
# UNIVERSITÀ DEGLI STUDI DI CATANIA

Dipartimento di Matematica e Informatica

Corso di Laurea Triennale in Informatica

*Leonardo Cantarella*


# Benchmarking Stateful Fuzzers over Lighttpd

Final Project Report


Supervisior: Giampalolo Bella

Advisor: Marcello Maugeri

External Advisor: Cristian Daniele



Academic Year 2023 - 2024

# Abstract

This thesis compare three stateful fuzzers—Fallaway, AFLNet, and ChatAFL—by employing them against Lighttpd, a high-performance web server. The significance of stateful fuzzing lies in its ability to identify vulnerabilities in applications characterized by intricate internal states, which may be overlooked by conventional fuzzing techniques. This research compares these instruments with regard to code coverage, crash detection, and the different kinds of eventually vulnerabilities. These results provide enlightening insights into the strengths and weaknesses of each fuzzer, hence guiding selection and improvements of stateful fuzzing approaches for modern software systems.

# Contents

# 1 Introduction

As the software systems are getting increasingly complex, ensuring their robustness and security has turned out to be a serious challenge. In this scenario, fuzzing has emerged as a powerful technique in the identification of security vulnerabilities and defects, which may remain elusive for traditional testing techniques. Fuzzing involves the generation of random test inputs in order to see how the software reacts to unexpected or malformed input data, looking for problems, such as crashes, unexpected behaviour, or security vulnerabilities.
The evolution of fuzzing methodologies has significantly enhanced their effectiveness for a wide range of applications. Traditional fuzzers typically focus on generating random inputs and observing the software responses. However, for applications that maintain internal states across several interactions, such as web servers or networked applications, this approach can be insufficient. These stateful applications call for more sophisticated fuzzing techniques that take into account the interaction between different states and transitions.
Stateful fuzzing is an advanced approach for solving the problems of applications that rely on state management. Whereas in stateless fuzzing, each input is considered a unique event, stateful fuzzing emulates the flow of activities along with the succeeding changes in the state of an application. It includes the generation of inputs which consider previous interactions and what these have done to the state of the application, hence providing a more realistic and deeper testing process.
An instance used in this thesis is Lighttpd, an open-source web server recognized for its effectiveness and ability to scale. Lighttpd is engineered to manage a substantial number of concurrent connections and accommodates multiple network protocols. Assessing Lighttpd offers a chance to scrutinize stateful fuzzing methodologies.
This thesis focuses on the benchmarking of stateful fuzzers to ascertain their efficiency in coverage and detecting eventually vulnerabilities in Lighttpd. Our evaluation stateful fuzzers are: Fallaway, AFLNet, and ChatAFL. Each of them has a its own approach toward stateful fuzzing. By analyzing the

performance of all these tools, we aim to report on various strengths and weaknesses of each, which gives necessary suggestions for improving stateful fuzzing techniques and enhancing the security of modern software systems. This introductory section creates a base for deep research in stateful fuzzing methodologies and their realization in real systems, setting the base for further chapters examining tests and results.

# 2 Background

## 2.1 Introduction to Fuzzing

Fuzzing, or fuzz testing, is a software testing technique that includes feeding a huge amount of random data into the system, called SUT (System Under Test), to find unprecedented responses and reveal major programming errors, along with key security vulnerabilities. The primary objective of fuzzing is to identify vulnerabilities such as buffer overflows, memory leaks, and other security weaknesses that can be exploited by attackers.
The success of fuzzing is based on its capabilities for automatic test case generation and for focusing its attention on portions of programs that otherwise would not have been tested by other more traditional testing technique.
Indeed, it is particularly effective for applications with complex input grammars, where manual test case creation would be impracticable.

### 2.1.1 Types of Fuzzing Techniques

Fuzzing methodologies vary, and there exist a lot for different applications and purposes:

- **Black-box Fuzzing**: This is a technique of generating inputs without prior knowledge of the internal structure of an application. It is easy to deploy but often less efficient as there is no internal feedback.
- **White-box Fuzzing**: This is one of those techniques that rely heavily on source code intuition, such as control flow and data flow, to provide maximum code coverage with test case generation. The approach often employs some sort of complex static and dynamic analysis methodologies.
- **Grey-box Fuzzing**: It's a strategy that combines the various merits of black-box and white-box fuzzing. It brings in partial knowledge about internal application details, and code coverage feedback guiding

the generation of inputs. It balances simplicity with effectiveness, and an example might be the family of tools called AFL (American Fuzzy Lop).

