

HCP Runbook 1

Triage Squid Alerts Using Typosquatting Algorithm

Date of Publish: 2018-12-21



<https://docs.hortonworks.com/>

Contents

Triage Squid Events.....	3
Triage Squid Using the Typosquatting Algorithm.....	3
Improve Scoring with a Domain Whitelist.....	10

Triage Squid Events

Security event triage rules determine which events require further follow up and which events can be archived without further investigation. HCP processes many events every day so effective triage helps analysts focus on the most important events.

The two components of security event triage are:

- Determine if the event is an alert.
- If the event is an alert, assign a score. If the event is not an alert, it is not scored.

Triage Squid Using the Typosquatting Algorithm

For this example, we use a simple triage rule to detect typosquatting. Typosquatting uses common domain misspellings to install malicious web content.

Procedure

1. Determine the number of possible typosquat permutations.

To configure the Bloom filter you need to specify roughly how many elements are going into the Bloom filter and what kind of false positive probability you want. You can use the CONSOLE output mode of the flatfile_summarizer.sh to count the number of typosquatted domains across the entire document.

- a) Create an extractor_count.json file at \$METRON_HOME/config and populate it with the following:

```
{
  "config" : {
    "columns" : {
      "rank" : 0,
      "domain" : 1
    },
    "value_transform" : {
      "domain" : "DOMAIN_REMOVE_TLD(domain)"
    },
    "value_filter" : "LENGTH(domain) > 0",
    "state_init" : "0L",
    "state_update" : {
      "state" : "state + LENGTH( DOMAIN_TYPOSQUAT( domain ) )"
    },
    "state_merge" : "REDUCE(states, (s, x) -> s + x, 0)",
    "separator" : ",",
  },
  "extractor" : "CSV"
}
```

where

columns	Indicates the schema of the CSV. There are two columns, rank at the first position and domain at the second position.
separator	Use a comma to separate the columns.
value_transform	For each row, transform each domain column by removing the TLD.
value_filter	Only consider non-empty domains.
state_init	Initialize the state, a long integer, to 0.

state_update	For each row in the CSV, update the state, which is the running partial sum, with the number of typosquatted domains for the domain.
state_merge	For each thread, we have a partial sum, we want to merge the partial sums into the total.

b) Run the extractor_count.json file:

```
$METRON_HOME/bin/flatfile_summarizer.sh -i ~/top-10k.csv -e ~/
extractor_count.json -p 5 -om CONSOLE
```

The output should look similar to the following:

```
WARN extractor.TransformFilterExtractorDecorator: Unable to setup
zookeeper client - zk_quorum url not provided. **This will limit some
Stellar functionality**

Processing /root/top-10k.csv
17/12/22 17:05:20 WARN resolver.BaseFunctionResolver: Using System
classloader
Processed 9999 - \
3496552
```

2. Generate the Bloom filter on HDFS.

a) Create an extractor_filter.json file at \$METRON_HOME/config and populate it with the following:

```
{
  "config" : {
    "columns" : {
      "rank" : 0,
      "domain" : 1
    },
    "value_transform" : {
      "domain" : "DOMAIN_REMOVE_TLD(domain)"
    },
    "value_filter" : "LENGTH(domain) > 0",
    "state_init" : "BLOOM_INIT(3496552, 0.001)",
    "state_update" : {
      "state" : "REDUCE( DOMAIN_TYPOSQUAT( domain ), (s, x) ->
BLOOM_ADD(s, x), state)"
    },
    "state_merge" : "BLOOM_MERGE(states)",
    "separator" : ", "
  },
  "extractor" : "CSV"
}
```

Most of the parameters are same as the extractor_count.json file, but there are three different parameters:

state_init	We have changed our state to be a bloom filter, initialized with: 3496552 - The size calculated in the previous step 0.001 - The false positive probability (0.1%)
state_update	Update the bloom filter (the state variable) with each typosquatted domain,
state_merge	Merge the bloom filters generated per thread into a final, single bloom filter to be written.

