



Exploring VirusTotal for security operations alert triage automation

Teemu Kantola

Bachelor's thesis

February 2022

Information and Communication Technologies

Bachelor's Degree Programme in Information and Communications Technology

Kantola Teemu

Exploring VirusTotal for security operations alert triage automation

Jyväskylä: JAMK University of Applied Sciences, February 2022, 38 pages

Information and Communication Technologies. Bachelor's Degree Programme in Information and Communications Technology.

Permission for web publication: Yes

Language of publication: English

Abstract

Security operations centres (SOC) of today must deal with an ever-increasing number of security alerts. Junior security analysts are often first in line to triage security alerts. Alert triage often requires performing similar tasks quickly and repeatedly, which makes the job stressful and monotonous. This results in high turnover rate in SOCs. One solution to mitigate this issue is to increase automation in this process.

Thesis was assigned by Nixu Oyj. The objective of the thesis was to examine how VirusTotal service can be used as a part of Security Orchestration, Automation, and Response (SOAR) platform's alert triage automation workflow. The requirements for the solution were that it reduces manual tasks security analysts must perform when triaging alerts and improve both incident response times and threat recognition accuracy. The objective was approached using a constructive method. A fictitious security alert was first triaged manually, and this was then used as basis for designing a workflow for SOAR to automate the same process.

The proposed SOAR automation first enriches indicators of compromise (IOC) associated with a security alert with data provided by VirusTotal. The enriched data is then analysed to assign a reputation value to the IOCs. The analysis results are presented to security analysts in a concise human-readable way. IOCs determined to be likely related to malicious activity are automatically searched from customer environment, and if any are found, SOAR suggests incident response escalation to security analysts.

The proposed automation was evaluated using a survey. The respondents found that it has potential to take over their task of gathering and analysing information from VirusTotal during alert triage, and it would have a positive effect in incident response times and their ability to recognise threats. Based on the survey, the automation workflow would add noticeable benefits to the assignor's SOC when implemented.

Keywords/tags (subjects)

alert triage, cyber security, incident analysis, indicator of compromise, ioc, security automation, security operations, security operations centre, soar, soc, virustotal

Miscellaneous (Confidential information)

Kantola Teemu

VirusTotal-verkkopalvelun hyödyntäminen tietoturvahälytysten ensikäsittelyn automatisoinnissa

Jyväskylä: Jyväskylän Ammattikorkeakoulu, helmikuu 2022, 38 sivua

Tietojenkäsittely ja tietoliikenne. Tieto- ja viestintätekniikka (AMK).

Verkkojulkaisulupa myönnetty: Kyllä

Julkaisun kieli: englanti

Tiivistelmä

Nykypäivän tietoturvalavomot joutuvat vastaanottamaan yhä enemmän ja enemmän tietoturvahälytyksiä. Hälytysten ensikäsittelystä vastaavat usein vasta-alkaneet tietoturva-analyttikot. Ensikäsittely vaatii nopeasti ja toistuvasti samojen tehtävien suorittamista, mikä tekee työstä stressaavaa ja monotoonista. Tämä johtaa siihen, että tietoturva-analyttikoiden vaihtuvuus valvomoissa on hyvin yleistä. Ratkaisuna tähän on ehdotettu ensikäsittelyn automatisointia.

Opinnäytetyön toimeksiantajana oli Nixu Oyj. Työn tavoitteena oli tutkia, kuinka VirusTotal-verkkopalvelua voisi hyödyntää tietoturvahälytysten ensikäsittelyn automatisoinnissa Security Orchestration, Automation, and Response (SOAR) -alustalla. Ratkaisuna kehitettiin suunnitelma SOAR-automaatiosta, jonka tarkoituksena on korvata tietoturva-analyttikoiden manuaalista työtä ensikäsittelyvaiheessa, vähentää tietoturva-poikkeamien hallinnan vasteaika, ja parantaa kykyä tunnistaa uhkia. Tavoitetta lähestyttiin menetelmällä, jossa SOAR-automaatio suunniteltiin fiktiivisen tietoturvahälytyksen ensikäsittelyn pohjalta.

Suunniteltu automaatio rikastaa tietoturvahälytyksiin liittyviä uhkaindikaattoreita VirusTotal-verkkopalvelusta saatavilla tiedoilla, jonka jälkeen analysoi rikastetun datan. Uhkaindikaattoreille asetetaan analyysin perusteella arvo, mikä kertoo, kuinka todennäköisesti se liittyy haitalliseen toimintaan. Analyysien tulokset esitetään tietoturva-analyttikoille selkokielellä. SOAR hakee mahdollisesti haitalliseen toimintaan liittyvät uhkaindikaattorit automaattisesti asiakkaan verkkoympäristöstä, ja näitä löytäessään ehdottaa tietoturva-analyttikolle hälytyksen eskalointia tietoturvapoikkeaman hallintaan.

Suunnitellun automaation toimivuutta arvioitiin kyselyn avulla. Vastaajat olivat sitä mieltä, että ratkaisulla on mahdollista korvata tietoturva-analyttikon manuaalinen tiedonkeruu VirusTotal-palvelusta ensikäsittelyvaiheessa. Vastaajat ajattelivat ratkaisun vähentävän tietoturvapoikkeamien reagointiin kuluva aikaa sekä parantavan uhkientunnistuskkyä. Kyselyn tulosten perusteella järjestelmän käyttöönotto toisi huomattavaa lisäarvoa tilaajan tietoturvalavomolle.

Avainsanat (asiasanat)

ensikäsittelyn automatisointi, hälytysten ensikäsittely, hälytysten luokittelu, ioc, kompromissin indikaattori, kyberturvakeskus, soar, soc, tietoturva, tietoturvan automatisointi, tietoturvalavonta, virustotal

Muut tiedot (salassa pidettävät liitteet)

Contents

1	Introduction	3
2	Research objective.....	5
3	Indicators of compromise	5
3.1	Appearance within a cyberattack	6
4	VirusTotal.....	9
5	Manual alert triage	13
6	Automated alert triage	18
6.1	Enriching IOCs.....	18
6.2	Analysing IOCs	20
6.2.1	Dependency problem in analysis rules	22
6.2.2	Displaying analysis results	23
6.3	Triaging alert	25
7	Results.....	26
8	Conclusion	29
8.1	Further development	30
	References	31
	Appendices	34
	Appendix 1. Python script to test effectiveness of dependency solving algorithm	34
	Appendix 2. Questionnaire form.....	35

Figures

Figure 1.	Pyramid of pain for attackers and defenders alike (adapted from Bianco, 2014)	6
Figure 2.	Amount of items processed ([VirusTotal statistics page], 2021)	10
Figure 3.	Sandbox and antivirus detections ([VirusTotal page on file 113569686], n.d.)	14
Figure 4.	Capabilities and indicators section ([VirusTotal page on file 113569686], n.d.)	15
Figure 5.	Snippet from VBA macros ([VirusTotal page on file 113569686], n.d.).....	15
Figure 6.	Document preview ([VirusTotal page on file 113569686], n.d.)	15
Figure 7.	Antivirus engine detections ([VirusTotal page on file M1cro.exe], n.d.).....	16
Figure 8.	Matching YARA rules ([VirusTotal page on file M1cro.exe], n.d.)	17
Figure 9.	Process tree ([VirusTotal page on file M1cro.exe], n.d.)	17
Figure 10.	Automation workflow high-level view.....	18
Figure 11.	Automation workflow for IOC enrichment.....	19
Figure 12.	Issue with cyclic dependency from relationships	22

Figure 13. Automation workflow for IOC analysis	22
Figure 14. Algorithmic efficiency of dependency solving algorithm for IOC analysis	23
Figure 15. Formatted analysis results for domain nikohulis[.]com	24
Figure 16. Formatted analysis results for file 113569686	24
Figure 17. Formatted analysis results for file M1cro.exe	25
Figure 18. Analysis rule match with a positive effect on an IOCs reputation.....	25
Figure 19. Automation workflow for triage automation	26
Figure 20. Search results for domain nikohulis[.]com in customer environment	26
Figure 21. How often VirusTotal is used for alert triaging?	27
Figure 22. How fatiguing VirusTotal usage is for alert triage?.....	27
Figure 23. Automation's potential to take over manual work	27
Figure 24. Automation's potential to improve incident response times.....	28
Figure 25. Automation's potential to improve threat recognition.....	28
Figure 26. How likely information produced by automation is to be manually confirmed?	29

Tables

Table 1. Historical DNS data ([VirusTotal page on IP address 45.151.144.162], n.d.).....	13
Table 2. Related objects for M1cro.exe	20
Table 3. Sample IOC analysis rules.....	21
Table 4. Analysis rule weight sum to reputation value mapping	21

1 Introduction

Malicious cyber activity is becoming more common. SonicWall (2021) research found ransomware attacks to have risen 66% between 2019 and 2020. IBM Security and Ponemon Institute (2020) joint study estimated the average cost of a data breach to be \$3.86 million based on interviews of 524 organisations suffering a data breach between August 2019 and April 2020. A cyberattack may have a devastating financial and reputational impact on an organisation. Full financial effect is often not immediate but gets exacerbated over time if partner, customer, or public trust is broken. Disruption in operations, investigation and recovery processes, increased insurance costs, potential fines, layoffs, and security solutions which are applied to prevent future attacks, all add to the costs of remediation. Organisations should take pre-emptive measures and start building a comprehensive cyber security strategy before an attack has taken place. Time and money need to be invested to ensure that company, employee, and customer information is protected.

Censuswide and Cybereason (2021) research found that a security operations centre (SOC) was the second most invested security solution for organisations post-ransomware attack. A SOC is an essential component in an organisation's cyber security strategy (Gartner, n.d.). Responsibilities of a SOC include training of staff, management of security tooling, log data collection and analysis for potential cybersecurity threats, and incident response (European Union Agency for Cybersecurity [ENISA], 2020). With these responsibilities, the primary function of a SOC is to reduce the risk of a successful cyberattack by preventing, detecting, and analysing cybersecurity threats, and minimise the damage upon breaches. A SOC may be internal or outsourced to a managed security services provider, or a combination of these with shared responsibilities (IANS Faculty, 2021). Outsourcing may be especially compelling option for small to medium-sized organisations as it most likely comes at a lesser cost, since the expertise, processes, and technologies are already well established (IANS Faculty, 2021).

