




Article

Key Requirements for the Detection and Sharing of Behavioral Indicators of Compromise

Antonio Villalón-Huerta ¹, Ismael Ripoll-Ripoll ² and Hector Marco-Gisbert ^{2,*}

¹ S2 Grupo, Ramiro de Maeztu 7, 46022 Valencia, Spain; antonio.villalon@s2grupo.es

² Department of Computing Engineering, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; iripoll@disca.upv.es

* Correspondence: hecmargi@disca.upv.es

Abstract: Cyber threat intelligence feeds the focus on atomic and computed indicators of compromise. These indicators are the main source of tactical cyber intelligence most organizations benefit from. They are expressed in machine-readable formats, and they are easily loaded into security devices in order to protect infrastructures. However, their usefulness is very limited, specially in terms of time of life. These indicators can be useful when dealing with non-advanced actors, but they are easily avoided by advanced ones. To detect advanced actor's activities, an analyst must deal with behavioral indicators of compromise, which represent tactics, techniques and procedures that are not as common as the atomic and computed ones. In this paper, we analyze why these indicators are not widely used, and we identify key requirements for successful behavioral IOC detection, specification and sharing. We follow the intelligence cycle as the arranged sequence of steps for a defensive team to work, thereby providing a common reference for these teams to identify gaps in their capabilities.

Keywords: cyber threat intelligence; indicator of compromise; IOC; TTP; MITRE ATT&CK



Citation: Villalón-Huerta, A.; Ripoll-Ripoll, I.; Marco-Gisbert, H. Key Requirements for the Detection and Sharing of Behavioral Indicators of Compromise. *Electronics* **2022**, *11*, 416. <https://doi.org/10.3390/electronics11030416>

Academic Editors: Changhoon Lee, Yu Chen and Jake (Jaeik) Cho

Received: 28 December 2021

Accepted: 26 January 2022

Published: 29 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Indicators of compromise (IOC) are key to cyber threat intelligence (CTI), as they enable and speed up the detection of malicious activities in technological infrastructures. They allow one to specify both the usage of technological capabilities, such as tools or artifacts, and the tactics, techniques and procedures (TTP) developed by threat actors. However, this last use case, the specification of TTP, is not extended among threat intelligence providers. These providers focus on the sharing of basic indicators, which provide immediate results when loaded into security platforms but which present an important problem: their lifespans. As they are easily modified by hostile actors, their usefulness is limited. In other words, most of the IOC shared today in threat intelligence sharing platforms are not the best ones, but the easiest to use ones.

We need cyber threat intelligence sharing to detect and respond to hostile actors' activities. We usually get indicators that allow us to achieve this goal on two main fronts: compromised hosts, with indicators such as hashes, filenames or mutexes; and networks, with indicators such as IP addresses or domain names. In fact, it is usual to differentiate indicators based on where they are seen [1]: network and host-based ones. However, apart from the problem of false positives with these atomic and computed indicators of compromise [2], those simple IOC, as we have stated before, have limited usefulness. For this reason, we must focus on the effective detection and sharing of behavioral IOC to face advanced threats, as these indicators are harder for a hostile actor to modify.

In this work we analyze this situation and we identify the key requirements for the effective detection and sharing of behavioral indicators of compromise; IOC of this type represent the tactics, techniques and procedures of threat actors, and their values are much higher than those of the basic indicators. By not exploiting and sharing them, defensive teams present an important gap in the detection of malicious activities.

The contributions of this paper are as follows:

- Analyzing the problems of specification, detection and sharing of behavioral indicators of compromise.
- Extracting the key features of cyber operations from advanced threat actors.
- Identifying and structuring the key requirements, from an intelligence perspective, for the detection and sharing of behavioral indicators of compromise.
- Identifying current efforts to fulfill those requirements and the failures of those efforts.

The rest of the paper is organized as follows. The background, Section 2, provides concepts regarding indicators of compromise and the intelligence cycle. In Section 3, we assess the problem of the detection and sharing of indicators of compromise. Section 4 analyzes the different approaches identified to specification and sharing of tactics, techniques and procedures. In Section 5, we identify the key requirements for specification and sharing among actors, and we discuss those requirements and the current status in Section 6, where future lines of research are also identified. Finally, Section 7 highlights the main results of our work.

2. Background

2.1. Indicators of Compromise

In CTI, an indicator of compromise is defined [3] as a piece of information that can be used to identify a potentially compromised system. This piece of information can range from a simple IP address to a complex set of tactics, techniques and procedures. In all cases, this information meets the definition of IOC: it can be used to identify a potentially compromised system.

Most researches [4–6] follow the classification of IOC stated by [7,8]. This classification defines the following three categories for IOC, based on their levels of complexity and related to the granularities of data represented by them:

- Atomic. Atomic indicators are those which cannot be broken down into smaller parts and retain their meanings in the context of an intrusion. Examples of atomic indicators include IP addresses and domain names.
- Computed. Computed indicators are those which are derived from data involved in an incident. Examples of computed indicators include hash values and regular expressions.
- Behavioral. Behavioral indicators are collections of computed and atomic indicators, often subject to qualification by quantity and possibly combinatorial logic. An example of a complex behavioral indicator could be repeated social engineering attempts of a specific style via email against low-level employees to gain a foothold in the network, followed by unauthorized remote desktop connections to other computers on the network delivering specific malware [4]; a simpler example could be a document file creating an executable object. Such indicators are captured as tactics, techniques and procedures, representing the modus operandi of the attacker [6].

While behavioral indicators of compromise are related to operational threat intelligence, atomic and computed ones are related to tactical threat intelligence. All those indicators are relevant to detecting compromises, but tactical intelligence has a shorter lifespan than operational intelligence, and it can also be more easily evaded, so in general terms it is less useful. In Figure 1, the relationship between indicators of compromise and intelligence levels is shown.

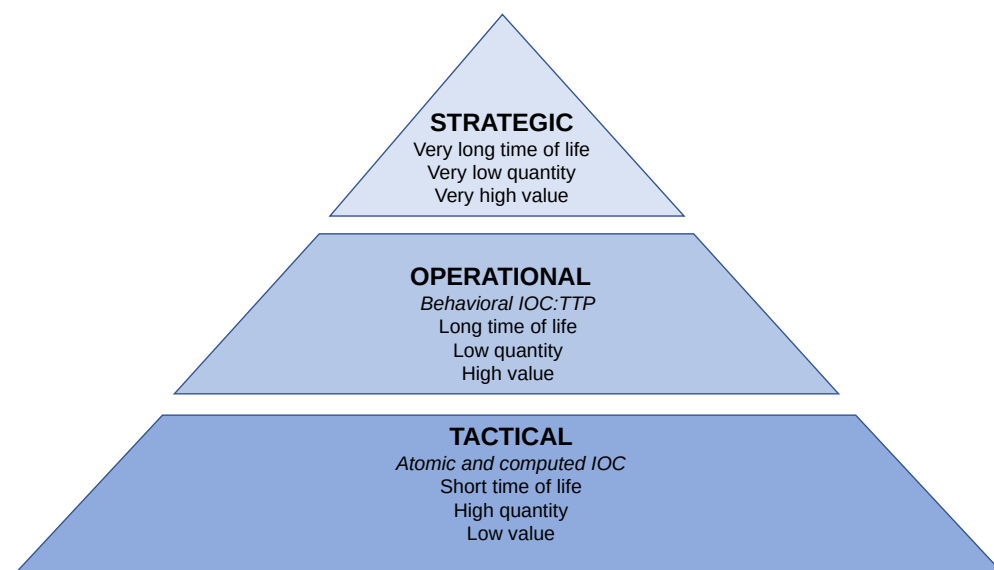


Figure 1. Indicators of compromise and intelligence levels.

Although less useful than behavioral ones, atomic and computed indicators of compromise are considered by many organizations the most valuable pieces of threat intelligence [9]. The main justification for this perception is related to the fact that the indicators representing tactical intelligence are usually expressed in machine-readable formats, so they can be easily loaded into security devices, providing immediate results. On the other hand, operational or strategic intelligence feeds in most cases require manual processing.

