THE UNIVERSITY OF MELBOURNE

EDEFuzz: A Web API Fuzzer for Excessive Data Exposures

Lianglu Pan

Shaanan Cohney

Toby Murray

Thuan Pham

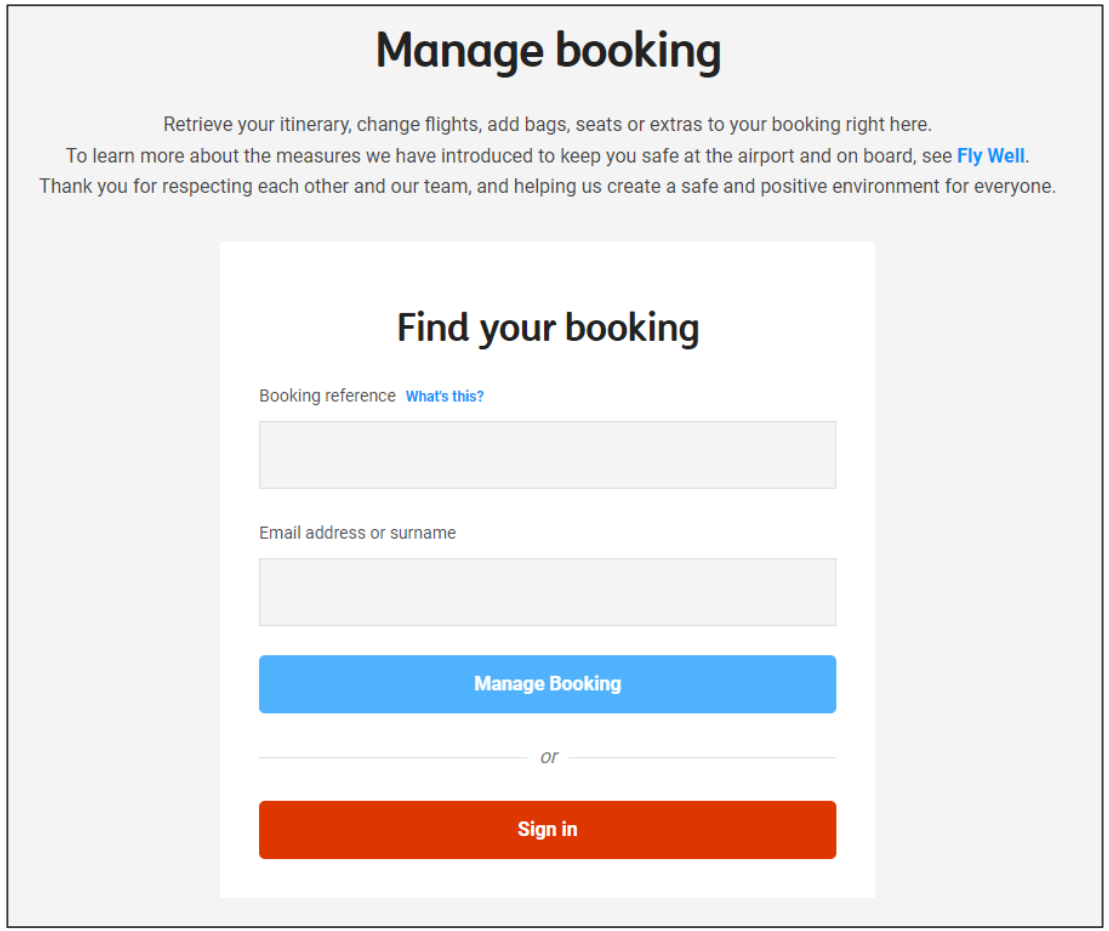
ACM SIGSOFT Distinguished Paper Award

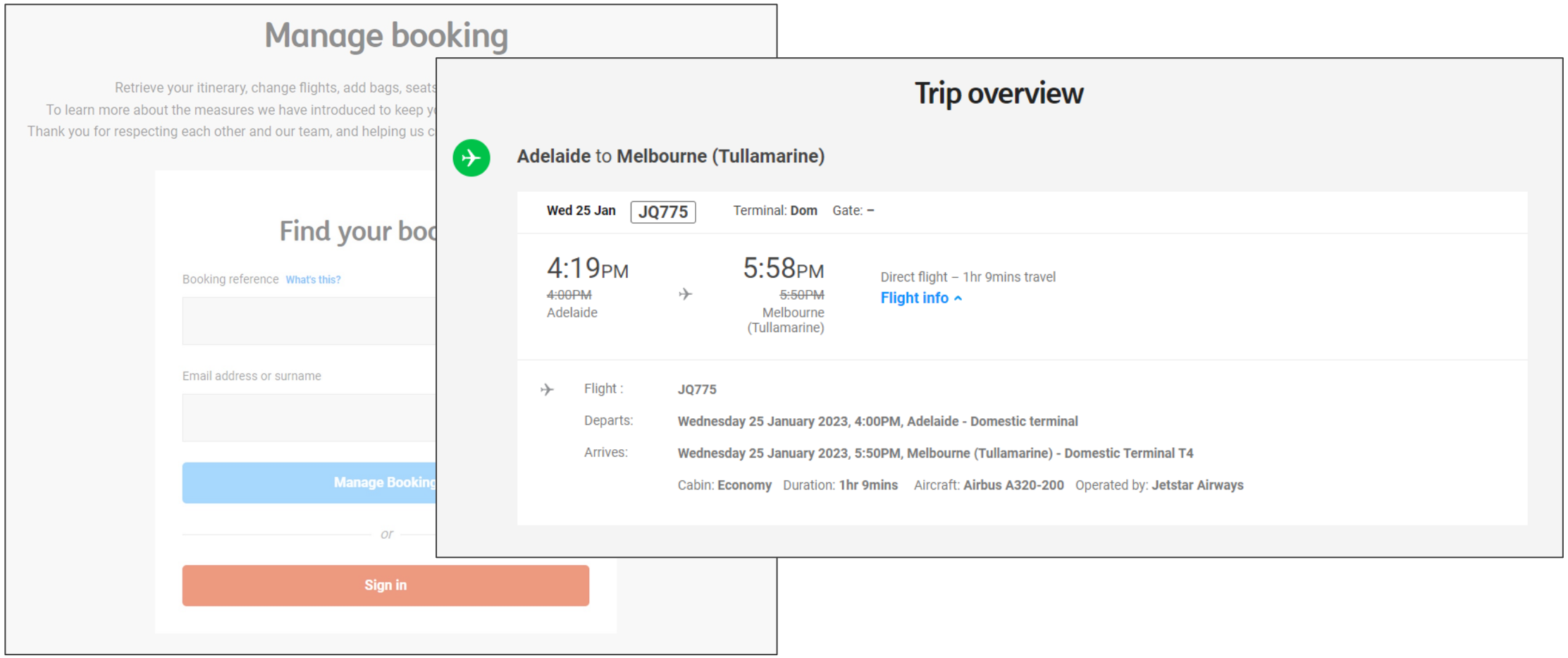
THE UNIVERSITY OF MELBOURNE

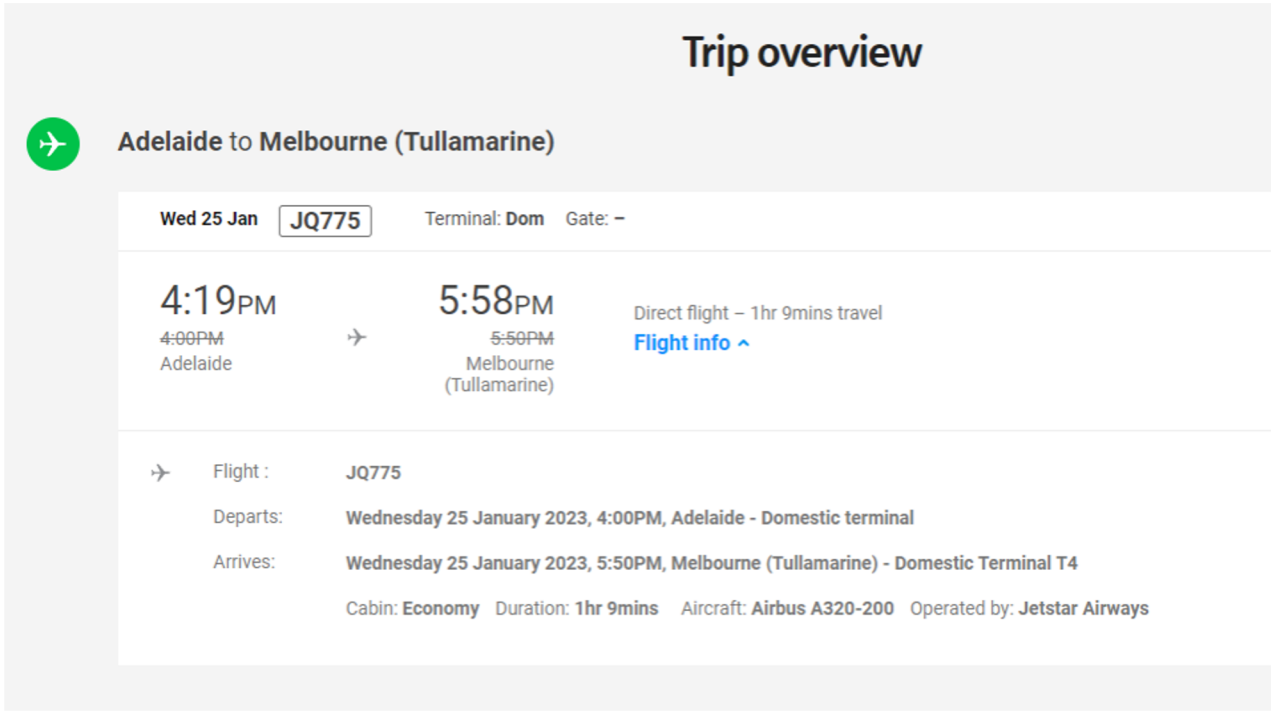
Vulnerabilities are everywhere!

Former Australian PM Tony Abbott's passport details and phone number obtained by hacker

An Australian hacker obtained Tony Abbott’s passport details and personal phone number using a photo of a plane boarding pass the former prime minister posted on social media.


Manage booking
Retrieve your itinerary, change flights, add bags, seats or extras to your booking right here.
To learn more about the measures we have introduced to keep you safe at the airport and on board, see Fly Well.
Thank you for respecting each other and our team, and helping us create a safe and positive environment for everyone.
Find your booking
Booking reference What's this?
Email address or surname
Manage Booking
or
Sign in


Manage booking
Retrieve your itinerary, change flights, add bags, seats
To learn more about the measures we have introduced to keep y
Thank you for respecting each other and our team, and helping us c
Find your boo
Booking reference What's this?
Email address or surname
