j1 holds the address where first implicit flow terminates $\sigma 1$ holds the level of the environment that must be restored

```
if \rho = \lambda
there are no open implicit flow
```

c[i] = load x ,
$$M[x] = (k, \tau)$$
, not_top(i, ρ)

load _

$$< \sigma$$
, i, M, S, $\rho > \rightarrow$
 $< \sigma$, i+1, M, (k, $\sigma \cup \tau$) · S, $\rho >$

$$c[i] = store x$$
 , $not_top(i, \rho)$

store.

 $< \sigma$, i, M, (k, τ) \cdot S, $\rho > \rightarrow$ $< \sigma$, i, M[(k, $\sigma \cup \tau$)/x], S, $\rho >$

$$\rho = (i, \tau) \cdot \rho'$$

ipd ____

 $<\sigma$, i, M, S, (i, τ) . ρ '> \rightarrow $<\tau$, i , M, S, ρ '>

i is the ipd of a control instruction

$$c[i] = goto j$$
, $not_top(i, \rho)$

goto _____

 $<\sigma$, i, M, S, ρ > \rightarrow $<\sigma$, j, M, S, ρ >

i is the ipd of a control instruction

$$c[i] = if j$$
, not_top(i, ρ)

if-false _____

<
$$\sigma$$
 , PC, M, (0, τ) · S, ρ > \rightarrow
< $\sigma \cup \tau$, PC+1, up(M),up(S), (σ , ipd(i)) ρ >

An implicit flow begins, whose level is the least upper bound between the security environment (σ) and the security level of the condition of the if (τ). The new security environment is ($\sigma \cup \tau$) (ipd(pc), σ) is pushed on the ipd stack ρ

up(M) upgrades the value of the variables assigned in the scope of the implicit flow beginning at PC

up(S) upgrades all elements in the stack

$$c[i] = if j$$
, $k!=0$, $not_top(i, \rho)$

if-true_

$$< \sigma$$
, i, M, (k, τ) · S, $\rho > \rightarrow$
 $< \sigma \cup \tau$, j, up(M),up(S), (ipd(i), σ). $\rho >$

An implicit flow begins, whose level is the least upper bound between the security environment (σ) and the security level of the condition of the if (τ). The new security environment is ($\sigma \cup \tau$) (ipd(pc), σ) is pushed on the ipd stack ρ

up(M) upgrades the value of the variables assigned in the scope of the implicit flow beginning at PC

up(S) upgrades all elements in the stack

Concrete rules

$$\begin{split} & \begin{array}{l} \operatorname{ipd} \quad \frac{\rho = (i, \tau) \cdot \rho'}{\langle \sigma, i, M, S, \rho \rangle \longrightarrow \langle \tau, i, M, S, \rho' \rangle} \\ & \operatorname{op} \quad \frac{c[i] = \operatorname{op} \quad not_top(i, \rho)}{\langle \sigma, i, M, (k_1, \tau_1) \cdot (k_2, \tau_2) \cdot S, \rho \rangle \longrightarrow \langle \sigma, i + 1, M, (k_1 \ op \ k_2, \tau_1 \sqcup \tau_2) \cdot S, \rho \rangle} \\ & \operatorname{pop} \quad \frac{c[i] = \operatorname{pop} \quad not_top(i, \rho)}{\langle \sigma, i, M, (k, \tau) \cdot S, \rho \rangle \longrightarrow \langle \sigma, i + 1, M, S, \rho \rangle} \\ & \operatorname{push} \quad \frac{c[i] = \operatorname{push} k \quad not_top(i, \rho)}{\langle \sigma, i, M, S, \rho \rangle \longrightarrow \langle \sigma, i + 1, M, (k, \sigma) \cdot S, \rho \rangle} \\ & \operatorname{load} \quad \frac{c[i] = \operatorname{load} x \quad M(x) = (k, \tau) \quad not_top(i, \rho)}{\langle \sigma, i, M, S, \rho \rangle \longrightarrow \langle \sigma, i + 1, M, (k, \tau \sqcup \sigma) \cdot S, \rho \rangle} \\ & \operatorname{store} \quad \frac{c[i] = \operatorname{store} x \quad not_top(i, \rho)}{\langle \sigma, i, M, (k, \tau) \cdot S, \rho \rangle \longrightarrow \langle \sigma, i + 1, M[(k, \tau)/x], S, \rho \rangle} \\ & \operatorname{goto} \quad \frac{c[i] = \operatorname{goto} j \quad not_top(i, \rho)}{\langle \sigma, i, M, S, \rho \rangle \longrightarrow \langle \sigma, j, M, S, \rho \rangle} \\ & \operatorname{if}_{false} \quad \frac{c[i] = \operatorname{if} j \quad not_top(i, \rho)}{\langle \sigma, i, M, (mod^P(F^P(i)), \tau), up_S(S, \tau), (ipd(i), \sigma) \cdot \rho \rangle} \\ & \operatorname{if}_{true} \quad \frac{c[i] = \operatorname{if} j \quad not_top(i, \rho)}{\langle \tau, j, up_M(M, mod^P(F^P(i)), \tau), up_S(S, \tau), (ipd(i), \sigma) \cdot \rho \rangle} \\ \end{array}$$

