Interpreter

Operational recursive interpreter created in OCaml with operations on strings, parser for reflection and dynamic information control flow through taint analysis.

Index

- Project
- Files
- Environment
- Storage
- Reflect
- Parser
- Taint Analysis
- Credits
- License

Project

Our interpreter is written in Ocaml, characterized with an operational semantic and includes functions, block and procedures. We have a dynamic environment and a static environment. In our dynamic environment we associate identifiers with denotable values, in the static environment we associate a boolean value (true,false) with the type Tainted or Untainted. We use in our environment a static scoping,this means that in each point of the program the environment is identified by the lessical structures of the code. In our dynamic domain each eval, dval, mval and tval type is a tuple: the first element is the value, the second is a boolean. If the last is true, the element is tainted, otherwise it is untainted.

The new environment gives the possibility to make static analysis on the code of the interpreter. We have created a parser in order to include in our semantic the command Reflect. This command takes as input a string that contains a command list. Through the parser the string is evaluated in order to recreate the command list. If the reflect try to modify existing values in the store, our dynamic controll look into the taint of these values. If there are some tainted values an exception will be launched and the program will be terminated.

Files

- /src/: source code of interpreter
 - /mod/: modules of project
 - environment.ml: environment of interpreter to link an identifier to a location in storage
 - storage.ml: storage of interpreter to store a value as an element of a list
 - domain.ml: domains of types, type's conversion and exceptions
 - operations.ml: operations of semantic's interpreter and typechecking
 - o parser.ml: parser for reflect
 - o semantic.ml: semantic of interpreter
 - syntax.ml: syntax of expressions, commands, etc.
- /test/: directory about tests
 - library.ml: library of tests to be executed in test.ml
 - test.ml: execution of tests' implementation from library.ml

Environment

Exception of wrong bind list of bindlist

```
exception WrongBindList
```

New polyformic type about environment, that take as input a string, the reference of environment, and return the type of that because an environment can be of different types.

```
type 't env =
  string -> 't
```

Definition of function that creates new empty environment taking as input a type and returning a new environment of that type.

```
let emptyenv (x) =
  function y -> x
```

Function created to apply a new environment.

```
let applyenv (x,y) =
    x y
```

Bind take as input three parameters to add a new value in environment: if 1 not exists then return e else apply r to the environment.

```
let bind (r,l,e) =
  function lu ->
   if lu = l
      then e
  else applyenv(r,lu)
```

As for bind, bindlist takes three parameters and matches il and el with two empty list returning the new reference; if il and el aren't two empty list, then bind them with relative function with the new reference r; else raise WrongBindList because il and el aren't lists.

Storage

Function created to apply a new storage.

```
type loc = int
```

New polyformic type about storage, that takes as input a loc, the reference of storage, and returns the type of that because a storage can be of different types.

```
type 't store =
  loc -> 't
```

In this snippet of code there are two definitions of function: newloc and initloc. Then I assign to name count the intreference of -1 used in the following two functions: the first assigns to count its value added by +1 and then, with the operator !, it dereferences its to get out the contents; while the second function assigns to count the value of -1.

Definition of function that creates new empty storage taking as input a new reference, executing initloc() and returning a new reference as output.

```
let emptystore (x) =
  initloc ();
  function y -> x
```

Function created to apply a new storage.

```
let applystore (x,y) =
  x y
```

In this function you allocate a new value through <code>newloc()</code> function and assign its location to name

1. Then if 1 already exists you return <code>e</code>, that is initially the new value to allocate casted to a polyformic type, else if it not exists run <code>applystore(r,lu)</code> function with the storage reference and the new loc created as parameters.

In this function if 1 casted as loc, the location passed as parameter, already exists then return e, that is the value to store; else you run applystore(r,lu)

```
let update ((r:'a store),(l:loc),(e:'a)) =
  function lu ->
   if lu = l
      then e
  else applystore(r,lu)
```

Parser

A string will be converted in an expression, command, an expressions' list or commands' list through a recursive analysis, character by character, in order to identify the first terminal and its relative parameters. During each search operation, the string is divided in two parts. The first part to which it is associated the semantic meaning and the remaining substring.

parseExp

Here each terminal including its parameters is reconstructed. parseTerminal returns the terminal (Int, String, Char, Bool, etc.). parseParam returns all the characters between (). With a pattern matching for each terminal, the semantic meaning is associated in order to create each expressions. In paramRemain we have the remaining substring not yet analyzed.