Manage Booking
or
Sign in
Trip overview
Adelaide to Melbourne (Tullamarine)
Wed 25 Jan JQ775 Terminal: Dom Gate: –
4:19PM
4:00PM
Adelaide
5:58PM
5:50PM
Melbourne (Tullamarine)
Direct flight – 1hr 9mins travel
Flight info
Flight : JQ775
Departs: Wednesday 25 January 2023, 4:00PM, Adelaide - Domestic terminal
Arrives: Wednesday 25 January 2023, 5:50PM, Melbourne (Tullamarine) - Domestic Terminal T4
Cabin: Economy Duration: 1hr 9mins Aircraft: Airbus A320-200 Operated by: Jetstar Airways




```
{
    "mmbRetrieveBookingResponse": {
        "bookingData": {
            "contactInfo": {
                "emailAddress": "[redacted]",
                "firstName": "Tony",
                "lastName": "Abbott",
                "phoneNumber": "04[redacted]"
            },
            "passengers": [
                {
                    "dOB": "[redacted]",
                    "docNumber": "R[redacted]",
                    "expirationDate": "[redacted]",
                    "fullName": "MR Anthony John Abbott",
                    ...
                }
            ],
            ...
        },
        ...
    }
}
```


Sensitive but not used by the page

Excessive Data Exposures

#3 in OWASP API Security Risks (2019)

Google

500. That's an error.

The server encountered an error and could not complete your request.