In today's interconnected world, it is not possible to deal with security in an isolated box; incidents are not unique, and organizations inhabit security ecosystems with common threats, vulnerabilities, risks and capabilities. Thus, in order to enhance one's own security it is mandatory to share cyber threat intelligence with other parties, such as private companies, interest groups or law enforcement agencies. Furthermore, of course, inside this scheme, IOC are a key piece: they are in fact the most shared type of threat intelligence. However, the major part of available shared data relates to atomic and computed indicators [10–12], such as IP addresses, file hashes or domain names. Data related to higher level threat intelligence, the most useful data, are by far less shared, so they are less used. In fact, when dealing with indicators of compromise, it is usual to refer only to atomic and computed ones [13].

2.2. Intelligence Cycle

NATO [14] defines intelligence as the product resulting from the directed collection and processing of information regarding the environment and the capabilities and intentions of actors, in order to identify threats and offer opportunities for exploitation by decision-makers. The same work also defines the intelligence cycle as the sequence of activities whereby information is obtained, assembled, converted into intelligence and made available to users. Although there are different versions of this cycle, we can summarize them in the following five steps:

- **Direction.** Determination of intelligence requirements, planning the collection effort, issuance of orders and requests to collection agencies and maintenance of continuous checking on the productivity of such agencies.
- **Collection.** The exploitation of sources by collection agencies and the delivery of the information obtained to the appropriate processing unit for use in the production of intelligence.
- **Processing.** The conversion of information into usable data suitable for analysis.
- **Analysis.** Tasks related to integration, evaluation or interpretation of information to turn it into intelligence: a contextualized, coherent whole.

- Dissemination. The timely conveyance of intelligence, in an appropriate form and by any suitable means, to those who need it.

A sixth step for the intelligence cycle [15] would be the evaluation and feedback: all steps, but specially the dissemination of the final product, feed a new iteration of the cycle, as shown in Figure 2.

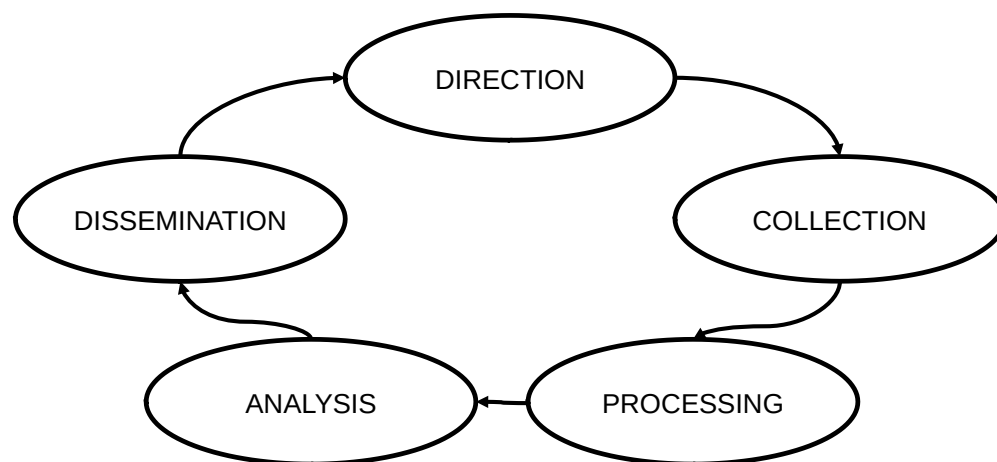


Figure 2. Intelligence cycle.

Intelligence as a product is regarded as the final result of a set of actions, which are sequentially launched; this set of actions, the intelligence cycle, is a simple explanation of a complex intelligence process. It starts when someone, such as an authority or a government, has particular information needs in order to make the best decision about a subject. At this point the cycle starts, identifying the requirements and planning the acquisition of the information that will be later processed and analyzed, in order to generate intelligence.

Once planned, the next stage is to acquire information, and this acquisition can be performed through different intelligence collection disciplines [16] commonly referred as “the INTs”: Signals Intelligence (SIGINT), open-source intelligence (OSINT), measurement and signature intelligence (MASINT), human intelligence (HUMINT) and geospatial intelligence (GEOINT). The essential elements of these INTs are not formally defined [17], neither are they agreed on between authors, but they define the families of sources the information can be gathered from: for example, a simple public website, a satellite, an intercepted artifact or a mole.

With the information gathered, processing and exploitation turn the information previously collected into a form suitable for the production of finished intelligence [18]. This stage includes tasks such as decryption, translation or data conversion, and as a part of the cycle, it is mandatory to analysis, in which the intelligence, the final product, is generated. This analysis must include the information gathered and processed, no matter which collection discipline it comes from. In this sense, we can refer to all-source intelligence, defined by [19] as, “The intelligence products, organizations and activities that incorporate all sources of information and intelligence, including open-source information, in the production of intelligence”.

Finally, once the intelligence as a product has been generated, it is delivered to the customer, the entity which had the information needs stated before, in a suitable form for its use and by a variety of means. This product will be used to aid the decision making process, and possibly, to start a new iteration of the intelligence cycle.

3. The Issue

3.1. Threat Specification

In CTI, many efforts have been made in order to characterize threats, campaigns and particular attacks by indicators of compromise, especially to establish a structured

and standardized information sharing scheme between actors. Back in 2007, the Internet Engineering Task Force (IETF) defined [20] IODEF (Incident Object Description Exchange Format), an XML data representation that provides a framework for sharing information commonly exchanged about computer security incidents. Although it is fairly static [21], over the years IODEF has been extended for different needs, such as the reporting of phishing events [22].

Private companies have also developed well-known standards to enable threat information sharing. Mandiant's OpenIOC is an extensible XML schem designed to describe the technical characteristics of evidence of compromise [23]. It provides indicators about files (such as full paths, imports and exports, or compile times), hosts and networks (such as DNS or URI), processes (such as handles or paths), registry entries (such as names or text), services (such as name or DLL) and signatures (such as Snort or Yara), among others.

Structured Threat Information Expression (STIX) is a language and serialization format used to exchange cyber threat intelligence, whose goal is to identify and represent all the elements of cyber threats in a flexible, automatable and human-readable way. Upon a standardized language, in XML format, the standard provides a common mechanism for addressing structured cyber threat information, improving consistency, efficiency, interoperability, and overall situational awareness through a unified architecture that structures and links all those elements of a threat, from lower level, observables or indicators, to higher level, campaigns and actors [24]. STIX was defined in 2012; it was sponsored by US Department of Homeland Security. In 2015, all the intellectual property and trademarks associated with STIX were licensed to OASIS, a nonprofit organization focused on the development and integration of open technological standards.

As of its release of version 2.0, STIX integrates Cyber Observable eXpression (CybOX), a structured language for cyber observables also developed by MITRE. In STIX 2.1, the latest version at the time of writing, the standard defines three types of core objects to represent cyber threat intelligence: one of them, SCO (STIX Cyber-observable Object), is used to characterize host-based and network-based information. SCO was introduced in STIX 2. In previous versions cyber-observables could only exist as objects within an Observed Data object. SCO represents observed facts about a network or host that may be used and related to higher level intelligence to form a more complete understanding of the threat landscape. STIX 2.1 defines the cyber-observable objects shown in Table 1, each of them with its corresponding properties.

Table 1. STIX 2.1 Cyber-observable objects.

Artifact	Autonomous System (AS)	Directory
Domain name	Email Address	Email Message
File	IPv4 Address	IPv6 Address
MAC Address	Mutex	Network Traffic
Process	Software	URL
User Account	Windows Registry Key	X.509 Certificate

Although both of them are still available and extensions can be made, OpenIOC and IODEF are today considered legacy formats [25]. STIX has been widely adopted, its particular applications having been explored in various fields, from malware detection [26] to critical infrastructure and industrial control system protection [27,28]. It has been extended to detect more complex patterns [29], and it can be used to provide improvements to other formats, being able to embed not only IODEF extensions but also other formats, such as OpenIOC or Yara rules [29,30]. In this way, STIX can be considered the most accepted CTI standard among the security community and the de facto one for describing threat intelligence data [31,32]. The European Union Agency for Cybersecurity (ENISA) [33] has recommended European Union states to implement STIX as a globally accepted standard.

