microsurf

Nicolas Dutly

May 10, 2022

CONTENTS:

tentative ! eventually the package should be published to pypi at release. The repository will probably have to be renamed too.

CHAPTER ONE

REQUIREMENTS

Microsurf has been tested on python 3.9 and python 3.6. It might work on other python version. If you want to install the package in a virtual environment, you will need a tool that allows you to do so:

```
pip3 install virtualenv
```

CHAPTER

TWO

INSTALLING MICROSURF

1. Clone the repository:

```
git clone https://github.com/Jumpst3r/msc-thesis-work.git
cd msc-thesis-work
```

2. Create a virtual environment (optional, highly recommended)

```
virtualenv env
source env/bin/activate
```

3. Install the microsurf package:

```
pip install -e .
```

Note: This installs the framework in *editable* mode, meaning you can edit the source code without having to reinstall it after making changes.

CHAPTER

THREE

QUICKSTART - OPENSSL CAMELLIA-128 EXAMPLE

This page will walk you through the basics of using Microsurf with an applied example, testing for side channel vulnerabilities in OpenSSL's Camellia-128 implementation.

Note: Make sure you *installed* the framework if you wish to follow along.

In the general case, testing a target binary for side channels requires a number of items which are highly dependent on the binary, these include:

- The arguments to be passed to the binary
- Which input is considered to be secret
- How to generate secret inputs

As a user of the `microsurf` library, you have to specify these items. Fortunately, doing so is straightforward. The general workflow is as follows:

1. Create a new `BinaryLoader` instance, this will allow you to configure general settings relating to the target binary:

Most optional arguments can be left as is, we'll dive into more details later on. Let's see how you would use the `BinaryLoader` class if you would like to test the OpenSSL implementation of *Camellia-128*.

Note: A full working example can be found here.

3.1 Emulation root directory (`rootfs` argument)

For dynamic binaries you will have to provide a root directory (the `rootfs` argument), in which the binary will be emulated. The structure of the directory might look as follows:

```
jail-openssl-x8632/
  lib/
    ld-linux.so.2
    libc.so.6
    libcrypto.so.1.1
    libssl.so.1.1
  input.bin
  openssl
```

Note: As a user of the framework, you have to ensure that all required shared objects are present in the correct location. Also create make sure that any input files your binary expects are present. For dynamic x86 binaries you can check which shared objects are expected to be where by running `ldd mybinary.bin`.

Hint: If a particular shared library is not found in the emulation root, microsurf will issue a warning with the name of the concerned library.

3.2 Specifying arguments (`args` argument)

If we want to encrypt a file `input.bin` using Camellia-128, we would run

```
openssl camellia-128-ecb -e -in input.bin -out output.bin -nosalt -K hexdata
```

where `hexdata` would be a 128bit hexadecimal key (for example: `96d496ea1378bf4f6e1f377606013e25`).

In the command above, the data we pass to the `-K` argument is secret. We can signal this to the microsurf framework by replacing the key with an '@' character:

```
opensslArgs = [
        "camellia-128-ecb",
        "-e",
        "-in",
        "input.bin",
        "-out",
        "output.bin",
        "-nosalt",
        "-K",
        "@",
    ]
scd = SCDetector(
        ...
        args=opensslArgs
        ...
    )
```

The '@' will be replaced with the data produced by the `randGen` function. If `isFile` is true, then the framework will assume that the target binary loads the secret from a file. In that case it will replace the '@' with the path to a temporary file, whose content is generated by the `randGen` function.

You can also mark partial arrguments as secret dependent, for example in the mbedTLS aes driver program, the expected arguments are:

```
./crypt_and_hash 0 input.bin output.bin AES-128-ECB SHA512␣
↪hex:5e1defa4a22621eca5ab3ec051feb3a8
```

In that case, the argument list to pass to microsurf would be:

```
args = [
        "0",
        "input.bin",
```

(continues on next page)

(continued from previous page)

```
        "output.bin",
        "AES-128-ECB",
        "SHA512",
        "hex:@",
    ]
```

3.3 Producing secrets (`randGen` argument)

A secret often has to adhere to some specific format in order to be processed by the target binary. Since microsurf cannot guess that, it is the end user's job to specify such a function. In our example, the `-K` flag expects a 128bit key specified as a hex string. Since this is a fairly common requirement, it is already implemented in the `microsurf` framework:

Note: The `randGen` parameter takes a **callable** object. The framework will validate whether it produces sufficiently random output when called.