If the problem persists, please report your problem and mention this error message and the query that caused it.
That's all we know.

Forbidden

You don't have permission to access this resource.

Apache/2.4.41 (Ubuntu) Server at beta.example.com Port 80

BANK

’ or ‘a’=‘a

Enter your password

Log in

Trust us with your money

Our website is totally secure and almost never gets hacked.

Manual Testing?

Keyword matching (“password”, “phone”, etc.)

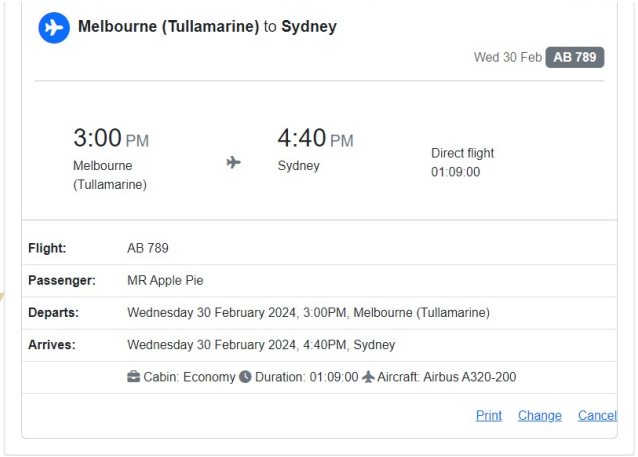
Value inferencing (credit card number, email address, phone number, etc.)

Excessive Data Exposures

How to detect if a web API exposes unnecessary data?

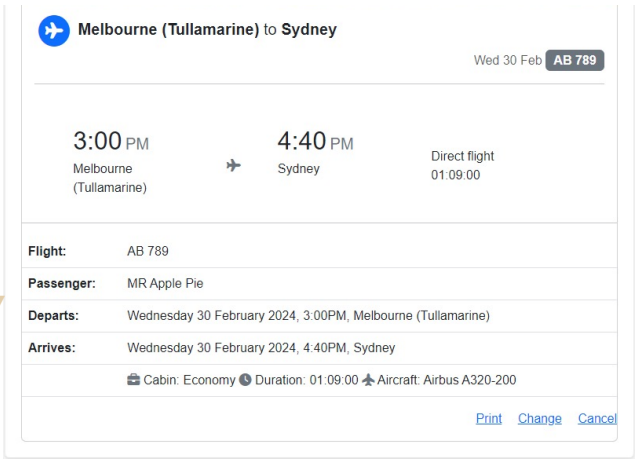
EDEFuzz – Idea

```
{
  "bookingData": {
    "flightNumber": "AB789",
    "departTime": "15:00",
    "arrivalTime": "16:40",
    "departAirport": "YMML",
    "arrivalAirport": "YSSY",
    "seatNumber": "12A",
  },
  ...
}
```

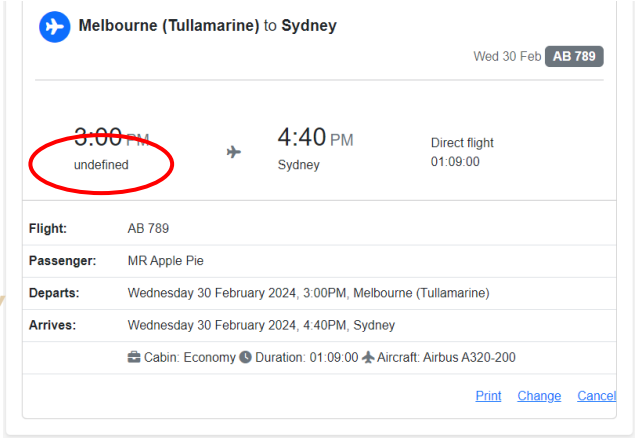


EDEFuzz – Idea


```
{
  "bookingData": {
    "flightNumber": "AB789",
    "departTime": "15:00",
    "arrivalTime": "16:40",
    "departAirport": "YMML",
    "arrivalAirport": "YSSY",
    "seatNumber": "12A",
  },
  ...
}
```

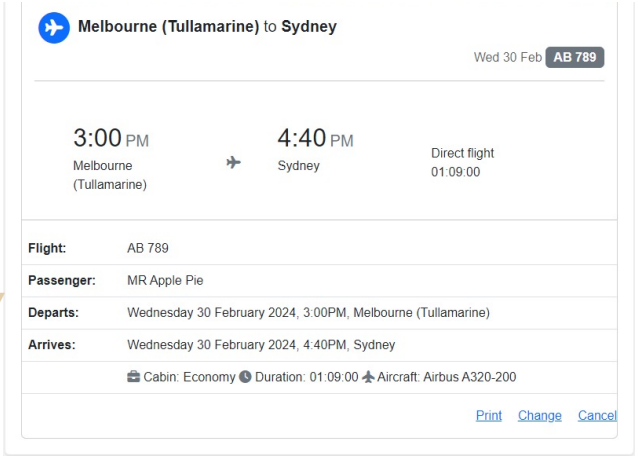




```
{
  "bookingData": {
    "flightNumber": "AB789",
    "departTime": "15:00",
    "arrivalTime": "16:40",
    "departAirport": "YMML",
    "arrivalAirport": "YSSY",
    "seatNumber": "12A",
  },
  ...
}
```