3.2. Real-World IOC

Despite all the efforts toward the characterization of threats exposed in the previous section, the most shared indicators of compromise are still the simplest ones. STIX is a complex standard, but it is mainly used to share atomic and computed indicators. The identification of the most used and shared types of indicators of compromise is a key question, as they will be the ones that hostile actors will try to evade in first place. IP addresses addresses and file hashes are considered the most shared indicators of compromise in the current literature [34]. Other authors progressively expanded this list to domain names [10], URLs [35] or malware signatures [11]. Reference [36] also includes in their list file names, dynamic link libraries, registry keys, email addresses, message objects and attachments or links inside messages.

We have processed all the information (available events) from some private MISP (Malware Information Sharing Platform) instances used in public and private security operations centers. Hashes such as MD5, SHA1 and SHA256 represent the most used types of indicator (23.23%), followed by IP addresses (21.10%) and domains and hostnames (19.75%). All the other types of indicators analyzed represent the remaining 35.92%, and their presence is far rarer (as an example, mutexes represent only 0.03% of global indicators). These results are consistent with the hypothesis stated in [37] about the types of indicators generally available, and of course they are aligned with our own experience.

Hashes are mainly linked to implants, whereas IP addresses and domain names are linked to command and control (C2) or exfiltration servers. This means that, theoretically, only with these kinds of indicators can we detect most activities in the persistence stage of an attack. In spite of the fact that most of the shared indicators among the community belong to these three classes, this security perception is not real. All of them are easily changed by a hostile actor, so the discovery of hashes, IP addresses and domain names causes little pain to the attacker, as the “Pyramid of Pain” [38] states. Any hostile actor who wants to evade detection will defeat, at least, these three types of IOC, as they are the most used ones. If an actor is able to cheat these indicators, it will be able to evade approximately 65% of the defender detection capabilities.

As main IOC types are almost useless when facing advanced threats, analysts have to look for alternatives to detect intrusions. Different approaches to provide other atomic and computed indicators of compromise have been developed, but they are not commonly shared among intelligence groups, so their usage is not as extended as it should be. In many cases this is mainly due to the lack of automatic tools to load intelligence and to get immediate results. For example, to identify minor changes between objects, which of course produce different hashes, fuzzy hashes [39,40] have been used. Through algorithms such as ssdeep and sdhash, this approach is able to detect similarities between files, helping the analyst to identify those changing objects. However, in any case, this similarity hashing provides just another computed indicator that again can be easily evaded by an advanced actor.

To provide more accurate detection, CTI must deal with the detection and sharing of behavioral indicators. This approach would allow analysts to detect the tactics and techniques of attackers no matter which atomic or computed indicators they use in a particular campaign (this is, no matter which hash, which IP address or which domain name). However, CTI sharing is mainly focused on those simple indicators of compromise that are easily evaded. In [41], the authors state that hashes, IP addresses and domain names are the easiest indicators to trace, to identify and to exploit quickly. As all of them can be easily expressed in machine-readable formats, for example, a simple blacklist, many security devices can be configured to load them automatically, so they provide immediate results. In fact, such types of indicators focus on immediacy [42], whereas complex ones, those related to goals or TTP, provide richer analysis but take more time to process.

4. Approaches and Limitations

The detection of behavioral indicators, instead of atomic or computed ones, has been largely studied in the malware field [43,44]; the so called behavioral signatures are applied in dynamic malware analysis and rely on the malware's behavior to identify patterns through multiple means, from the monitoring of system calls to temporal logic formulae [45]. These methods extend the classical static signature detection, complementing or superseding it. However, if we do not focus on such a specific piece of an attack, the malware, and we try to expand this detection approach to the tactics, techniques and procedures of an attacker, the work gets complicated.

The STIX "Attack Pattern" SDO describes the tactics, techniques and procedures that adversaries develop to compromise targets: this is, just what behavioral indicators represent. This SDO contains textual descriptions of the patterns, along with references to external objects; it relates to other SDO, such as "Indicator", but these ones are, again, pure atomic or computed indicators of compromise. Therefore, although STIX allows the specification of tactics, techniques and procedures, it does not provide a common vocabulary for describing TTP, making STIX not suitable for a machine-readable specification of behavioral indicators of compromise.

In [46], TTPDrill was presented, a tool to extract threat actions from CTI unstructured text; it is the first approach to represent TTP in a structured form, in this case from published threat intelligence reports. However, although TTPDrill provides a novel model to identify threat actions in a machine-readable format from unstructured text, it relies on specific observables, not providing a general capability for the use of behavioral signatures.

Representing how an adversary works in an operation is not standardized among the CTI community, so this information has to be manually handled in most cases. As the relevant security information is usually consolidated in a security information event management (SIEM) platform, these technologies are the place where this information must be analyzed to detect indicators of compromise. Microsoft has developed Kusto Query Language, used in Azure both to monitor and to perform threat hunting [47], which is a non-standard language that cannot be used *as is* outside the Microsoft ecosystem; in addition, providers such as Elastic have defined their own open rules and language (EQL, Event Query Language) to query Elastic SIEM. Both examples are proprietary ones, and their particular specifications cannot be shared with other technologies.

An important effort towards the normalization of a language that allows analysts to query SIEM technologies in order to detect all kinds of IOC has been made with SIGMA rules. SIGMA <https://github.com/Neo23x0/sigma> (accessed on 23 December 2021) is a generic and open signature format used to describe relevant log events in a straightforward manner. In other words, SIGMA is to SIEM events what Snort is to network traffic or Yara is to files. Although SIGMA's goal is not to standardize a format to describe behavioral indicators of compromise, the language can be used to query SIEM events, and provides full coverage for all kind of indicators, from atomic to behavioral. For example, the Turla Advanced Persistent Threat group executes different lateral movement techniques in a compromised Windows system, identified as T1059, T1077, T1083 and T1135 by MITRE ATT&CK. ATT&CK—Adversarial Tactics, Techniques, and Common Knowledge—is a globally accessible knowledge base of adversary tactics and techniques based on real-world observations. The following SIGMA rule allows one to query an SIEM for these lateral movement techniques, looking for the execution of particular commands in Windows systems that are registered by sysmon and sent to the SIEM:

```
action: global
title: Turla Group Lateral Movement
id: c601f20d-570a-4cde-a7d6-e17f99cb8e7f
status: experimental
description: Detects automated lateral movement by Turla group
references:
```

```
- https://securelist.com/the-epic-turla-operation/65545/
tags:
- attack.g0010
- attack.execution
- attack.t1059
- attack.lateral_movement
- attack.t1077
- attack.discovery
- attack.t1083
- attack.t1135
author: Markus Neis
date: 2017/11/07
logsource:
category: process_creation
product: windows
falsepositives:
- Unknown
---
detection:
selection:
CommandLine:
- 'net use \\%DomainController%\C$ "P@ssw0rd" *'
- 'dir c:\\*.doc* /s'
- 'dir %TEMP%\*.exe'
condition: selection
level: critical
---
detection:
netCommand1:
CommandLine: 'net view /DOMAIN'
netCommand2:
CommandLine: 'net session'
netCommand3:
CommandLine: 'net share'
timeframe: 1m
condition: netCommand1 | near netCommand2 and netCommand3
level: medium
```

This simple example does fit into the category of behavioral IOC, and of course the language allows one to specify more complex rules. SIGMA has become the de facto standard to query SIEM events, but it does not provide full coverage for the specification of all behavioral procedures. This standard must be improved and complemented with post processing capabilities or equivalent over the stored data to be able to specify a full range of behavioral indicators of compromise. In addition, although SIGMA is supported by most SIEM technologies, such as QRadar and Splunk, it is not the native query language in these technologies. This fact, in addition to the limitations of the language, it forces analysts to maintain queries in different platform-dependent query languages if they want to utilize its SIEM capabilities maximally.

User behavior and entity analytics, UEBA, sometimes referred simply as user behavior analytics (UBA, an older concept superseded by UEBA), is also an effort to track behavioral indicators with tools and through integration into SIEM technologies. UEBA offers [48] profiling and anomaly detection based on a range of analytic approaches, usually using a combination of basic (e.g., rules that leverage signatures, pattern matching and simple statistics) and advanced analytics methods (e.g., supervised and unsupervised machine learning). UEBA's strengths are related to its advanced methods, but although these

could be efficient approaches in many cases, machine learning has not been proved to be a practical solution for the hunting of advanced actors. The problems in the anomaly detection field are clear and were identified years ago [49]. For the detection of advanced actors, many of them are linked to the stealth attacks these actors usually perform, and to the complexity of identifying anomalies by establishing a user or entity behavior baseline. Although UEBA is a promising approach, it still lacks maturity regarding advanced actors, and in any case, it does not provide a suitable standard with which to share behavioral indicators between different technologies or platforms.