Issue in security operations centres

Organisations' digital estate is growing rapidly due to digitalisation, expanding their attack surface (Gartner, n.d.). This means increasing amounts of log data and overwhelming amounts of (often similar) alerts for SOC. First response to alerts is often the responsibility of junior security analysts working in shifts around the clock (Zimmerman, 2014). Long-term work on monotonous tasks is

demotivating, may cause feelings of stagnation in personal development, reduced efficiency, and burnout, which leads to high turnover rate (Sundaramurthy, 2017). Thus, there is a short supply of people with years of expertise in this role. This is an issue for SOC. One approach to mitigate this issue is to automate monotonous tasks. This would leave security analysts with more interesting and challenging tasks where human ingenuity is required.

Automation in security operations centres

Ponemon Institute (2019) survey-based research on SOC effectiveness touched the subject of security analyst burnout, where 67% of the respondents consider increased workflow automation to alleviate issues leading to burnout. In addition to solving personnel issues, automation has been found to greatly improve monetary value of SOC. IBM Security & Ponemon Institute (2020) study presents correlation between lesser costs in data breach remediation and higher maturity in automation of threat identification and containment. The study found that the costs were 31.8% less for organisations with fully deployed automation than those with only partial automation, and 59.4% less than for those without automation. These results were based on rough estimations from the interviewees, and thus are only indicative.

Modern SOC have a security orchestration, automation, and response (SOAR) platform as the central component of their operations. SOAR is a technology that integrates and coordinates different security solutions and is used to create highly automated workflows that use the solutions to perform SOC tasks. The purpose of SOAR is to streamline security operations, such as security analysts' daily tasks. SOAR enables performing complex tasks across IT systems and security solutions with minimal human interaction. (Imam, 2019)

When SOC receives an alert, the first response is to perform so called alert triage (Zimmerman, 2014). In security operations terms triage means to determine if an alert should be dismissed as false positive or escalated to incident response and at what priority. Triage often requires performing the same repetitive tasks continuously. Many tasks related to alert triage can potentially be either fully or partly automated using SOAR. Automating as much as possible of this process is crucial for SOC: reduced fatigue for security analysts, faster response times to security incidents, and improvement in threat recognition accuracy.

2 Research objective

The thesis was assigned by Nixu Oyj, a cybersecurity company based in Finland. Objective of the thesis is to examine VirusTotal service and how to use it as part of SOAR alert triage automation for their SOC. Can the service be utilised automatically at such level that it can take over security analyst actions during triage? Can it improve incident response times and threat recognition accuracy? Research emphasises on efficient utilisation of the service; how to get the most value out of the data it provides from the perspective of a security analyst? As a result of the thesis, a SOAR automation workflow is proposed that attempts to answer these questions. The proposal is evaluated by surveying security analysts of the assignor's SOC. No identifying information is collected to ensure respondent anonymity.

Thesis first introduces indicators of compromise (IOC), an important concept for alert triage and VirusTotal. This is followed by an introduction to VirusTotal service itself and its application programming interface (API) that is used to programmatically interact with the service. Thesis objective is approached using a constructive method. A fictitious security alert is first triaged manually using VirusTotal, and that is then used as a basis for a proposal to automate the same process using SOAR. The plan should, when implemented, reduce monotonous tasks from security analysts, and allow them to triage alerts faster and recognise threats more accurately.

3 Indicators of compromise

Alert triage often involves collection and analysis of forensic evidence. IOCs are a type of forensic evidence which, as their name suggest, **might** indicate compromise of an asset or assets (Brata Deb, et al. 2018). They may be anything from simple artefacts such as file hashes and domain names, to an abnormal chain of processes on a workstation started from opening an email attachment. They are an essential piece in alert triage, allowing security analyst to form a more complete picture of the potential threat. Successfully recognising IOCs and their relation to a threat allow for prompt and proper incident response.

3.1 Appearance within a cyberattack

Bianco (2014) coined an aptly named pyramid of pain concept. The pyramid illustrates how challenging it is for adversary operations when defenders detect certain types of IOCs, but at the same time how difficult it is for defenders to apply detection logic for them as well (Figure 1). Brata Deb et al. (2018) categorise IOCs into network, system, static and behavioural. However, their distinction is confusing as system and network IOCs are behavioural by their definition. There appears to be no set standards for IOC categories and terms in the cybersecurity space. For example, cybersecurity company CrowdStrike uses the term indicator of attack to denote behavioural IOCs (IOA VS IOC, 2021). This thesis uses IOC definitions from the pyramid of pain, and further separates them into behavioural and static categories.

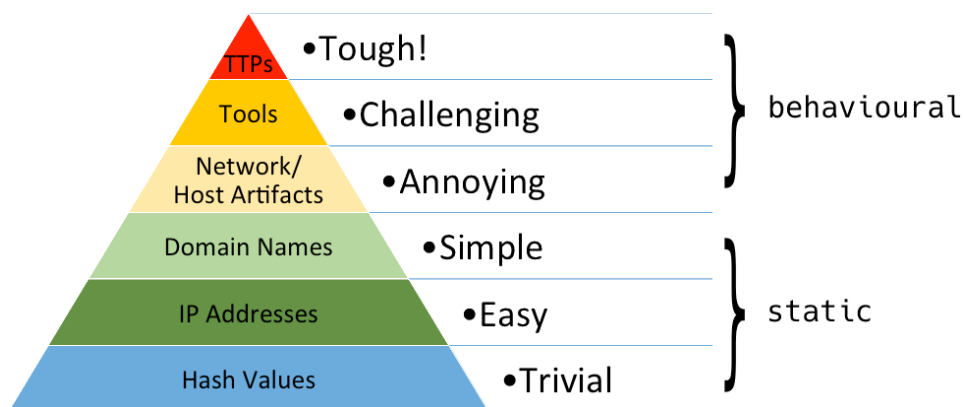


Figure 1. Pyramid of pain for attackers and defenders alike (adapted from Bianco, 2014)

The pyramid draws a clear distinction between behavioural and static IOCs. Behavioural IOCs can be chain of commands an adversary entered on an endpoint after initial access, tools they might use, artefacts written to disk or registry modifications by those tools, or traces of activity left in web proxy logs. Detecting (and especially preventing) these is difficult but may force adversaries to give up or research new strategies. Static IOCs such as hash values, IP addresses and domain names on the other hand are very easy to detect and prevent but are as trivial for adversaries to change (Bianco, 2014).

To demonstrate how different types of IOCs from the pyramid appear in an attack, a real-life scenario is used as a reference. In the beginning of 2021, Microsoft observed threat actor HAFNIUM

exploiting critical zero-day vulnerabilities against on-premises Microsoft Exchange Servers (Microsoft 365 Defender Threat Intelligence Team et al., 2021).

TTPs

At the peak of the pyramid of pain are tactics, techniques, and procedures (TTP) which describe adversary behaviour within a kill chain. National Institute of Standards and Technology guideline on cyber threat information sharing define TTP as:

The behavior of an actor. A tactic is the highest-level description of this behavior, while techniques give a more detailed description of behavior in the context of a tactic, and procedures an even lower-level, highly detailed description in the context of a technique. (Badger et al., 2016)

Once HAFNIUM had gained foothold on their target, they used a legitimate Windows Sysinternals utility ProcDump to dump memory of LSASS process (Microsoft 365 Defender Threat Intelligence Team et al., 2021). LSASS is a Windows system process which stores credentials in memory – a great resource for adversaries looking to escalate privileges or move laterally within the environment. The cyber security community often maps these procedures into tactics and techniques defined in MITRE ATT&CK® framework. This procedure can be mapped to tactic Credential Access and technique OS Credential Dumping with sub-technique LSASS Memory (Enterprise Matrix, 2021). In short, HAFNIUM gathered credentials by dumping LSASS process memory.

Tools

The same event could be viewed from another perspective – usage of a specific tool within a procedure i.e., the second topmost level of the pyramid. This is much narrower point of view, covering less attack surface. HAFNIUM could have switched from ProcDump to some obscure process dumping tool, modify an existing one, or even create a new one if motivated enough. Monitoring the usage of specific process dumping tools can be worthwhile – it can produce very accurate alerts with low false positive rates. Detecting the most common process dumping tools may be far more cost and time effective than attempting to create detection logic that covers it all.

Network artefacts

Adversaries often require network communications to achieve their goal. This means that their actions leave traces in logs of network devices such as firewalls, web servers, proxies, domain name system (DNS) servers, and intrusion detection systems. HAFNIUM used reverse shells written in PowerShell as their method of establishing a command-and-control channel (Microsoft 365 Defender Threat Intelligence Team et al., 2021). Reverse shell usage leaves network artefacts which may be detected by monitoring network traffic for anomalous patterns.

Host artefacts

Adversary actions leave traces on endpoints as well, and their behaviour is likely different from normal end user activity. These artefacts may be processes, registry modifications, created files and directories, scheduled tasks and services, etc. HAFNIUM used a portable version of 7-Zip file archiver tool from the root of C:\ProgramData\ directory to prepare data for exfiltration (Microsoft 365 Defender Threat Intelligence Team et al., 2021). It is highly unusual (but not unheard of) for legitimate processes or scripts to be executed from ProgramData directory root.

Static IOCs

The lowest levels of the pyramid have static IOCs. In attacks where the adversary tools and infrastructure are previously unknown, it's not possible to detect their actions based solely on hash, domain, and IP address values. Knowledge on these IOCs is always gained after the fact, and thus are only useful in detecting ongoing and future attacks. If these IOCs are known before an attack takes place, they can produce very precise alerts with low false positive rates.

Static IOCs can be used as a part of behavioural IOCs. HAFNIUM used PowerShell to download powercat tool from GitHub domain raw.githubusercontent.com (Microsoft 365 Defender Threat Intelligence Team et al., 2021). This domain is likely too common to alert on its own, but a detection rule can potentially be created if combined with additional logic. For example, detection rule could check that initiating process must be a Windows command interpreter.