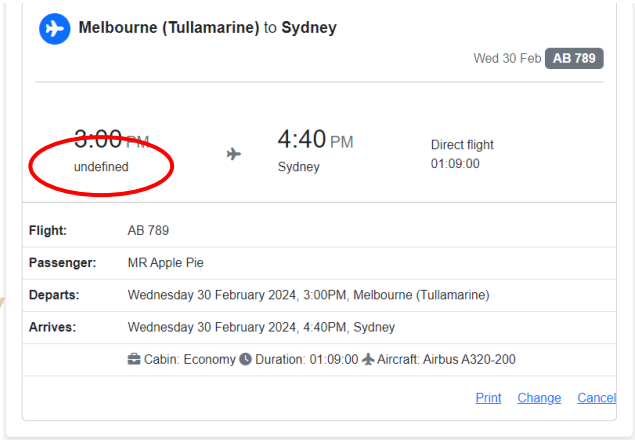



EDEFuzz – Idea

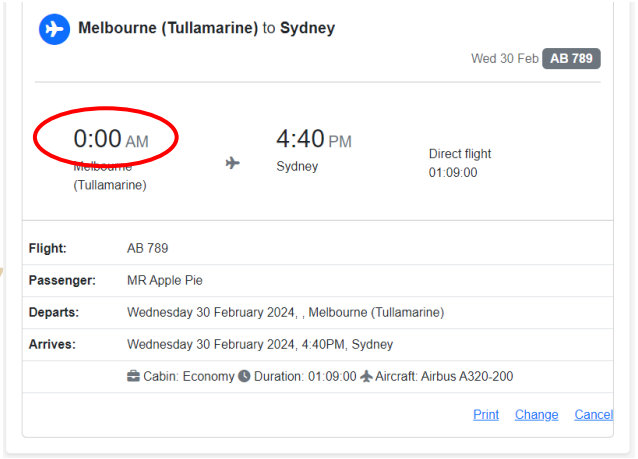

```
{
  "bookingData": {
    "flightNumber": "AB789",
    "departTime": "15:00",
    "arrivalTime": "16:40",
    "departAirport": "YMML",
    "arrivalAirport": "YSSY",
    "seatNumber": "12A",
  },
  ...
}
```

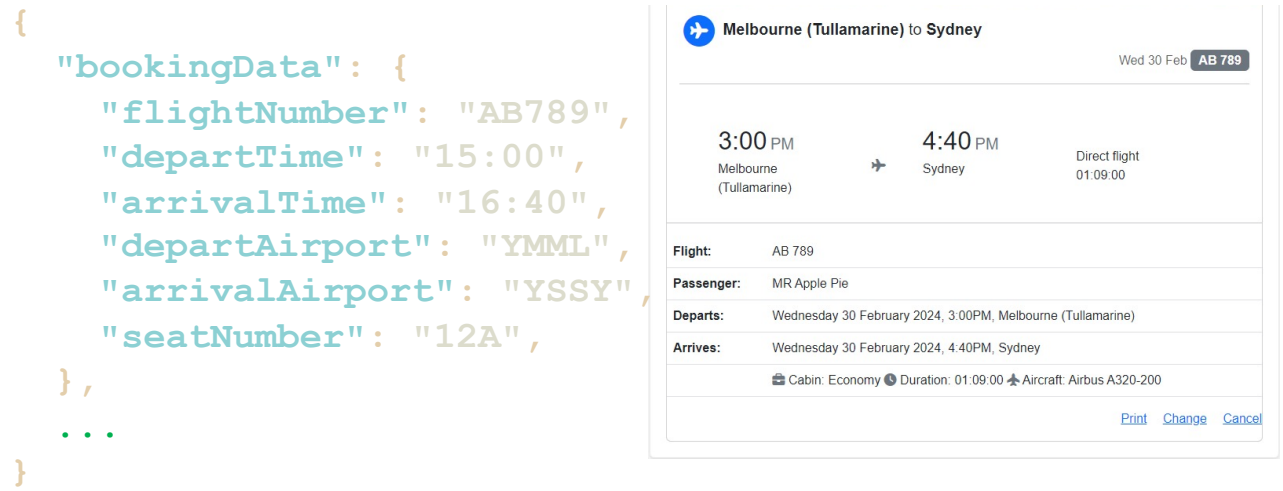

```
{
  "bookingData": {
    "flightNumber": "AB789",
    "departTime": "15:00",
    "arrivalTime": "16:40",
    ~~"departAirport": "YMML",~~
    "arrivalAirport": "YSSY",
    "seatNumber": "12A",
  },
  ...
}
```