5. Key Requirements for Behavioral IOC Detection and Sharing

By adopting the intelligence cycle as a working model, we have identified key requirements for the detection and sharing of behavioral IOC. In this section, we provide those requirements in each of the steps of the cycle. They all should be considered in the planning stage, to establish the direction of the process from the collection of data to the dissemination of intelligence in the form of behavioral IOC. Our final goal is to generate behavioral IOC that can be shared and used in an effective way to detect threat actors TTP. These requirements are focused on the detection and sharing of behavioral IOC: they must be considered together with other requirements for a security operations center to perform threat hunting activities in an effective way. For example, we do not emphasize requirements such as automatizing or false positive reduction, as they are not focused on behavioral IOC detection but on general detection capabilities.

During our research, we have analyzed threat actors' cyber operations, information on which is available both in frameworks such as MITRE ATT&CK, in intelligence reports about particular threat actors such as [50–52] and in campaign reports such as [53]. Particularly, we have analyzed the activities of different advanced persistent threats from Russia (APT28, APT29, Turla, etc.), China (APT1, APT17, Ke3chang, etc.) and Iran (APT33, Clever Kitten, etc.). This analysis allowed the identification of key features of advanced offensive cyber operations, including not only their techniques, but also characteristics related to their goals, targets and artifacts. The most relevant identified features are shown in Table 2.

Table 2. Key features in offensive operations.

Feature	Description
Multiple targets	Advanced threat actors target a wide spectrum of victims, including sectors such as military, government, technology, energy or even non-profit organizations
Broad range of techniques	Advanced threat actors achieve their goals through a broad range of techniques. These techniques are usually stealth, in order to go unnoticed, and one single threat actor can execute different techniques linked to the same tactic, even in a single operation against a particular target
Tailored tools and artifacts	Advanced threat actors can use multiple tools and artifacts in their operations. These tools and artifacts range from specifically developed malware to legitimate system tools, and in many cases the threat actor is aware of the deployed counter measures in the target and knows how to evade them
Potential indicators	Hostile activities leave traces in targeted systems and internal network traffic. In addition, the target perimeter security must be monitored in order to detect connections to command and control or exfiltration servers
Compromises spread over time	Once a target is compromised, this compromise spreads over time in most operations, thereby giving the threat actor the ability to control its target for months or years

The fact that a single threat actor targets multiple victims from different sectors is directly linked to the targeting of multiple infrastructures, protected by multiple security technologies. These technologies are provided by different vendors, and each of them uses its custom logging formats and data. In order to guarantee accurate detection and response capabilities, these data must be normalized, regardless of the technological data source, to a common format that allows analysts to search for hostile activities regardless of the technology used. These data must be analyzed through a common platform-independent language that can be shared among different defensive teams and exploited no matter which technologies are monitored and which specific analysis platform is used in each case.

As advanced threat actors can achieve their tactics through different techniques, even in the same campaign, defensive teams must be able to find hostile behaviors regardless of the mechanisms used in each case. This means that the analysis tool, usually the SIEM [54,55], must provide these teams the ability to specify all previously identified techniques and the new ones that are discovered during a particular analysis. This process must be quick for the defensive team, in order to provide agility to the detection of potential compromises.

The employment of tailored tools and artifacts, including legitimate tools provided by operating systems, is related to the stealth techniques executed by threat actors. Covertness being a must in hostile operations, most movements will not generate any alerts in security systems such as antivirus software or firewalls. This situation forces the defensive team to identify not only misuses, but especially anomalies, so it is mandatory to analyze, so to acquire and process, normal activities in systems and networks, and in most cases, to establish a baseline or a reference to define the normal behavior of the users and infrastructures.

Hostile activities leave traces in different points of the targeted infrastructure: the compromised systems and the network traffic. In fact, as we have stated before, atomic and computed indicators of compromise are usually divided into host and network based. A particular subset of network indicators are those related to domain names and IP addresses, which are usually seen on the network's perimeter. This distribution of indicators forces the defensive team to analyze data from multiple sources, establishing relationships between them in a central repository where these data are received and stored. Although not particularly focused on the detection of behavioral IOC, works such as [56,57] reflect the requirement to analyze, and so to acquire, these multi source data. In fact, ATT&CK matrices available online from MITRE define the mandatory data sources for the detection of each technique, most of them being multisource.

Once a threat actor compromises its target, this compromise spreads over time. Persistence periods range from months to years in many cases. For example, a threat actor such as APT1 can maintain access to victim networks for an average of 356 days. Four years and ten months is the longest persistence period reported [58]. For this reason, to make a whole picture of the operation, the defensive team must be able to analyze, and so to store, historical data in order to identify the initial entry point, the hostile activities performed and the internal systems that have been compromised.

To deal with the identified features of advanced offensive cyber operations, we have identified the key requirements for effective behavioral IOC detection and sharing. Following each of the steps of the intelligence cycle, in Table 3 we summarize these requirements. Please note that they are applicable not only in information technology infrastructure, but also in industrial control environments, where cyber threat intelligence is also a must [27,59] and where all the problems we have identified in our work are also present.

Table 3. Key requirements for TTP detection.

IC Stage	Key Requirements
Acquisition	Acquire data from multiple, relevant sources Acquire not only alerts, but regular events
Processing	Central data repository where relationships can be established Common format for stored data Long term retention
Analysis	Platform-agnostic implementation Full native coverage for all techniques Correlation of data from multiple sources Comparison of correlated data against a reference
Dissemination	Machine readable and exportable format Standard query language among providers

5.1. Acquisition

For effective TTP detection, it is mandatory to acquire information from multiple data sources, those where main TTP can be identified. Taking as a reference the MITRE ATT&CK framework [60], where tactics and techniques are analyzed, we found the different data sources that enable the detection of each particular technique. Summarizing these data sources, we identified three main points to acquire data from:

- Endpoint, including not only user endpoints but also servers, where processes are created, files are opened and threat activities are performed at last; this data source includes global infrastructures for endpoints, such as Windows Active Directory.
- Network, including payload and net flow, where threat movements, both lateral and external, are performed.
- Perimeter, where input and output of data between the threat actor and its target is performed, including network devices such as firewalls, data loss prevention systems and virtual private network servers.

In all cases, the mandatory information to acquire is that related to regular activities, not only that related to alerts. Although we identify this key requirement, this global, regular data acquisition is not widely extended [25], thereby impacting the quality of later steps of the intelligence cycle and the final intelligence product. A suspicious behavior is an event or a sequence of events; in most cases, no one event is suspicious by itself. In other words, a behavioral indicator of compromise is not simply a set of atomic or computed ones. Thus, while atomic and computed indicators of compromise can be detected by rule-based systems such as Snort (for network indicators) and Yara (mainly for host indicators), whereas many of the behavioral ones can not be identified this way. These systems provide specific misuse detection capabilities, so it would be hard for them, especially for Yara, to allow an analyst to identify tactics and techniques represented by behavioral indicators of compromise. This is an important point, as most tactics cannot detect security alerts using only a data source: as stated before, most threat actors will not generate security alerts in their regular activities. In fact, from a classical intrusion detection systems perspective, alerts could be considered “misuse detection”, which has to be complemented with “anomaly detection” through the processing and analysis of the acquired regular events.

5.2. Processing

We have also identified particular requirements for the detection of behavioral IOC in the processing stage, once the analysts have acquired specified data. In the first place, to achieve this detection, it is mandatory to have a centralized point where data can be collected; this point is usually a SIEM, where logs from multiple sources are stored in a common format. Related to the data acquisition from multiple sources, as detailed before,

in the processing stage it is a key requirement to receive and store these data, as they will be later correlated in the analysis stage.

Although this should be also a requirement for the detection of all kinds of IOC, in the case of behavioral data, the requirement of having a centralized point with a common format for logs from multiple sources is especially relevant, as it allows the correlation between events from these different sources. As we will detail in the next section, unlike the detection of atomic or computed indicators, techniques to detect behavioral IOC are usually based on the correlations of data from multiple sources, so we must consider this particular requirement as a key one.