### 2.1.2 Fuzzing Inputs Generation

There are also different ways to generate inputs for fuzzing:

- **Mutation-based Fuzzing**: This generates new inputs through random mutations of existing inputs (for example modifying bits or bytes of existing test cases). It requires no knowledge about the structure of the inputs but is often a lot weaker compared with other generations for applications requiring highly structured inputs.
- **Generation-based Fuzzing**: This builds the inputs from scratch, based on a formal characterization of the input format, grammar or protocol specification. It has proved quite effective in applications where the inputs have to be complex or systematically structured.

### 2.1.3 Coverage-Guided Fuzzing

**Coverage-guided fuzzing** is a subtype of grey-box fuzzing that leverages code coverage information to drive the generation of test inputs. It aims to explore as many code paths as possible by continuously generating inputs that maximize the coverage.
For example, the American Fuzzy Lop (AFL) fuzzer is a popular coverage-guided fuzzer that uses a feedback loop to guide the generation of new test cases. AFL instruments the binary to track the code coverage during execution and uses this information to guide the mutation of test cases. The fuzzer maintains a queue of test cases and iteratively selects, mutates, and executes them to maximize the code coverage.
In this context is also important to describe the concept of **edge coverage**, that is a metric that measures the number of unique edges traversed by the program during execution. An edge is a transition between two basic blocks in the control flow graph of the program (a basic block is a sequence of instructions not containing any jumps or branches). For example, consider the following code snippet:

```
if (x > 0) {
    y = 1;
} else {
    y = 2;
}
```

In this case, there are two edges 2.1: one from the condition to the true branch and one from the condition to the false branch. The basic blocks are the condition, the true branch, and the false branch.
This edges are used in the **coverage map**, that is a binary compiled with edge coverage information, where each edge is mapped to a bit in the coverage map.

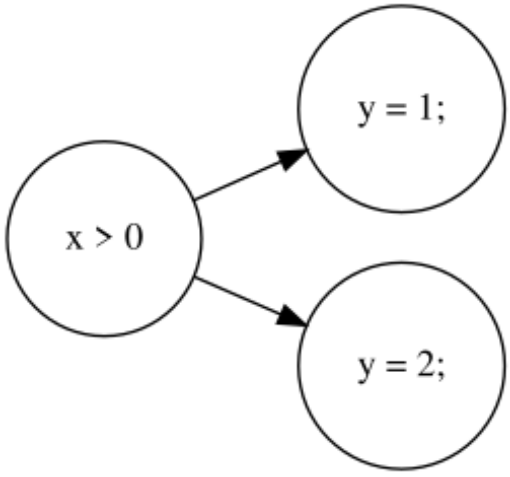


**Figure 2.1:** Example of edge between basic blocks

When an edge is executed, the corresponding bit in the coverage map is set to a probability value. The fuzzer uses this information to guide the generation of new test cases that maximize the coverage.
Within this action a coverage-guided fuzzer maintains a **queue** of inputs and adds new inputs that had allowed it to increase the coverage of the program. In particular, it adds to the queue those inputs that have allowed to increase the coverage of the program, and it uses a scheduler to select which input to mutate next.
The **scheduler** can be based on different strategies, such as the Coverage-Yield (CY) strategy, which uses a round-robin approach to select the next input to mutate, or the Outgoing Edges (OE) strategy, which prioritizes inputs that exercise new edges in the program (this is done by using metadata that keeps track of the outgoing edges of each state in the program and selecting the input that exercises the most new edges).
Talking about queue and inputs, we have to talk about corpus and how it is structured.
The **corpus** is a collection of inputs that the fuzzer uses to seed the initial test cases and to guide the mutation process. The mix of corpus, coverage

map, queue and scheduler is the core of the coverage-guided fuzzing strategy. These are some examples of coverage-guided strategies:

- **Multiple Corpus Single Map (MCSM)**: This strategy uses a single coverage map to track the coverage of the program and multiple queues to store the inputs. Each queue contains inputs that have allowed the fuzzer to reach a specific state of the program. The fuzzer uses a scheduler to select which input to mutate next based on the coverage achieved by the input.
- **Multiple Corpus Multiple Map (MCMM)**: This strategy uses multiple coverage maps to track the coverage of the program and multiple queues to store the inputs. Each queue contains inputs that have allowed the fuzzer to reach a specific state of the program. The fuzzer uses a scheduler to select which input to mutate next based on the coverage achieved by the input.