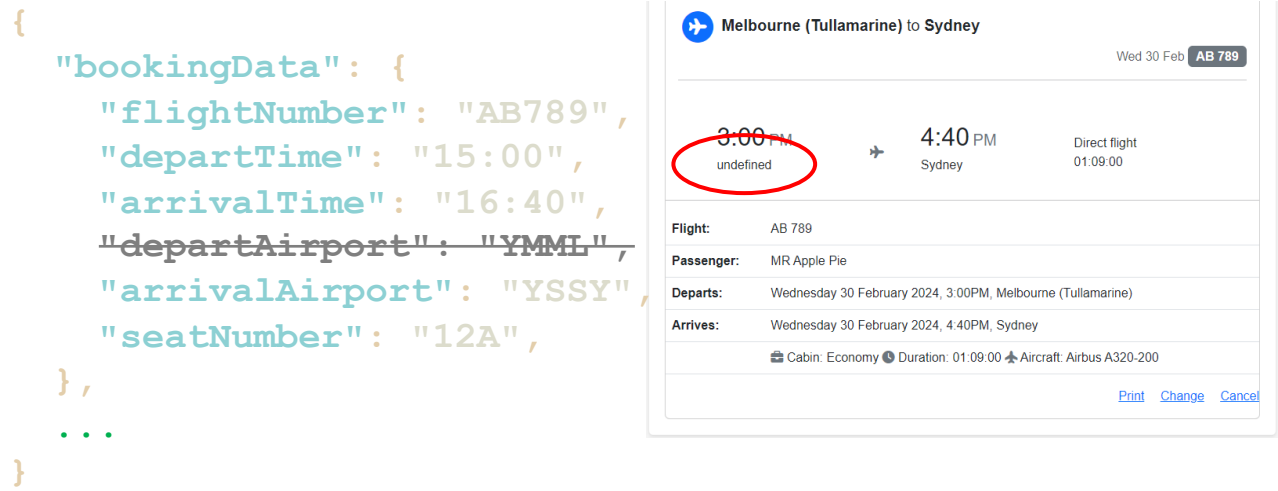



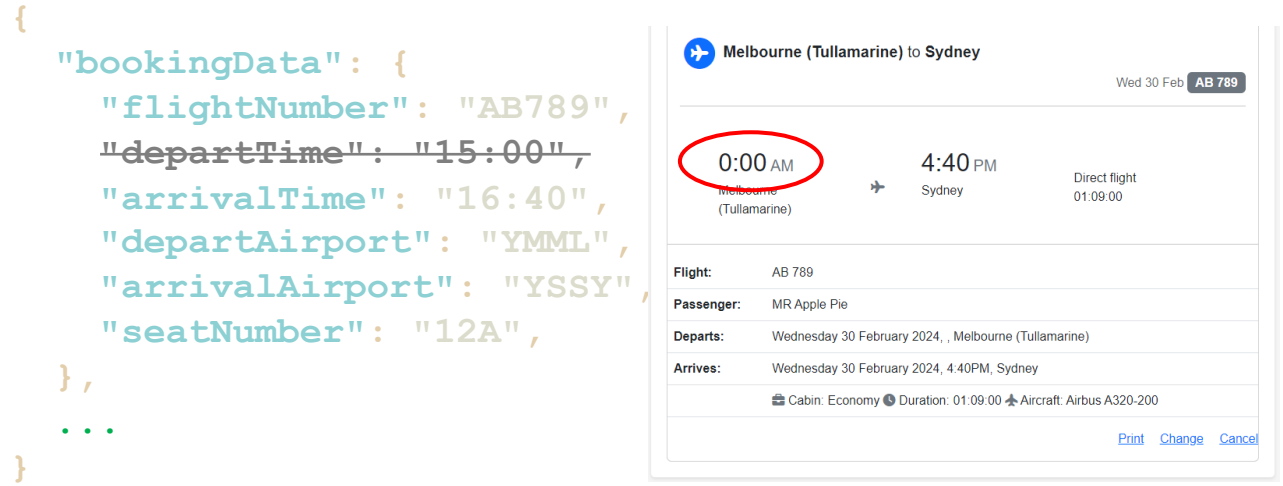

```
{
  "bookingData": {
    "flightNumber": "AB789",
    ~~"departTime": "15:00",~~
    "arrivalTime": "16:40",
    "departAirport": "YMML",
    "arrivalAirport": "YSSY",
    "seatNumber": "12A",
  },
  ...
}
```

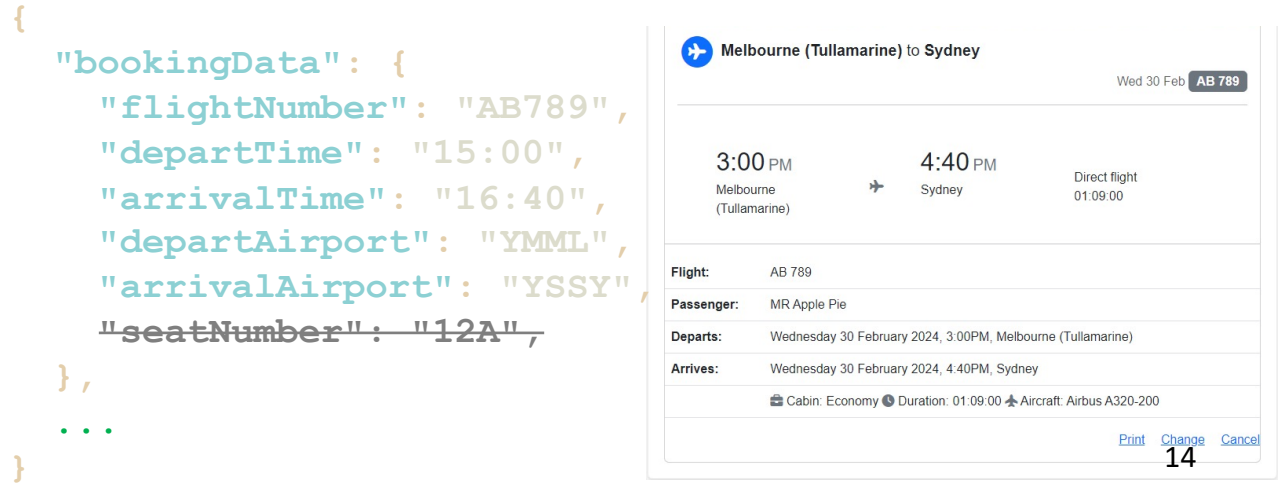



EDEFuzz – Idea









EDEFuzz – Metamorphic relation

A data field, d, is considered non-excessive, if

$$\mathrm{diff}(D_{origin}, D_{mutated}) \neq 0$$

D_{origin} : the web page rendered using R_{origin}, represented by document object model (DOM) tree