Static IOCs can also prove to be very useful when determining the extents of security incidents. For example, device X may have alerted SOC due to malicious PowerShell script which has contacted

some unknown IP address. Devices Y and Z may be similarly compromised but lack endpoint visibility unlike device X, and thus do not alert. It's possible that these devices can be identified from firewall logs for contacting the same unknown IP address.

Behavioural IOCs are clearly far more efficient than static IOCs for detection purposes. However, alerts based on behavioural IOCs are often accompanied with related static IOCs. During alert triage, security analysts often use third-party services such as VirusTotal to gather information on them to contextualise alerts.

4 VirusTotal

VirusTotal is a web service for analysing files and uniform resource locators (URL) for threats and a crowdsourced threat intelligence platform. The service was founded in 2004 by cybersecurity company Hispasec Sistemas, acquired by Google in 2012, and transferred to Chronicle in 2018 (Frequently Asked Questions, n.d.). Since then, the service has grown to be a massive collaborative platform between Chronicle, security product vendors, security researchers, and the open-source community (Contributors, n.d.).

Today VirusTotal is a feature-packed tool utilized in various cyber security related tasks. Threat intelligence analysts are likely to find it useful for mapping threat campaigns and threat actor infrastructures, security analysts can use it to contextualise often bare-bones alerts, forensic analysts may be able to find essential artefacts which are no longer available on compromised devices, and malware researchers have it as a library of malware from which new and old samples can be downloaded for offline studies. It is a valuable tool for the cyber security community and companies.

VirusTotal collects information on files using contributed antivirus products, sandbox products and characterisation tools (Contributors, n.d.). Intrusion detection systems (IDS) equipped with crowdsourced rules are used to monitor network traffic during sandbox execution (Crowdsourced IDS Rules, n.d.). Sysmon logs produced during sandbox execution are scanned with crowdsourced Sigma rules (Sigma rules, n.d.). Files are also matched against crowdsourced YARA rules to identify malware attributes and binary patterns (Crowdsourced YARA rules, n.d.).

VirusTotal collects information on URLs using contributed URL scanners (Contributors, n.d.). URLs, domains, and IP addresses are matched against various contributed threat feeds and datasets (Contributors, n.d.).

In addition to collecting lots of information on files, URLs, domains, and IP addresses, VirusTotal also stores information about how all these items are related. This is a key feature. It allows users uncover much more information about threats. For example, a submitted malware sample may be linked with information about its command-and-control server locations, other malware it may have dropped during its execution, or URLs it has been served from.

It is advantageous to be a part of a cloud computing giant when providing a service of this scale.

Figure 2 displays statistics on the amounts of daily files and URLs processed by VirusTotal. Around 1 million new unique files and URLs are submitted per day.

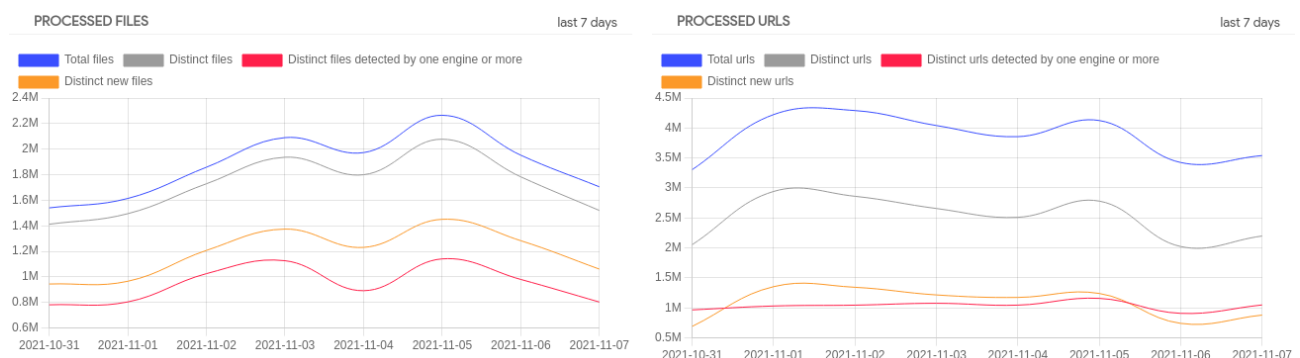


Figure 2. Amount of items processed ([VirusTotal statistics page], 2021)

VirusTotal exposes an extensive public API that allows users and developers to programmatically interact with the web service using hypertext transfer protocol. Its feature-set reflects most of the web interfaces capabilities. A key is required to authenticate API requests; thus, it is only available to registered users. Some other features are available only for paying customers via VirusTotal Enterprise subscription. ([VirusTotal API documentation], n.d.)

The core functionality in the API revolves around three concepts: **objects**, **collections**, and **relationships**. Objects are items in the VirusTotal database such as files and URLs, and collections are

sets of objects of common type. Relationships are two-way links that describe how objects are related to each other. For example, a file object might be linked to a URL object by “in the wild URLs” relationship, meaning that this file has been downloaded from that URL at some point. In turn, the URL object is linked to the file object by “downloaded files” relationship, meaning that this URL has served that file at some point. ([VirusTotal API documentation], n.d.)

VirusTotal API supports many features, but for the purposes of this thesis the scope is restricted to a small subset. Only endpoints that return data about objects representing IOCs are considered.

Detailed information about individual object is requested from endpoint ([VirusTotal API documentation], n.d.):

```
/api/v3/{collection name}/{object id}

{
  "data": {
    "attributes": {
      ..
    },
    "type": "{object type}",
    "id": "{object id}",
    "links": {
      "self": "https://www.virustotal.com/api/v3/{collection name}/{object id}"
    }
  }
}
```

Most VirusTotal API responses are in JavaScript Object Notation (JSON) format. The *attributes* property contains the object data, and the *id* and *type* properties are the object identifier and its type. For example, a file object has such attributes as different types of hashes, the latest antivirus scan results, crowdsourced YARA and IDS rule matches. An optional *relationships* query parameter can be supplied to retrieve information about an object’s relations as well. This adds an additional *relationships* property within the *data* property ([VirusTotal API documentation], n.d.):

```
/api/v3/{collection name}/{object id}?relationships={relationship 1},{relationship 2}

"relationships": {
  "{relationship 1}": {
    "meta": {
      "cursor": "{cursor value}"
    },
    "data": [
      {
        "type": "{related object type}",
        "id": "{related object id}"
      },
      ..
    ],
    "links": {
      "self": "https://www.virustotal.com/api/v3/{collection name}/{object id}"
    }
  }
}
```

```

id}/relationships/{relationship 1}?limit=20",
  "related": "https://www.virustotal.com/api/v3/{collection name}/{object
id}/{relationship 1}",
  "next": "https://www.virustotal.com/api/v3/{collection name}/{object
id}/relationships/{relationship 1}?cursor={cursor value}&limit=20"
},
"{relationship 2}": {
  ..
}
}

```

The number of relationships that can be supplied is not limited – all relationships of interest can be requested at once. However, the *relationships* property contains only **descriptors** (identifier and type) for related objects, and at most 20 are returned per relationship. All supplied relationships must be applicable to the object type, otherwise an error is returned. Detailed information for related objects is requested from individual relationship endpoints ([VirusTotal API documentation], n.d.).

```

/api/v3/{collection name}/{object id}/{relationship}
{
  "meta": {
    "count": {total number of objects},
    "cursor": "{cursor value}"
  },
  "data": [
    {
      "attributes": {
        ..
      },
      "type": "{related object type}",
      "id": "{related object id}",
      "links": {
        "self": "https://www.virustotal.com/api/v3/{collection name}/{related
object id}"
      }
    },
    ..
  ],
  "links": {
    "self": "https://www.virustotal.com/api/v3/files/{object
id}/{relationship}?limit=20",
    "next": "https://www.virustotal.com/api/v3/files/{object
id}/{relationship}?cursor={cursor value}&limit=20"
  }
}

```

Relationship endpoints return maximum of 20 objects by default, but an optional *limit* query parameter can be supplied to adjust this ([VirusTotal API documentation], n.d.). The maximum value that this can be set to is 40 based on testing. More objects can be requested iteratively by supplying a *cursor* query parameter in subsequent requests ([VirusTotal API documentation], n.d.). A value for *cursor* is provided in the responses within the *meta* property.

An undocumented feature was discovered for relationship endpoints. The *relationships* query parameter is valid for relationship endpoints as well. This adds a *relationships* property for all returned objects. Exception to this are certain relationship endpoints that return different types of objects. For example, *network_location* relationship for URL objects can return either a domain object or an IP address object. In this case only relationships common to both domain and IP address object types should be requested, or the API may respond with an error.

5 Manual alert triage

This chapter examines how a security analyst might use VirusTotal web interface for alert triage. A fictitious security alert that a SOC might receive is used for this purpose. The security alert concerns an anomaly in DNS traffic within customer network: on 17th of September 2021, multiple internal hosts have attempted to resolve a newly registered domain **nikohulis[.]com** in a short span of time. Domain registration is easy, cheap, and quick, and because of this threat actors tend to abuse the system profusely. Chen et al. (2019) from Unit 42 team of Palo Alto Networks have researched newly registered domains and found over 70% to be malicious or at least suspicious.

Investigating domain nikohulis[.]com in VirusTotal shows that it has been registered just over a week ago on 7th of September, and no domain scanners or datasets consider it malicious or suspicious ([VirusTotal page on domain nikohulis[.]com], n.d.). However, the domain registrar is Eranet International Limited. This is a registrar providing the previously mentioned easy, cheap, and quick domain registration service ([Eranet domain registration page], n.d.). DNS records show that it is currently hosted in Russia at IP address 45.151.144[.]162. VirusTotal stores passive DNS replication data, historical information about domains an IP address has resolved to (Martinez, 2013). Table 1 lists domains that have been observed resolving to 45.151.144[.]162. It certainly doesn't seem to be a stranger to phishing.