## 2.2 Stateful Fuzzing: Concepts and Challenges

*Stateless fuzzing* is a traditional fuzzing technique that generates random inputs to test the behavior of an application. However, this approach is not always effective for applications that maintain internal states across multiple interactions.
For example cosidering an FTP server, like LightFTP, until the user is not authenticated, all the inputs are meaningless. In this case, the fuzzer should be able to generate a sequence of inputs that first authenticate the user and then test the behavior of the application. *Stateful fuzzing* adds state awareness to traditional fuzzing methods. It considers an application's internal state and how that state might affect subsequent inputs handling. This becomes particularly critical for applications that handle complex state information, such as network servers, databases, and interactive applications.

### 2.2.1 Understanding Stateful Applications

Stateful applications maintain state across multiple interactions or sessions. Examples include network servers that manage connection states, authentication states, session identifiers, or other state information specific to an application. These states significantly influence the processing of inputs and the behavior of the application over time.

Good practices in state transition management are crucial for both security and reliability: bugs related to state transitions can lead to vulnerabilities such as unauthorized access, denial of service (DoS), or data corruption. Stateful fuzzers attempt to model and explore these state transitions by generating input sequences that mimic valid usage scenarios while concurrently monitoring state changes to ensure comprehensive coverage of all possible transitions. To better understand this concept, consider a simple state model shown in Figure 2.2.

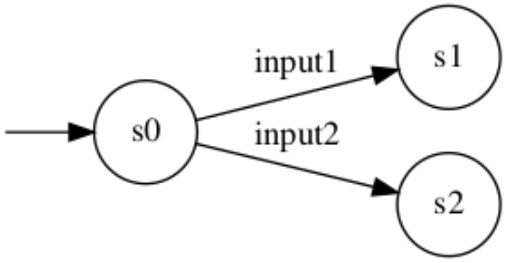


**Figure 2.2:** A simple state model illustrating state transitions in a stateful application

### 2.2.2 Key Techniques in Stateful Fuzzing

Stateful fuzzing involves several advanced techniques that distinguish it from traditional fuzzing approaches:

- **State Modeling**: The process of building a model representing an application state machine by reverse engineering source code, observing real interactions, or guide test case generation in conjunction with machine learning techniques.
- **State Tracking**: This involves tracking the state of the application across successive inputs; indeed, tracking of network traffic, system calls, or internal state variables.
- **Feedback Mechanisms**: With feedback mechanisms, one can prioritize those test cases that tend to explore new states or code paths; hence, the general efficiency of fuzzing can be improved.
- **Sequence Generation**: It is the need to generate input sequences to properly model actual use, since the findings of vulnerabilities often depend on specific sequences or state transitions.

- **Learning-Based Approaches**: Certain fuzzers utilize machine learning or heuristic methodologies to dynamically ascertain the structural configuration of the application's state machine, thereby enabling the fuzzer to adjust and enhance its efficacy progressively.

### 2.2.3 Challenges in Stateful Fuzzing

Successfully performing testing is fraught with several challenges in stateful fuzzing:

- **State Explosion**: As in real life, an application itself may have a number of possible states, and with more states, a risk for exponential growth in process complexity increases. In this case, state abstraction, pruning, or prioritization counters the *state explosion* in an essential way.
- **Protocol Complexity**: Generating meaningful input sequences can involve deep knowledge of complex protocols or state machines. This often includes much domain-specific knowledge or even advanced algorithms.
- **Performance Overhead**: To date, state tracking performed by the application and input sequence generation can cause significant computational costs, hence slowing down the fuzzing process.
- **Handling Non-Deterministic Behavior**: The nondeterministic behavior of stateful applications often results from concurrency, differences in external inputs, or even timing variations. These factors therefore make the reproduction of bugs and receiving consistent fuzzing results usually difficult.

