A Cost-Efficient Threat Intelligence Platform Powered by Crowdsourced OSINT

Alexander Khalil Daou
School of Computing
University of Portsmouth
alexanderdaou@outlook.com

Fudong Li
Department of Computing and Informatics, Bournemouth University,
Bournemouth, UK
fli@bournemouth.ac.uk

Stavros Shiaeles
Centre for Cybercrime and Economic Crime
University of Portsmouth
Stavros.Shiaeles@port.ac.uk

***Abstract*—Cyberattacks are a primary concern for organisations of all kinds, costing billions of dollars globally each year. As more businesses begin operating online, and as attackers develop more advanced malware and evolve their modus operandi, the demand for effective cyber security measures grows exponentially. One such measure is the threat intelligence platform (TIP): a system which gathers and presents information about current cyber threats, providing actionable insight to aid security teams in employing a more proactive approach to thwarting attacks. These platforms and their accompanying intelligence feeds can be costly when purchased from a commercial vendor, creating a financial barrier for small and medium-sized enterprises. This paper explores the use of crowdsourced open-source intelligence (OSINT) as an alternative to commercial threat intelligence. A model TIP is developed using a combination of crowdsourced OSINT, freeware, and cloud services, demonstrating the feasibility and benefits of using OSINT over commercial solutions. The developed TIP is evaluated using a dataset containing 16,713 malware samples collected via the MalwareBazaar repository.**

Keywords—Cyber Threat Intelligence, Open Source, OSINT, Threat Intelligence Platform, Data Analytics, Freeware, Cloud, Indicators of Compromise

I. Introduction

Cybercrime has rapidly become one of the greatest threats to organisations worldwide, and most large-scale attacks require the use of malware. According to Accenture Security & Ponemon Institute [1], the average annual number of security breaches per organisation had increased by 67% in the span of five years from 2013 to 2018. IBM Security & Ponemon Institute [2] identified the global average total cost of a single data breach to be $4.24 million in 2021, the highest average total cost in the history of their Cost of Data Breach Report and a 9.8% increase from the $3.86 million average total cost for 2020. Such an event can be enough to critically impact small and medium-sized enterprises (SMEs).

In order to effectively counter the threat posed by malware, organisations should have appropriate countermeasures in place. One such countermeasure is the threat intelligence platform (TIP). TIPs collect, aggregate and organise information related to cyber threats, supplying security teams with the details they require to identify and thwart attacks in a proactive manner. An effective TIP will empower security analysts, providing them with a clear understanding of current threats and their indicators [3]. By visualising this threat intelligence data in a TIP, the various charts produced can uncover relationships, anomalies, trends and patterns which provide a clearer view of an organisation's threat landscape, better positioning them to secure their staff and infrastructure. Visualisations of the collected data can provide actionable insight, which is of great value to security teams [4].

Nonetheless, purchasing a vendor-provided TIP and various commercial threat intelligence feeds can be costly [5], putting SMEs and independent researchers at a disadvantage. This paper aims to demonstrate the feasibility of developing an effective TIP with minimal operational expenditure, utilising open-source intelligence (OSINT). Using OSINT, the cost of commercial threat intelligence feeds can be eliminated completely. The purpose of the paper is to illustrate how OSINT can be leveraged in combination with certain tools to develop a successful TIP, removing the large capital expense of a commercial solution. The intention is to make TIPs more accessible to SMEs and independent researchers, also providing organisations already using commercial platforms and feeds with a more cost-efficient alternative.

The remainder of the paper is structured as follows: Section 2 presents the literature review. Section 3 explains the model TIP's implementation. Section 4 offers the evaluation and discussion. Section 5 concludes the paper and proposes future directions for further development.

II. Literature Review

The following literature review examines cyber threat intelligence, crowdsourced open-source intelligence in particular, and the MalwareBazaar OSINT repository.

A. Cyber Threat Intelligence

Shackleford [6, p. 1] defines cyber threat intelligence (CTI) as "data collected, assessed and applied regarding security threats, threat actors, exploits, malware, vulnerabilities and compromise indicators". The power of CTI comes from the collected indicators of compromise (IOCs), such as IP addresses or malware signatures. IOCs may be fed into firewalls or SIEMs, or ingested into a searchable index or visualisation tool to produce a dashboard for data analytics [7, 8]. CTI greatly benefits security staff and IT managers in various aspects, including incident avoidance and mitigation, prioritisation of vulnerability management, implementation of proactive countermeasures against emerging threats, and the development of case studies for internal use [9]. Moreover, organisations may share CTI amongst each other, developing a "herd immunity" against common threats [10]. Having a collaborative intelligence sharing platform allows for new threats to be more quickly identified, and for effective responses to be coordinated throughout the community [11]. However, attackers often personalise attacks for their specific target organisations, meaning new IOCs discovered and shared by one organisation may be of little to no use to other organisations [7].