3.4 Selective tracing (`sharedObjects` argument)

To selectively trace shared objects, a list of names can be passed to the `sharedObjects` argument. Note that this only works for dynamic libraries. For example:

```
sharedObjects = ['libssl', 'libcrypto']
```

will ignore every other shared library (`libc` etc). Only canonical library names are needed, no need to pass the exact file name.

Note: The binary specified in `binPath` will always be traced, no need to pass it to the `sharedObjects` parameter.

For OpenSSL, the BinaryLoader object would look as follows:

```
from microsurf.pipeline.Stages import BinaryLoader
from microsurf.utils.generators import getRandomHexKeyFunction

rootfs = 'path-to-rootfs'
binpath = rootfs + "openssl"

opensslArgs = [
    "camellia-128-ecb",
    "-e",
    "-in",
    "input.bin",
    "-out",
    "output.bin",
    "-nosalt",
    "-K",
    "@",
```

(continues on next page)

(continued from previous page)

```
]
sharedObjects = ['libcrypto'] # only trace libcrypto.so

binLoader = BinaryLoader(
    path=binpath,
    args=opensslArgs,
    rootfs=rootfs,
    rndGen=getRandomHexKeyFunction(128),
    sharedObjects=sharedObjects
)
```

3.5 Specifying Detection Modules

There are currently two detection modules which can be used:

1. The secret dependent memory read detection *module*
2. The secret dependent control flow detection *module*

They both take a number of arguments - through most can be left to default values. For further information consult the pages dedicated to the two modules.

The only required argument is the `binaryLoader`, which is used to pass the previously created `BinaryLoader` object:

```
from microsurf.pipeline.DetectionModules import CFLeakDetector, DataLeakDetector

binLoader = BinaryLoader(...)
data_leak_detection = DataLeakDetector(binaryLoader=binLoader)
cf_leak_detection = CFLeakDetector(binaryLoader=binLoader)
```

3.6 Executing the analysis

Having created our required detection modules, we are now ready to execute the side channel detection pipeline. To do so, we can create a `SCDetector` object and pass along the list of detection modules:

```
from microsurf.pipeline.DetectionModules import CFLeakDetector, DataLeakDetector
from microsurf.pipeline.Stages import BinaryLoader
from microsurf import SCDetector

# Create the binary loader as described before
binLoader = BinaryLoader(...)
data_leak_detection = DataLeakDetector(binaryLoader=binLoader)
cf_leak_detection = CFLeakDetector(binaryLoader=binLoader)

scd = SCDetector(modules=[
        data_leak_detection,
        cf_leak_detection,
    ])
```

(continues on next page)

(continued from previous page)

```
scd.exec()
```

This will search for any data and control flow side channels in the target application.

Note: Per default, the `SCDetector` will execute a quick analysis. Key bit dependency estimation is not performed by default. To

Warning: If you know that your target binary has non-deterministic behavior, pass `deterministic=True` when creating the `BinaryLoader` object. Not doing so might trigger false positives. For more common pitfalls, refer to the *FAQ*

By default a report will be created in the `reports` directory. If not present, it will be created. If you wish to continue working and processing results with python, you can access the underlying pandas dataframe like so:

```
result_dataframe = scd.DF
```

A detailed documentation for all high level modules can be found *here*.

CHAPTER FOUR

SECRET DEPENDENT CONTROL FLOW

This page gives an introduction to secret dependent control flow, assuming no prior technical knowledge.

4.1 Introduction

4.2 Why should I care ?

4.3 The `CFLeakDetector` module

class `microsurf.pipeline.DetectionModules.`**`CFLeakDetector`**(*, *binaryLoader*, *miThreshold=0.2*, *flagVariableHitCount=False*)

The CFLeakDetector class is used to collect traces for analysis of secret dependent conctrol flow.

Args: binaryLoader: A BinaryLoader instance miThreshold: The treshold for which to produce key bit estimates.
Values lower than 0.2 might produce results which do not make any sense (overfitted estimation).

flagVariableHitCount: Include branching instruction which were hit a variable number of times in the report.
Doing so will catch things like secret dependent iteration counts but might also cause false positives. Usually these are caused by a secret dependent branch earlier in the control flow, which causes variable hit rates for subsequent branching instructions. Fixing any secret dependent branching and then running with flagVariableHitCount=True is advised.

4.4 Further References

CHAPTER

FIVE

SECRET DEPENDENT MEMORY ACCESSES

This page gives an introduction to secret dependent memory accesses, assuming no prior technical knowledge.

5.1 Introduction

5.2 Why should I care ?

5.3 The `DataLeakDetector` module

`class microsurf.pipeline.DetectionModules.`**`DataLeakDetector`**`(`*, *binaryLoader*, *miThreshold=0.2*`)`

The DataLeakDetector class is used to collect traces for analysis of secret dependent memory accesses.