parseTerminal

Returns the terminal from expression and command.

```
e.g.
INPUT: XXX(YYY(CCC(1), ZZZ),3
OUTPUT:
XXX
(YYY(CCC(1), ZZZ),3
```

```
let parseTerminal s =
  let rec extractTerminal stringToExtract terminalName =
    let trimmedStringToExtract = String.trim(stringToExtract) in
    (* if the length of str is 0 -> finish *)
    if String.length(trimmedStringToExtract) = 0
    then
        (* returns the terminal and the remaining string *)
        terminalName, trimmedStringToExtract
    else
        let charToElaborate = String.sub trimmedStringToExtract 0 1 in (*takes the
```

parseParam

Returns the params included in the string.

```
e.g.
INPUT: (YYY(CCC(1), ZZZ),3
OUTPUT:
YYY(CCC(1), ZZZ ,3
```

```
let parseParam stringToParse =
    (* t = str remain, e = str extracted, o = open '('*)
    let rec extractParams t e o =
        if String.length(t) = 0
        then
            e, t (*finish*)
        else
            let c = String.sub t 0 1 in
            let 1 = String.length(t) in
            let r = String.sub t 1 (1-1) in
            match c with
            (*if c=')' and o=1 means that I have found the all parameter*)
            | ")" -> if o = 1
                then e, r
                (*call recoursively extractParams ^ c*)
                else extractParams r (e ^ c) (o -1)
            | "(" \rightarrow extractParams r (e ^ c) (o +1)
            (*if no '(' opened than return r*)
            |  |  \rightarrow if o = 0
                then extractParams r e o
(*if o =! 0 keep calling extractParams and save c*)
                else extractParams r ( e ^ c) o
    in extractParams stringToParse "" 0;;
```

parselde

Reads the characters inclueded in "" and returns the correspondant ide.

```
e.g. INPUT: "CIAO", XXX(YYY(3))
OUTPUT:
CIAO
,XXX(YYY(3)
```

parseInt

```
Reads the int.
```

```
e.g.
INPUT: (3), CCC(ZZZ(3))
OUTPUT:
3
, CCC(ZZZ(3))
```

parseChar

```
Reads the char.
```

```
e.g.
INPUT: 'c', CCC(ZZZ(3))
OUTPUT:
c
, CCC(ZZZ(3))
```

```
let parseChar stringToParse =
   let fullLenght = String.length stringToParse in
   let findSingleQuote = String.index stringToParse '\'' in
   let remain = String.sub stringToParse (findSingleQuote+3) (fullLenght-findSingleQu
   (*returns the char in the array string at the index of the first ' found *)
   stringToParse.[findSingleQuote+1], remain
```

parseList

Takes a string and returns a substring created with the first list found and all the remaining substring.

```
let parseList stringToParse =
  let rec extractList t e o =
    if String.length t = 0
    then
       e, t
    else
    let c = String.sub t 0 1 in
```

```
let l = String.length(t) in
       let r = String.sub t 1 (1-1) in
       match c with
        | "]" -> if o = 1
           then e, r
           else extractList r (e ^ c) (o -1)
        | "[" ->
           if o = 0
           then extractList r \in (o +1)
           else extractList r (e ^ c) (o +1)
        | ";" -> extractList r e o
        (*if there's no more [ returns r ed e*)
           then extractList r e o
           else extractList r ( e ^ c) o
in extractList stringToParse "" 0;;
```

parseldeList

In the function we need to be able to identify a list of ide. This function takes a string and create that list of ide.

```
let parseIdeList stringToParse =
  let idelist, remain = parseList stringToParse in
  let rec parseIdes (toExtract : string) (extracted : string list) =
     let parsed, remain = parseIde toExtract in
     if parsed = ""
     then extracted
     else parseIdes remain (parsed :: extracted)
  in parseIdes idelist [], remain
```

parseBool

Identify the taint in the primitive value (Int, String, Char, Bool, etc).

```
let rec parseBool stringToParse =
    let trimmedString = String.trim(stringToParse) in
    if (String.length trimmedString) < 4
        then failwith ("cannot parse bool from: " ^ stringToParse)
    else if trimmedString.[0] = ','
        then parseBool (String.sub trimmedString 1 (String.length(trimmedString)-1
    else if trimmedString.[0] = '\n'
        then parseBool (String.sub trimmedString 1 (String.length(trimmedString)-1
    else
        let tt = String.sub trimmedString 0 4 in
              if (tt="true") then let rr = String.sub trimmedString 4 (String.length
        else
        let ff = String.sub trimmedString 0 5 in</pre>
```

```
if (ff="false") then let rr = String.sub trimmedString 4 (String.l
else failwith ("cannot find true or false")
```

