Oddities of finding and files (from an implementer's view)

CSAF Community Days 2024

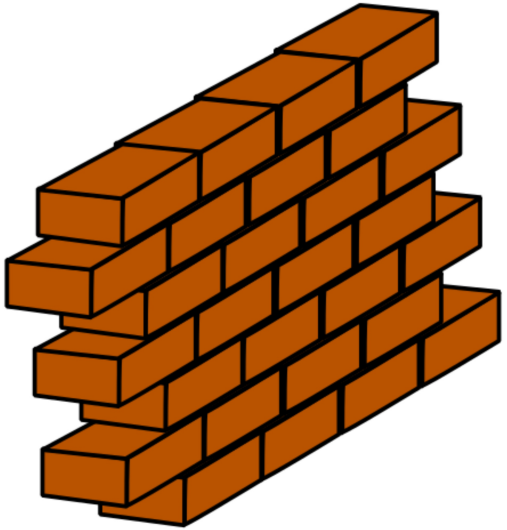
<bernhard@intevation.de>
Bernhard E. Reiter, Dipl. Systemwiss., MSc

TLP:CLEAR

no curl

Too fast

Browser only



no go-client

Wrong origin

Only one at a time

TLP:CLEAR

Use a server ready for automation

- Serve static files or cache well
- Avoid Akamai (or demand from your CDN to accept all user-agents)
- Allow browsers to access it (CORS Headers)
- Use .well-known/csaf/
- (use the checker on your provider from external)

TLP:CLEAR

Presenting work from Alexej Antoni

Documents downloaded mainly in September 2024

CSAF document analysis

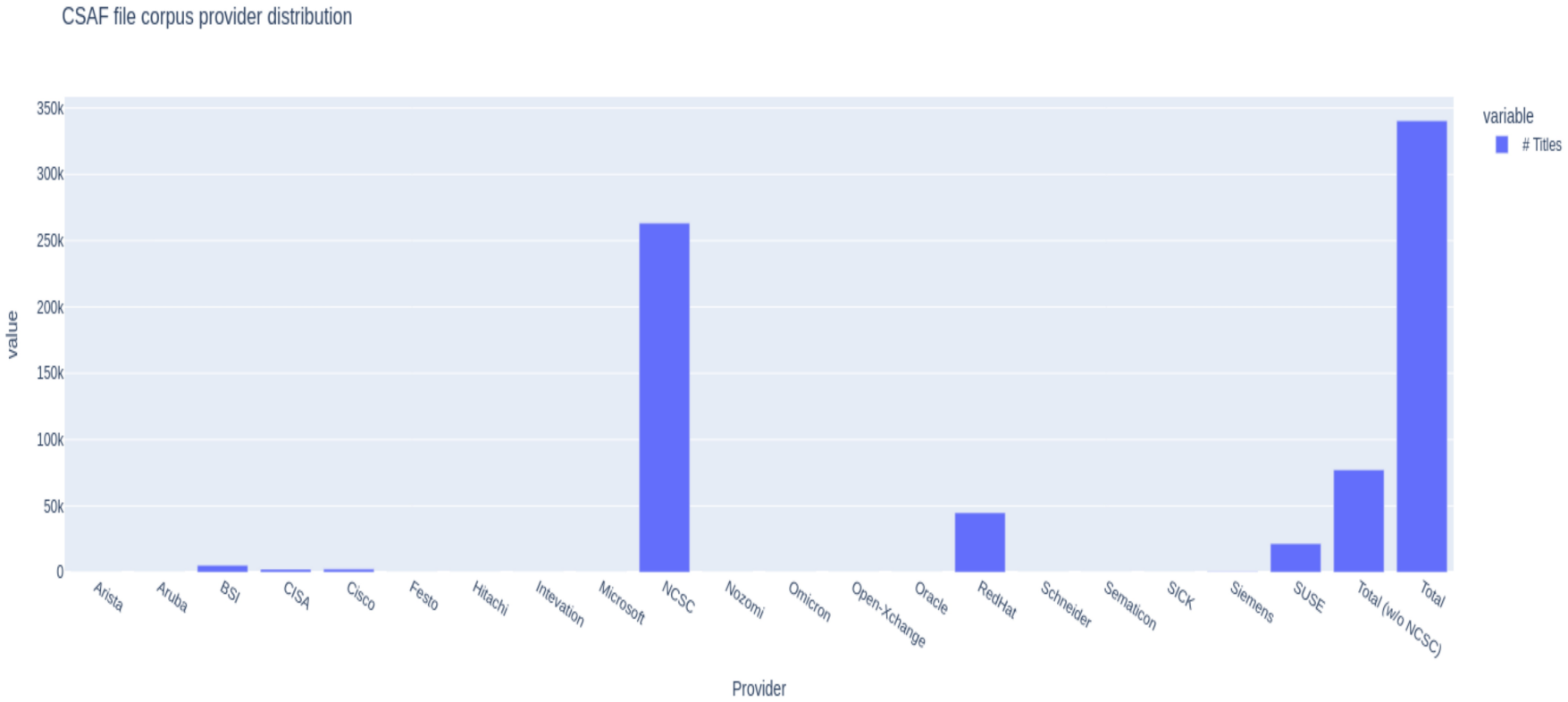
Corpus:

- Currently total of **340,000** JSON files
- 20 CSAF document „providers“
- Skewed dataset (e.g.: 260,000 NCSC files vs 200 Siemens files)
- Source: BSI Lister and list of potential candidates (from T. Schmidt)
- Using csaf_downloader or custom scripts
- Stored gzipped, expanded: 20 GByte