In addition, as we detail when speaking on the analysis stage, long-term retention is a must for a successful behavioral IOC detection. In fact, it is a must, from a forensic point of view, to detect all kinds of IOC: when a threat intelligence feed is received, analysts must look backwards for its presence in the stored events. However, when dealing with behavioral IOC, apart from this forensic approach, long-term retention is mandatory to identify stealthy behaviors. The detection of these stealthy techniques requires the analysis of events far in time, to compare them and to establish relationships to identify the behavior of a threat actor. Without this long-term retention, it may not be possible to identify techniques linked to advanced threat actors. Of course, to enable this kind of retention, and considering that an identified key requirement is to gather not only alerts, but also regular events from different sources, big data architectures that can handle all this information are a must, not only for the processing stage, but for all of the activities of the intelligence cycle [61,62].

5.3. Analysis

The analysis stage is perhaps the most important one in the intelligence cycle, although it cannot be accomplished without proper acquisition and processing activities. The requirements we have stated in previous steps of the intelligence cycle are mandatory for a successful analysis, thereby enabling different capabilities for analysts to work, especially through the SIEM.

The first identified key requirement for the analysis is to be able to specify the behavioral IOC in a technology-agnostic way. This requirement implies that the IOC can be used regardless of the SIEM deployed in each case, but also that the same IOC specification regarding particular data sources can be used regardless of the technology of these data sources. For example, given a particular firewall (data source) and an IOC detecting a malicious behavior on the firewall, by analyzing the data on this SIEM, the results should be able to be loaded in any other SIEM to assist firewall technology in the industry. As we will see later in this section, it is a must to be able to load shared behavioral IOC from third parties.

This IOC specification capability has to provide native full-coverage for all identified techniques. This coverage can be achieved through a common query language, such as SIGMA, or through a common format for stored data and a suitable API to query this data. As we have detailed in Section 4, the first option is the most extended among SIEM providers. It would need to overcome the incompatibility of various languages. Following this identified requirement, in the case of SIGMA it is mandatory to expand the language to provide full coverage for all techniques, or to give SIGMA post-processing capabilities to generate results in a format suitable to be managed by common programming languages.

Most techniques cannot be identified by analyzing events from a single data source [63]; in fact, only about ten particular MITRE ATT&CK techniques (out of 185) can be detected using a single data source. Thus, as we have stated before, the ability to establish relationships between events from different sources is a must. To achieve this requirement, SIEM technologies have to provide this capability, especially through the normalization to a common format of the information received from different data sources.

To identify most behavioral indicators, it is also mandatory, in the analysis stage, to establish a relationship between events or alerts by comparing them against a specific refer-

ence. This relationship is usually a temporal one, but it can also be based on dependencies or simply on a comparison against a normality model. For example, a parent–child process match that can be considered suspicious.

SIEM environments are the main repository for storing events from relevant sources from a security perspective. They collect not only misuse alerts, for example, from systems such as intrusion detection systems and antivirus systems, but also "normal" events from all kinds of data sources, from endpoints to firewalls. They also provide the ability to establish relationships between events, so as we have stated before, SIEM should be the right place to identify all types of indicators of compromise.

5.4. Dissemination

At this point, successful detection of behavioral IOC could have been performed in a security operations center. However, as we have stated before, threat intelligence sharing between defensive centers is a must: without a proper sharing, no single center can detect most hostile operations, especially those performed by advanced actors. For this reason, organizations collaborate to define defensive actions against complex attack vectors by sharing information about threats [64]. Thus, now we will approach the last stage of the intelligence cycle that relates to the dissemination, to the sharing, of the final intelligence product.

As with atomic or computed indicators, behavioral ones must be usually shared by uploading them to a threat intelligence sharing platform such as MISP. To do so, behavioral IOC must be exportable as a single intelligence unit, commonly called "hunt". This implies that a single behavioral IOC generated in one platform, for example, in a specific SIEM, can be exported from this SIEM in the form of intelligence unit.

In this dissemination stage, another key requirement for sharing behavioral IOC is that the intelligence unit has to be machine readable: that is, it has to be actionable when loaded into a SIEM. As with atomic and computed indicators of compromise, whose extended use mainly relies on the fact that they can be loaded into security devices and generate immediate results, as we have stated before [41], with behavioral IOC, we identify the same requirement in order for effective sharing.

To achieve this last requirement, we also identify as a previous mandatory requirement a common accepted language for the specification and sharing of behavioral IOC among providers. Without such a capability, the effort that has to be made to translate TTP from natural language to specific, sometimes proprietary technical standards such as the ones referenced in this work, is not acceptable. With the current lack of a common language, analysts have to iterate the specifications among many platforms, with independent and non-compatible specification languages, thereby multiplying their work not only for the initial specification, but most importantly, to its maintenance and upgrade.

Please note that in this section we have focused on the technological requirements for behavioral IOC sharing; other relevant aspects, such as legal or human matters, are out of the scope of our work, and in fact are common to all CTI sharing approaches [65,66], not specifically to the dissemination of behavioral IOC.

5.5. A Practical Example

To provide a practical test for our proposal, we have analyzed a particular technique performed by threat actors and how the different requirements we have identified must be fulfilled. We have chosen the command and control (C2) tactic stated by MITRE ATT&CK, in particular, the T1071.001 technique, related to command and control through web traffic protocols. While using this technique, adversaries may communicate using application layer protocols associated with web traffic to avoid detection and network filtering by blending in with existing traffic. A malicious HTTP/S hit is hidden inside the legitimate web traffic. Commands to the remote system, and often the results of those commands, will be embedded within the protocol traffic between the client and server, as the MITRE ATT&CK framework states. For our example, we have chosen a particular implementation

of the technique in which the hostile actor buys and uses domains whose names contain strings related to legitimate sites, trying in this way to go unnoticed. We must highlight that this technique is behavioral, so it is not possible to detect it through the use of atomic or computed indicators of compromise.

The detection of this particular technique requires one in first place to acquire, process and analyze information from a main data source: the web proxy or equivalent where navigation logs are stored. These logs must be analyzed with a mechanism that is able to detect the use of legitimate strings inside malicious website names; a simple approach can be looking for preidentified legitimate strings related to existing companies, such as “google”, “microsoft” or “adobe”, in the domain names that have been navigated to by the organization.

Once a suspicious navigation event has been identified, a set of actions must be performed by the analysis team. These activities are shown in Figure 3 in the form of playbook. In first place, the analysts will query intelligence sources to check if the domain is suspicious or not—for example, by confirming whether it is registered by the original company or by a third party entity. In the latter, as the domain will be considered suspicious, the analysts will look for more hits in the historical navigation records (for example, in the stored proxy logs with the available retention period) with the goal of identifying the time period the domain has been connected to by the organization.