One issue an organisation may face when aggregating various CTI sources is the format of the data ingested. Intelligence sources often serve data in different formats or schemas, leaving recipients with difficulties when combining data from multiple sources into one unified platform [6]. Another concern is the matter of quantity over quality. Threat analysts are at risk of being overwhelmed with large amounts of low-quality IOCs offered by repositories. The volume and content of data should be evaluated before ingestion, ensuring a manageable quantity of timely, actionable intelligence is being taken in. Only a small proportion of available intelligence is relevant to a particular organisation [12].

B. *Crowdsourced Open-Source Intelligence*

The U.S. Department of Defense defines open-source intelligence as "intelligence that is produced from publicly available information and is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement" [13, Sec. 931]. Ultimately, OSINT is CTI collected via any "open" public source – intelligence available in the public domain for use by any and all entities.

OSINT has become a staple of cybersecurity threat awareness, considering the wealth of cost-free information and IOCs that can be obtained from it. Effective OSINT resources can provide early, useful intelligence that allows organisations to more proactively counter emerging threats [4]. Twitter is commonly cited as a reliable OSINT resource, providing early and often accurate security alerts before other sources [14, 15]. One of OSINT's key advantages is its accessibility. There are no legal concerns with regard to the collection and dissemination of OSINT data, as it is all readily available in the public domain. OSINT research can be conducted anywhere at any time, making it globally accessible [16]. OSINT is also generally free of charge, as opposed to vendor-sourced commercial CTI which may be expensive. OSINT is normally cost-free because it is often crowdsourced by volunteering individuals, typically researchers. Doan et al. [17, p. 87] define a crowdsourced system as one which "enlists a crowd of humans to help solve a problem defined by the system owners". In this case, a crowdsourced OSINT system would be one in which individuals submit their CTI findings (i.e. IOCs) to a platform which then aggregates the data and shares it.

OSINT also has some drawbacks. Despite the rise in popularity of crowdsourced intelligence, it could actually be hindering security teams by overwhelming analysts with false positives and intelligence gaps, leading to "paralysis by analysis" [18]. As CTI analysis is backward-looking, continuous monitoring of threat actors and underground sites is required to achieve timely, actionable intelligence. Knowledge of multiple languages may also be needed to analyse IOCs pertaining to foreign entities. Whilst unfiltered OSINT may provide a view of the bigger picture, tailor-made classified intelligence would be required for a more focused analysis [16].

C. *MalwareBazaar*

MalwareBazaar is a project by abuse.ch which collects and shares malware samples, with the purpose of aiding security analysts and researchers to better protect their organisations and customers [19]. It is a malware sample repository which is free of restrictions such as daily download limits or paywalls, allowing users to submit and download as many samples as they desire without incurring fees. The platform tracks only malware files, no potentially unwanted applications or benign files. The project's website clearly highlights to users the sort of contributions to be made, and has a rule for uploading samples no older than 10 days in order to keep fresh malware in the repository's recent submissions [20].

Several research studies have used MalwareBazaar to great benefit. When developing the Malware Analysis and Intelligence Tool (MAIT), Yucel et al. [10] used the MalwareBazaar API to retrieve advanced persistent threats (APTs) associated with particular malware samples, and to determine when the samples were first identified and uploaded to the MalwareBazaar database. This allows MAIT to produce detailed CTI reports for submitted malware samples, including information such as associated APT actors and first seen dates. Lunghi [21] also used the MalwareBazaar API to perform queries for samples using specific hashes. They highlight the availability of Import and TLSH hashes in MalwareBazaar, as well as digital certificate information, showing the great variety of IOC types available in the repository. This metadata was subsequently used to successfully determine the origin of an unknown malware sample. Groenewegen & Janssen [22] used MalwareBazaar for the entirety of their malware sample dataset when evaluating TheHive Project. Their study demonstrates how MalwareBazaar assists in rapidly producing reports on malware submitted to TheHive. Mohandas et al. [23] used a combined total of 70 samples from MalwareBazaar for training and testing datasets in their successful development of a new method for detecting unknown malware. In their study on learning-based portable executable malware family classification methods, Ma et al. [24] used 3,971 samples from MalwareBazaar for their dataset, favouring the platform for having "the latest malware samples". In their article detailing a comparative analysis of CTI sources, Ramsdale et al. [25] credit MalwareBazaar as a comprehensive intelligence feed, and reference its usage and retransmission by other CTI providers.