## 2.3 Lighttpd: A Case Study for Stateful Fuzzing

*Lighttpd* is an open-source web server optimized for performance with very low memory usage. It is designed to handle huge volumes of parallel connections with minimal overhead, making it particularly useful on systems with limited resources or those requiring a high degree of concurrency. Its modular design and support for advanced web protocols make it a popular choice for embedded systems, cloud computing platforms, and high-traffic websites. It was used by popular websites like Wikimedia, YouTube, and SourceForge, nowadays is used in general for ...

### 2.3.1 Overview of Lighttpd Architecture

Lighttpd operates on an event-driven architecture, which enables it to serve many requests concurrently. An asynchronous I/O framework is employed to minimize overhead in network connections, allowing the server to scale efficiently under varying workloads. The key features of Lighttpd include:

- **Modular Design**: Provides a series of modules for implementing functions like URL rewriting, HTTP compression, SSL/TLS, and WebSockets. The modular design allows for customization based on specific needs.
- **Protocol Support**: Out of the box, it supports HTTP/1.1, HTTPS, FastCGI, SCGI, and HTTP/2, making it suitable for a wide range of web applications and services.
- **Security Attributes**: Advanced integrated security features include TLS/SSL encryption, prevention of denial-of-service attacks, and multiple authentication options.

### 2.3.2 Relevance of Lighttpd for Fuzzing

Lighttpd is an important SUT for fuzzing due to its common use behind various internet applications. These characteristics make it a suitable candidate for evaluating fuzzing techniques. By default, Lighttpd is not stateful, but it maintains transient states during the processing of requests 2.3.

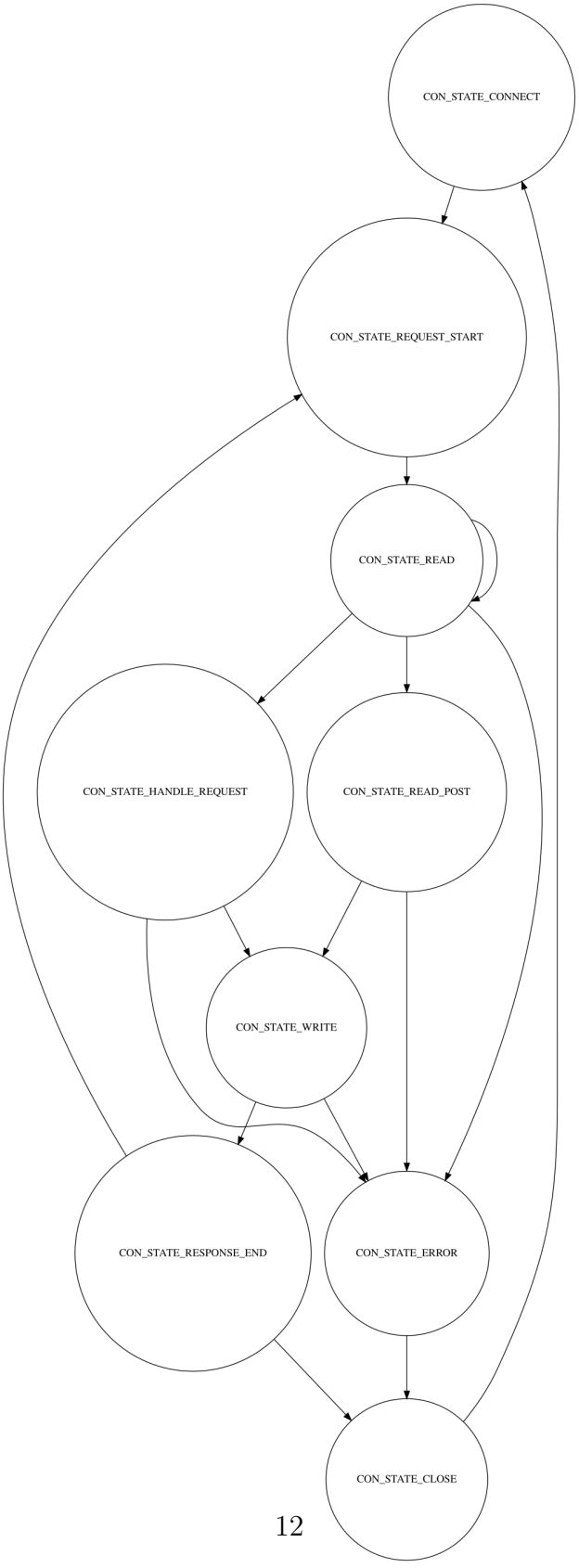




**Figure 2.3:** State model of Lighttpd

The state model seems to be complex, but the state are effectively transitory. A sample flow of states in Lighttpd is as follows:

- **CON_STATE_CONNECT**: The initial default state before a connection is established.
- **CON_STATE_REQUEST_START**: The state right after a connection is established and wait for request.
- **CON_STATE_READ**: The state when the server is reading the request (here the server can be in a loop to read the request).
- **CON_STATE_HANDLE_REQUEST**: The state when the server is processing the request, if body length is null.
- **CON_STATE_READ_POST**: The state when the server is processing the request, if body length is more than 0.
- **CON_STATE_WRITE**: The state when the server is writing the response.
- **CON_STATE_REQUEST_END**: The state after the request has been fully received.
- **CON_STATE_CLOSE**: The state when the connection is closed.

In the *CON_STATE_CONNECT* we have no control of input, because it only changes when the connection is established.
The only stationary state could be the *CON_STATE_READ*, because it is a loop that reads the request (i.e. if we send line by line the request, we will loop into it).
The other states are transitory because, at the end, we will go back to *CON_STATE_REQUEST_START* or *CON_STATE_CONNECT*.
For this thesis, we have chosen to model the state of the server based on the existence or non-existence of a resources 2.4. A resource can be a file, a directory, or any other entity that can be accessed via HTTP. By considering two types of requests—those that attempt to access an existing resource and those that attempt to access a non-existing resource—we can effectively explore different states of the server.

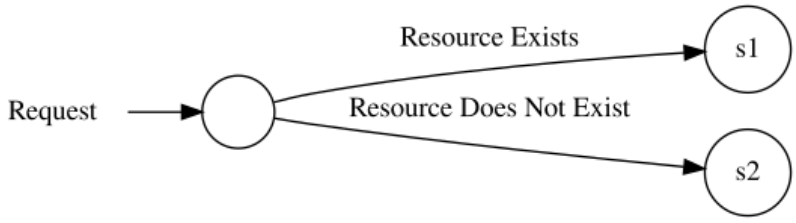


**Figure 2.4:** State model of Lighttpd based on the existence of resources

## 2.4 Fuzzers Overview: AFLNet, ChatAFL, and Fallaway

For this thesis, three stateful fuzzers—AFLNet, ChatAFL, and Fallaway—will be benchmarked over Lighttpd.

### 2.4.1 AFLNet

*AFLNet* [1] is a greybox fuzzer for protocol implementations. Unlike existing protocol fuzzers, it takes a mutational approach and uses statefeedback besides code-coverage feedback in order to guide the fuzzing process. AFLNet is seeded with a corpus of recorded message exchanges between the server and an actual client. No protocol specification or message grammars are required. It acts like a client, replaying variants of the original sequence of messages ever sent to the server, retaining only the variants which managed to increase the coverage of either code or state space. The response codes supplied by the server are utilised by AFLNet for server states that a message sequence manages to exercise. From this feedback, AFLNet identifies progressive regions in the state space and systematically steers towards such regions.

### 2.4.2 ChatAFL

*ChatAFL* [2] is a large language model-guided protocol fuzzer. It is based on AFLNet but integrates three concrete components. First, the fuzzer uses the LLM to extract a machine-readable grammar for a protocol that is used for structure-aware mutation. Second, the fuzzer uses the LLM to increase the diversity of messages in the recorded message sequences that are used as initial seeds. The fuzzer uses the LLM in an attempt to escape a coverage plateau, where LLM is prompted to generate messages to reach new states. The ChatAFL artifact is configured within ProfuzzBench-a well-accepted

benchmark for stateful fuzzing of network protocols. This way, this allows smooth integration with an already established format.

### 2.4.3 Fallaway

*Fallaway* [3] is a stateful, code coverage-based fuzzer and combines techniques from AFL* and AFLnet. It is implemented by extending the library LibAFL, which is a library to build modular fuzzers written in Rust. Its performance gain can partly be explained by the fact that the SUT process is reused for more than one test case. The state-awareness in test cases and feedback seems to have no effect on performance regarding code coverage. Nevertheless, there are a number of limitations in the approach followed by Fallaway which, after their resolution, might further increase the benefit that will be gained from having state-aware test cases and feedback.