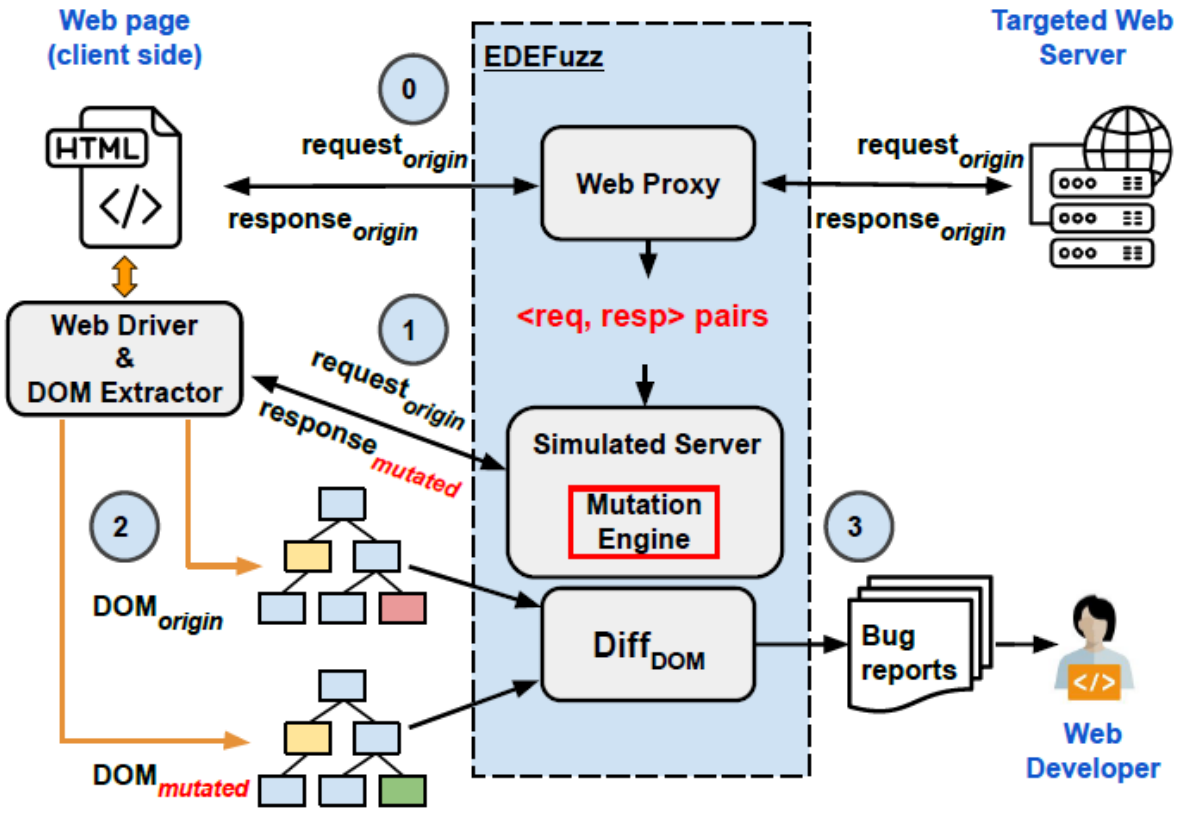
$D_{mutated}$: the web page rendered using $R_{mutated}$, represented by DOM tree

R_{origin} : the original API response produced by the web server

$R_{mutated}$: the mutated API response (by removing one data field, d, from R_{origin})

EDEFuzz – Overview

- Save a copy of the rendered webpage after the API call
- Try deleting each of the fields in the response
- Check if the newly rendered page is the same as the saved copy



EDEFuzz – Preparation Phase

The configuration file
(prepared manually by tester)

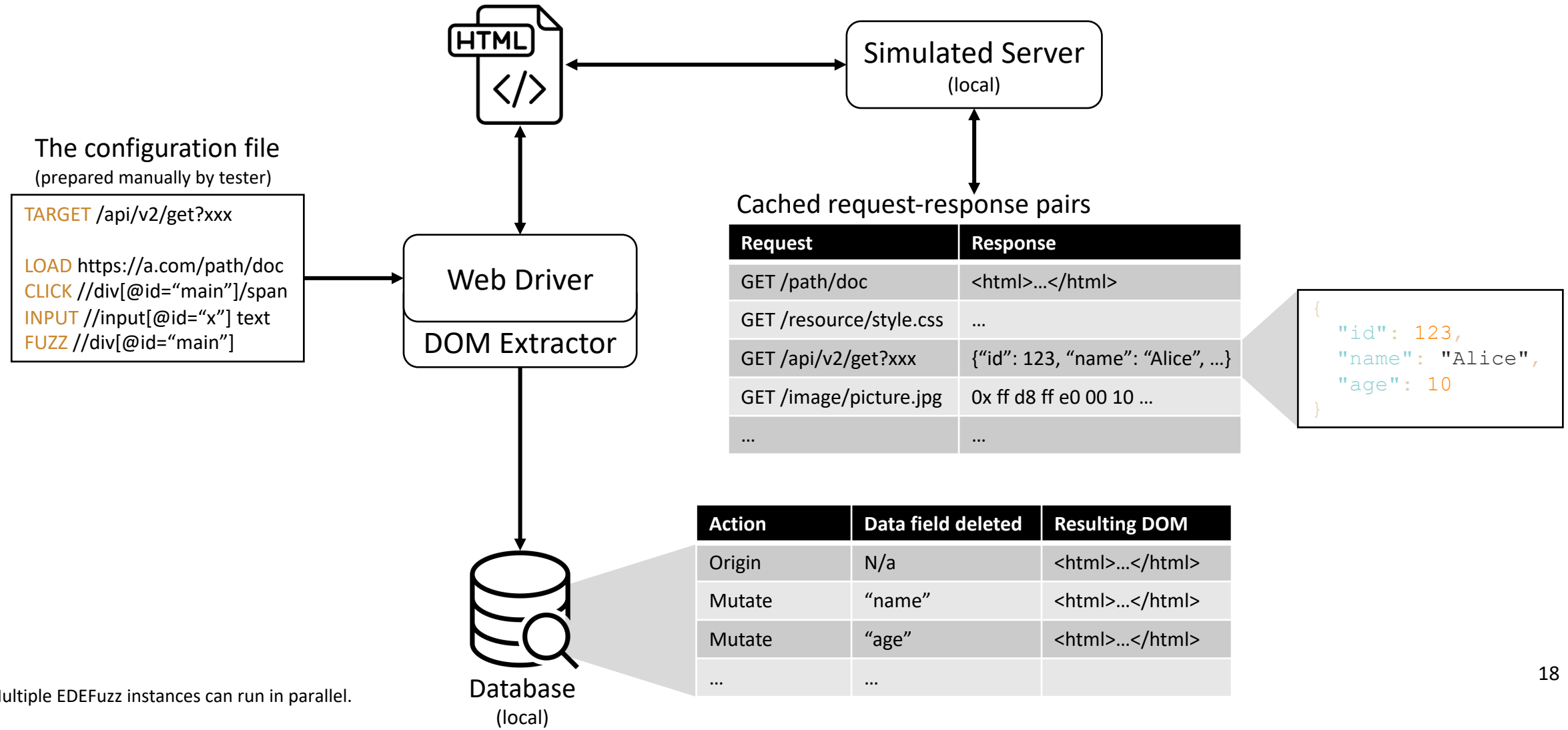
```
TARGET /api/v2/get?xxx

LOAD https://a.com/path/doc
CLICK //div[@id="main"]/span
INPUT //input[@id="x"] text
FUZZ //div[@id="main"]
```