TLP:CLEAR

Corpus provider distribution

Provider	# Titles
Arista	9
Aruba	54
BSI	5317
CISA	2319
Cisco	2398
Festo	15
Hitachi	58
Intevation	1
Microsoft	96
NCSC	263251
Nozomi	27
Omicron	10
Open-Xchange	17
Oracle	11
RedHat	44861
Schneider	128
Sematicon	1
SICK	42
Siemens	212
SUSE	21654
Total (w/o NCSC)	77230
Total	340481

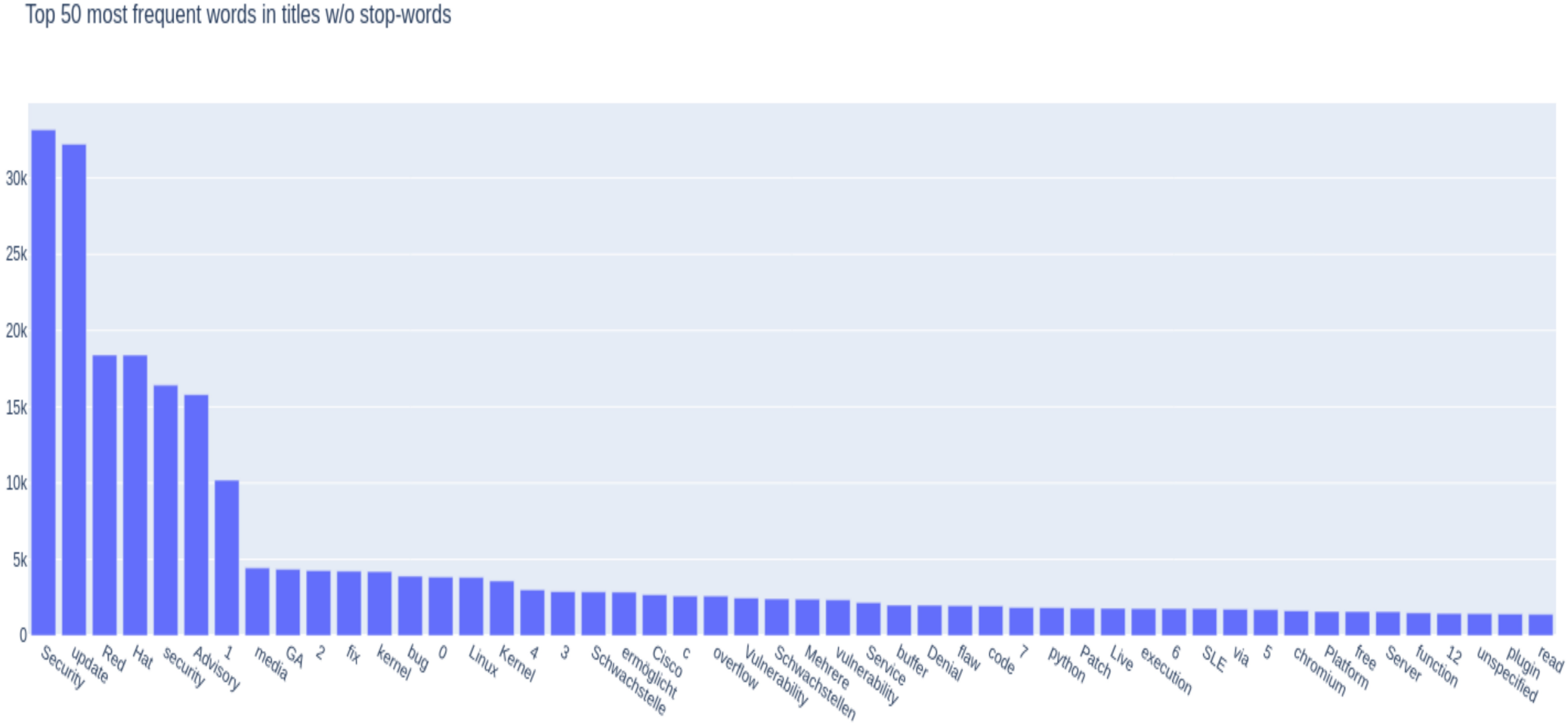


Title Analysis: Word Frequency – without NCSC