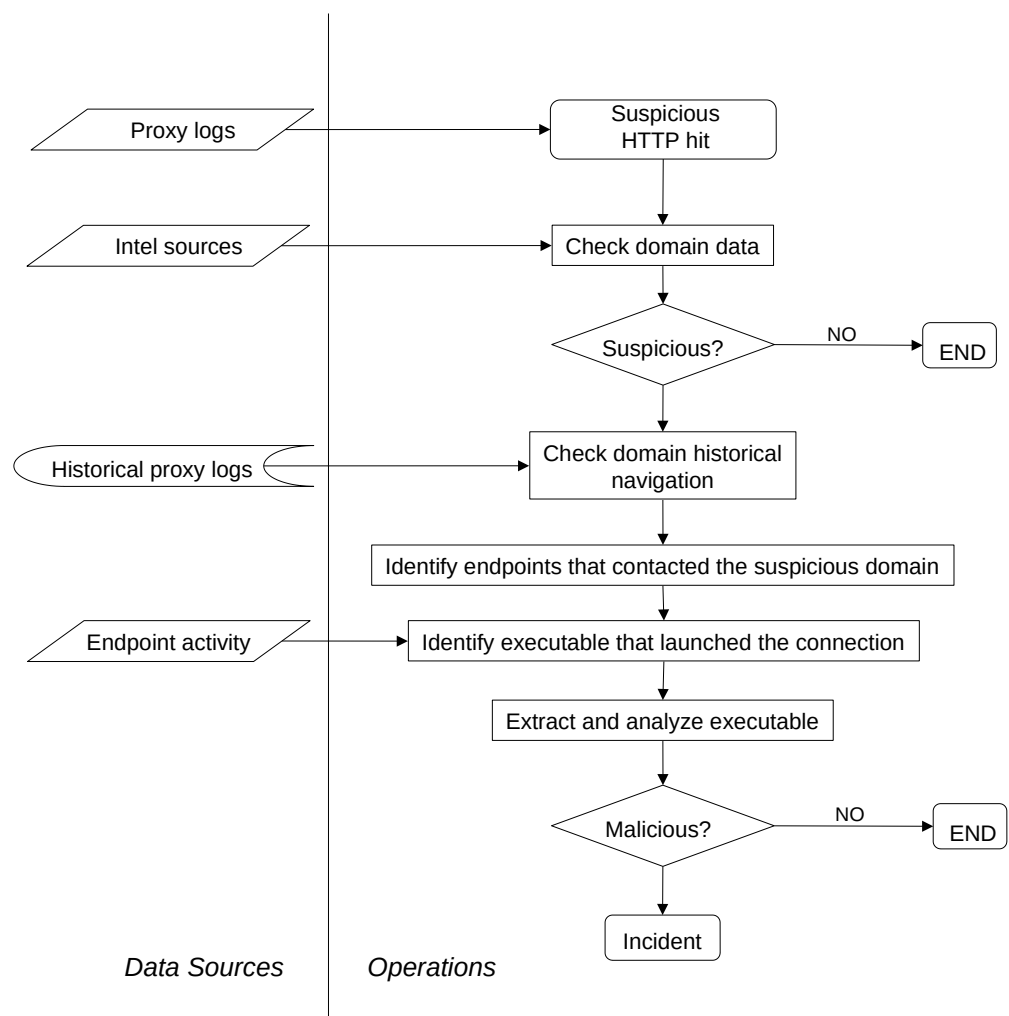


Figure 3. Playbook for the analysis of suspicious HTTP hits.

By analyzing data from these historical records, the analyst will identify the set of endpoints that contacted the suspicious domain; for all of them, and by the analysis of endpoint activity logs, the executable that performed the suspicious connection will be identified, in order to extract and analyze it in a malware laboratory. If this last analysis confirms that the executable is malicious, an incident will be raised; otherwise, it will be considered a false positive and the investigation will be closed. If an incident is detected, it is mandatory to specify the detection of the technique in the form of behavioral IOC, to automate its detection not only in the organization, but also to share it with other interest groups.

With this simple example, we can confirm the mandatory requirements for the detection of a behavioral IOC. In first place, it is clear that data from multiple sources must be acquired to enable the detection of the technique: in this case, from the navigation logs and from the endpoint activity. None of these sources can provide by itself the full picture of the situation to the analyst, as they log different types of information that must be considered together for the whole detection scheme. In addition, as an anomalous domain web hit is not a security violation by itself, it does not generate an alert in any security mechanism. In this way, if we register only alerts, these anomalous hits will go unnoticed.

Regarding the processing step, we must remember in first place the need for retention of the data. We must have a long term retention scheme to identify historical records that help us to draw the global picture of our investigation: the total number of endpoints contacting the suspicious domain, the occasion at which this domain was first seen in the organization or the frequency of hits. In addition, when dealing with medium or large organizations, it is mandatory to centralize information from multiple sources in a single repository and with a common format. This requirement will enable analysts to establish relationships between events from these sources; otherwise, these relationships will be manually established, which is not possible in most organizations.

The analysis step is performed in the SIEM platform. This platform must provide mechanisms for the full analysis of all data from multiple sources, as it has to centralize all the information from these sources and provide the capability to exploit it. In this exploitation, correlation is mandatory for the analysts to establish relationships between different datasets, as they must establish a common reference (usually a temporal one) for all the data. A platform-agnostic implementation is also a must for two main reasons: The first one, to compare data from different technologies across the organization. A single organization can have proxies, firewalls or EDR (endpoint detection and response) from multiple vendors, each of them generating logs in its own proprietary format. The second one is perhaps the most important—to be able to automate and share. As we will see in next step, this means the use of one technique among multiple organizations that will surely have different security technologies.

Finally, in the dissemination step, to be able to share the specifications of the technique with other security groups, and also to receive and exploit specifications from these security groups, the two key requirements we have identified must be fulfilled. The technique must be specified in a standard language among technology providers, particularly SIEM ones, as not all organizations use the same SIEM. In addition, this specification must be exportable, to allow sharing, and most importantly, it must be machine readable to load it automatically in security tools to detect the identified technique.

We have provided a practical example for the detection of a particular technique and identified the fulfillment of all the key requirements we have proposed. Although we have chosen a simple technique used by advanced threat actors, our findings can be applied to all kind of techniques. If we consider more complex techniques coming up against large organizations, the fulfilling of all the requirements is still mandatory.

6. Discussion

The most used types of indicators in CTI sharing are atomic and computed ones: in particular, file hashes, IP addresses and network domains. However, as these indicators of

compromise are easily changed by a threat actor, they can be evaded with little or no effort. While they are definitely not the most useful ones, they are the most used and shared, mainly because they can be expressed in machine-readable formats: this feature allows these indicators to be automatically loaded into security devices, thereby providing an immediate result for the customer. However, this situation represents a problem for security analysts, as most of the shared intelligence is easily evaded by hostile actors, rendering it useless. This problem must be addressed with the detection and sharing of behavioral indicators of compromise. In this work we have identified the key requirements for both of these activities and discussed them.

All but one of the key requirements we have identified in our work are today fulfilled with most SIEM technologies. The main unfulfilled requirement is the definition of a standard query language among providers. There are some efforts to fill this gap, which are mainly aligned with MITRE ATT&CK and especially in SIEM technologies, the most relevant being the SIGMA language. However, until now there is not a commonly-accepted standard suitable for sharing behavioral signatures between analysts using different technologies. This motivates two problems: the inability to share, and thus the difficulty of detecting; and the effort needed to translate TTP from natural language to specific vendor-dependent formats.

These two main problems can only be addressed by the specification of a commonly accepted standard that provides full behavioral detection capabilities to SIEM or to any technologies and products focused on the detection of advanced threat actors. The definition of such a standard will enable the use of these indicators among analysts using different technologies, and as a direct consequence, this usage will enable information sharing between analysts. As this standard becomes accepted and used, and behavioral signatures can be shared and automatically loaded into security systems, the detection capabilities of security teams will increase significantly.

Although there are many efforts to characterize threats among the CTI community, including their behavior, little progress has been achieved toward define a machine-readable format accepted as the standard. In this sense we have identified two main references to be considered: STIX and SIGMA. STIX provides capabilities for defining TTP as SDO, but it does not allow one to specify behavioral signatures; the specification of TTP is based on atomic and computed indicators of compromise, so if an attacker changes them, the particular TTP will not be detected. SIGMA language is the nearest thing to such a standard, but it has to be improved to expand its capabilities. In addition, at the time of writing, it is not natively supported by many SIEM technologies, so SIGMA queries must be turned into specific SIEM signatures through a converter. This fact introduces two main problems: complexity in the management of SIGMA queries among multiple SIEM providers and dependency on the capabilities of those converters.

As we have stated before, from a technical point of view, the only unfulfilled requirement for detecting and sharing behavioral indicators of compromise is the definition of a standard query language among providers. However, the existence of the means does not mean that all defensive teams are able to perform this detection and sharing.

The complexity of our work relied on the identification of a structured methodology suitable for the analysis of the lessons learned after an incident. We adopted the intelligence cycle, as we believe this identification of hostile operations is a counter intelligence activity, so it has to be structured and analyzed in this way. The highlighting of requirements and their arrangement based on a consistent model provides a homogeneous framework to identify gaps in the detection capabilities provided by CERT teams. The analysis of the lessons learned to identify the mandatory requirements for the detection and sharing of behavioral indicators of compromise has its own complexity. This kind of analysis after an incident is not usually public, so we have partially based our research in our own experience in incident handling.

We identify as a future line of research measuring the quality of indicators of compromise. Although behavioral indicators have longer lifetimes and provide more accurate

detection, the analysis of the quality of IOC, in order to decay them, is an ongoing work. We defend that, in general terms, a behavioral IOC is more useful than an atomic or a computed one, but this generality has exceptions, and the quantification of this usefulness must be analyzed in all cases.