III. Implementation

This section highlights the steps taken to create the model TIP. The platform's requirements are discussed, followed by an overview of its design. Its development is then broken down into key components and described.

A. *Requirements*

Before commencing the design and development of the TIP, the most suitable product for each component of the platform had to be selected. Research was conducted to determine which products would provide the best features and functionality to meet the requirements of an effective TIP, at the lowest possible cost. The platform required the following components: infrastructure, an OSINT repository, a database, and a data analytics tool.

1) Infrastructure

In order for the TIP to remain constantly accessible, and for data collection to remain consistent and timely, the server hosting the platform must be fully operational at all times. A fault-tolerant solution is required to make the most of a TIP. In order to remain proactive, the platform must continue ingesting fresh data and should remain highy available. As a result, cloud-based infrastructure would be the ideal solution due to the low-cost and highly-scalable nature of the cloud.

Amazon Web Services (AWS) was the cloud service provider of choice.

2) *OSINT Repository*

MalwareBazaar was chosen as a prime example of a suitable crowdsourced OSINT repository because it does an excellent job in overcoming some of the limitations highlighted in Sections 2A and 2B. It is, however, only a single example of the many available OSINT repositories available for use. The ideal TIP would aggregate feeds from various CTI sources. MalwareBazaar was chosen as a proof of concept, specifically for IOCs pertaining to malware samples. Other OSINT repositories which could be used in addition to MalwareBazaar include those found in the MISP Threat Sharing platform, such as blocklist.de and further offerings by abuse.ch such as URLhaus [11, 25].

3) *Database*

As virtually all CTI feeds store data in either CSV or JSON format, a schemaless NoSQL database would be the appropriate solution for the storage and retrieval of malware sample metadata. This database would act as a central repository for all retrieved data, fresh and historical. MongoDB was selected as it was determined to be the most suitable NoSQL offering. MongoDB is ACID-compliant, which is especially important for bulk-importing/deleting large historical datasets. The Community Edition is free to install. A BI connector and ODBC driver are also freely available, allowing for compatibility with a variety of business intelligence tools. This is a key requirement for the TIP this research sets out to develop, as a database with limited compatibility would reduce the selection of analytics tools available to the developer – each researcher or organisation may choose to utilise different platforms for visualising their collected data, a highly-compatible database provides this convenience.

4) *Data Analytics Tool*

An effective analytics tool will allow users to create charts which clearly highlight the relationship between various fields of the CTI dataset, e.g. from which region most samples of a particular malware family were reported. Having a large variety of chart types available is important, as some charts better suit certain types of data (e.g. geographical, statistical, temporal), and some stakeholders may prefer to view data in a particular way. Having a solution capable of catering to a wide audience is key for delivering presentations and reports, which are common use cases for TIPs. Microsoft's Power BI, the go-to for many organisations, was the clear choice of data analytics tool for this research. Power BI Desktop is a cost-free solution which allows for detailed reports to be easily produced, with a great range of charts available for use.

B. *Design*

The AWS EC2 cloud computing service allows customers to spin up and operate virtual machines (VMs) in the cloud. As Power BI Desktop is only available for Windows, a Windows Server 2019 instance was deployed to host the TIP. The EC2 instance communicates with the MalwareBazaar API to retrieve the malware sample metadata, saving each retrieval as a JSON document, then uses the "mongoimport" tool to ingest the data into the MongoDB database collection. This is all handled by a PowerShell script. Optionally, historical data may also be retrieved and formatted for ingestion. An ODBC connection is configured between the MongoDB database and Power BI. The data is then imported into Power BI, and visualisation can be produced from there. Figure 1 presents a high-level diagram of this architecture.

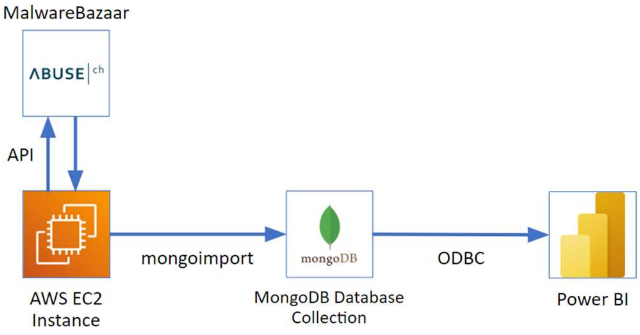

Fig. 1. Simple model TIP design.