splitListElement

From a string that should represent a list, extracts the elements. Needed for the Block.

```
let splitListElement stringToParse =
   let rec extractParamList t e o rlist =
           if String.length t = 0
(* finish but I need to check if I need to return the list with the element appended o
            then if String.length e = 0 then rlist else e :: rlist
             else
               let c = String.sub t 0 1 in
               let 1 = String.length(t) in
               let r = String.sub t 1 (1-1) in
            match c with
            | ";" ->
            if o = 0
(* if I have found ; and the o = 0 means that I'm inside the list between to parameter
If the '()' are opened, means that I'm inside one parameter inside one list of lists.
case 1:[(.....); (.....)], ......
case 2:[([..;..;]), .....*)
               then if String.length e = 0
(*if there is no story to save, skip the char*)
                   then extractParamList r "" o (rlist) (* skip *)
                   else extractParamList r "" o (e :: rlist)
               else extractParamList r (e^c) o rlist (* element not complete, save th
            "(" -> extractParamList r (e^c) (o+1) rlist
            | ")" -> extractParamList r (e^c) (o-1) rlist
            _ -> extractParamList r (e^c) o rlist
in extractParamList stringToParse "" 0 [];;
```

parseExpList

Returns a list of string and the remaining string to be evaluated.

```
and parseExpList stringToParse =
  let explist, remainExplist = parseList stringToParse in
  let rec parseExps (toExtract : string) (extracted : exp list) =
    if String.trim toExtract = ""
    then
       extracted
  else
    let parsed, remain = parseExp toExtract in
       parseExps remain ( parsed :: extracted)
```

```
in parseExps explist [], remainExplist
```

parseStringIdexExp

From a string extract ide*exp. Needed in the function below parseldexExpList.

```
let parseStringIdexExp stringToParse =
  let ide, remain0 = parseIde stringToParse in
  let exp0, remain1 = parseExp remain0 in
  ide, exp0;;
```

parseldexExpList

From a list of string extracts (ide*exp) list.

```
let parseIdexExpList (listideexp:string list) =
   List.rev_map (fun x -> parseStringIdexExp x) listideexp;;
```

parseCom

As parseExp above, the string that contains a command or a list of commands is evaluated and the commands reconstructed.

parseComList

Returns a list of commands and the remaining string to be evaluated.

```
and parseComList stringToParse =
  let comlist,remain = parseList stringToParse in
  let rec parseCommands (toExtract : string) (extracted : com list) =
```

```
if String.trim toExtract = ""
    then
        extracted
    else
        let parsed, remain = parseCom toExtract in
        parseCommands remain (parsed :: extracted)
in parseCommands comlist [], remain
```

Reflect

Reflect is a command that takes as input a string that contains a command list. Through the parser, the string is evaluated in order to recreate the command list.

```
parser.ml:

let rec parseCom s =
    let terminal, terminalRemain = parseTerminal s in
    let params, paramRemain = parseParam terminalRemain in
        match terminal with

semantic.ml:

| Reflect(x) ->
    let comList, str = parseComList x in
        semcl(comList,r,s,t)
```

Look at here below for the details of parseCom.

If the reflect try to modify existing values in the store, as written before, our dynamic controll look into the taint of these values. The original fuction called update in the store has been modified in order to prevent possible code injection. A new function preupdate has been created. If the value is already in the store and the taint of e is true an exception will be launched.

```
and preupdate ((r: 'a store),(1: loc),(e: mval)): ('a store) =
   if (mvaltobool e = false)
        then update (r,1,e)
   else if (mvaltobool (applystore (r,1)) = false)
        then update (r,1,e)
   else raise Untrusted
```

e.g. of reflect tainted:

```
let r = emptyenv(Unbound);;
```

```
let s = emptystore(Undefined);;
let t = emptyenv(Untyped);;

let d = [("x",Newloc(Int(4,false)));("y",Newloc(Int (1,false)))] ;;

let r1, s1 = semdv(d,r,s,t);;

let str = "[Block([],[],
    [While(Not(Equ(Val(Den(\"x\")),Int(0,false))),[Assign(Den(\"y\"),Prod(Val(Den(\"y\")),let s1 = semc(Reflect(str),r1,s1,t);;

val str : string =
    "[Block([],[],\n[While\n(Not\n(Equ\n(\nVal(Den(\"x\")),\nInt(0,false)\n)\n\n),\n[Assiglet s1 = semc(Reflect(str),r1,s1,t)

Exception: Domain.Untrusted.
```

Taint Analysis

An expression, command or a declaration between a tainted and an untainted input, returns an untainted output if the result is indipendent by the tainted input. Otherwise, if output depends by a tainted input, it returns a tainted output.