# 3 Fallaway Setup

Fallaway distinguishes itself from other fuzzers like AFLNet or ChatAFL by using a **persistent mode**. This mode allows the fuzzer to maintain the server's state across multiple requests, which is especially useful in scenarios where the server does not reset its state between requests, such as when managing user sessions or maintaining authentication states in a web application.
The persistent mode is implemented by modifying the Lighttpd server to maintain its state between requests. The server operates in a separate process, and the fuzzer interacts with it via a socket. The fuzzer sends requests to the server and receives responses, using the results to guide the generation of subsequent requests. This process continues in a loop until the fuzzing session is complete.
To enable this, we must modify the Lighttpd code to ensure that the server continuously receives, processes, and responds to requests without shutting down. The changes are made to the function `server_main_loop` in the `src/server.c` file, and to the connection handling functions in `src/connections.c` of the Lighttpd source code. The specific code changes are shown in the next section, providing a comparison between the original and modified code.

## 3.1 Lighttpd Code Modifications for Persistent Mode

Table 3.1 presents a comparison of the original and modified code for the `connections.c` file. The modifications to this file are crucial for maintaining an open connection state, ensuring that the fuzzer can interact continuously with the server. We also had to clean all buffers and old data for that connection.
Table 3.2 shows a comparison of the original and modified code for the `server.c` file. The changes made here are essential for enabling persis-

tent server operation, allowing the fuzzer to manage and maintain server state across multiple requests, looping into the __AFL_LOOP.

**Original Code**

```
static void connection_handle_shutdown(connection *con) {
    ...
    connection_reset(con);

    /* close the connection */
    if (con->fd >= 0
        && (con->is_ssl_sock
        || 0 == shutdown(con->fd, SHUT_WR))) {
        con->close_timeout_ts = log_monotonic_secs;

        request_st * const r = &con->request;
        connection_set_state(r, CON_STATE_CLOSE);
        if (r->conf.log_state_handling) {
            log_error(r->conf.errh, __FILE__, __LINE__,
                "shutdown_for_fd_%d", con->fd);
        }
    } else {
        connection_close(con);
    }
}
```

**Modified Code**

```
static void connection_handle_shutdown(connection *con) {
    ...
    connection_reset(con);

    /* keep the connection open and reset it */
    request_reset_ex(&con->request);
    chunkqueue_reset(con->read_queue);
    con->request_count = 0;
    con->is_ssl_sock = 0;
    con->revents_err = 0;
    connection_set_state(&con->request,CON_STATE_REQUEST_START);
}
```

**Table 3.1:** Comparison of Original and Modified Code for 'src/connections.c'

**Original Code**

```
static void server_main_loop (server * const srv) {
    ...
    server_load_check(srv);

    #ifndef _MSC_VER
    static
    #endif
    connection * const joblist = log_con_jqueue;
    log_con_jqueue = sentinel;
    server_run_con_queue(joblist, sentinel);

    if (fdevent_poll(srv->ev, log_con_jqueue != sentinel ? 0 : 1000) > 0)
        last_active_ts = log_monotonic_secs;
}
```

**Modified Code**

```
static void server_main_loop (server * const srv) {
    ...
    server_load_check(srv);

    while (__AFL_LOOP(INT64_MAX)) {
        fdevent_poll(srv->ev, -1);

        #ifndef _MSC_VER
        static
        #endif
        connection * const joblist = log_con_jqueue;
        log_con_jqueue = sentinel;
        server_run_con_queue(joblist, sentinel);
    }

    srv_shutdown = 1;
}
```

**Table 3.2:** Comparison of Original and Modified Code for 'src/server.c'

## 3.2 Mutator and Corpus

Another crucial aspect of the fuzzing process involves the corpus and the mutator. In this experiment, we define the state of the server based on the

existence 3.1 or non-existence 3.2 of resources. Specifically, we consider two types of requests: one that attempts to access a resource that exists and another that attempts to access a resource that does not exist.