Table 1. Historical DNS data ([VirusTotal page on IP address 45.151.144.162], n.d.)

Date resolved	Domain
2021-09-02	nikohulis[.]com
2021-09-02	milojeriks[.]com
2021-09-02	thecomerica[.]com

2021-09-02 lcomerica[.]com
 2021-09-02 bankicomerica[.]com
 2021-09-02 bankingcomerica[.]com

Domain nikohulis[.]com file relations show that an Office document with antivirus detections has communicated with it ([VirusTotal page on domain nikohulis[.]com], n.d.). The document was submitted with name **113569686** and it has a SHA256 hash value of a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f. It is detected as malware by VMRay sandbox and 19 out of 63 antivirus engines (Figure 3).

Dynamic Analysis Sandbox Detections	
⚠ The sandbox VMRay flags this file as: MALWARE	
Security vendors' analysis on 2021-09-01T18:27:07 ▾	
Ad-Aware	⚠ VBA:Amphitryon.3724
Arcabit	⚠ VBA:Amphitryon.DE8C
Avira (no cloud)	⚠ W97M/Dldr.Sload.zbqps
BitDefender	⚠ VBA:Amphitryon.3724
ClamAV	⚠ Doc.Malware.Valyria-6749505-0
Cynet	⚠ Malicious (score: 99)
Elastic	⚠ Malicious (high Confidence)
Emsisoft	⚠ VBA:Amphitryon.3724 (B)
eScan	⚠ VBA:Amphitryon.3724
FireEye	⚠ VBA:Amphitryon.3724
GData	⚠ VBA:Amphitryon.3724
Kaspersky	⚠ UDS:DangerousObject.Multi.Generic
MAX	⚠ Malware (ai Score=89)
McAfee-GW-Edition	⚠ BehavesLike.Downloader.dc
NANO-Antivirus	⚠ Trojan.Ole2.Vbs-heuristic.druvzi
Sangfor Engine Zero	⚠ Malware.Generic-VBA.Save.Obfuscated
SentinelOne (Static ML)	⚠ Static AI - Malicious OPENXML
Tencent	⚠ Heur.Macro.Generic.e.c5a067d
Zoner	⚠ Probably Heur.W97Obfuscated

Figure 3. Sandbox and antivirus detections ([VirusTotal page on file 113569686], n.d.)

VirusTotal gathers noteworthy properties from files and aggregates them into an easy-to-understand form in *capabilities and indicators* section (Figure 4). For example, “May execute code from Dynamically Linked Libraries” capability likely originates from static analysis on the document’s VBA macros, where a reference is declared to function LoadLibraryA located in kernel32.dll (Figure

5). VirusTotal also allows previewing certain types of files in browser - in this case, the document uses classic phishing techniques to coerce users to enable macro execution (Figure 6).

Capabilities And Indicators ⓘ

- ⚠ May execute code from Dynamically Linked Libraries.
- ⚠ Seems to contain deobfuscation code.
- ⚠ Makes use of macros.
- ⚠ May write to a file.
- ⚠ May create OLE objects.

Figure 4. Capabilities and indicators section ([VirusTotal page on file 113569686], n.d.)

OpenXML Document Info ⓘ

Macros And VBA Code Streams

+ ThisDocument.cls

Option Explicit

```
Private Declare PtrSafe Function vsk9 Lib "kernel32" Alias "LoadLibraryA" (ByVal x As String) As LongPtr
```

Figure 5. Snippet from VBA macros ([VirusTotal page on file 113569686], n.d.)

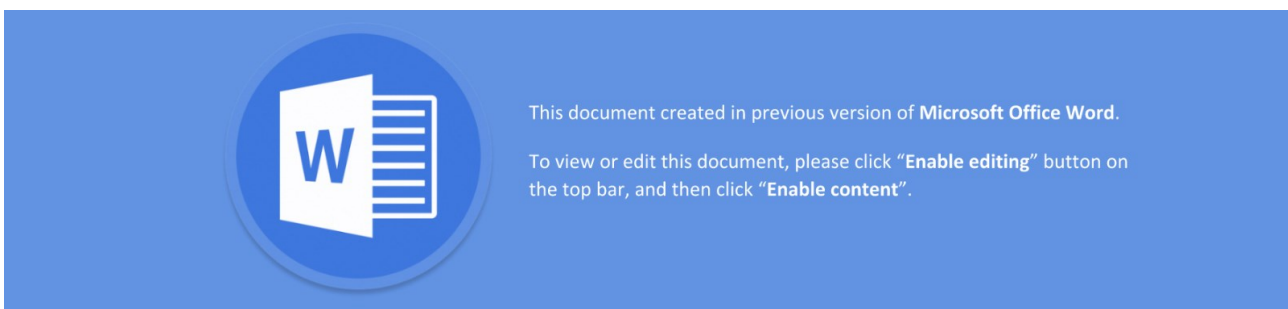


Figure 6. Document preview ([VirusTotal page on file 113569686], n.d.)

Document 113569686 and domain nikohulis[.]com have a related portable executable file **M1cro.exe** in common which has only a handful of antivirus detections ([VirusTotal page on file 113569686], n.d.; [VirusTotal page on domain nikohulis[.]com], n.d.). It is dropped on disk by 113569686 and communicates with nikohulis[.]com. It has a SHA256 hash value of f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64 ([VirusTotal page on

file M1cro.exe], n.d.). There are no detections from sandbox products, and only 3 out of 67 antivirus engines that can analyse this type of file detect it but have rather non-descriptive threat labels (Figure 7).

Security vendors' analysis on 2021-09-04T21:17:02	
BitDefenderTheta	Gen:NN.ZexaF.34126.Uy3@a8NB5sni
eGambit	Trojan.Generic
FireEye	Generic.mg.ffda00e195cf0445
Acronis (Static ML)	Undetected
Ad-Aware	Undetected

Figure 7. Antivirus engine detections ([VirusTotal page on file M1cro.exe], n.d.)

Antivirus products in VirusTotal are command line versions configured by the vendors as they see fit and perform only static analysis (Quintero, 2012). For this reason, it is important to understand that antivirus engine detections in VirusTotal may vary greatly from a real-world scenario. Depending on how long it has been since the last analysis, submitting files for reanalysis may improve the results as vendors may have updated their signatures.

VirusTotal privacy policy mentions that antivirus vendors receive submitted files for review if they are labelled as malicious by at least one other vendor (Privacy Policy, 2021). M1cro.exe is a recent submission ([VirusTotal page on file M1cro.exe], n.d.). It may take some time for the other vendors to pick it up.

Legitimacy of unknown files should not be judged based solely on the amount of antivirus detections in VirusTotal. For example, there are two matching YARA rules for M1cro.exe that identify malware behaviour (Figure 8). The first rule attempts to catch files that use so called heaven's gate technique by matching byte sequences corresponding to series of x86 instructions. Heaven's gate is used by malware as an evasion technique and has been observed in recent malware such as Emotet (Nagy, 2019). The second rule looks for strings and exported functions related to reflective dynamic-link library (DLL) loading technique. This technique allows DLL loading directly from memory instead of disk and is often used by malware for defence evasion and anti-analysis purposes (Seifert, 2017).


 Matches rule HeavensGate by kevoreilly from ruleset HeavensGate at https://github.com/kevoreilly/CAPEv2 ↳ Heaven's Gate: Switch from 32-bit to 64-mode
 Matches rule ReflectiveLoader by Florian Roth from ruleset gen_loaders at https://github.com/Neo23x0/signature-base ↳ Detects a unspecified hack tool, crack or malware using a reflective loader - no hard match - further investigation recommended

Figure 8. Matching YARA rules ([VirusTotal page on file M1cro.exe], n.d.)

M1cro.exe has been analysed by VirusTotal's inhouse ZenBox sandbox ([VirusTotal page on file M1cro.exe], n.d.). Its process tree shows that it queries system information via systeminfo.exe, a behaviour not unusual for malware (Figure 9). At this point there is more than enough information for security analyst to complete triage.

Processes Tree

```

↳ 1268 - 'C:\Users\user\Desktop\M1cro.exe'
    ↳ 6444 - C:\Windows\SysWOW64\systeminfo.exe C:\Windows\System32\systeminfo.exe
        ↳ 7024 - C:\Windows\System32\conhost.exe C:\Windows\system32\conhost.exe 0xffffffff -ForceV1

```

Figure 9. Process tree ([VirusTotal page on file M1cro.exe], n.d.)

Summary

Based on a single static IOC it is already possible to hypothesise a scenario for the events using information from VirusTotal alone. Customer was possibly targeted in a phishing campaign and a couple of employees fell for it. The employees had opened malicious documents in Microsoft Word and unknowingly allowed execution of malicious code by enabling macros. This may have started an infection chain where malware with yet unknown capabilities was downloaded and executed on their devices, unless endpoint protection software successfully intercepted it. The security alert is very likely true positive and should be escalated to incident response for further investigation and remediation.

6 Automated alert triage

This chapter presents a plan to automate alert triage using SOAR. The plan is based on the manual triage process of the previous chapter and the observations made during it. Automation workflow is triggered when SOAR receives a security alert that has associated static IOCs. The workflow has three phases: IOC enrichment, IOC analysis, and alert triage (Figure 10). After the phases are over it's up to the security analysts to complete the triage.