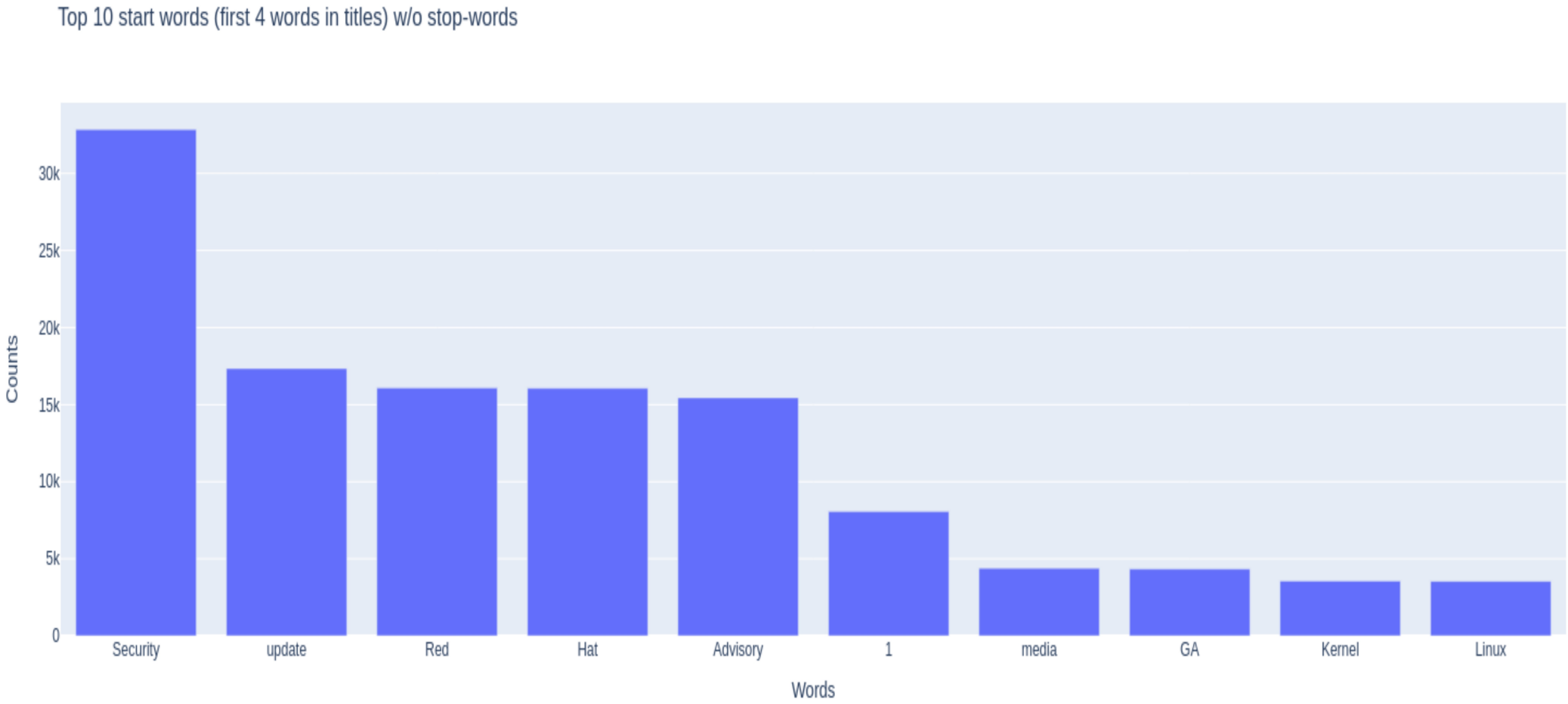
Preliminary:

- NCSC title data is very uniform, always in the pattern CVE-YEAR-IDNUM, title is the same as JSON file name and document tracking ID-Number
- we left NCSC out of this analysis step: (most frequent word would be „CVE“ occuring 263251 times, followed by year sub-strings)

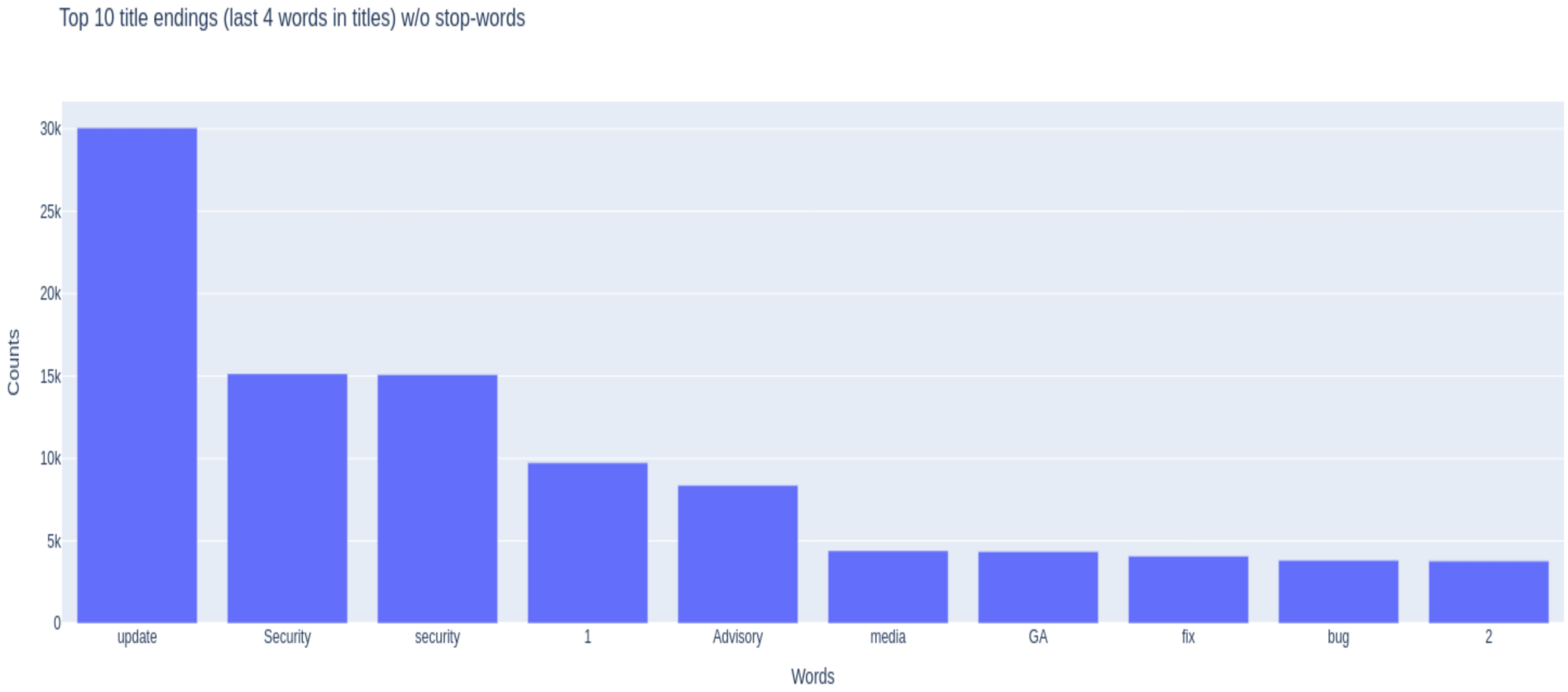
Title Analysis: Word Frequency – Top 50 most frequent words