Note that this TIP design is a simple model to demonstrate the value of crowdsourced OSINT, freeware, and cloud services for the development of a cost-efficient TIP solution. It is not intended for use in a professional setting. A well-architected design suitable for a production deployment is referenced in Section 5.

C. *Development*

1) *Fresh Data Ingestion*

Before commencing data ingestion, a MongoDB database and collection were created in order to receive and store data. The MongoDB GUI, Compass, was installed and used. A PowerShell script was developed which communicates with the MalwareBazaar API each hour to download a JSON file containing the metadata of each malware sample (e.g. hashes, file size, malware family) uploaded to the repository within the past hour. Upon download, each file is automatically imported into the MongoDB database collection.

2) *Historical Data Ingestion*

Analysing historical malware data alongside fresh data provides a more holistic perspective of the cyber threat landscape, enabling analysts to view trends over time. MalwareBazaar allows users to download a CSV file containing all ingested malware samples from the repository's inception to the current point in time. This file, however, requires enrichment as it lacks certain metadata fields present in samples retrieved through the MalwareBazaar API. Using a separate PowerShell script for historical sample retrieval, the samples in the historical data CSV file are enriched by querying their missing fields through the API, and formatted for output to JSON files in the same structure as the hourly samples. This ensures coherence for when the fields are read from MongoDB and imported into Power BI.

Hourly sample files often contain more than one sample in each JSON file because multiple samples are likely to have been submitted to MalwareBazaar within the span of each hour. However, with the historical data, each sample is a separate JSON document. Because the structure of both the fresh and historical sample files are identical, they are imported as one dataset into Power BI.

3) *Data Visualisation*

After connecting the MongoDB database to Power BI via ODBC, various charts can be produced to visualise all the ingested intelligence. Figure 2 presents an example of charts that can be generated.

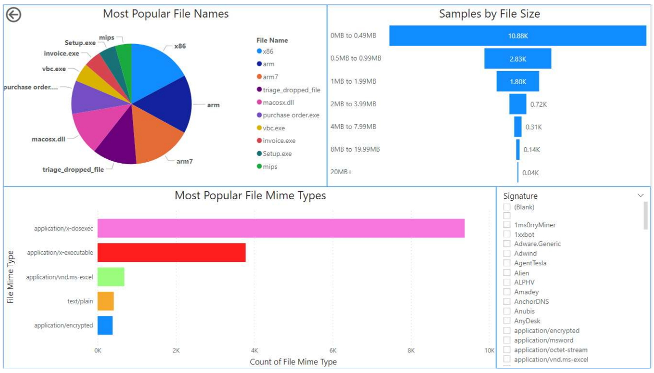

Fig. 2. Example of charts in a Power BI report.

IV. Evaluation

This section assesses key outcomes of the paper's undertaking. The findings obtained from the TIP are discussed, delving into the actionable insight gained from the platform with examples. The cost of operating the TIP is examined, highlighting its affordability and cost advantage over commercial CTI alternatives. Finally, the platform's limitations are identified and discussed.

A. *Chart Analysis*

The charts produced in this paper are merely examples which aim to demonstrate the versatility of the TIP and its capabilities. Any data field(s) may be used in combination with any of the chart types available in a given analytics tool. Live ingestion of fresh samples commenced on 09/01/2022, and lasted until 25/01/2022. The most recent 10,000 historical samples were ingested, up to the commencement of live data ingestion. The oldest of these 10,000 samples dates back to 04/12/2021. Therefore, all insight obtained from these visualisations pertains to the time period of 04/12/2021 to 25/01/2022. A total of 10,350 documents were ingested into MongoDB, amounting to 16,713 malware samples (as seen in Figure 3).

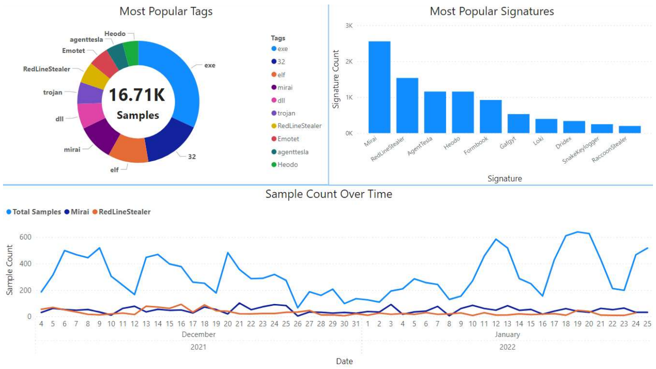

Fig. 3. Further example of charts in a Power BI report.