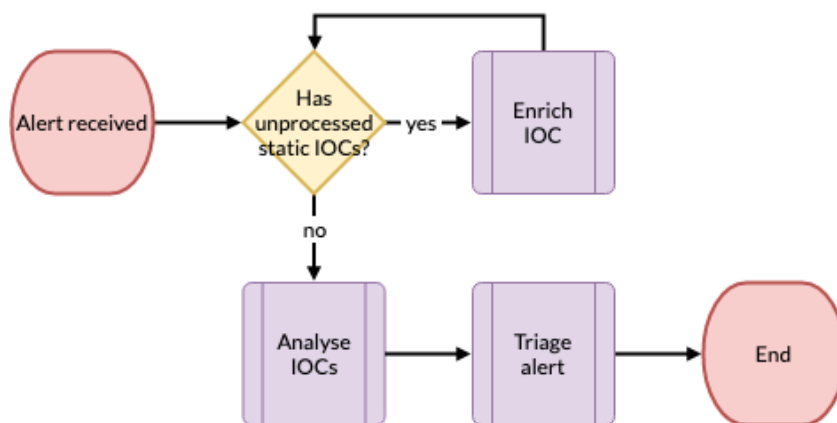


Figure 10. Automation workflow high-level view

6.1 Enriching IOCs

Static IOCs associated to a security alert are enriched with detailed information requested from VirusTotal API (Figure 11). There may be conditions that exclude certain IOCs from enrichment and analysis, for example if the same IOC has already been processed a while ago. It is also possible that an IOC does not exist in VirusTotal's database. In this case VirusTotal responds with an error, and the process continues onto the next IOC. SOAR indicates to security analysts if an IOC could not be enriched by VirusTotal.

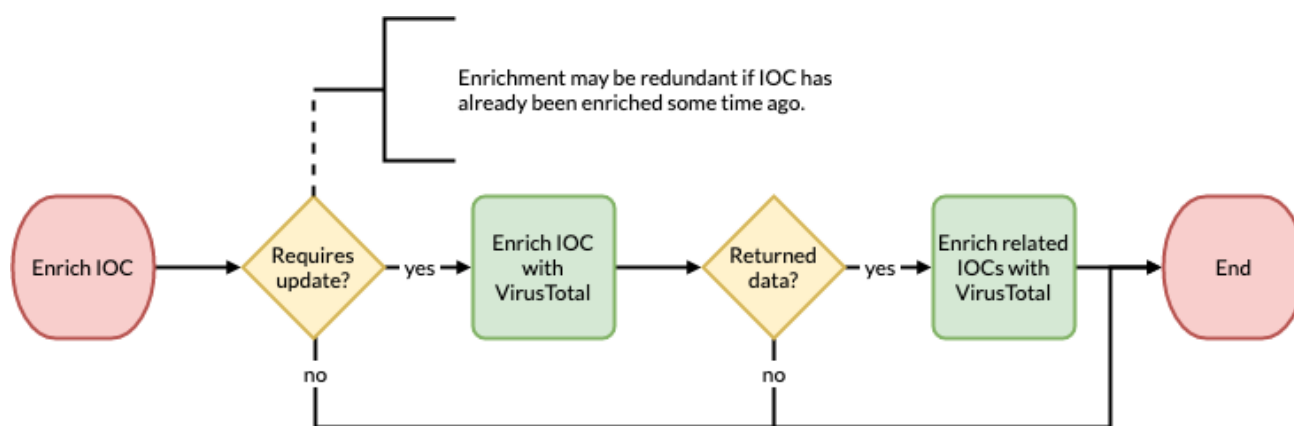


Figure 11. Automation workflow for IOC enrichment

The only static IOC associated to the security alert was domain nikohulis[.]com. SOAR requests detailed information on the domain along with object descriptors for interesting relationships:

```
https://www.virustotal.com/api/v3/domains/nikohulis[.]com?relationships=communicating_files
```

SOAR stores attributes from the response that might be of interest to security analysts or useful for automated analysis. IOC relations were the key element in having a successful manual triage. It is important to use this information in automation as well. SOAR requests detailed information on the files that have communicated with nikohulis[.]com:

```
https://www.virustotal.com/api/v3/domains/nikohulis[.]com/communicating_files?relationships=compressed_parents,contacted_domains,contacted_ips,contacted_urls,dropped_files,embedded_domains,embedded_ips,embedded_urls,execution_parents,itw_domains,itw_ips,itw_urls
```

VirusTotal responds with a list of file objects, including the malicious document 113569686 and the unrecognized potential malware M1cro.exe. Each file object has lists of object descriptors for all relationships specified with the *relationships* query string parameter. All objects and object descriptors are ingested by SOAR as IOCs and are associated to the security alert.

SOAR enriches automatically only IOCs directly associated to a security alert and their immediate relations. This means that IOCs related to 113569686 and M1cro.exe are not enriched automatically as they are not directly associated to the alert or immediate relations to nikohulis[.]com. Iteratively querying data for related IOCs could cause very long request chains or end in an infinite

loop. However, SOAR allows security analysts to manually initiate enrichment and analysis for any ingested IOC as this information may sometimes be imperative for successful triaging.

Having lists of object descriptors for relationships is useful. This information allows certain relationship endpoints to be excluded per IOC since it is known beforehand if they return data. This saves API quota and SOAR network resources. Table 2 shows how eight out of twelve relationship endpoints wouldn't have to be queried if security analyst initiated manual enrichment for M1cro.exe, since it is known they would not return objects.

Table 2. Related objects for M1cro.exe

Relationship	Number of related objects
compressed_parents	0
contacted_domains	3
contacted_ips	19
contacted_urls	0
dropped_files	0
embedded_domains	0
embedded_ips	0
embedded_urls	0
execution_parents	1
itw_domains	1
itw_ips	0
itw_urls	1

VirusTotal API usage is limited by quota that is specified per customer service agreement. A plan should be created to handle the situation where that quota is exceeded. SOAR should make it clear to security analysts if an IOC has not been fully enriched and analysed due to exceeded quota limit.

6.2 Analysing IOCs

SOAR assigns IOCs with one of the following reputation values from analysis: benign, suspicious, or malicious. These values indicate the likelihood that an IOC is related to malicious activity. To deter-

mine reputation value, SOAR matches analysis rules against IOCs (Table 3). Rules are weighted according to their assumed severity. The sum of the matching rules weight is mapped to a reputation value according to Table 4. Security analysts should be able to override reputation value set by automation.

Table 3. Sample IOC analysis rules

Rule	IOC type	Condition	Weight
Contains obfuscated macros	File	Has <i>macros</i> and <i>obfuscated</i> tags.	2
Detected by antivirus products	File	Has detections from more than 2 preferred or more than 6 total antivirus vendors.	2
Detected by sandbox products	File	Has detections from sandbox products.	2
Matches on multiple YARA rules	File	Has matches on more than 1 YARA rule.	2
Contacted by malicious files	Domain	Has <i>communicating_files</i> relationship malicious files.	1
Contacts newly registered domains	File	Has <i>contacted_domains</i> relationship to domains registered under 6 months ago.	1
Dropped by malicious files	File	Has <i>execution_parents</i> relationship to malicious files.	1
Drops malicious files	File	Has <i>dropped_files</i> relationship to malicious files.	1
Hosted malicious files	Domain	Has <i>downloaded_files</i> relationship to malicious files.	1
Invalid certificate	File	Has <i>invalid-signature</i> tag.	1
Newly registered	Domain	Registered under 6 months ago.	1
Has a trusted AS owner	IP address	Autonomous system owner is trusted.	-6
Has a trusted distributor	File	File distributor is trusted.	-6

Table 4. Analysis rule weight sum to reputation value mapping

Weight sum	Reputation value
$x > 3$	Malicious
$0 < x \leq 3$	Suspicious
$x \leq 0$	Benign

6.2.1 Dependency problem in analysis rules

IOCs and their relationships form directed cycle graphs, where IOCs make the vertices and relationships the edges. Analysis rules change them into dependency graphs, as analysis rules are allowed to have conditions that depend on other IOCs and their properties. In effect, this means that it may not be sufficient to analyse each IOC only once. Figure 12 exemplifies the situation – once IOCs A and B have been analysed, their weight sums may be incorrect. There may be rules applying to A that depend on the updated properties of B. If the weight sum of an IOC changes on analysis, all its related IOCs must be analysed again. Figure 13 provides an algorithm to resolve this issue.