Title Analysis: Word Frequency – Top 10 start words



Title Analysis: Word Frequency – Top 10 endings



Title Analysis: Word Frequency

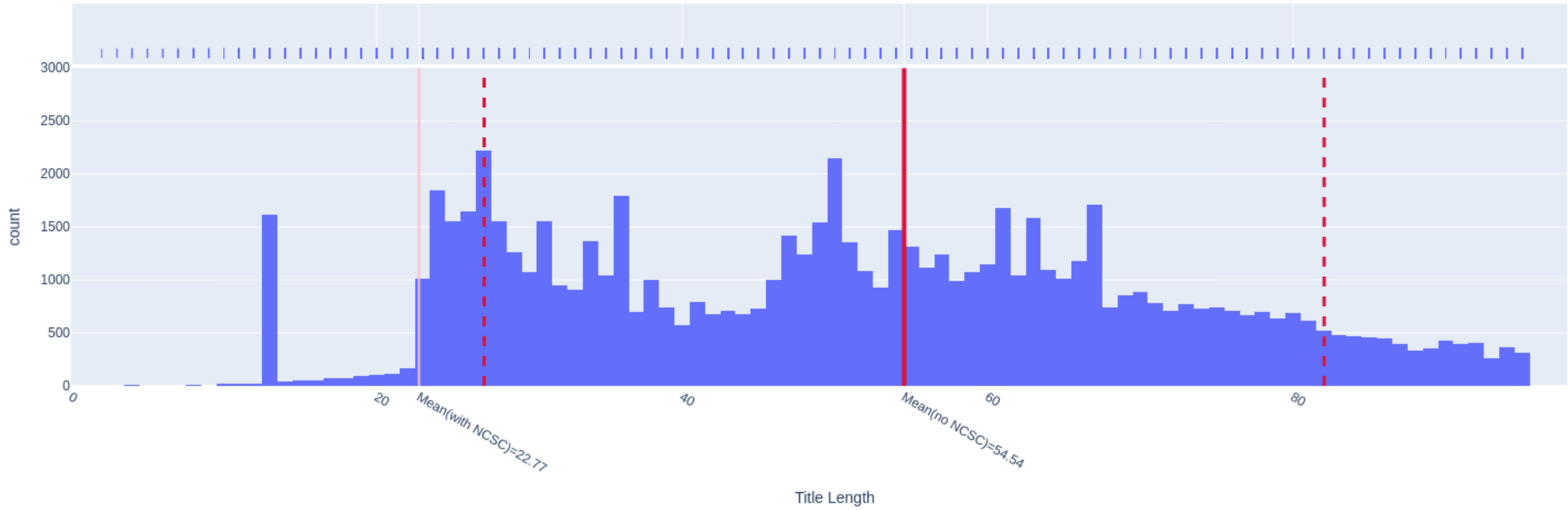
Findings:

- 34 % of Red Hat Security Advisory titles start with "Red Hat Security Advisory: "
- 3,5% of Red Hat Security Advisory titles are the 13-character string "security flaw", which amounts to 1572 of the 44861 files (this explains the outlier 13-character bar in the upcoming title-length histogram that exludes NCSC)
- most frequent words in titles are „Security“ and „update“, the most referenced operating system is „Linux“