Within Figure 3, a "Most Popular Tags" donut chart shows that .exe files and 32-bit programs are the most popular types of reported malware in this dataset, followed by .elf files. This signifies that Windows systems are more targeted than Linux ones. A "Most Popular Signatures" column chart shows that Mirai is the most common malware family by a considerable amount, followed by RedLine Stealer, then Agent Tesla. Mirai being the most prominent malware family is coherent with the results observed in the "Most Popular Tags" chart. This better informs security teams of the most current malware threats, allowing for the appropriate mitigations to be employed. A "Sample Count Over Time" line chart shows that the highest number of malware samples was submitted on 19th January, 2022, with significantly less activity around late December 2021. Measures can be created in Power BI for specific malware signatures, showing the number of samples of a particular malware family reported each day compared to other malware. This provides insight into overall malware activity over time, showing which types are most prominent and when.

Clicking on a category in a visualisation updates the other visualisations on that same page, so that they reflect statistics relevant to the selected category. For example, by selecting "elf" from the donut chart in Figure 3, it is revealed that a total of 5,408 samples in the dataset are .elf files (as shown in Figure 4). The majority of these .elf malware files are Mirai and Gafgyt, and the largest number of .elf samples were submitted in late January.

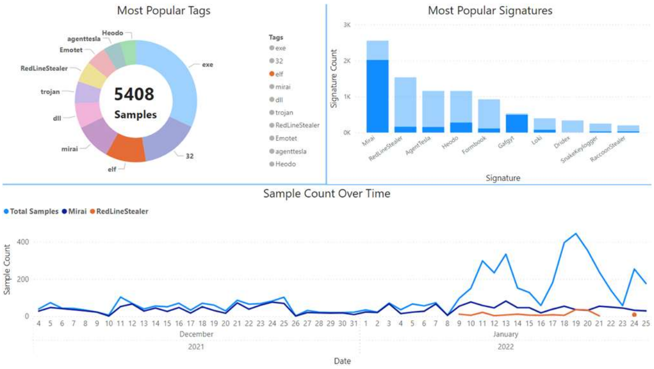

Fig. 4. Filtering for .elf malware samples in the report.

In addition to those presented in this paper, a variety of other charts were also developed, including a "Most Popular File Types" tree map, a "Sample Report Count by Country" filled map, and a "Reporter" table.

B. *Investigation of Agent Tesla*

There is plenty to be learned about a particular malware family by analysing the broad set of statistics pertaining to it. For example, selecting "AgentTesla" on the Power BI report's slicer immediately reveals that this type of malware is most often a Windows executable file no larger than 2 MB in size (see Figure 5). Some of the most common file names for this malware, such as "purchase order.exe" and "shipping documents.exe", suggest that threat actors may attempt to deliver Agent Tesla by disguising it as a legitimate document and emailing it to an organisation's staff. It can be deduced that this malware is commonly used to target the financial sector, judging by the popular names which make reference to "SOA" and "SWIFT".

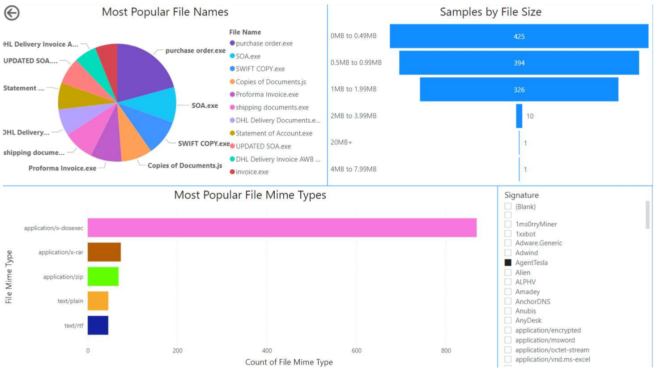

Fig. 5. Slicing for the "AgentTesla" malware signature.

Reviewing the other pages with this slicing reveals further information about Agent Tesla. Of all samples in the dataset, 1,157 were identified as Agent Tesla, and 1.38% of these samples were reported anonymously. The malware is associated with 32-bit executables. It has been reported mostly from France, followed closely by the US, then other European countries. Agent Tesla sample submissions were at their highest in early December 2021. The insight obtained from the TIP regarding Agent Tesla is consistent with previously known information about this malware family. Agent Tesla is a .NET-based trojan usually delivered through phishing emails, mostly targeting the utilities and financial services industries [26, 27].