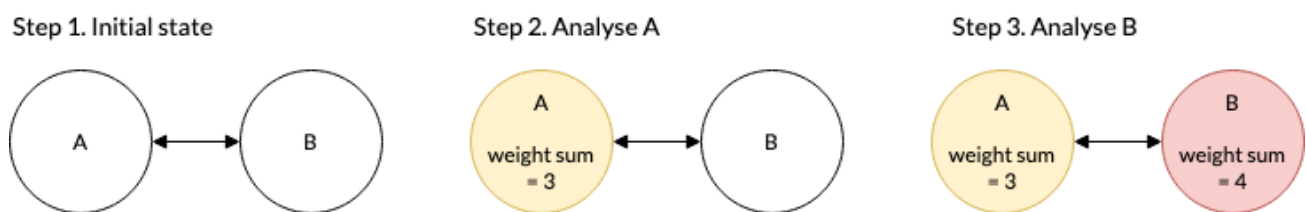


Figure 12. Issue with cyclic dependency from relationships

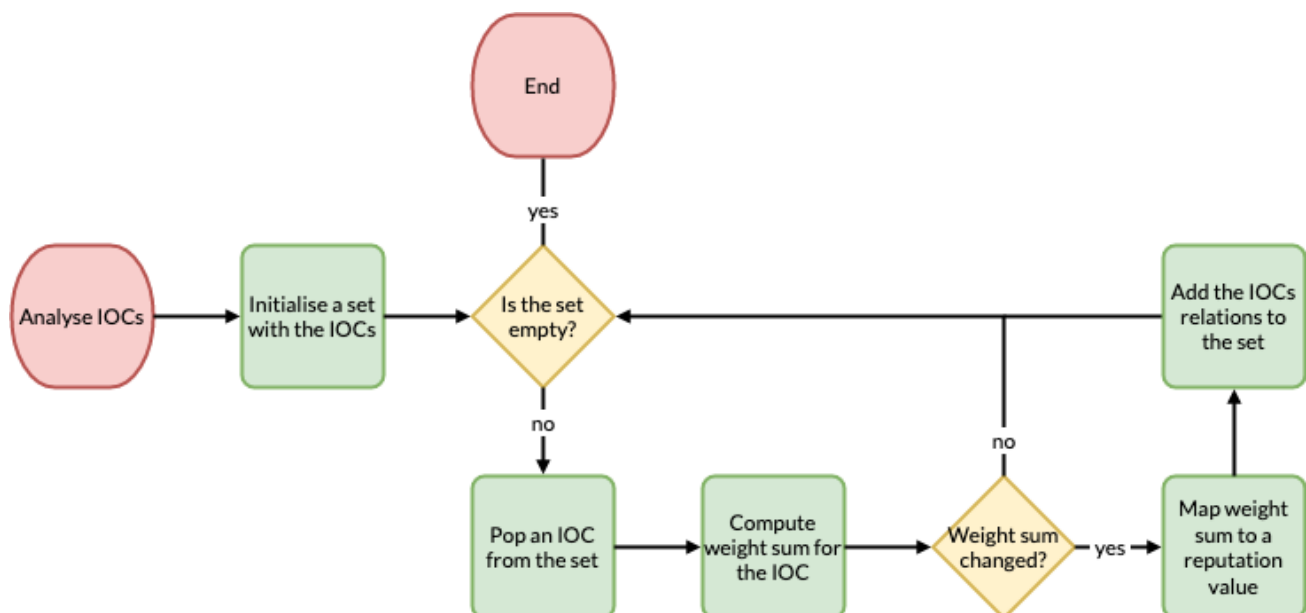


Figure 13. Automation workflow for IOC analysis

Appendix 1. has a Python script that implements the algorithm and tests its efficiency in terms of required loop iterations per number of IOCs. The test picks an initial IOC and creates a relationship between it and the rest. The rest are set to have a second, random relationship with one other IOC. The initial IOC has a pre-set weight of 4, and there are two analysis rules: relation to at least one malicious IOC with weight of 1, and relation to at least 3 suspicious IOCs with weight of 2. Figure 14 shows the output from the script. The algorithmic efficiency is around $O(n^2)$ in Big-O notation, meaning it gets worse quickly as the number of IOCs grows.

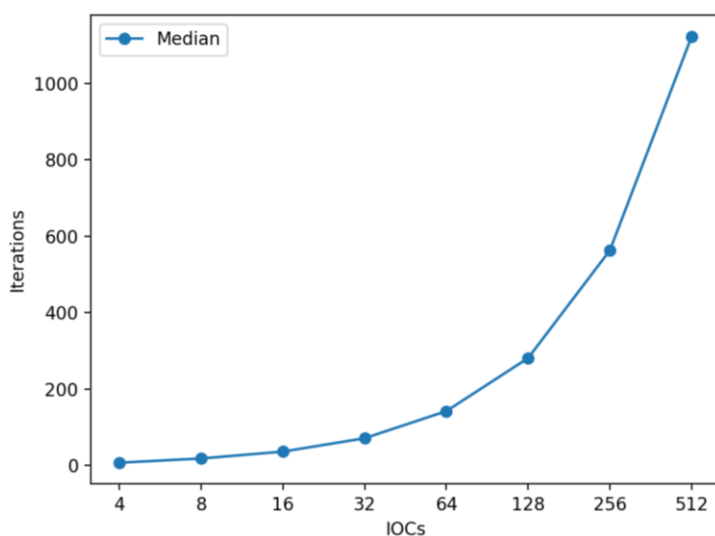


Figure 14. Algorithmic efficiency of dependency solving algorithm for IOC analysis

6.2.2 Displaying analysis results

It is important that security analysts understand how the reputation value was determined for IOCs when handling security alerts. The data SOAR uses to compute a reputation value must be converted into human-readable format. SOAR should present this information concisely in an easily interpretable format for security analysts. The following mock-up images are based on applying the sample analysis rules defined in Table 3 to IOCs from the security alert.

Figure 15 shows a mock-up of analysis results on domain nikohulis[.]com. It shows that the domain has been registered only recently and it has associations to two malicious files. Its reputation value is set to suspicious. There are matches from three analysis rules and their weight sum of three does not cross the threshold for reputation value malicious.

Contacted by malicious files

- ↳ a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f
- ↳ f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64

Hosted malicious files

- ↳ f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64

Newly registered

- ↳ Registered on 2021-07-09

Figure 15. Formatted analysis results for domain nikohulis[.]com

Figure 16 shows a mock-up for the Office document 113569686. It contacts two newly registered domains and has a match on document file specific rule that looks for malicious macros. It is detected by both antivirus and sandbox products. Its reputation value is set to malicious due to its weight sum of 8.

Contains obfuscated macros**Detected by antivirus products**

- ↳ 19 products of which 5 are preferred

Detected by sandbox products**Contacts newly registered domains**

- ↳ milojeriks[.]com registered on 2021-07-09
- ↳ nikohulis[.]com registered on 2021-07-09

Drops malicious files

- ↳ f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64

Figure 16. Formatted analysis results for file 113569686

Figure 17 has a mock-up for the portable executable file M1cro.exe. Similar to the manual triage, the automation finds enough notable properties to consider it malicious, even though it is detected by only minimal number of antivirus engines. This IOC gets a weight sum of 5 due to YARA matches, relations to other IOCs, and an invalid certificate.

Matches on multiple YARA rules

- ↳ HeavensGate (<https://github.com/kevoreilly/CAPEv2>)
- ↳ ReflectiveLoader (<https://github.com/Neo23x0/signature-base>)

Contacts newly registered domains

- ↳ nikohulis[.]com registered on 2021-07-09

Dropped by malicious files

- ↳ a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f

Invalid certificate

Figure 17. Formatted analysis results for file M1cro.exe

IOCs may have matches on rules that affect their reputation positively, such as known distributors for files. These are displayed as well and take precedence over malicious and suspicious observations in the list (Figure 18).

Has a trusted distributor

- ↳ Distributed by Microsoft (source: Microsoft Corporation)

Figure 18. Analysis rule match with a positive effect on an IOCs reputation

6.3 Triaging alert

SOAR proceeds to the triage phase after completing IOC analysis. The goal of this phase is two-fold: Determine the extents of the potential compromise and if the alert should be escalated to incident response (Figure 19). To determine the extents, SOAR searches for suspicious and malicious IOCs in customer environment. If customer assets are discovered with at least one malicious or at least 3 suspicious IOCs, SOAR suggests escalation to incident response. Figure 20 has a mock-up for how these potentially compromised assets are listed for security analysts to review. The list should be expanded to include information about the context in which the IOCs were found.



Figure 19. Automation workflow for triage automation

Asset	First seen	Last seen
10.0.102.16	2021-09-17T11:01:01Z	2021-09-17T11:01:05Z
10.0.102.71	2021-09-17T11:13:50Z	2021-09-17T11:13:50Z
10.0.12.10	2021-09-17T10:19:11Z	2021-09-17T11:13:10Z
w10host	2021-09-17T10:19:11Z	2021-09-17T11:13:10Z

Figure 20. Search results for domain nikohulis[.]com in customer environment

7 Results

A questionnaire was given to security analysts of the assignor's SOC to evaluate the proposed automation workflow. Twelve responses were received. Appendix 2. contains the questionnaire form. The questionnaire first surveyed the analysts' familiarity with the VirusTotal service in general. Following questions addressed the actual research objectives:

1. Is the proposed solution sufficient to take over security analyst tasks during alert triage?
2. Can it improve incident response times and threat recognition accuracy?

The questionnaire first surveyed how popular VirusTotal service is for alert triage among the respondents, and how fatiguing they feel its usage is (Figure 21, Figure 22). The questions intended to find out if VirusTotal is part of their alert triage routine, and how familiar they are with it as that might be useful data when analysing the results of later questions. The answers show that the respondents use the service regularly and suggests that they are familiar with its capabilities but find its usage for alert triage somewhat to very fatiguing.

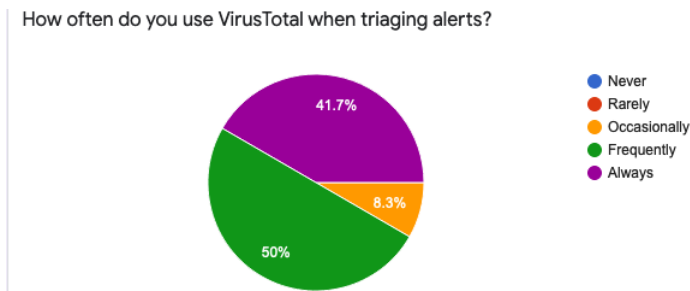


Figure 21. How often VirusTotal is used for alert triaging?

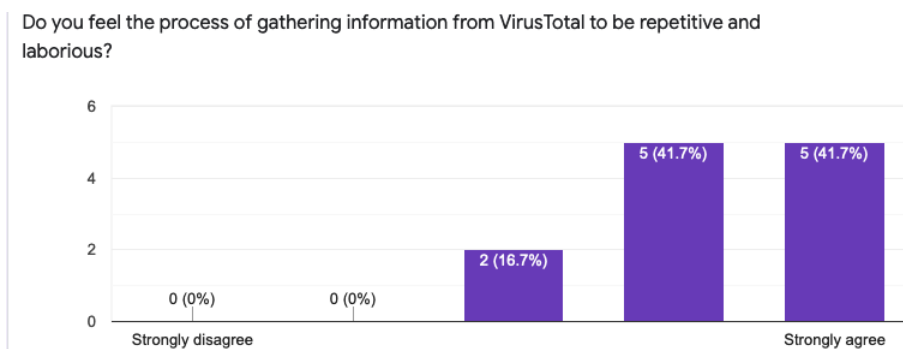


Figure 22. How fatiguing VirusTotal usage is for alert triage?

The third question asked opinion for the automation's potential to replace manual usage of VirusTotal for alert triage (Figure 23). A large majority of the respondents feel that the solution it is good enough to take over their tasks in this regard. A few respondents were not as convinced, and two were undecided.