Abstract semantics

Abstract constants Abstract security levels

Abstract Values Abstract Addresses Abstract Memories Abstract Stacks Abstract Environments Abstract States: $\nabla^{\#} = \{ \bullet \}$ $\mathcal{L}^{\#} = \mathcal{L}$ $\mathcal{V}^{\#} : \nabla^{\#} \times \mathcal{L}^{\#} = \mathcal{L}$ $\mathcal{I}^{\#} = \mathcal{I}$

 $\mathcal{A}^{\#} = \mathcal{A}$ $\mathcal{M}^{\#} : \operatorname{var} \to (\mathcal{L} \cup \mathcal{A})$ $S^{\#} : (\mathcal{L} \cup \mathcal{A})^{*}$ $\mathcal{E}^{\#} = \mathcal{E} = \mathcal{L}$ $\mathcal{L} \times A \times \mathcal{M}^{\#} \times S^{\#} \times (A \cup \{0\})$

Abstract operational semantics

the abstract semantics:

- abstracts concrete values into their security level:
 α (k,σ)=σ
- uses the same rules of the concrete semantics on the abstract domains

Both rules for if are always applied -

if_{true} if_{false}

- A(P) : abstract transition system for P
- finite
- multiple path
- each path of C(P) is correctly abstracted onto a path of A(P)

Results

Theorem 1

```
A program P satisfies SIF if for each state of A(P) such that c[i] = halt, then for each x : L it is:
```

```
M[x] = L (value)
or
M[x]=(i, L) for some i (address)
```

Results

Theorem 2

A program P satisfies TERM if each state of A(P)

< σ , i, M, S, ρ > such that σ = H

does not belong to a cycle.

Results

Theorem 3

A program P satisfies TIME if:

 all paths in A(P) starting from a state satisfying top(S)=H and c[i] = if and ending with a state satisfying PC=ipd(i) have the same length.

Another example: concrete semantics

x:(0,H) y:(1,L) ipd(2) = 5, ipd(6)=10

1	load y
2	if5
3	push 3
4	store x
5	load x
6	if9
7	push 1
8	goto 10
9	push 0
10	store y
11	halt

<ENV, PC, [M(x), M(y)], Stack, IPDstack>

 $\langle L, 1, [(0, H)(1, L)], \lambda, \lambda \rangle$

```
\langle L, 2, [(0, H)(1, L)], (1, L), \lambda \rangle
```

```
\langle L, 5, [(0, H)(1, L)], \lambda, (5, L) \rangle
```

 $\langle L, 5, [(0, H)(1, L)], \lambda, \lambda \rangle$

 $\langle L, \mathbf{6}, [(0, H)(1, L)], (0, H), \lambda \rangle$

if_{false}

load

if_{true}

ipd

load

 $\langle H, 7, [(0, H)(1, L)], \lambda, (10, L) \rangle$

y push

 $\langle H, 8, [(0, H)(1, L)], (1, H), (10, L) \rangle$

 \downarrow goto $\langle H, 10, [(0, H)(1, L)], (1, H), (10, L) \rangle$

🕴 ipd

 $(L, 10, [(0, H)(1, L)], (1, H), \lambda)$

store

 $\langle L, 11, [(0, H)(1, H)], \lambda, \lambda \rangle$

An example: abstract semantics



Correctness of the analysis

In the definition of the abstract semantics, we have applied Abstract Interpretation.

Abstract interpretation is a widely applied method for designing approximate semantics of programs.

P. Cousot, R. Cousot. Abstract interpretation frameworks. Journal of Logic and Computation, 2, 1992

Example: PINcloner malicious app

The app catches the private information of the user's Personal Identification Number (PIN) without directly assigning of it to the public variable clone. It achieves this by using a mask that reveals the value of each bit of the PIN.