C. *Cost Analysis*

This TIP model requires no capital expenditure, and runs solely on minor operational expense. All components of the TIP are cost-free other than the AWS services used for infrastructure. AWS Cost Explorer was used to determine how much it costs to operate the TIP on a daily basis. Table 1 shows the daily cost to be $3.19. It can be seen that the t3.large EC2 instance itself was responsible for the majority of the total daily cost, with the Elastic Block Store (EBS) volume costing a minor amount in comparison. The cost of commercial CTI feeds can vary, often ranging from $1,500 to $10,000 per month for a single feed [5]. The monthly cost of this TIP is approximately (3.12 x 30) $93.60, around 6% of the lower bound cost of a commercial feed.

TABLE I. DAILY OPERATIONAL EXPENSE

Usage Type	**Usage Type Total ($)**
EUW2-BoxUsage:t3.large	2.93
EUW2-EBS:VolumeUsage.gp3	0.19
Total Cost	3.12

D. *Limitations of the Model*

Despite the major benefits of this TIP model, there are some limitations worth noting.

A larger number of historical samples were going to be ingested initially; however, this took far too long so a historical dataset of 10,000 samples was settled for. In an attempt to speed up the processing of historical samples, the EBS volume was temporarily upgraded to io2 storage, and the instance was temporarily upgraded to the r5b.8xlarge type. The io2 storage type has dramatically increased performance, with up to 64,000 IOPS and a throughput of up to 1 GiB/s. The r5b.8xlarge instance type offers 32 vCPUs, 256 GiB of RAM, and 10 Gbps network performance. Despite these major enhancements to computation, storage, and network speed, historical sample files were still taking a long time to produce. When writing files with the same number of lines, but populated with dummy data instead of pulling from the API, 1000 lines were written in under a second. This demonstrates that the bottleneck is in the API requests. The abuse.ch servers hosting the MalwareBazaar API are a limiting factor with regard to historical sample enrichment. It took almost four hours to retrieve and write the data for 10,000 samples; this could be an issue for those who wish to enrich a much larger number of historical samples from this repository.

The quality of the visualisations produced may be impacted by a lack of reporting. Even if the quality of each sample ingested is high, if not enough samples are submitted to an OSINT repository then the results observed in the visualisations may be misleading. For example, the "Sample Count Over Time" line chart may not accurately represent the peak times a malware sample is present if it is not reported enough during this timeframe. MalwareBazaar makes sample submission convenient, however there is no way any OSINT repository can guarantee a consistently accurate report count – uncontrollable factors may cause contributors not to report as frequently at certain times. This can result in unclear results when analysing a particular field. For example, it may be hard to determine the most common malware signature for the .elf file type if there is only a small number of reported samples with this file type. A larger dataset paints a clearer picture.

Though MalwareBazaar attempts to identify the malware family of submitted malware samples, the "signature" field is not always populated instantly. This field for some samples remains "null" until MalwareBazaar has had enough time to identify it. This can cause trouble for the ingestion of fresh data. If a sample's signature has not been identified within an hour, it will not contribute to the "signature" field data. This can be resolved by allowing for a larger time interval when ingesting fresh data. Samples may be retrieved daily for example, as opposed to hourly, in order to allow for signature identification. Whilst this may reduce the number of "null" signatures by allowing more processing time, it will in turn reduce the freshness of the ingested data.

V. CONCLUSION AND FUTURE WORK

Using a combination of crowdsourced OSINT, freeware, and cloud services, a cost-efficient TIP with comprehensive data visualisations was produced, serving as a potential alternative to commercial solutions. The insight obtained from these visualisations can better inform security teams of the most relevant and current threats to their organisations, allowing them to employ more focused technical and administrative controls to mitigate such threats.

A production-suitable implementation of the model TIP has been developed as an AWS CloudFormation template [28]. This design conforms to the AWS Well-Architected Framework, ensuring a secure and resilient TIP appropriate for professional usage, whilst remaining a fraction of the cost of commercial solutions. Figure 6 presents a high-level diagram of this architecture. See the referenced GitHub repository for further information.

