TU Delft High-level Mutations for JSON Typed Data in Big Data Fuzz Testing

Author: Lars Rhijnsburger Supervisor: Burcu Kulahcioglu Ozkan

CSE3000 Research Project 29 June 2021

01 Background

Big Data is growing and present in almost all aspects of our life.

BigFuzz is a new tool used for tabular Big Data fuzzing, but it is still limited.

JSON schema's exist to summarise and validate JSON instances.

02 Problem

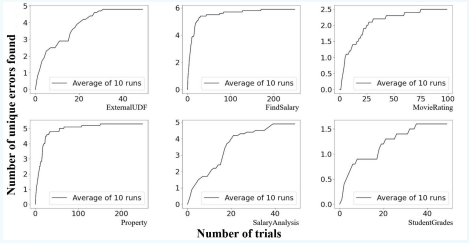
JSON is the main exchange format over the World Wide Web and has no automated big data testing tool.

"How can we provide users to enter input specifications and implement mutations in a generic way for all kinds of JSON typed data".

03 Method

Input structure assumptions.

Extend BigFuzz instead of making a new version of it.



Schema

```
[
  {
    "$id": "An example schema",
    "$schema": "https://json-schema.org/draft/2020-12/schema",
    "title": "Example",
    "type": "object",
    "properties": {
      "AnyProperty": {
        "type": "[string, boolean, integer, number]",
        "description": "A description of the property"
      }
    }
  }
]
```

Valid Data Example

```
[
  {
    "AnyProperty": "[string, boolean, integer, number]"
  }
]
```

04 Results & Conclusions

Finds bugs in less number of trials.

Loses descriptive exception information due to type inferred parsing.

Proof of concept: Big Data JSON programs can be fuzzed effectively.

05 Future Work

Research the use of different ways for mutations.

More options for fuzzing with guidance.