The PINcloner (pseudocode)

input : PIN output : clone clone := 0x0000; mask := 0x0001; while(mask > 0) b :=PIN & mask; if (b ! = 0) clone:=clone || mask; mask := mask << 1;

PINcloner malicious app

Initially the mask has all the bits set to 0 except for the least significant bit, which is set to 1: this bit shifts one step to the left after each loop cycle and clones the value of one bit of the PIN during each cycle.

In particular, the direct assignment of the PIN bits to the clone variable is avoided by using a variable b. This last variable is different from 0 if and only if the i-th bit of the PIN and of the mask are both equal to 1: the value of b can be used to set (or not set) the i-th bit of the clone variable.

Once the app has gained the access to the user's private data, the access control mechanism is not able to reveal the illicit flow

J_PINCloner: pseudocode

PIN read from a file and copied into another file

Files:

- PIN: H

- Clone: L

```
public class J_PINcloner{
 final int pinSize = 6;
  FileReader pin_file;
 FileWriter cloned_pin_file;
 public J_PINcloner(){
   pin_file = new FileReader("PIN");
    cloned_pin_file = new FileWriter("clone");
 public static void main(String|| args){
    J_PINcloner p = new J_PINcloner();
   for(int i = 0; i < p.pinSize; i + +){
      int PIN = p.pin_file.read();
     int clone = _clone(PIN);
     p.cloned_pin_file.write(clone);
     p.cloned_pin_file.flush();
 private static int _clone(int PIN){
    intmask = 0x0001:
    int clone = 0x0000;
   while(mask > 0){
     int b = (PIN \& mask);
     if(b! = 0)
        clone = clone | mask;
     mask = mask << 1:
   return clone:
```

J_PINCloner: SIF analysis

```
SIF violated in method public static void main(String) arg
DETAILS :
 SIF Violated on Instruction :
 41 : invokejava.io.FileWriter.write(I)V
DEPENDENCES :
 DEP(41) = 0.18;
 SL[0] = L; SL[18] = L;
 env(41) = L
BEFORE STATE OF 41
  Registers : L, L, L, H, H
  Operand Stack : [H · FL]
SECURITY POLICY :
 PIN = FH
  clone = FL
SECURITY CONTEXT :
  J_PINcloner.pinSize = L
  J_PINcloner.pin_file = FH
  J_PINcloner.cloned_pin_file = FL
  J_PINcloner. < init > ()V.arg0 = L
  J_PINcloner. < init > ()V.CALLER = L
  J_PINcloner._clone(I)I.arg1 = H
  J_PINcloner._clone(I)I.return = H
  J_PINcloner._clone(I)I.CALLER = L
  J_PINcloner.main(...)V.arg1 = L
  J_PINcloner.main(...)V.CALLER = L
```

J_PINCloner: SIF analysis

```
- - - - LIBRARY FUNCTIONS - - - --
  java.lang.Object. < init > ()V.arg0 = L
  java.lang.Object. < init > ()V.CALLER = L
EXECUTION PATH :
  0 : new J_PINcloner
  3 : dup
  4 : invoke J_PINcloner. < init > ()V
  7:astore1
  21 : aload 1
  22 : getfield J_PINcloner.pin_file
  25 : invoke java.io.FileReader.read()I
  28: istore 3
  29: iload 3
  30 : invoke J_PINcloner._clone(I)I
  33: istore 4
 35 : aload 1
  36 : getfield J_PINcloner.cloned_pin_file
 39: iload 4
  41 : invoke java.io.FileWriter.write(I)V
```

FAILED, SIF VIOLATED in J_PINcloner

Exceptions are special events used for signaling errors during the execution of a program.