- b) Generate the Bloom filter in HDFS at /tmp/reference/alexa10k_filter.ser:

```
$METRON_HOME/bin/flatfile_summarizer.sh -i ~/top-10k.csv -o /tmp/
reference/alexa10k_filter.ser -e ~/extractor_filter.json -p 5 -om HDFS
```

3. Apply your new filter to domains from the squid telemetry.

- Display the Management UI.
- Select the Squid sensor from the list of sensors on the main window.
- Click the pencil icon in the list of tool icons



for the sensor.

The Management UI displays the Squid sensor panel.

- Click the **Advanced** button.
- Click



(expand window) next to the **RAW JSON** field.

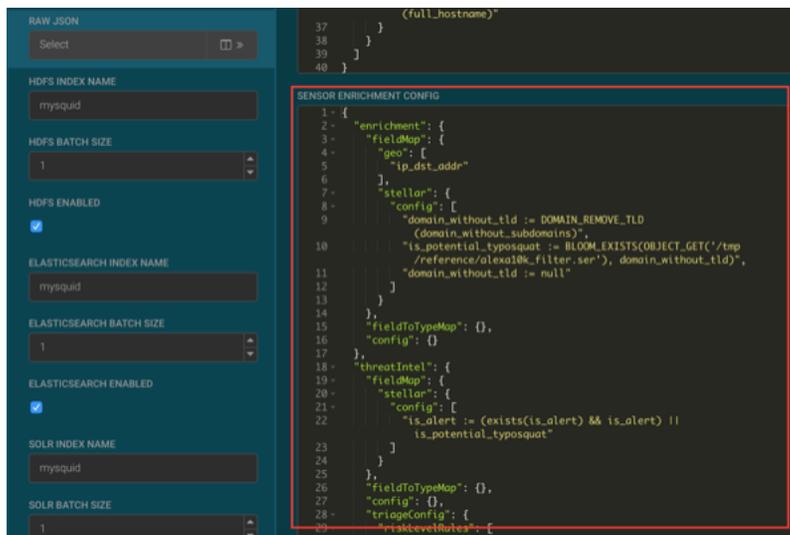
- Replace the JSON information in the **SENSOR ENRICHMENT CONFIG** section with the following JSON information:

```
{
  "enrichment": {
    "fieldMap": {
      "geo": [
        "ip_dst_addr"
      ],
      "stellar": {
        "config": [
          "domain_without_tld :=
DOMAIN_REMOVE_TLD(domain_without_subdomains)",
          "is_potential_typosquat := BLOOM_EXISTS(OBJECT_GET('/tmp/reference/
alexa10k_filter.ser'), domain_without_tld)",
          "domain_without_tld := null"
        ]
      }
    },
    "fieldToTypeMap": {},
    "config": {}
  },
  "threatIntel": {
    "fieldMap": {
      "stellar": {
        "config": [
          "is_alert := (exists(is_alert) && is_alert) ||
is_potential_typosquat"
        ]
      }
    },
    "fieldToTypeMap": {},
    "config": {},
    "triageConfig": {
      "riskLevelRules": [
        {
          "name": "Alexa 10k Typosquat Bloom",
```

```

    "comment": "Inspect a bloom filter with potentially typosquatted domains from the top Alexa 10k",
    "rule": "is_potential_typosquat != null && is_potential_typosquat",
    "score": 50,
    "reason": "FORMAT('%s is a potential typosquatted domain from the top 10k domains from alexa', domain_without_subdomains)"
  }
],
"aggregator": "MAX",
"aggregationConfig": {}
}
},
"configuration": {}
}

```



- g) Click **SAVE** below the JSON information.
 - h) Click **SAVE** at the bottom of the Squid sensor configuration panel.
4. After you identify a potential typosquatted domain, investigate it, and determined that it is legitimate, you can stop future alerts by using a domain whitelist enrichment.

- a) In the Management UI, click the pencil icon next to the mysquid sensor.
The Management UI displays the sensor configuration form.
- b) Click the **Advanced** button.
- c) Click



(expand window button) next to the **RAW JSON** field.