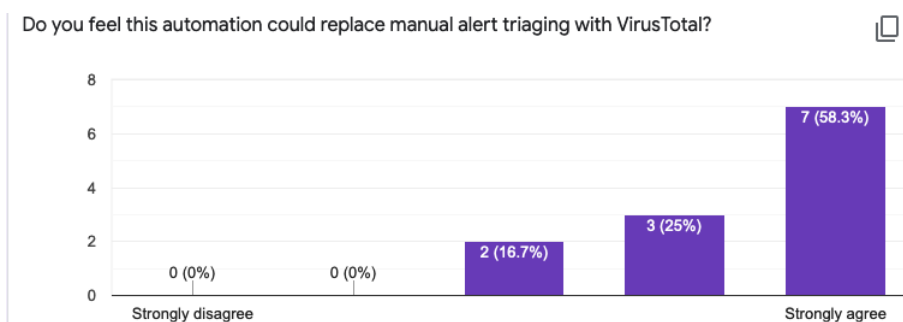


Figure 23. Automation's potential to take over manual work

The fourth question asked if the proposed automation workflow would improve incident response times (Figure 24). Large majority of the respondents feel that there would be a strong improvement, and a few believe there would be least a slight improvement. Minority felt that it would have no difference at all.

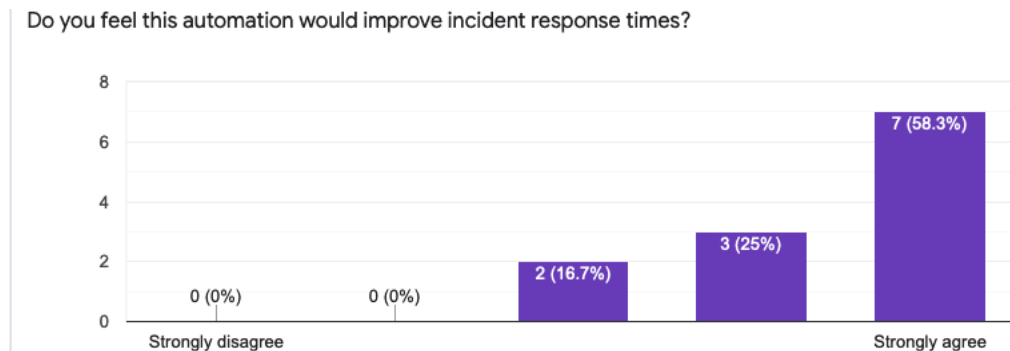


Figure 24. Automation's potential to improve incident response times

The fifth question asked if the proposed automation workflow would improve accuracy in threat recognition (Figure 25). The respondents feel that threat recognition would be improved, but not as much as incident response times. Additionally, the respondents who felt that there would be no difference to incident response times, felt that there would be improvement in threat recognition.

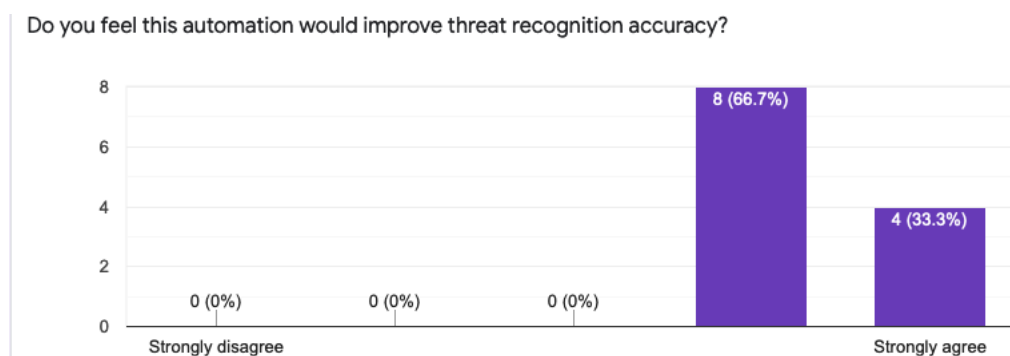


Figure 25. Automation's potential to improve threat recognition

The final question surveyed how much trust the respondents would put on automation (Figure 26). The answers show that the respondents trust the automation to a degree. However, most respondents feel like they need to occasionally go and manually confirm information from VirusTotal that the automation has produced.

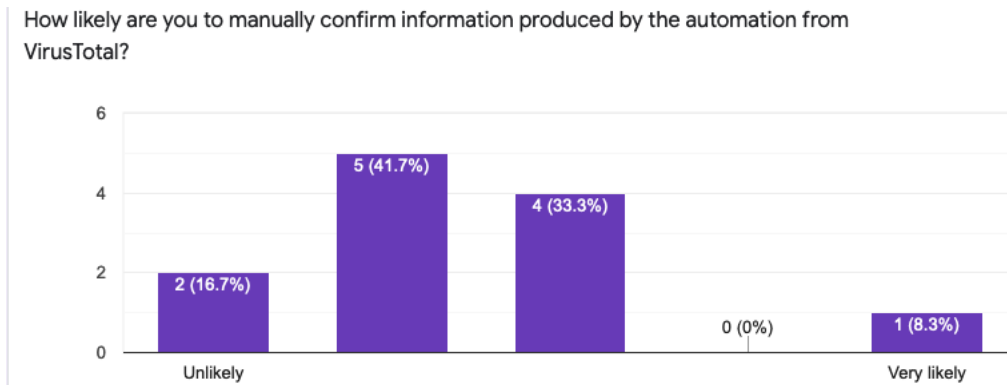


Figure 26. How likely information produced by automation is to be manually confirmed?

8 Conclusion

The research objective of this thesis was to examine how alert triage process could be automated with SOAR using VirusTotal. The thesis attempted to answer the following questions: can automation utilise VirusTotal at a sufficient level to take over security analyst's tasks for alert triage, and if such automation would improve incident response times and threat recognition accuracy. A SOAR automation workflow was planned that:

1. Programmatically analyses IOCs associated to security alerts using data from VirusTotal.
2. Assigns a reputation value (benign, suspicious, malicious) to IOCs based on the analysis results.
3. Presents the analysis results to security analysts in a concise, human-readable way.
4. Searches for suspicious and malicious IOCs in customer environment, and if found, suggests security analysts to escalate to incident response.

Security analysts of the assignor's SOC were given a survey to evaluate the proposed automation workflow in terms of the research questions. The responses were positive. The respondents felt that the automation has potential to take over their tasks in triaging alerts using VirusTotal, reducing monotonous manual tasks. The respondents thought that it would provide noticeable improvements in incident response times, but not as much in threat recognition accuracy. The difference may be explained by the respondents' familiarity with VirusTotal. The survey shows that the respondents use the service regularly, and thus are likely competent in using it to recognise threats, but the process may take quite a while which adds to response times.

To get better insight into why the respondents felt that these would be improved, additional questions should have been added. Best results would've been achieved using free-form fields. The survey was based on the results of a single example of how the automation could triage an alert. There is no concrete data to verify the reliability of this research. Its reliability can only be properly verified once the system is implemented and has produced data from real-world usage.

8.1 Further development

The survey showed that there was some distrust in the automation, as many analysts felt that they are somewhat likely to confirm its results manually from VirusTotal. This has a noticeable impact on the effectiveness of the automation. The survey does not tell if there is distrust in automated systems in general or this system specifically. How to build trust in automated systems?

Thesis focused on alert triaging using data from VirusTotal. The resulting automation workflow could be expanded to include data from other external sources as well, such as abuse.ch. It would be useful to research how data from VirusTotal and other sources could be aggregated, and if the same analysis rules could be then applied to all sources.

References

[Eranet domain registration page]. (n.d.). Eranet International. Retrieved November 27, 2021, from <https://www.eranet.com/domain/index.php>

[VirusTotal API documentation]. (n.d.). VirusTotal. Retrieved August 17, 2021, from <https://developers.virustotal.com/reference>

[VirusTotal page on domain nikohulis[.]com]. (n.d.). VirusTotal. Retrieved September 17, 2021, from [https://www.virustotal.com/gui/domain/nikohulis\[.\]com](https://www.virustotal.com/gui/domain/nikohulis[.]com)

[VirusTotal page on file 113569686]. (n.d.). VirusTotal. Retrieved September 17, 2021, from <https://www.virustotal.com/gui/file/a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f>

[VirusTotal page on file M1cro.exe]. (n.d.). VirusTotal. Retrieved September 17, 2021, from <https://www.virustotal.com/gui/file/f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64>

[VirusTotal page on IP address 45.151.144.162]. (n.d.). VirusTotal. Retrieved September 17, 2021, from <https://www.virustotal.com/gui/ip-address/45.151.144.162>

[VirusTotal statistics page]. (n.d.). VirusTotal. Retrieved November 9, 2021, from <https://www.virustotal.com/gui/stats>

Badger, L., Johnson, C., Skorupka, C., Snyder, J., & Waltermire, D. (2016, October). *Guide to Cyber Threat Information Sharing* (NIST Special Publication 800-150). U.S. Department of Commerce, National Institute of Standards and Technology. <http://dx.doi.org/10.6028/NIST.SP.800-150>

Bianco, D. J. (2014, January 17). The Pyramid of Pain. *Enterprise Detection & Response*. <http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html>

Brata Deb, S., Gupta, A., Kumarguru, P., & Verma, M. (2018, November 9-11). *Analysing Indicator of Compromises for Ransomware: Leveraging IOCs with Machine Learning Techniques* [Conference paper]. IEEE International Conference on Intelligence and Security Informatics, Miami, USA. <https://ieeexplore.ieee.org/document/8587409>

Censuswide, & Cybereason. (2021). *Ransomware: The True Cost to Business*. Cybereason. https://www.cybereason.com/hubfs/dam/collateral/ebooks/Cybereason_Ransomware_Research_2021.pdf

Chen, Z., Kwan, K., & Wang, J. J. (2019, August 20). Newly Registered Domains: Malicious Abuse by Bad Actors. *Palo Alto Networks blog*. <https://unit42.paloaltonetworks.com/newly-registered-domains-malicious-abuse-by-bad-actors/>

Contributors. (n.d.). VirusTotal. Retrieved November 27, 2021, from <https://support.virustotal.com/hc/en-us/articles/115002146809-Contributors>

Crowdsourced IDS Rules. (n.d.). VirusTotal. Retrieved November 27, 2021, from <https://support.virustotal.com/hc/en-us/articles/360018436078-Crowdsourced-IDS-Rules>