The rising of an exception is referred as throwing

Every time an exception is thrown, the Java runtime system breaks the standard execution flow of the program and calls the handler that catches and manages the exception

The correct handler for a given exception type can be found searching backwards through the call stack of the method

If no appropriate handler is found the program terminates

Java Exception Hierarchy (incomplete)



A simple example with exceptions

```
1
2 import java.io.IOException;
  public class Hello {
    public static void main(String[] args) {
5
6
    public void stampa(String s){
7
            System.out.println("Outside");
8
9
      try { System.out.println("Inside try");}
10
      catch (ArithmeticException e) {System.out.println("Print catch");}
11
12
     System.out.println("Print ....");
13
14
15
```

```
Compiled from "Hello.java"
  public class Hello {
2
    public Hello();
3
      Code :
4
          0: aload 0
5
          1: invokespecial #1 // Method java/lang/Object."<init >":()V
6
          4: return
7
8
    public static void main(java.lang.String[]);
9
      Code :
10
          0: return
11
12
    public void stampa(java.lang.String);
13
      Code:
14
          0: getstatic #2 //Field_java/lang/System.out:Ljava/io/PrintStream;
15
                        #3 // String Outside
          3: 1dc
16
          5: invokevirtual #4 // Method java/io/ PrintStream.println:(Ljava/lang/String;)V
17
                            #2 // Field java/lang/System.out: Ljava/io/PrintStream;
          8: getstatic
18
                            #5 // String Inside try
         11: 1dc
19
         13: invokevirtual #4 // Method java/io/PrintStream.println:(Ljava/lang/String;)V
20
         16: goto
                            \mathbf{28}
21
         19: astore_2
22
                            #2 // Field_java/lang/System.out:Ljava/io/PrintStream;
         20: getstatic
\mathbf{23}
         23: 1dc
                            #7 // String Print catch
24
         25: invokevirtual #4 // Method java/io/PrintStream.println:(Ljava/lang/String;)V
25
                            #2 // Field java/lang/System.out: Ljava/io/PrintStream;
         28: getstatic
26
         31: 1dc
                            #8 // String Print ...
27
         33: invokevirtual #4 // Method java/io/PrintStream.println:(Ljava/lang/String;)V
28
         36: return
29
30
      Exception table:
31
                  to target type
          from
32
                         19 Class java/lang/ArithmeticException
              8
                    16
33
34
```

From the exception table: instructions from 8 to 16 are executed in a protected way. Moreover, 19 is the first instruction of the exception handler. Instruction 8 is protected and it may throw an exception, that can be captured by an exception handler (executing the code at 19).

Assume that a protected instruction is a control instruction that throws an exception depending on a high condition.

- The handler of the exception must be executed under a security environment that is high
- The execution of the exception handler that captures the exception may reveal information on the value of the high condition, thus causing a leakage of information

SOLUTION:

- 1. The body of the exception handlers can be thought as particular extensions of the methods code
- 2. An extended control flow graph is built the extended control flow graph is defined as the graph obtained from the method graph augmented with edges starting from protected instructions to the first instruction of the protecting handlers.



PINCIONER malicious applet

Let us consider the PINCloner applet, where PIN_FILE file and Clone_File file are the input and the output files, respectively.

PIN_FILE is a private file containing a secret PIN (a sequence of 0/1 characters, for simplicity).

Let us suppose that the PINCloner application can read from the private file. The applet clones the user secret PIN with the exception mechanism.

After every character has read, it will be written in a public file by the handler of the exception.

The PINCloner clones the characters of the PIN_FILE by throwing different kind of exceptions depending on the value read

- NullPointerException
- ArithmeticException

PINCloner bytecode (an excerpt)

```
import java.io.IOException;
2
  public class PinCloner {
3
     public static void main(String[] args) {
5
6
    public void PinCloner() {
7
           try {
8
               int p;
9
                for (...) { // reads a char from PIN_FILE and store it in p;
10
11
                . . . . .
                if (p == 0) { throw new ArithmeticException(); }
12
                else { throw new NullPointerException (); }
13
                }
14
         }
15
       catch (ArithmeticException e) { // write 0 in Clone_File
16
17
       catch (NullPointerException e) {// write 1 in Clone_File
18
19
20
          . . . . .
       }
21
22
```

PINCIoner bytecode extended CFG



PINCloner bytecode extended CFG

- instructions from 0 to 22 are protected by two exception handles starting at instruction 22 and instruction 34, respectively. Instruction 3 is an if with four successors:
 - the natural successors,
 - plus the two entry points of the exception handlers
- The control region of 3 includes the instructions of the exception handlers, and consequently these instructions are executed in a security environment given by the condition of the ifne.
- Since the condition depends on the 0/1 value of PIN character read from a high security file, the implicit flow is high. The handler of the exception, write such value into the low security Clone_File file.

PINCloner malicious applet



PIN_FILE H Clone_file L

- > The analysis starts with L assigned to all the other resources
- The application violates the secure information flow because high security data are written on a public file
- > The leakage is detected by the static analysis