Args: binaryLoader: A BinaryLoader instance miThreshold: The treshold for which to produce key bit estimates.
Values lower than 0.2 might produce results which do not make any sense (overfitted estimation).

5.4 Further References

CHAPTER

SIX

ADVANCED USAGE

Note: Make sure you *installed* the framework if you wish to follow along.

In the *Quickstart* section, we went over how to use the microsurf library in an end-to-end manner. This is very useful for the average user, as it only requires two basic steps:

1. Create a `SCDetector` object.
2. Call the `.exec()` function on the created object.

However, from a research perspective, it is sometimes required to be able to have fine-grained control on some aspects of the process. Fortunately, microsurf is a *framework*, and as such, provides you with the tools you need to build your own multi-arch, custom side channel detection pipeline.

To do so, you will still need to create a `SCDetector` object (refer to the *Quickstart* for more information). Once the object is created, you will have access to a number of lower level functions, which are documented below, along with the `SCDetector` constructor:

6.1 SCDetector module

`class microsurf.`**`SCDetector`**`(`*modules*, *itercount=1000*, *addrList=None*`)`

The SCDetector class is used to perform side channel detection analysis.

Parameters

- **`modules`** (`List[Detector]`) – List of detection modules to run.
- **`itercount`** (`int`) – Number of traces per module to collect when estimating key bit dependencies.
- **`addrList`** (`Optional[List[int]]`) – List of addresses for which to perform detailed key bit dependency estimates. If None, no estimates will be performed. If an empty list is passed, estimates will be generated for all leaks. To selectively perform estimates on given leaks, pass a list of runtime addresses as integers. The runtime addresses can be taken from the generated reports (Run first with addrList=None and then run a second time on addresses of interest as found in the report.)

`exec()`

Perform the side channel analysis using the provided modules, saving the results to ‘results’.

6.2 Elf tools module

The microsurf framework also exposes a number of function which allow you to parse ELF files (using pyelftools under the hood). These can be useful to add context to detected side channels, if the binary is compiled with debug symbols (or not stripped).

`microsurf.utils.elf.`**`getCodeSnippet`**`(`*file*, *loc*`)`

Returns a list of source code lines, 3 before and 3 after the given offset for a given ELF file.

Note that only DWARF <= v4 is supported.

Parameters

- **`file`** (`str`) – Path to the ELF file
- **`loc`** (`int`) – Offset value

Return type `List[str]`

Returns source code lines, 3 before and three after

Raises **`FileNotFoundError`** – If the given ELF file does not exist

`microsurf.utils.elf.`**`getfnname`**`(`*file*, *loc*`)`

Returns the symbol of the function for a given PC

Parameters

- **`file`** (`str`) – Path to elf file (relative or absolute)
- **`loc`** (`int`) – PC value

Returns the symbol name of the function for a given PC if the symbols are not stripped, None otherwise.

Raises **`FileNotFoundError`** – If the given file does not exist.

6.3 Traces module

In microsurf, traces are represented as objects:

`class microsurf.pipeline.tracetools.Trace.`**`MemTrace`**`(`*secret*`)`

Represents a single Memory Trace object.

Initialized with a secret, trace items are then added by calling the .add(ip, memaddr) method.

Parameters

- **`secret`** – The secret that was used when the trace
- **`recorded.`** (*`was`*) –

`add``(`*ip*, *memaddr*`)`

Adds an element to the current trace. Note that several target memory addresses can be added to the same PC by calling the function repeatedly.

Parameters

- **`ip`** – The instruction pointer / PC which caused
- **`read`** (*`the memory`*) –
- **`memaddr`** – The target address that was read.

`remove`(*keys*)
Removes a set of PCs from the given trace

Parameters `keys` (`List[int]`) – The set of PCs to remove

`class microsurf.pipeline.tracetools.Trace.`**`MemTraceCollection`**(*traces*, *possibleLeaks=None*)
Creates a Memory trace collection object. The secrets of the individual traces must be random.

Parameters `traces` (`list[`*`MemTrace`*`]`) – List of memory traces

`buildDataFrames`()
Build a dictionary of dataframes, indexed by leak adress. T[leakAddr] = df with rows indexing the executions and columns the addresses accessed. The first column contains the secret.

`class microsurf.pipeline.tracetools.Trace.`**`PCTrace`**(*secret*)
Represents a single Program Counter (PC) Trace object.

Initialized with a secret, trace items are then added by calling the .add(ip) method.