Crowdsourced YARA Rules. (n.d.). VirusTotal. Retrieved November 27, 2021, from <https://support.virustotal.com/hc/en-us/articles/360015658497-Crowdsourced-YARA-Rules>

Enterprise Matrix. (2021, April 29). The MITRE Corporation. Retrieved November 27, 2021, from <https://attack.mitre.org/versions/v9/matrices/enterprise/>

European Union Agency for Cybersecurity. (2020). *How to set up CIRT and SOC*. <https://www.enisa.europa.eu/publications/how-to-set-up-csirt-and-soc/@download/fullReport>

Frequently Asked Questions. (n.d.). VirusTotal. Retrieved November 27, 2021, from <https://developers.virustotal.com/docs/frequently-asked-questions>

Gartner. (n.d.). *Securing the Digital Business*. RSA. <https://www.rsa.com/content/dam/en/analyst-report/gartner-securing-the-digital-business.pdf>

IANS Faculty. (2021, February 16). How to Choose the Right SOC Model. *IANS blog*. <https://www.iansresearch.com/resources/all-blogs/post/security-blog/2021/04/30/how-to-choose-the-right-soc-model>

IBM Security, & Ponemon Institute. (2020). *Cost of a Data Breach Report 2020*. IBM. <https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/>

Imam, F. (2019, March 12). *Security Orchestration, Automation and Response (SOAR)*. Infosec Resources. <https://resources.infosecinstitute.com/topic/security-orchestration-automation-and-response-soar/>

IOA VS IOC. (2021, May 13). CrowdStrike. Retrieved November 27, 2021, from <https://www.crowdstrike.com/cybersecurity-101/indicators-of-compromise/ioa-vs-ioc/>

Martinez, E. (2013, April 1). VirusTotal += Passive DNS replication. *VirusTotal blog*. <https://blog.virustotal.com/2013/04/virustotal-passive-dns-replication.html>

Microsoft 365 Defender Threat Intelligence Team, Microsoft 365 Security, & Microsoft Threat Intelligence Center. (2021, March 16). HAFNIUM targeting Exchange Servers with 0-day exploits. *Microsoft blog*. <https://www.microsoft.com/security/blog/2021/03/02/hafnium-targeting-exchange-servers/>

Nagy, L. (2019, October 2-4). *Exploring Emotet, an elaborate everyday enigma* [Conference session]. Virus Bulletin International Conference, London, United Kingdom. <https://www.youtube.com/watch?v=Oldilbsbblc>

Ponemon Institute. (2019). *Improving the Effectiveness of the Security Operations Center*. Devo Technology. <https://www.devo.com/wp-content/uploads/2019/07/2019-Devo-Ponemon-Study-Final.pdf>

Privacy Policy. (2021, January 27). VirusTotal. Retrieved November 27, 2021, from <https://support.virustotal.com/hc/en-us/articles/115002168385-Privacy-Policy>

Quintero, B. (2012, August 29). AV Comparative Analyses, Marketing, and VirusTotal: A Bad Combination. *VirusTotal blog*. <https://blog.virustotal.com/2012/08/av-comparative-analyses-marketing-and.html>

Seifert, C. (2017, November 13). Detecting reflective DLL loading with Windows Defender ATP. *Microsoft blog*. <https://www.microsoft.com/security/blog/2017/11/13/detecting-reflective-dll-loading-with-windows-defender-atp/>

Sigma rules. (n.d.). VirusTotal. Retrieved November 27, 2021, from <https://support.virustotal.com/hc/en-us/articles/360015738658-Sigma-rules>

SonicWall. (2021). *2021 SonicWall Cyber Threat Report*. <https://www.sonicwall.com/mediabrary/en/white-paper/2021-cyber-threat-report.pdf>

Sundaramurthy, S. C. (2017). *An Anthropological Study of Security Operations Centers to Improve Operational Efficiency* [Doctoral dissertation, University of South Florida]. University of South Florida archive. <http://scholarcommons.usf.edu/etd/6958>

Zimmerman, C. (2014, October). *Ten Strategies of a World-Class Cybersecurity Operations Center*. The MITRE Corporation. <https://www.mitre.org/sites/default/files/publications/pr-13-1028-mitre-10-strategies-cyber-ops-center.pdf>

Appendices

Appendix 1. Python script to test effectiveness of dependency solving algorithm

```
import matplotlib.pyplot as plt
from matplotlib.ticker import ScalarFormatter
import pandas as pd
import random
import statistics

def solve(n: int):
    # Create N IOCs, first in the list is the "initial IOC".
    iocs = list(range(n))

    # Make the initial IOC have relation with all other IOCs (two-way).
    rels = {ioc: {iocs[0]} for ioc in iocs[1:]}
    rels[iocs[0]] = set(iocs[1:])

    # Make other IOCs have relations to each other randomly (two-way).
    for ioc in iocs[1:]:
        # Get a random IOC that is not the initial IOC or "self".
        rnd = random.choice([tmp for tmp in iocs if tmp != iocs[0] and tmp != ioc])
        # Setup relations.
        rels[ioc].add(rnd)
        rels[rnd].add(ioc)

    # Initialise a set with all IOCs.
    need_analysis = set(iocs)
    weights = {}

    c = 0
    while need_analysis:
        ioc = need_analysis.pop()

        weight = 0
        # Initial IOC has a pre-set weight.
        if ioc == iocs[0]:
            weight += 4
        # Relationship with at least one malicious IOC.
        if any([weights[rel] > 3 for rel in rels[ioc] if rel in weights]):
            weight += 1
        # Relationship with more than 3 suspicious IOCs.
        if sum(map(lambda rel: weights[rel] > 0 if rel in weights else False, rels[ioc])) > 3:
            weight += 2

        # If IOC weight has changed, add its related IOCs to the set.
        if ioc not in weights or weight != weights[ioc]:
            need_analysis.update(rels[ioc])
        # Update IOC weight.
        weights[ioc] = weight

        c += 1

    return c

# Get the medians of 1000 iterations of solve per 2^2 .. 2^10 IOCs.
num_iocs = [2**i for i in range(2, 10)]
medians = [statistics.median([solve(n) for _ in range(1000)]) for n in num_iocs]

pd.DataFrame(
    data=medians,
    columns=['Median'],
    index=num_iocs).plot(marker='o', label="Median")

ax = plt.gca()
ax.set_xscale('log')
ax.set_xticks(num_iocs)
ax.set_xlabel('IOCs', fontsize=10)
ax.set_ylabel('Iterations', fontsize=10)
ax.get_xaxis().set_major_formatter(ScalarFormatter())

plt.minorticks_off()
plt.show()
```

Appendix 2. Questionnaire form

Thesis questionnaire

I am writing a Bachelor's thesis on the subject: Exploring VirusTotal for security operations alert triage automation. The result of the thesis is a proposal for a SOAR automation workflow, that:

1. Programmatically analyses IOCs associated to security alerts using data from VirusTotal.
2. Assigns a reputation value (benign, suspicious, malicious) to IOCs based on the analysis results.
3. Presents the analysis results to security analysts in a concise, human-readable way.
4. Searches for suspicious and malicious IOCs in customer environment, and if found, suggests security analysts to escalate to incident response.

Below are mock-up images of how the automation would present its analysis results in SOAR regarding a security alert. The mock-ups are based on an analysis of a fictitious security alert, described:

"An anomaly in DNS traffic within customer network: on 17th of September 2021, multiple internal hosts have attempted to resolve a newly registered domain nikohulis[.]com in a short span of time."

Please answer the questions in the form. The answers will be used to evaluate the proposed automation workflow.

Notes

- The automated analysis takes into consideration IOCs directly related to IOCs associated with security alerts. There are results for files 113569686 and M1cro.exe as well, as these are directly related to domain nikohulis[.]com.
- The analysis mock-ups were designed on 17th of September 2021. Information around the IOCs have since been updated on VirusTotal.

Analysis results for domain nikohulis[.]com, reputation value: suspicious

Contacted by malicious files

- ↳ a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f
- ↳ f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64

Hosted malicious files

- ↳ f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64

Newly registered

- ↳ Registered on 2021-07-09

Analysis results for Office document 113569686, SHA256:

a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f, reputation value: malicious

```
Contains obfuscated macros
Detected by antivirus products
  ↳ 19 products of which 5 are preferred
Detected by sandbox products
Contacts newly registered domains
  ↳ milojeriks[.]com registered on 2021-07-09
  ↳ nikohulis[.]com registered on 2021-07-09
Drops malicious files
  ↳ f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64
```

Analysis results of portable executable file M1cro.exe, SHA256

f60edb47f56f0bfa141093ecec2d655988b201d9ec5ee55e967726513a8fae64, reputation value malicious

```
Matches on multiple YARA rules
  ↳ HeavensGate (https://github.com/kevoreilly/CAPEv2)
  ↳ ReflectiveLoader (https://github.com/Neo23x0/signature-base)
Contacts newly registered domains
  ↳ nikohulis[.]com registered on 2021-07-09
Dropped by malicious files
  ↳ a466e70b09bbe9019c9f57d220d64d4b2cbc0e3b592d61d067ac54889a46831f
Invalid certificate
```

Search results from customer environment for domain nikohulis[.]com.

Asset	First seen	Last seen
10.0.102.16	2021-09-17T11:01:01Z	2021-09-17T11:01:05Z
10.0.102.71	2021-09-17T11:13:50Z	2021-09-17T11:13:50Z
10.0.12.10	2021-09-17T10:19:11Z	2021-09-17T11:13:10Z
w10host	2021-09-17T10:19:11Z	2021-09-17T11:13:10Z

How often do you use VirusTotal when triaging alerts? *

1. Never
2. Rarely
3. Occasionally
4. Frequently
5. Always

Do you feel the process of gathering information from VirusTotal to be repetitive and laborious? *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

Do you feel this automation could replace manual alert triaging with VirusTotal? *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

Do you feel this automation would improve incident response times? *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

Do you feel this automation would improve threat recognition accuracy? *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strong agree

How likely are you to manually confirm information produced by the automation from VirusTotal?

1 2 3 4 5

Unlikely ☐ ☐ ☐ ☐ ☐ Very Likely