TLP:CLEAR

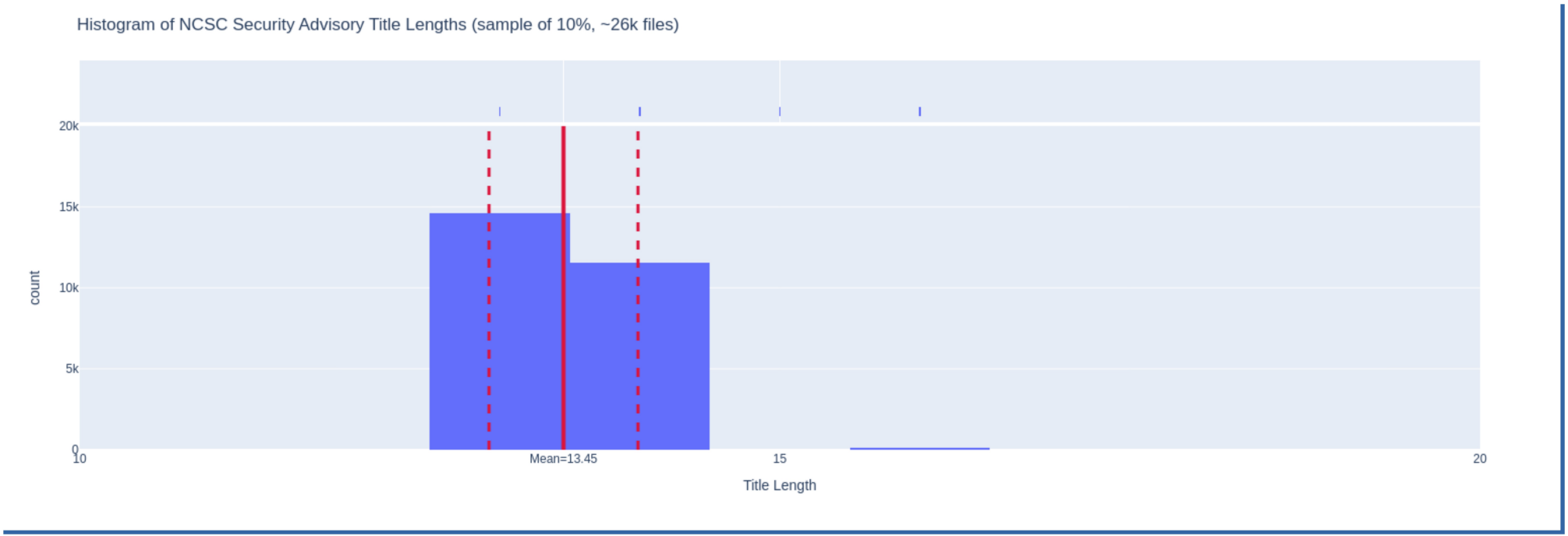
Title Analysis: Title Length

Histogram of selected Security Advisory Title Lengths below 95th percentile length threshold (4107 longest titles filtered out for histogram visibility).



TLP:CLEAR

Title Analysis: Title Length



TLP:CLEAR

Title Analysis: Title Length

- 95% of title lengths (without NCSC) are within a 96 character limit; 99% within 122 characters
- The average title string length excluding NCSC data is ~55 characters
- Including NCSC would move the total average length of title strings to around 23 characters
- Shortest title only has 2 characters „**2)**“, belongs to RedHat security advisory document cve-2014-3609.json
- Longest title of 1465 characters (and 3 other titles with 1000+ characters) belong to SUSE (suse-su-2020_0642-1.json: 1115, suse-su-2020_0640-1.json: 1187, suse-ru-2019_1161-1.json: 1456, suse-su-2019_2267-1.json: 1465)

TLP:CLEAR

Title Analysis: Title Length – the longest

Title of SUSE document „**suse-su-2019_2267-1.json**“ (1465 chars):

"Security update for ardana-ansible, ardana-barbican, ardana-cinder, ardana-cluster, ardana-cobbler, ardana-db, ardana-designate, ardana-extensions-nsx, ardana-glance, ardana-heat, ardana-horizon, ardana-input-model, ardana-installer-ui, ardana-ironic, ardana-keystone, ardana-logging, ardana-magnum, ardana-monasca, ardana-mq, ardana-neutron, ardana-nova, ardana-octavia, ardana-opsconsole, ardana-opsconsole-ui, ardana-osconfig, ardana-service, ardana-ses, ardana-swift, ardana-tempest, crowbar-core, crowbar-ha, crowbar-openstack, crowbar-ui, java-monasca-common, java-monasca-common-kit, openstack-ceilometer, openstack-cinder, openstack-designate, openstack-heat, openstack-horizon-plugin-neutron-fwaas-ui, openstack-horizon-plugin-neutron-lbaas-ui, openstack-horizon-plugin-neutron-vpnaas-ui, openstack-ironic, openstack-ironic-python-agent, openstack-keystone, openstack-magnum, openstack-manila, openstack-monasca-notification, openstack-monasca-persister, openstack-monasca-persister-java, openstack-monasca-persister-java-kit, openstack-neutron, openstack-neutron-gbp, openstack-neutron-lbaas, openstack-nova, openstack-octavia, openstack-tempest, python-ardana-configurationprocessor, python-cinder-tempest-plugin, python-ironicclient, python-keystonemiddleware, python-monasca-tempest-plugin, python-openstackclient, python-openstacksdk, python-proliantutils, python-python-engineio, python-swiftlm, python-vmware-nsx, python-vmware-nsxlib, yast2-crowbar"