- d) Replace the **is_potential_typosquat** field value with the following:

```

"is_potential_typosquat := not (ENRICHMENT_EXISTS('domain_whitelist', domain_without_tld, 'enrichment', 't')) && BLOOM_EXISTS(OBJECT_GET('/tmp/reference/alexal0k_filter.ser'), domain_without_tld)",

```

RAW JSON

Select

HDFS INDEX NAME

mysquid

HDFS BATCH SIZE

1

HDFS ENABLED

ELASTICSEARCH INDEX NAME

mysquid

ELASTICSEARCH BATCH SIZE

1

ELASTICSEARCH ENABLED

SOLR INDEX NAME

mysquid

SENSOR ENRICHMENT CONFIG

```

37     }
38   }
39 }
40 }

1 - {
2 -   "enrichment": {
3 -     "fieldMap": {
4 -       "geo": [
5 -         "ip_dst_addr"
6 -       ],
7 -     "stellar": {
8 -       "config": [
9 -         "domain_without_tld := DOMAIN_REMOVE_TLD
10 -        (domain_without_subdomains)",
11 -        "is_potential_typosquat := not (ENRICHMENT_EXISTS
12 -        ('domain_whitelist', domain_without_tld, 'enrichment', 't'))
13 -        && BLOOM_EXISTS(OBJECT_GET('/tmp/reference/alexa10k_filter
14 -        .ser'), domain_without_tld)",
15 -        "domain_without_tld := null"
16 -      ]
17 -    }
18 -   },
19 -   "fieldToTypeMap": {},
20 -   "config": {}
21 - },
22 - "threatIntel": {
23 -   "fieldMap": {
24 -     "stellar": {
25 -       "config": [
26 -         "is_alert := (exists(is_alert) && is_alert) ||
27 -         is_potential_typosquat"

```

- e) Click **SAVE** below the JSON information.
 - f) Click **SAVE** at the bottom of the Squid sensor configuration panel.
5. Ensure that the results appear in the Alerts UI.
- a) Enter `cnn.com` or `nsp.com` in the browser connected to the HCP proxy.
 - b) Display the Alerts UI.

In the Score column, you should see events with non-zero scores and the `is_alert` field set to `true`.

APACHE METRON

Logged in as metron - Logout

Searches +

All time

Alerts (4667)