Another relevant line of research is the use of behavioral indicators of compromise in emerging or changing applications and technologies in which security is an especially relevant issue. In this sense, we should mention research fields such as smart contracts [67], smart cities [68], Internet of Things [69–71], 5G communications [72] and cryptocurrencies [73], as well as the combined applications for these topics. In all of them, the applications of all kinds of IOC, including behavioral, are a key issue for the detection of security compromises.

7. Conclusions

A relevant problem in cyber threat intelligence is the sharing of behavioral indicators of compromise, those that specify the tactics, techniques and procedures of hostile actors. The fact that atomic and computed indicators of compromise are the most shared and exploited ones reduces the detection capabilities for defensive teams: these indicators are easy to evade, which leaves a window of opportunity for threat actors.

In this work we have analyzed this problem and identified the key requirements for the detection and sharing of behavioral indicators of compromise. To structure our requirements, we have followed the intelligence cycle, so we identified these requirements in each of the stages of the cycle. This identification was based on the analysis of threat actors and their campaigns, from which we extracted the main features of these hostile operations. The detection of any type of IOC has to be performed through a SIEM, where the information gathered from different sources is normalized and centralized. Nowadays, most SIEM technologies provide the mandatory technical capabilities and fulfill all the identified requirements. However, the main barrier to the detection and sharing of behavioral indicators of compromise is the lack of a common machine-readable format to specify behavioral signatures and share them among different SIEM providers.

This lack must be addressed in the short-term through the definition and acceptance of such a standard, a common format to describe behavioral indicators of compromise. While the security community does not define and accept this standard, these indicators will not be massively used, so they will not be massively shared, and the ability to detect the TTP of advanced actors will remain residual. In this sense, we identify SIGMA as the main current effort, but its acceptance among technology providers has to be increased.

Once this common standard has been designed and accepted among the community, the rest of the work will rely on each particular team's defensive capabilities, from acquisition to dissemination. The identified key requirements are mandatory for effective detection of behavioral IOC, but all of them can be fulfilled with current technologies. In this sense, our proposal, structured in the form of intelligence cycle, can be used to measure and compare the capabilities of different defensive teams.

Although the value of a behavioral indicator of compromise is usually higher than the value of a computed or atomic one, in some cases a simple IOC can provide high quality detection. Regarding the measurement and quantification of the quality of an indicator of compromise, the type of IOC (atomic, computed or behavioral) is of course relevant, but many other parameters must be considered, such as the source reliability or the last time the IOC was seen. We identify this measurement as a relevant line of research for future works.

Author Contributions: Writing—original draft, A.V.-H., I.R.-R. and H.M.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abu, S.; Selamat, S.R.; Yusof, R.; Ariffin, A. An Enhancement of Cyber Threat Intelligence Framework. *J. Adv. Res. Dyn. Control. Syst.* **2018**, *10*, 96–104.
2. Rantos, K.; Spyros, A.; Papanikolaou, A.; Kritsas, A.; Ilioudis, C.; Katos, V. Interoperability challenges in the cybersecurity information sharing ecosystem. *Computers* **2020**, *9*, 18. [CrossRef]
3. Harrington, C. Sharing indicators of compromise: An overview of standards and formats. *Emc Crit. Incid. Response Cent.* **2013**.
4. Rid, T.; Buchanan, B. Attributing cyber attacks. *J. Strateg. Stud.* **2015**, *38*, 4–37. [CrossRef]
5. Niakanlahiji, A.; Safarnejad, L.; Harper, R.; Chu, B.T. IoCMiner: Automatic Extraction of Indicators of Compromise from Twitter. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 4747–4754.
6. Skopik, F.; Filip, S. Design principles for national cyber security sensor networks: Lessons learned from small-scale demonstrators. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–8.
7. Cloppert, M. *Security Intelligence: Attacking the Cyber Kill Chain*; SANS Institute: Bethesda, MA, USA, 2009; Volume 26.
8. Hutchins, E.M.; Cloppert, M.J.; Amin, R.M. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Lead. Issues Inf. Warf. Secur. Res.* **2011**, *1*, 80.
9. Brown, R.; Lee, R.M. *The Evolution of Cyber Threat Intelligence (CTI): 2019 SANS CTI Survey*; SANS Institute: Bethesda, MA, USA, 2019.
10. Iqbal, Z.; Anwar, Z. Ontology Generation of Advanced Persistent Threats and their Automated Analysis. *Nust J. Eng. Sci.* **2016**, *9*, 68–75.
11. Vakilinia, I.; Cheung, S.; Sengupta, S. Sharing susceptible passwords as cyber threat intelligence feed. In Proceedings of the MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM), Los Angeles, CA, USA, 29–31 October 2018; pp. 1–6.
12. Kazato, Y.; Nakagawa, Y.; Nakatani, Y. Improving Maliciousness Estimation of Indicator of Compromise Using Graph Convolutional Networks. In Proceedings of the 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 10–13 January 2020; pp. 1–7.
13. Wu, Y.; Huang, C.; Zhang, X.; Zhou, H. GroupTracer: Automatic Attacker TTP Profile Extraction and Group Cluster in Internet of Things. *Secur. Commun. Netw.* **2020**, *2020*. [CrossRef]
14. Office, N.S. *NATO Glossary of Terms and Definitions (English and French)*; NATO Standardization Agency (NSA): Brussels, Belgium, 2018.
15. Joint Chief of Staff. Joint Publication 2-0. Joint Intelligence. Technical Report. 2013. Available online: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf (accessed on 21 December 2021).
16. Boury-Brisset, A.C.; Frini, A.; Lebrun, R. *All-Source Information Management and Integration for Improved Collective Intelligence Production*; Technical Report; Defence Research and Development Canada Valcartier: Quebec, QC, Canada, 2011.
17. Clark, R.M.; Oleson, P.C. Cyber Intelligence. *J. U.S. Intell. Stud.* **2018**, *24*, 11–23.
18. Richelson, J. *The US Intelligence Community*, 7th ed.; Routledge: England, UK, 2016.
19. U.S. Army. *Field Manual 2-0 Intelligence, The US Army*; Headquarters Department of the Army: Washington, DC, USA, 2004.
20. Danyliw, R.; Meijer, J.; Demchenko, Y. The Incident Object Description Exchange Format. RFC 5070. 2017. Available online: <https://datatracker.ietf.org/doc/html/rfc5070> (accessed on 21 December 2021).
21. Burger, E.W.; Goodman, M.D.; Kampanakis, P.; Zhu, K.A. Taxonomy model for cyber threat intelligence information exchange technologies. In Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security, Vienna, Austria, 23–25 October 2014; pp. 51–60.
22. Cain, P.; Jevans, D. Extensions to the IODEF-Documents Class for Reporting Phishing. RFC 5901. 2010. Available online: <https://datatracker.ietf.org/doc/html/rfc5901> (accessed on 17 December 2021).
23. Mavroeidis, V.; Bromander, S. Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. In Proceedings of the 2017 European Intelligence and Security Informatics Conference (EISIC), Athens, Greece, 11–13 September 2017; pp. 91–98.
24. Barnum, S. Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX). *Mitre Corp.* **2012**, *11*, 1–22.
25. Ramsdale, A.; Shiaeles, S.; Kolokotronis, N. A Comparative Analysis of Cyber-Threat Intelligence Sources, Formats and Languages. *Electronics* **2020**, *9*, 824. [CrossRef]
26. Mirza, Q.K.A.; Mohi-Ud-Din, G.; Awan, I. A cloud-based energy efficient system for enhancing the detection and prevention of modern malware. In Proceedings of the 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, Switzerland, 23–25 March 2016; pp. 754–761.
27. Abe, S.; Uchida, Y.; Hori, M.; Hiraoka, Y.; Horata, S. Cyber Threat Information Sharing System for Industrial Control System (ICS). In Proceedings of the 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Nara, Japan, 11–14 September 2018; pp. 374–379.