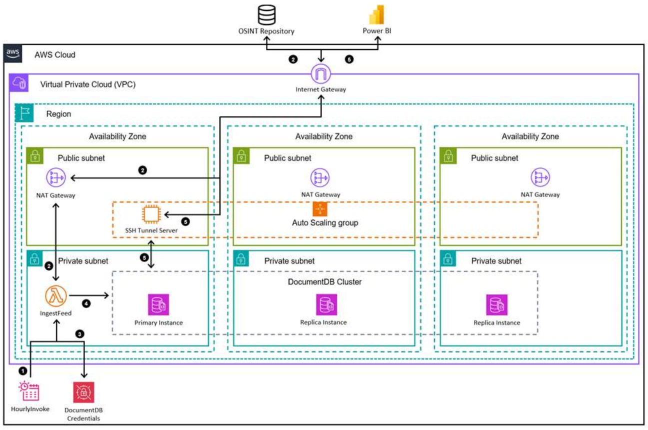

Fig. 6. Well-architected TIP design.

An equivalent infrastructure as code template of the above architecture may be developed for other cloud service providers, such as Microsoft Azure or Google Cloud Platform, potentially using Terraform.

The proposed TIP should be further developed through integration of additional high-quality OSINT repositories, such as those found in MISP, for analysis of a more comprehensive dataset spanning IOCs of all types. The TIP's user should tailor the OSINT feeds to ensure ingestion of IOCs relevant to their requirements, as opposed to ingesting all data available. This is commonly done by specifying parameters when querying a repository's API. Analysts may choose to ingest feeds in STIX format, for relational analysis of CTI components as graph data [11, 25].

For a more intelligent TIP, one may upgrade to the Pro or Premium versions of Power BI in order to access additional analytical features, such as AI-powered data modelling. Implementing anomaly detection combined with a notification service can provide analysts with early identification of emerging developments, enhancing their proactiveness.

REFERENCES

[1] Accenture Security & Ponemon Institute, "Ninth Annual Cost of Cybercrime Study", Mar. 2019. [Online]. Available: https://www.accenture.com/_acnmedia/PDF-96/Accenture-2019-Cost-of-Cybercrime-Study-Final.pdf

[2] IBM Security & Ponemon Institute, "Cost of a Data Breach Report 2021", 2021. [Online]. Available: https://www.ibm.com/downloads/cas/OJDVQGRY

[3] Palo Alto Networks, "What is a Threat Intelligence Platform". Accessed: Apr. 24, 2022. [Online]. Available: https://www.paloaltonetworks.com/cyberpedia/what-is-a-threat-intelligence-platform

[4] G. González-Granadillo, M. Faiella, I. Medeiros, R. Azevedo, and S. González-Zarzosa, "ETIP: An Enriched Threat Intelligence Platform for improving OSINT correlation, analysis, visualization and sharing capabilities" in *Journal of Information Security and Applications*, vol. 58(102715), pp. 1-15, May 2021. [Online]. Available: https://doi.org/10.1016/j.jisa.2020.102715

[5] E. Tittel. "Five criteria for purchasing from threat intelligence providers". TechTarget.com. https://www.techtarget.com/searchsecurity/feature/Five-criteria-for-purchasing-threat-intelligence-services (accessed Apr. 24, 2022).

[6] D. Shackleford, "Who's Using Cyberthreat Intelligence and How?", SANS Institute, Feb. 2015. [Online]. Available: https://cdn-cybersecurity.att.com/docs/SANS-Cyber-Threat-Intelligence-Survey-2015.pdf

[7] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks" in *Computers & Security*, vol. 72, pp. 212-233, Jan. 2018. [Online]. Available: https://doi.org/10.1016/j.cose.2017.09.001

[8] Verizon, "2015 Verizon Data Breach Investigations Report", Apr. 2015. [Online]. Available: https://doi.org/10.13140/RG.2.1.4205.5768

[9] iSIGHT Partners & Gartner, "Executive Perspectives on Cyber Threat Intelligence", 2015. [Online]. Available: https://scadahacker.com/library/Documents/Threat_Intelligence/iSight%20Partners%20-%20Executive%20Perspectives%20on%20Cyber%20Threat%20Intelligence.pdf

[10] C. Yucel, A. Lockett, I. Chalkias, D. Mallis, and V. Katos, "MAIT: Malware Analysis and Intelligence Tool" in *Information & Security: An International Journal*, vol. 50(1), pp. 49-65, 2021. [Online]. Available: https://doi.org/10.11610/isij.5024

[11] C. Wagner, A. Dulaunoy, G. Wagener, and A. Iklody, "MISP – The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform" in *WISCS '16: Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, 2016, pp. 49-56, doi: 10.1145/2994539.2994542.