Filters

enrichm_country 8

ip_dst_addr 10

ip_src_addr 1

source_type 2

Group By

2 source_type

10 ip_dst_addr

8 enrichm_country

1 ip_src_addr

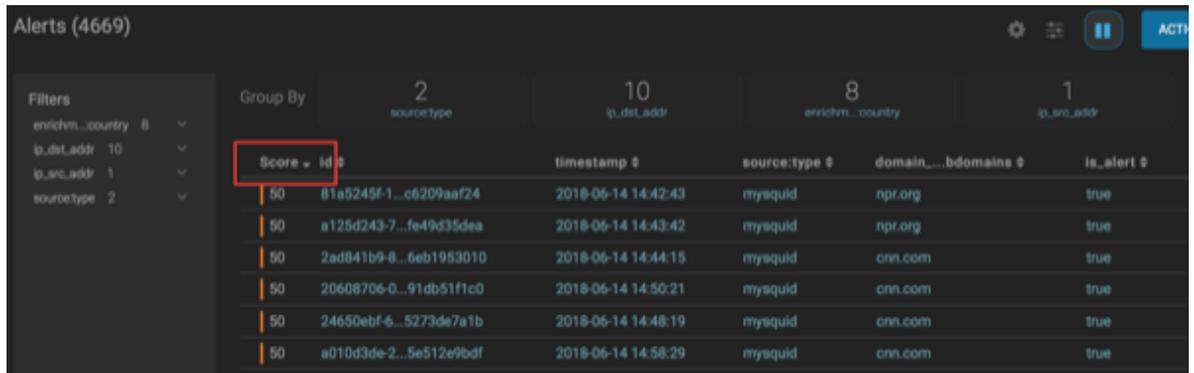
UnGroup

Score	id	timestamp	source-type	domain...bdomains	is_alert
50	f54a9b11-f...c1da799df5	2018-06-14 17:25:32	mysquid	google.com	true
50	4d170910-4...930f98ede7	2018-06-14 17:24:54	mysquid	cnn.com	true
50	5d70c10f-7...e176452002	2018-06-14 17:22:52	mysquid	cnn.com	true
50	244d5a3c-9...48be64503	2018-06-14 17:21:40	mysquid	google.com	true
50	6dbbe56c-7...8f68fba8dd	2018-06-14 17:20:50	mysquid	cnn.io	true
50	e6578da1-1...329d7b4af5	2018-06-14 17:20:50	mysquid	cnn.com	true
50	35b7debe-6...cae589e45d	2018-06-14 17:18:48	mysquid	cnn.com	true
50	182b0bb-d...74d072530c	2018-06-14 17:18:05	mysquid	scorecardr...search.com	true
50	f0ff801d-3...fda173fafa	2018-06-14 17:17:29	mysquid	google.com	true
50	8cbd8bdd-1...1f2d166394	2018-06-14 17:17:07	mysquid	google.com	true
50	03d5c1fc-d...93be2676bc	2018-06-14 17:16:59	mysquid	doubleclick.net	true

If you want to view the columns as they appear in the screen shot, click the gear icon to the left of the **Actions** button and unselect all fields except **Score**, **id**, **timestamp**, **source:type**, **domain_withoutsub_domains**, and **is_alert** fields, then click **Save**.

- c) Click the **Score** header to sort the events ascending by Score.

Click again to sort descending by Score. A downward arrow appears next to the **Score** header when sorted descending by Score.



The screenshot shows a table of alerts with the following columns: Score, id, timestamp, source.type, domain..., bdomains, and is_alert. The 'Score' column is highlighted with a red box. The table contains 6 rows of data, all with a score of 50 and is_alert set to true.

Score	id	timestamp	source.type	domain...	bdomains	is_alert
50	81a5245f-1...c6209aa24	2018-06-14 14:42:43	mysquid	npr.org		true
50	a125d243-7...fe49d35dea	2018-06-14 14:43:42	mysquid	npr.org		true
50	2ad841b9-8...6eb1953010	2018-06-14 14:44:15	mysquid	cnn.com		true
50	20608706-0...91db51f1c0	2018-06-14 14:50:21	mysquid	cnn.com		true
50	24650ebf-6...5273de7a1b	2018-06-14 14:48:19	mysquid	cnn.com		true
50	a010d3de-2...5e512e9bdf	2018-06-14 14:58:29	mysquid	cnn.com		true

- d) Click between the columns of one of the Scored alerts to view the alert details.

The fields beginning with **threat:triage:rules** show the results of all the triage rules. The **threat:triage:score** field is the aggregated score of the event. If there is more than one triage rule, this field will contain the score combining the results from all the rules. The **is_alert** field is set only if the triage rules indicate the event is an alert.

```
uat
method          CONNECT
source:type     mysquid
threat:trriage:rules:0
:comment       Inspect a bloom
               filter with
               potentially
               typosquatted
               domains from the
               top Alexa 10k
threat:trriage:rules:0
:name          Alexa 10k
               Typosquat Bloom
threat:trriage:rules:0
:reason       npr.org is a
               potential
               typosquatted
               domain from the
               top 10k domains
               from alexa
threat:trriage:rules:0
:score        50
threat:trriage:score
              50
timestamp      1528987363820
url            media.npr.org:443
```

- e) To see all the alerts for a particular domain, click the domain name.
The Alerts UI displays only the alerts with the selected domain name.

