Re: MLCS: final project

Hans Peter Reiser <hansr@ru.is>

Čt 12.10.2023 19:15

Komu:David Pažout <davidp23@ru.is>

Hi,

sure, its on a MinIO server (like Amazon S3), you need these access credentials:

```
"smartro": {
"url": "https://share-smartvmi.pads.fim.uni-passau.de",
"accessKey": "Y8wFB5LF3K3qhqt5",
"secretKey": "h3zGyChyu98ZdWPdytcsEg6q0spRChfZ",
"api": "s3v4",
"path": "auto"
},
```

shas_by_families.json is a json hash mapping classes (tags) to file hashes.

In the train folder, each file is the trace of a single program. The first part of the file name (up to the '-') is the file hash.

The zip files contain the complete trace, with one API call in a single line. like this line:

```
{"level":"info","ts":"2023-10-05T13:11:24Z","msg":"hookCallback hit","vmi_ts":"2023-10-05T13:06:06Z","vmi_logger":"ApiTracing_FunctionHook","vmi_Gla":"77212b20","vmi_Module":"ntdll.dll","vmi_Function":"NtQueryValueKey"}
```

corresponds to the API call "NtQueryValueKey"
If you prefer, I could run a script that transforms these verbose traces into just the API calls to reduce the size.

I just saw that more items have been added the data set yesterday and today. Its this size in total (compressed)

```
% mc du smartro/api-traces-smartvmi/train
3.6GiB 72371 objects api-traces-smartvmi/train
```

```
% mc ls smartro/api-traces-smartvmi
[2023-10-06 07:55:44 GMT]  72MiB STANDARD shas_by_families.json
[2023-10-12 17:06:17 GMT]     0B train/
% mc ls smartro/api-traces-smartvmi/train
[2023-10-11 10:21:09 GMT]  21KiB STANDARD 000003272077691ce8cbc781f727dda81177ef58face05682ec0aba935fe0d55-2WXDXS6B5C.11102023_102109.zip
[2023-10-12 17:03:30 GMT]  20KiB STANDARD 000003272077691ce8cbc781f727dda81177ef58face05682ec0aba935fe0d55-TSEFLX9V5W.12102023_170330.zip
```

[…]

Best,
Hans

> On 12 Oct 2023, at 16:57, David Pažout <davidp23@ru.is> wrote:
>
> Hello,
> could you provide me with the dataset (or where to find it) for the topic 2.1(2)?
>
> David P.