Each eval, dval, mval and tval type of interpreter is a tuple: the first element is the value, instead the second is a boolean. If the last is true, it is tainted, otherwise it is untainted.

```
type eval ==
    | Eint of int * bool
...
```

Strings

Len()

It get as input a stringand returns its length as an integer. If it's untainted, remains untainted, otherwise tainted.

```
| _ -> failwith("Error: match")
else failwith ("Error: type")
```

Upper()

It get as input a string and returns its updated with all characters in uppercase. If it's untainted, remains untainted, otherwise tainted.

Lower()

It get as input a string s and returns its updated with all characters in lowercase. If it's untainted, remains untainted, otherwise tainted.

Get()

It get as input a string x and an int y to return a character c of x at index y. If string x is untainted, the returning char c will be untainted, otherwise tainted.

Set()

It get as input a string \times , an int y and a char z to return the same string \times updated replacing its char at position y with the new char z. If string \times after Set() will be the same of that before, then its tait remains the same. Otherwise, if \times is changed, then will be made a new string through substring operation: you merge two string, the first and the last part, with in the middle the new char z. In this last case, will be done an OR between taints of string \times and char z to be setted.

Contains()

It get as input a char \times and a string y to return a boolean: true if \times is in y, otherwise false. If string y or char \times is tainted, then the returning bool value will be tainted, otherwise untainted.

Sub()

It get as input a string \times and two int, y and z, to return a new substring of \times that starts at position y and ends in z. If the substring of string \times after Sub() is empty "", then it's untainted, otherwise will be return the same taint of \times .

Concat()

It get as input two string, s1 and s2, to return a new string s3 composed by s1 concatenate with s2. If both string are empty "", then will be return an empty string "" untainted, independently of taints of two string s1 and s2; if one of two string is empty, you will get the other string with its taint; otherwise will be done an OR between taints of s1 and s2.

Operations

Minus()

It get as input an integer and returns it negative, with the same taint of before.

Iszero()

It get as input an integer and returns a boolean: true if it is equal to 0, otherwise false with the same taint of before.

Equ()

It get as inputs two integer and returns a boolean: true if the first is equal to second, otherwise false. If one of two is tainted, it returns tainted, otherwise untainted.

Sum()

It get as inputs two integer and returns the sum between them as result. If the second integer is 0, result is indipendent by its taint so interpreter return the first integer with its taint; otherwise it returns difference between them with OR operation between their taints.

Diff()

It get as inputs two integer and returns the difference between them as result. If the second integer is 0, result is indipendent by its taint so interpreter return the first integer with its taint; otherwise it returns difference between them with OR operation between their taints.

Prod()

It get as inputs two integer and returns the product between them as result. If one of two is 0, interpreter returns a 0 untainted because the result is indipendent by inputs. Instead, if one of two is a 1, the result will be the same of the other input with its taint because 1 doesn't change inputs. Otherwise, it returns their product as result and OR operation between their taints.

Div()

It get as inputs two integer and returns the division between them as result. If division will be made between an integer (!=0) and a 0, interpreter raises an exception: Exception: DivisionByZero. Instead, if 0 will be divided by an integer, it returns always a 0 untainted, because the result is

indipendent by inputs. Another case could be that when dividend is a 1, the interpreter returns the same divider (!=1) with its taint. Otherwise, this operation return division between the two values and the OR operation between their taints; so if one of two is tainted, returns tainted, otherwise untainted.

Logic

Not()

It get as input a boolean and returns its negated. If it's untainted, remains untainted, otherwise tainted.

Or()

It get as inputs two booleans and returns || operation between them as result. If one of two is tainted, returns tainted, otherwise untainted.

And()

It get as inputs two booleans and returns & operation between them as result. If both are tainted, returns tainted, otherwise untainted.

Credits

Andrea Bazerla - VR377556 Valentina Mantelli - VR072986 Lorenzo Bellani - VR??????

License

Copyright © 2017 Andrea Bazerla Released under The MIT License