Searches: domain_without_subdomains:npr.org

Alerts (118)

Filters: enrichm...country 1, ip_dst_addr 1, ip_src_addr 1, source.type 1

Group By: source.type, ip_dst_addr, enrichm...country, ip_src_addr

Score	id	timestamp	source.type	domain...bdomains	is_alert
50	81a5245f-1...c6209aaf24	2018-06-14 14:42:43	mysquid	npr.org	true
50	a125d243-7...fe49d35dea	2018-06-14 14:43:42	mysquid	npr.org	true
50	20226cf0-1...8450c06c4a	2018-06-14 15:13:46	mysquid	npr.org	true
50	5585a502-5...50b8c811ad	2018-06-14 16:13:58	mysquid	npr.org	true
50	72d8bc08-6...a0f381d4ae	2018-06-14 16:44:01	mysquid	npr.org	true
50	2d6a1b69-e...7aaac38e5	2018-06-14 17:14:06	mysquid	npr.org	true
50	b7acca5f-9...f6e34895a0	2018-06-14 17:43:13	mysquid	npr.org	true
50	c94815d7-6...4b4b241fdd	2018-06-14 14:42:43	mysquid	npr.org	true
50	84263126-8...c353ee117e	2018-06-14 15:42:52	mysquid	npr.org	true
50	ce51fcad-3...a4ecd33d10	2018-06-14 15:42:52	mysquid	npr.org	true
50	25d3169a-e...d8c9c71827	2018-06-14 16:12:58	mysquid	npr.org	true
50	66cd12c3-d...f0748ff8df	2018-06-14 17:43:13	mysquid	npr.org	true
50	e26bbf78-3...6ba57d8283	2018-06-14 17:43:13	mysquid	npr.org	true

- f) To remove a filter, click **x** next to the filter.
To view all events, click **x** on the Searches field.

Searches: domain_without_subdomains:npr.org

Alerts (119)

Filters: enrichm...country 1, ip_dst_addr 1, ip_src_addr 1, source.type 1

Group By: source.type, ip_dst_addr

Score	id	timestamp
50	81a5245f-1...c6209aaf24	2018-06-14 14:42:43
50	a125d243-7...fe49d35dea	2018-06-14 14:43:42

Improve Scoring with a Domain Whitelist

Once you have identified and investigated a potential typosquatted domain and found that it is legitimate, you can stop future alerts by using a domain whitelist enrichment.

Procedure

1. Display the Management module UI.
2. Select the Squid sensor from the list of sensors on the main window.

- Click the pencil icon in the list of tool icons



for the Squid sensor.

- Click **Advanced**.

- Click



(expand window button) next to the **RAW JSON** field.

The screenshot shows a configuration interface for a sensor. On the left, there are several configuration sections:

- RAW JSON**: A dropdown menu with 'Select' and an expand button (square with arrow) highlighted with a red box.
- HDFS INDEX NAME**: A text input field containing 'mysquid'.
- HDFS BATCH SIZE**: A dropdown menu with '1' selected.
- HDFS ENABLED**: A checked checkbox.
- ELASTICSEARCH INDEX NAME**: A text input field containing 'mysquid'.
- ELASTICSEARCH BATCH SIZE**: A dropdown menu with '1' selected.
- ELASTICSEARCH ENABLED**: A checked checkbox.