Parameters

- `secret` – The secret that was used when the trace
- `recorded.` (*was*) –

`add`(*range*)
Adds an element to the current trace. Note that several target memory addresses can be added to the same PC by calling the function repeatedly.

Parameters `range` – the start / end PC of the instruction block (as a tuple)

`class microsurf.pipeline.tracetools.Trace.`**`PCTraceCollection`**(*traces*, *possibleLeaks=None*, *flagVariableHitCount=False*)
Creates a PC trace collection object. The secrets of the individual traces must be random.

Parameters `traces` (`list[`*`PCTrace`*`]`) – List of memory traces

todo

CHAPTER

SEVEN

MODULE DOCUMENTATION

This page gives detailed documentation about the different modules.

7.1 General Modules

7.1.1 `BinaryLoader`

`class microsurf.pipeline.Stages.`**`BinaryLoader`**(*path*, *args*, *rootfs*, *rndGen*, *sharedObjects=[]*, *deterministic=False*, *resultDir='results'*)

The BinaryLoader class is used to tell the framwork where to located the target binary, shared libraries and to specify emulation and general execution settings.

Parameters

- **`path`** (`Path`) – The path to the target executable (ELF-linux format, ARM/MIPS/X86/RISCV).
- **`args`** (`List[str]`) – List of arguments to pass, '@' may be used to mark one argument as secret.
- **`rootfs`** (`str`) – The emulation root directory. Has to contain expected shared objects for dynamic binaries.
- **`rndGen`** (`SecretGenerator`) – The function which will be called to generate secret inputs.
- **`sharedObjects`** (`List[str]`) – List of shared objects to trace, defaults to tracing everything.
- **`deterministic`** (`bool`) – Force deterministic execution.
- **`resultDir`** (`str`) – Path to the results directory.

7.1.2 `SCDetector`

`class microsurf.`**`SCDetector`**(*modules*, *itercount=1000*, *addrList=None*)

The SCDetector class is used to perform side channel detection analysis.

Parameters

- **`modules`** (`List[Detector]`) – List of detection modules to run.
- **`itercount`** (`int`) – Number of traces per module to collect when estimating key bit dependencies.

- **`addrList`** (`Optional[List[int]]`) – List of addresses for which to perform detailed key bit dependency estimates. If None, no estimates will be performed. If an empty list is passed, estimates will be generated for all leaks. To selectively perform estimates on given leaks, pass a list of runtime addresses as integers. The runtime addresses can be taken from the generated reports (Run first with addrList=None and then run a second time on addresses of interest as found in the report.)

7.2 Detection Modules

7.2.1 `DataLeakDetector`

`class microsurf.pipeline.DetectionModules.`**`DataLeakDetector`**`(*, `*binaryLoader*, *miThreshold=0.2*`)`

The DataLeakDetector class is used to collect traces for analysis of secret dependent memory accesses.

Args: binaryLoader: A BinaryLoader instance miThreshold: The treshold for which to produce key bit estimates.
Values lower than 0.2 might produce results which do not make any sense (overfitted estimation).

7.2.2 `CFLeakDetector`

`class microsurf.pipeline.DetectionModules.`**`CFLeakDetector`**`(*, `*binaryLoader*, *miThreshold=0.2*, *flagVariableHitCount=False*`)`

The CFLeakDetector class is used to collect traces for analysis of secret dependent conctrol flow.

Args: binaryLoader: A BinaryLoader instance miThreshold: The treshold for which to produce key bit estimates.
Values lower than 0.2 might produce results which do not make any sense (overfitted estimation).

flagVariableHitCount: Include branching instruction which were hit a variable number of times in the report.
Doing so will catch things like secret dependent iteration counts but might also cause false positives. Usually these are caused by a secret dependent branch earlier in the control flow, which causes variable hit rates for subsequent branching instructions. Fixing any secret dependent branching and then running with flagVariableHitCount=True is advised.

7.3 Secret Generators

`class microsurf.pipeline.DetectionModules.`**`DataLeakDetector`**`(*, `*binaryLoader*, *miThreshold=0.2*`)`

The DataLeakDetector class is used to collect traces for analysis of secret dependent memory accesses.

Args: binaryLoader: A BinaryLoader instance miThreshold: The treshold for which to produce key bit estimates.
Values lower than 0.2 might produce results which do not make any sense (overfitted estimation).

CHAPTER

EIGHT

FREQUENTLY ASKED QUESTIONS

This page covers common pitfalls that could be encountered during the usage of the microsurf library.

8.1 Emulation Errors

todo

8.2 Secret Dependent Memory Accesses

todo

8.3 Secret Dependent Control Flow

todo

PYTHON MODULE INDEX

INDEX