TLP:CLEAR

Filesize Analysis -- processing

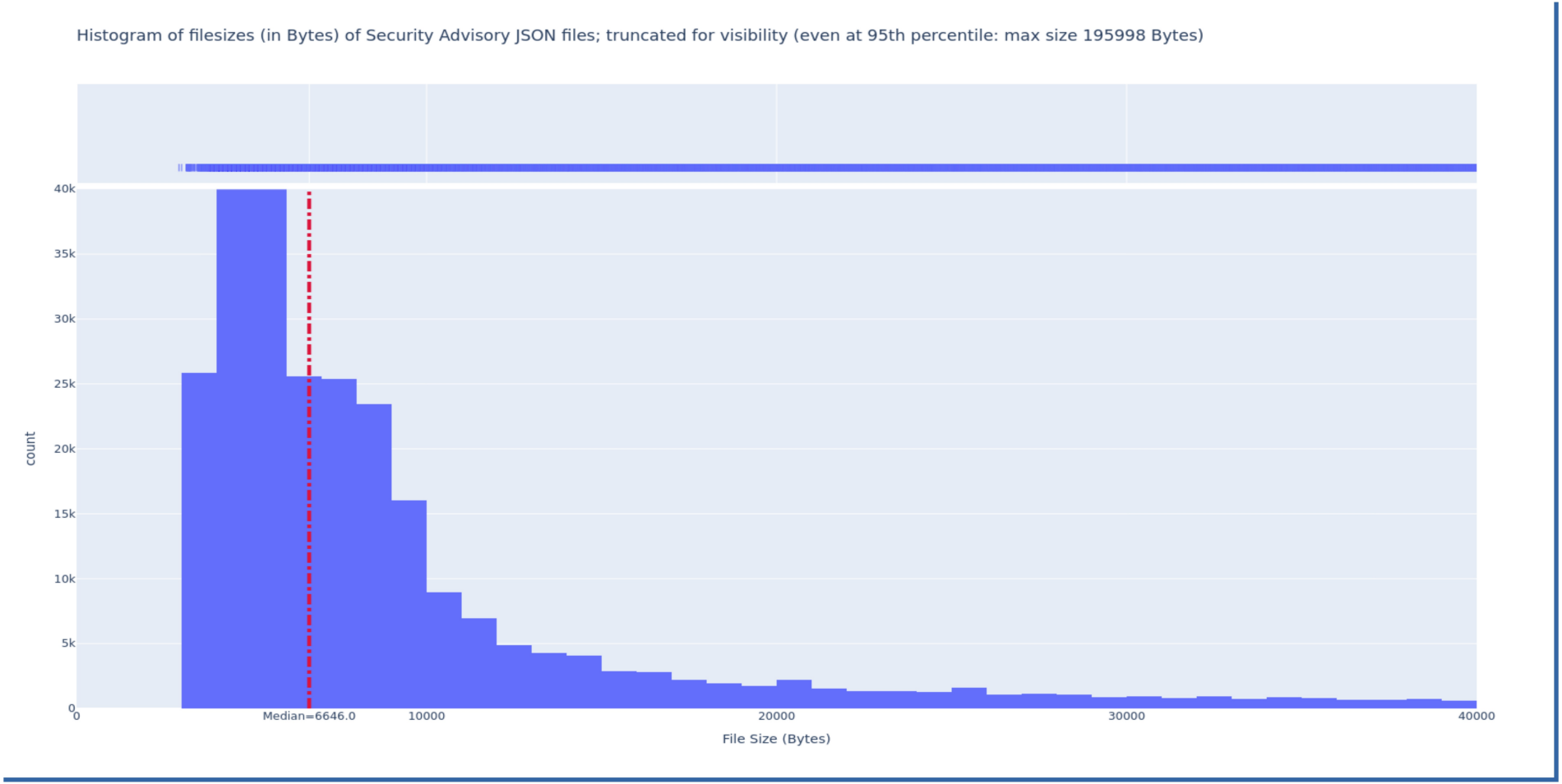
- Data Collection: We used the following Bash command to aggregate the filesize of each JSON file per provider:

```
for file in */*.json; do
  stat --format="%s" $file >> *filesizes.txt
done
```

- Outlier Detection: Additionally, this step identified large outlier files, which were later collected and further analyzed using a Python script.