The main area displays the **SENSOR ENRICHMENT CONFIG** field, which is highlighted with a red box. It contains the following JSON configuration:

```

1 - {
2 -   "enrichment": {
3 -     "fieldMap": {
4 -       "geo": [
5 -         "ip_dst_addr"
6 -       ],
7 -     },
8 -     "fieldToTypeMap": {},
9 -     "config": {}
10 -  },
11 -   "threatIntel": {
12 -     "fieldMap": {},
13 -     "fieldToTypeMap": {},
14 -     "config": {},
15 -     "triageConfig": {
16 -       "riskLevelRules": [],
17 -       "aggregator": "MAX",
18 -       "aggregationConfig": {}
19 -     }
20 -   },
21 -   "configuration": {}
22 - }

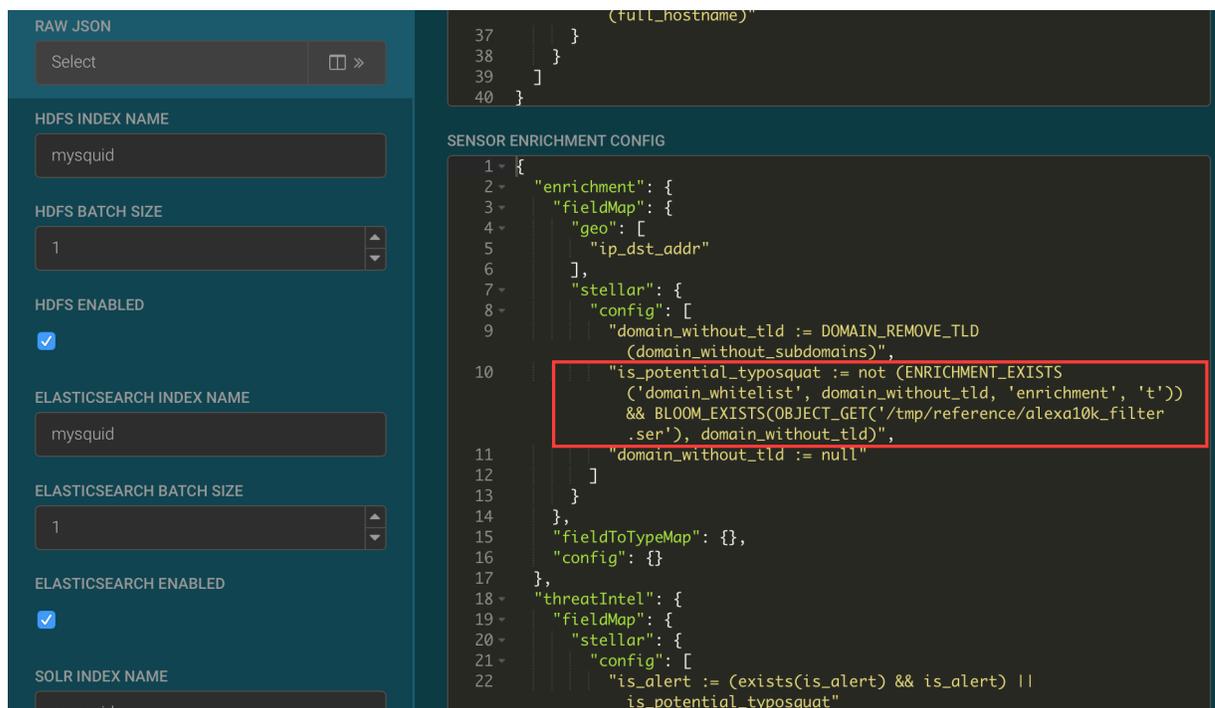
```

- Replace the `is_potential_typosquat` information with the following:

```

"is_potential_typosquat := not (ENRICHMENT_EXISTS('domain_whitelist',
domain_without_tld, 'enrichment', 't')) && BLOOM_EXISTS(OBJECT_GET('/tmp/
reference/alexal0k_filter.ser'), domain_without_tld)",

```



The screenshot displays the HCP configuration interface. On the left, there is a sidebar with configuration options for HDFS and Elasticsearch. The main panel shows the "SENSOR ENRICHMENT CONFIG" section, which contains a JSON configuration. A red box highlights the following configuration line:

```
10 "is_potential_typosquat := not (ENRICHMENT_EXISTS  
    ('domain_whitelist', domain_without_tld, 'enrichment', 't'))  
    && BLOOM_EXISTS(OBJECT_GET('/tmp/reference/alexa10k_filter  
        .ser'), domain_without_tld)",
```

7. Click **SAVE** below the JSON panel.
8. Click **SAVE** at the bottom of the Squid sensor configuration panel.
9. Open cnn.com or npr.com in the browser connected to the HCP proxy.
10. Open the Alerts UI.
11. Click on the **timestamp** column header until the events are sorted descending by timestamp.
Proxy events to cnn.com and npr.org are no longer alerts.