The **corpus** consists of a set of initial test cases that represent these two scenarios. By including requests for both existing and non-existing resources, we ensure that the fuzzer can effectively explore different states of the server.

```
PUT /hello.txt HTTP/1.1
Host: 127.0.0.1:8080
Content-type: text/plain
Content-length: 13

Hello, World!
```

**Figure 3.1:** Example of existent resource request

```
DELETE /hello.txt HTTP/1.1
Host: 127.0.0.1:8080
User-Agent: curl/8.0.1
Accept: */*
```

**Figure 3.2:** Example of non-existent resource request

The **mutator**, which is responsible for generating variations of the requests, is relatively straightforward. Its primary function is to modify the existing requests by appending a sequence of characters ('$\backslash r \backslash n \backslash r \backslash n$') to the end of each request. This modification is essential as it guarantees that the requests are well-formed and adheres to the HTTP protocol standards. By ensuring the requests are properly formatted, we enable the server to parse and process them correctly, which is vital for accurate fuzz testing.

## 3.3 Setting Up the Fuzzing Environment

To run the fuzzer, we need to build a Docker container that includes all the necessary dependencies and the modified Lighttpd server. The Dockerfile below shows the steps to set up this environment.

```
FROM libaflstar

WORKDIR /

# Copy the patch file
COPY ./lighttpd.patch /lighttpd.patch

ENV DEBIAN_FRONTEND=noninteractive

# Install lighttpd dependencies
RUN apt-get install -y \
    autoconf \
    automake \
    libtool \
    m4 \
    pkg-config \
    libpcre2-dev \
    zlib1g-dev \
    zlib1g \
    openssl \
    libssl-dev \
    scons

# Create the root directory for the server
RUN chmod 777 /tmp

# Install

# Set up environment variables for ASAN
ENV ASAN_OPTIONS='abort_on_error=1:symbolize=0:detect_leaks=0:
   detect_stack_use_after_return=1:detect_container_overflow=0:
   poison_array_cookie=0:malloc_fill_byte=0:max_malloc_fill_size
   =16777216'

# Download lighttpd
ENV CC=afl-cc
ENV CXX=afl-cc
RUN git clone https://git.lighttpd.net/lighttpd/lighttpd1.4.git
   lighttpd
WORKDIR /lighttpd
RUN git checkout 9f38b63cae3e2
RUN git apply /lighttpd.patch
```

```
RUN ./autogen.sh
RUN scons CC=/AFLplusplus/afl-cc CXX=/AFLplusplus/afl-cc -j 4
    build_static=1 build_dynamic=0
RUN mv /lighttpd/sconsbuild/static/build/lighttpd /lighttpd/
    lighttpd

# Copy the corpus
COPY ./corpus /corpus

# Copy the config file
COPY ./lighttpd.conf /lighttpd.conf

# Copy the run script
COPY ./run.sh /LibAFLstar/run.sh
# Make it executable
RUN chmod +x /LibAFLstar/run.sh

WORKDIR /LibAFLstar
```

The Docker container is configured with all the dependencies to run the fuzzer and build the modified Lighttpd server, providing a controlled environment to conduct the fuzzing experiment.

## 3.4 Fuzzing Execution and Results

We ran the fuzzer for 24 hours and a week, using the following script to execute it:

```
#!/bin/bash
bin="${1:-mcsm-cy}"
loops="${2:-1000}"

timeout 24h cargo run --release --bin libaflstar-http-$bin -- --
    in-dir /corpus --out-dir /output_lighttpd --target-port 8080
    --loops $loops -t 300 /lighttpd/lighttpd -D -f /lighttpd.conf
```

In particular we have:

- **bin**: the state scheduler strategy.
- **loops**: taken by the __AFL_LOOP, it is the number of iterations that the fuzzer will do.
- **timeout 24h**: the fuzzer will run for 24 hours (or a week).

The results of the fuzzing process will be discussed in the chapter 5.

# 4 Comparison with AFLNet and ChatAFL

How they works, how they are different, how they are similar, how they are better, how they are worse, etc. Script, dockerfile, etc. Replayer.

# 5 Results

Discussions and comparisions of graphs and results.

# 6 Conclusion

# 7 Future Works

# References


[1] Van-Thuan Pham, Marcel Böhme, and Abhik Roychoudhury. Aflnet: A greybox fuzzer for network protocols. In *Proceedings of the 13rd IEEE International Conference on Software Testing, Verification and Validation : Testing Tools Track*, 2020.

[2] Ruijie Meng, Martin Mirchev, Marcel Böhme, and Abhik Roychoudhury. Large language model guided protocol fuzzing. In *Proceedings of the 31st Annual Network and Distributed System Security Symposium (NDSS)*, 2024.

[3] TODO. Todo. In *TODO*, 2024.