TLP:CLEAR

Filesize Analysis

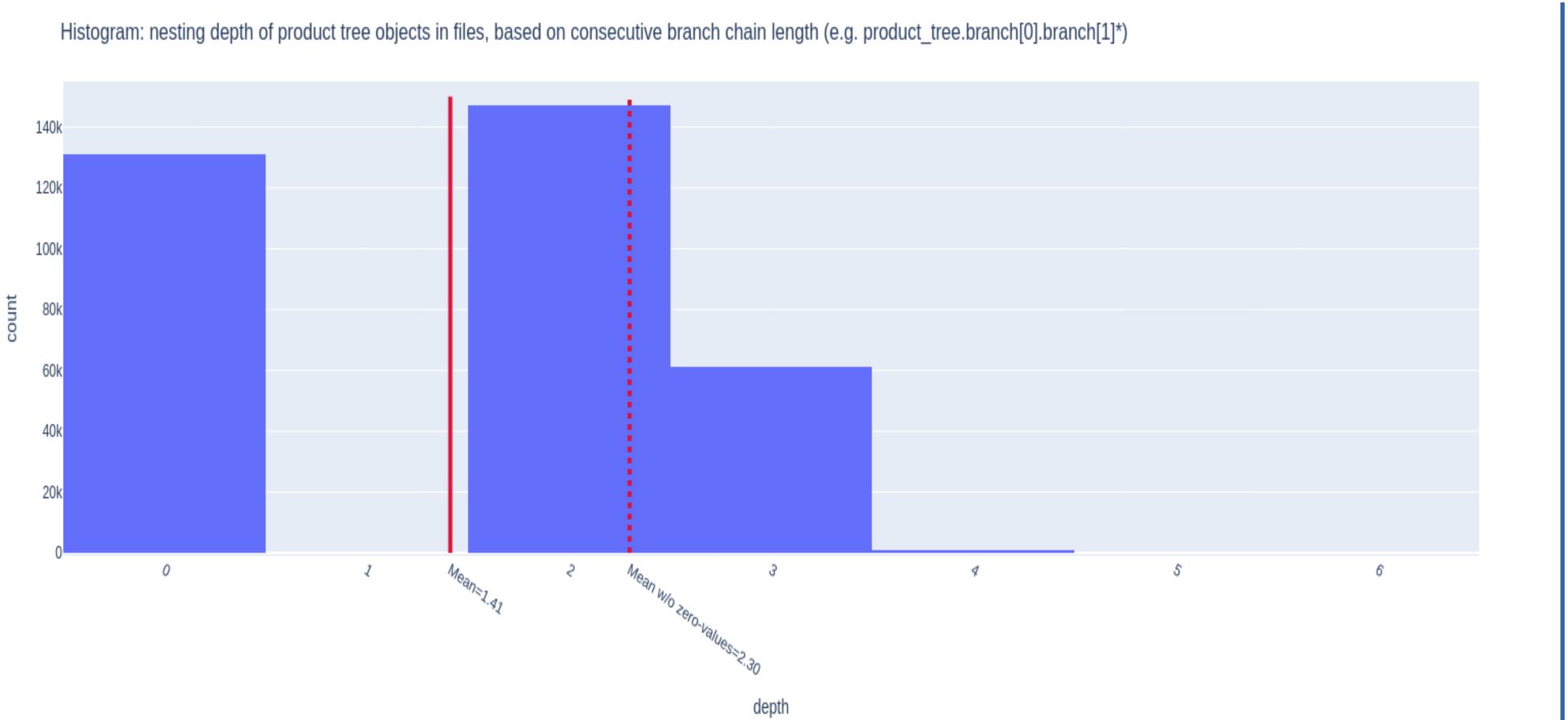


Filesize Analysis

- Largest file (by far) is a ~**103MB** Red Hat file (rhsa-2018_3140.json)
- Smallest file with filesize of only 2925 Bytes is the Intevation file (intevation-os-2024-0001.json)
- **Median filesize is 6646 Bytes** (strongly positively skewed distribution → mean does not represent central tendency too well)
- 95% of filesizes are below 200 kB; 99% of filesizes are below 1MB
- Red Hat has most files (3051 files) above or equal to 99th percentile threshold, followed by SUSE (312 files)