[12] D. Chismon and M. Ruks, "Threat Intelligence: Collecting, Analysing, Evaluating", MWR InfoSecurity, 2015. [Online]. Available: https://nsarchive.gwu.edu/sites/default/files/documents/5023734/United-Kingdom-Government-Threat-Intelligence.pdf

[13] United States Government. (2006, Jan. 6). *Public Law 109–163, National Defense Authorization Act For Fiscal Year 2006*. [Online]. Available: https://www.congress.gov/109/plaws/publ163/PLAW-109publ163.pdf

[14] R. Campiolo, L. A. Santos, D. M. Batista, and M. A., Gerosa, "Evaluating the Utilization of Twitter Messages as a Source of Security Alerts" in *SAC '13: Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 2013, pp. 942-943, doi: 10.1145/2480362.2480542.

[15] C. Sabottke, O. Suciu, and T. Dumitraş. (Aug. 2015). "Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits". Presented at Proceedings of the 24th USENIX Security Symposium. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke

[16] G. Hribar, I. Podbregar, and T. Ivanuša, "OSINT: A "Grey Zone"?" in *International Journal of Intelligence and CounterIntelligence*, vol. 27(3), pp. 529-549, May 2014. [Online]. Available: https://doi.org/10.1080/08850607.2014.900295

[17] A. Doan, R. Ramakrishnan, and A. Y. Halevy, "Crowdsourcing Systems on the World-Wide Web" in *Communications of the ACM*, vol. 54(4), pp. 86-96, Apr. 2011. [Online]. Available: https://doi.org/10.1145/1924421.1924442

[18] R. Trost. (Sep. 2015). "Crowdsourcing Intelligence: Friend or Foe?!". Presented at 2015 Cybersecurity Innovation Forum. [Online]. Available: https://csrc.nist.gov/Presentations/2015/Crowdsourcing-Intelligence-Friend-or-Foe

[19] "MalwareBazaar | About". Abuse.ch. https://bazaar.abuse.ch/about/ (accessed Dec. 08, 2021).

[20] "MalwareBazaar | FAQ". Abuse.ch. https://bazaar.abuse.ch/faq/ (accessed Dec. 08, 2021).

[21] D. Lunghi. (Jun. 2021). "Taking Advantage of PE Metadata, or How To Complete your Favorite Threat Actor's Sample Collection". Presented at SSTIC 2021. [Online]. Available: https://www.sstic.org/media/SSTIC2021/SSTIC-actes/Taking_Advantage_of_PE_Metadata_or_How_To_Complete/SSTIC2021-Article-Taking_Advantage_of_PE_Metadata_or_How_To_Complete_your_Favorite_Threat_Actor_Sample_Collection-lunghi.pdf

[22] A. Groenewegen and J. S. Janssen, "TheHive Project: The maturity of an open-source Security Incident Response platform", University of Amsterdam, Jun. 2021. [Online]. Available: https://www.researchgate.net/publication/352715439_TheHive_Project_The_maturity_of_an_open-source_Security_Incident_Response_platform

[23] P. Mohandas, S. K. Santhosh Kumar, S. P. Kulyadi, M. J. Shankar Raman, V. V. S and B. Venkataswami, "Detection of Malware using Machine Learning based on Operation Code Frequency" in *2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2021, pp. 214-220, doi: 10.1109/IAICT52856.2021.9532521.

[24] Y. Ma, S. Liu, J. Jiang, G. Chen, and K. Li, "A Comprehensive Study on Learning-Based PE Malware Family Classification Methods" in *ESEC/FSE 2021: Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2021, pp. 1314-1325, doi: 10.1145/3468264.3473925.

[25] A. Ramsdale, S. Shiaeles, and N. Kolokotronis, "A Comparative Analysis of Cyber-Threat Intelligence Sources, Formats and Languages" in *Electronics*, vol. 9(5), 824, May 2020. [Online]. Available: https://doi.org/10.3390/electronics9050824

[26] "Strategic Analysis: Agent Tesla Expands Targeting and Networking Capabilities". Cofense.com. https://cofense.com/strategic-analysis-agent-tesla-expands-targeting-and-networking-capabilities/ (retrieved Apr. 22, 2022).

[27] G. More. "Catching the RAT called Agent Tesla | Qualys Security Blog". Qualys.com. https://blog.qualys.com/vulnerabilities-threat-research/2022/02/02/catching-the-rat-called-agent-tesla (accessed Apr. 22, 2022).

[28] *A Cost-Efficient Threat Intelligence Platform Powered by Crowdsourced OSINT.* (2023). AlphaKiloDelta (GitHub account holder). [Online]. Available: https://github.com/AlphaKiloDelta/A-Cost-Efficient-Threat-Intelligence-Platform-Powered-by-Crowdsourced-OSINT