Web Driver

Cached request-response pairs

Request	Response
GET /path/doc	<html>…</html>
GET /resource/style.css	…
GET /api/v2/get?xxx	{"id": 123, "name": "Alice", …}
GET /image/picture.jpg	0x ff d8 ff e0 00 10 …
…	…

EDEFuzz – Execution Phase

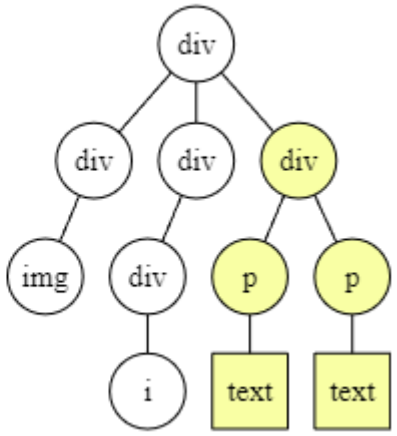


*Multiple EDEFuzz instances can run in parallel.

EDEFuzz – Reporting Phase

$$\text{diff}(D_{origin}, D_{mutated}) \neq 0$$

```
<div class="container">
  <div class="logo">
    <img src="..." />
  </div>
  <div class="location_icon">
    <div class="icon">
      <i class="icon_car"></i>
    </div>
  </div>
  <div class="driver_info">
    <p>Name: ...</p>
    <p>Phone: ...</p>
  </div>
</div>
```



Two DOM trees are considered identical if:

C1: Their root nodes have the same tag name (e.g.)

C2: Their root nodes have the same number of attributes

C3: Each corresponding pair of attributes have the same value

C4: Their root nodes have the same number of children

C5: Each corresponding pair of children representing a tag element is identical, with respect to conditions C1-C4

C6: Each corresponding pair of children representing a string is identical

* C3 and C6 are relaxed in certain conditions

EDEFuzz – Evaluation

- (RQ-1) **Accuracy.** Of the data fields flagged by EDEFuzz, what proportion are true excessive data exposures (i.e. unused by the web page). This evaluates the usefulness of our metamorphic relation.
- (RQ-2) **Applicability.** To what proportion of widely used web sites can EDEFuzz be applied successfully? This helps to understand limitations of our approach, both inherent and those that arise from EDEFuzz's current implementation.
- (RQ-3) **Efficiency.** How much human effort and computational time is required to apply EDEFuzz? This sheds light on the scalability of our approach.
- (RQ-4) **Prevalence of Sensitive Data Leakage.** Of those fields flagged by EDEFuzz as excessive, what proportion contain sensitive data? This helps us understand how prevalent sensitive data leakage is amongst excessive data exposure issues.

EDEFuzz – Evaluation

- (RQ-1) **Accuracy.**

Eight selected Australian websites

True Positive = 98.65%

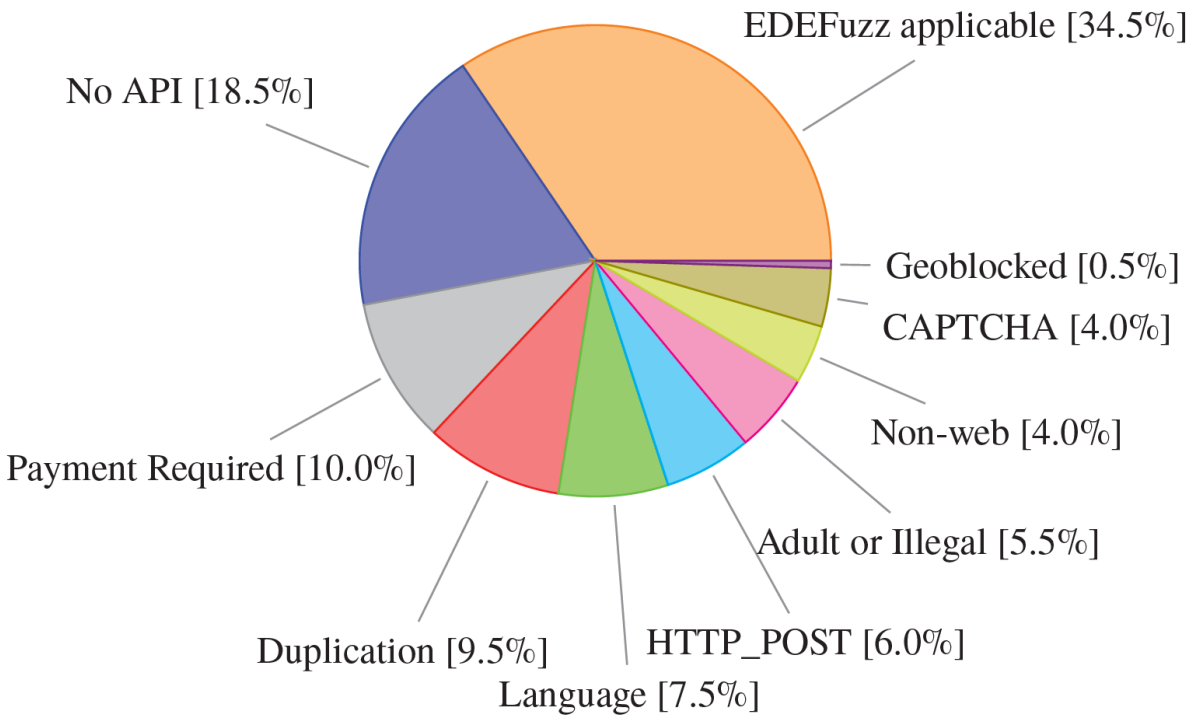
Target	Data Fields	Reported	Confirmed	TP
A	189	124	124	100.00%
B	18	16	14	87.50%
C	2600	2580	2504	97.05%
D	545	506	479	94.66%
E	4249	4147	4127	99.52%
F	778	749	749	100.00%
G	120	100	100	100.00%
H	1465	1066	1066	100.00%