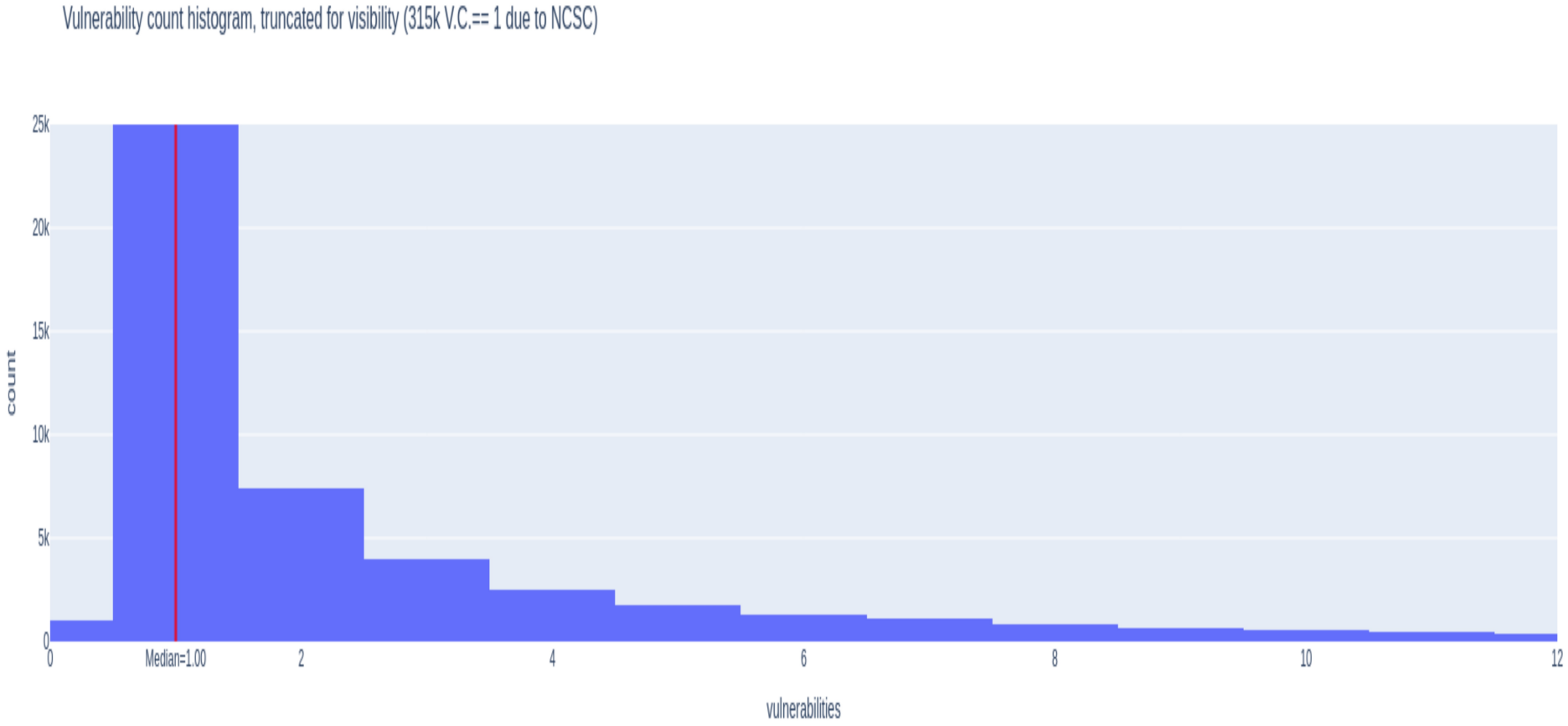
TLP:CLEAR

Product Tree „Depth“

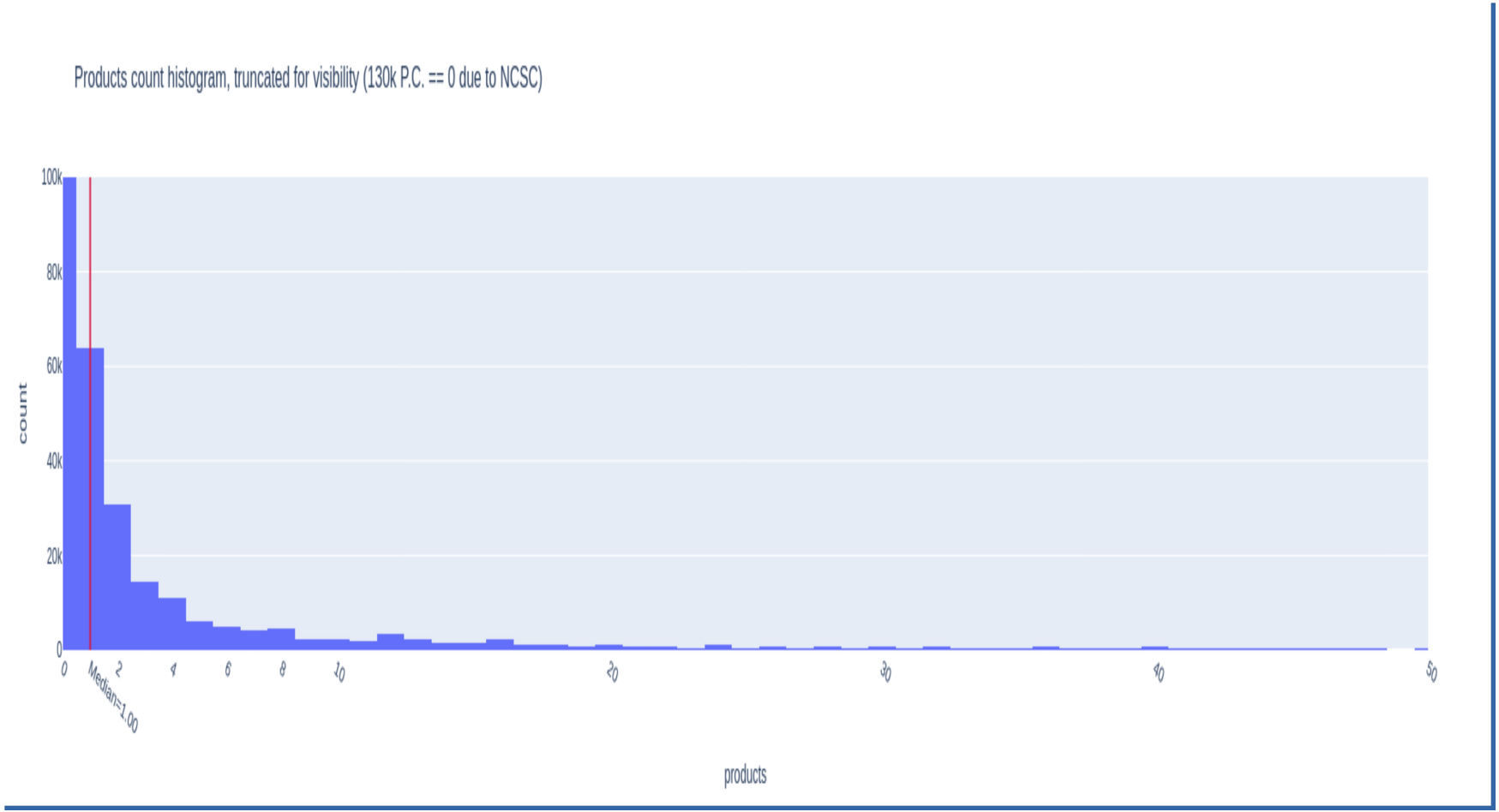


TLP:CLEAR

Vulnerabilities and Product Counts: V.C. Stats



Vulnerabilities and Product Counts: P.C. Stats



TLP:CLEAR

Vulnerabilities and Product Counts

- some very large outliers pulling the averages up
- **99%** of files have a **vulnerability** object **count below 12** with an average of ~1,20 vulnerability objects per file, if the top 1% outliers were to be removed
- **99%** of files have a uniqely referenced **product count below 796 with an average of ~21** referenced products per file, if the top 1% outliers were to be removed
- Caveat: Over 131k files from NCSC returned a P.C. == 0, which might be due to missing product_status key within Vulnerabilities objects, that we relied on to aggregate referenced products,
- **Highest Vulnerabilities** Count we found in a SUSE file (V.C. == **2310**: opensuse-su-2024_12948-1.json)
- **Highest Product Count** we found in a Red Hat file (P.C. == **30942**: cve-2023-39325.json)

TLP:CLEAR

Write or generate „nice“ CSAF Documents

- Remove unsignificant words from the title
- Do not re-generate on unsignificatn changes
- Don’t make one advisory for „everything“

TLP:CLEAR

Can't you allow a little deviation?

2024-04-18T12:42Z

2024-11-12T08:00:00.0000000

JSON latün-1

OpenPGP v3

Content-Type: application/json; charset=utf-8

TLP:CLEAR

When establishing a new standard: Be strict about what you accept!

~~Be strict when sending and tolerant when receiving.~~

„Tolerance of errors in early deployments is most likely to result in problems.“

RFC 9413

TLP:CLEAR