EDEFuzz – Evaluation

- (RQ-2) **Applicability.**

Alexa top 200 sites (131 excluded)

Applicable 53/69 sites (76.8%)



EDEFuzz – Evaluation

- (RQ-3) **Efficiency.**

  < 20 minutes to identify a target API, and to compose a configuration file.

  < 20 minutes to inspect the flagged fields, to determine which are sensitive.

EDEFuzz – Evaluation

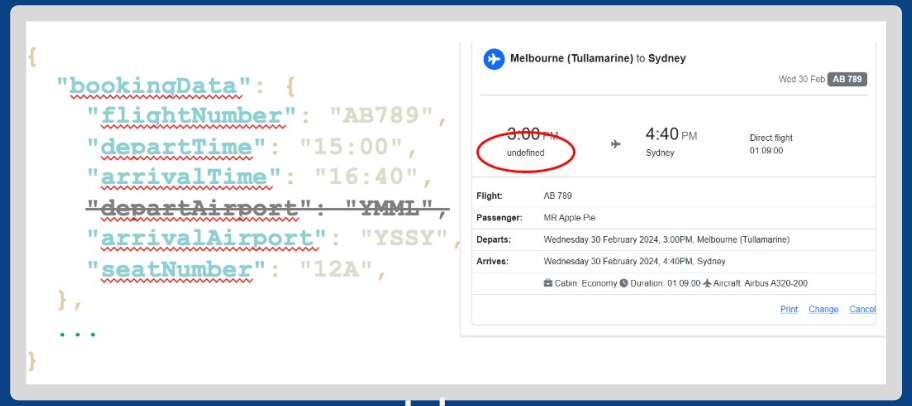
- (RQ-4) **Prevalence of Sensitive Data Leakage.**

Four (4) in Australian dataset (4 out of 8)

One (1) in Alexa Top 200 dataset (1 out of 53)

THE UNIVERSITY OF MELBOURNE

EDEFuzz Q & A

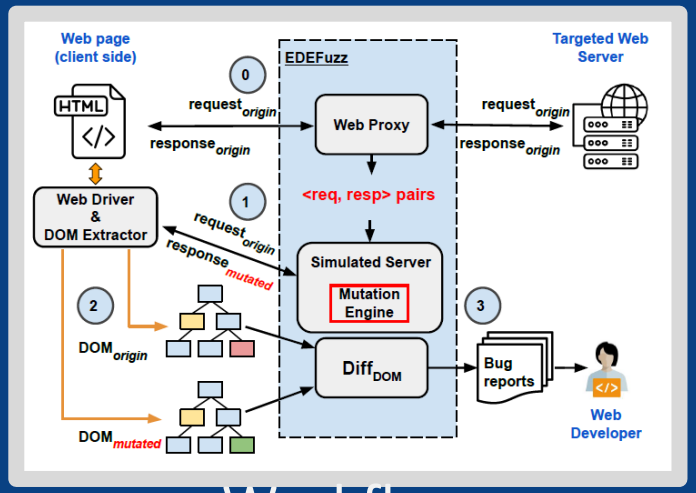


Idea

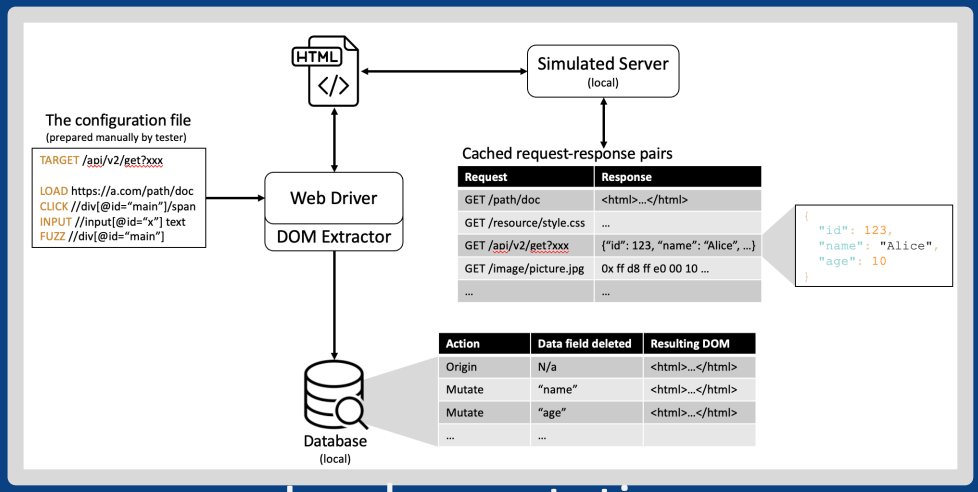
Evaluation

Accuracy
Applicability
Efficiency
Prevalence of EDEs

Evaluation



Workflow



Implementation