28. Troiano, E.; Soldatos, J.; Polyviou, A.; Polyviou, A.; Mamelli, A.; Drakoulis, D. Big Data Platform for Integrated Cyber and Physical Security of Critical Infrastructures for the Financial Sector: Critical Infrastructures as Cyber-Physical Systems. In Proceedings of the 11th International Conference on Management of Digital EcoSystems, Limassol, Cyprus, 12–14 November 2019; pp. 262–269.
29. Ussath, M.; Jaeger, D.; Cheng, F.; Meinel, C. Pushing the limits of cyber threat intelligence: Extending STIX to support complex patterns. In *Information Technology: New Generations*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 213–225.
30. Tounsi, W. *Cyber-Vigilance and Digital Trust: Cyber Security in the Era of Cloud Computing and IoT*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
31. Sauerwein, C.; Sillaber, C.; Musmann, A.; Breu, R. Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. In Proceedings of the Wirtschaftsinformatik 2017 Proceedings (Track 8—Information Privacy and Information Security), St. Gallen, Switzerland, 8–15 February 2017; pp. 837–851.
32. Gong, S.; Lee, C. Cyber Threat Intelligence Framework for Incident Response in an Energy Cloud Platform. *Electronics* **2021**, *10*, 239. [[CrossRef](#)]
33. Cyberwiser. *Implementation of the Network and Information Security (NIS) Directive*; Technical Report ETSI TR 103 456; European Telecommunications Standards Institute: Sophia Antipolis Cedex, France, 2017.
34. Rhoades, D. Machine actionable indicators of compromise. In Proceedings of the 2014 International Carnahan Conference on Security Technology (ICCSST), Rome, Italy, 13–16 October 2014; pp. 1–5.
35. Gong, S.; Cho, J.; Lee, C. A Reliability Comparison Method for OSINT Validity Analysis. *IEEE Trans. Ind. Informatics* **2018**, *14*, 5428–5435. [[CrossRef](#)]
36. Tounsi, W.; Rais, H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Comput. Secur.* **2018**, *72*, 212–233. [[CrossRef](#)]
37. Ghazi, Y.; Anwar, Z.; Mumtaz, R.; Saleem, S.; Tahir, A. A supervised machine learning based approach for automatically extracting high-level threat intelligence from unstructured sources. In Proceedings of the 2018 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 17–19 December 2018; pp. 129–134.
38. Bianco, D. The Pyramid of Pain. Technical Report. 2013. Available online: <http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html> (accessed on 13 November 2021).
39. Park, C.; Chung, H.; Seo, K.; Lee, S. Research on the classification model of similarity malware using fuzzy hash. *J. Korea Inst. Inf. Secur. Cryptol.* **2012**, *22*, 1325–1336.
40. French, D.; Casey, W. Fuzzy Hashing Techniques in Applied Malware Analysis. In *Results of SEI Line-Funded Exploratory New Starts Projects*; Software Engineering Institute, Carnegie Mellon University: Pittsburgh, PA, USA, 2012; pp. 2–19.
41. Almohannadi, H.; Awan, I.; Al Hamar, J.; Cullen, A.; Disso, J.P.; Armitage, L. Cyber threat intelligence from honeypot data using elasticsearch. In Proceedings of the 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), Krakow, Poland, 16–18 May 2018; pp. 900–906.
42. Kambara, Y.; Katayama, Y.; Oikawa, T.; Furukawa, K.; Torii, S.; Izu, T. Developing the Analysis Tool of Cyber-Attacks by Using CTI and Attributes of Organization. In *Workshops of the International Conference on Advanced Information Networking and Applications*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 673–682.
43. Jacob, G.; Debar, H.; Filiol, E. Behavioral detection of malware: From a survey towards an established taxonomy. *J. Comput. Virol.* **2008**, *4*, 251–266. [[CrossRef](#)]
44. Ligh, M.; Adair, S.; Hartstein, B.; Richard, M. *Malware Analyst's Cookbook and DVD: Tools and Techniques for Fighting Malicious Code*; Wiley Publishing: Indianapolis, IN, USA, 2010.
45. Beaucamps, P.; Gnaedig, I.; Marion, J.Y. Behavior abstraction in malware analysis. In *International Conference on Runtime Verification*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 168–182.
46. Husari, G.; Al-Shaer, E.; Ahmed, M.; Chu, B.; Niu, X. TTPDrill: Automatic and accurate extraction of threat actions from unstructured text of CTI sources. In Proceedings of the 33rd Annual Computer Security Applications Conference, New York, NY, USA, 4–8 December 2017; pp. 103–115.
47. De Tender, P.; Rendon, D.; Erskine, S. *Pro Azure Governance and Security. A Comprehensive Guide to Azure Policy, Blueprints, Security Center, and Sentinel*; Springer: Berlin/Heidelberg, Germany, 2019.
48. Bussa, T.; Litan, A.; Phillips, T. *Market Guide for User and Entity Behavior Analytics*; Technical Report; Gartner: Stamford, CT, USA, 2016.
49. Gates, C.; Taylor, C. Challenging the anomaly detection paradigm: A provocative discussion. In Proceedings of the 2006 Workshop on New Security Paradigms, Schloss Dagstuhl, Germany, 19–22 September 2006; pp. 21–29.
50. Intelligence, T. APT28: A Window into Russia's Cyber Espionage Operations; Technical Report. Available online: <https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-apt28.pdf> (accessed on 3 October 2021).
51. Mwiki, H.; Dargahi, T.; Dehghantanha, A.; Choo, K.K.R. Analysis and triage of advanced hacking groups targeting western countries critical national infrastructure: APT28, RED October, and Regain. In *Critical Infrastructure Security and Resilience*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 221–244.
52. Utterback, K. An Analysis of the Cyber Threat Actors Targeting the United States and Its Allies. Ph.D Thesis, Utica College, New York, NY, USA, 2021.
53. Shuya, M. Russian cyber aggression and the new Cold War. *J. Strateg. Secur.* **2018**, *11*, 1–18. [[CrossRef](#)]

54. Kotenko, I.; Chechulin, A. Attack modeling and security evaluation in SIEM systems. *Int. Trans. Syst. Sci. Appl.* **2012**, *8*, 129–147.
55. Kotenko, I.; Chechulin, A. Common framework for attack modeling and security evaluation in SIEM systems. In Proceedings of the 2012 IEEE International Conference on Green Computing and Communications, Besancon, France, 20–23 November 2012.
56. Miloslavskaya, N. Stream Data Analytics for Network Attacks’ Prediction. *Procedia Comput. Sci.* **2020**, *169*, 57–62. [[CrossRef](#)]
57. Ayoade, G.; Chandra, S.; Khan, L.; Hamlen, K.; Thuraisingham, B. Automated threat report classification over multi-source data. In Proceedings of the 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA, USA, 18–20 October 2018.
58. Center, T.M.I. APT1 Exposing One of China’s Cyber Espionage Units; Technical Report. 2014. Available online: <https://www.mandiant.com/media/9941/download> (accessed on 28 October 2021).
59. Wagner, T.D. Cyber Threat Intelligence for “Things”. In Proceedings of the 2019 International Conference on Cyber Situational Awareness, Data Analytics Furthermore, Assessment (Cyber SA), Oxford, UK, 3–4 June 2019; pp. 1–2.
60. Strom, B.E.; Applebaum, A.; Miller, D.P.; Nickels, K.C.; Pennington, A.G.; Thomas, C.B. Mitre ATT&CK: Design and Philosophy; Technical Report; The MITRE Corporation. Available online: <https://www.mitre.org/sites/default/files/publications/pr-18-094-4-11-mitre-attack-design-and-philosophy.pdf> (accessed on 5 December 2021).
61. Wheelus, C.; Bou-Harb, E.; Zhu, X. Towards a big data architecture for facilitating cyber threat intelligence. In Proceedings of the 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Larnaca, Cyprus, 21–23 November 2016; pp. 1–5.
62. Mtsweni, J.; Mutemwa, M.; Mkhonto, N. Development of a cyber-threat intelligence-sharing model from big data sources. *J. Inf. Warf.* **2016**, *15*, 56–68.
63. Pennington, A.; Applebaum, A.; Nickels, K.; Schulz, T.; Strom, B.; Wunder, J. Getting Started With ATT&CK. Technical Report; The MITRE Corporation. Available online: <https://www.mitre.org/publications/technical-papers/getting-started-with-attack> (accessed on 5 December 2021).
64. Preuveneers, D.; Joosen, W. Sharing Machine Learning Models as Indicators of Compromise for Cyber Threat Intelligence. *J. Cybersecur. Priv.* **2021**, *1*, 8. [[CrossRef](#)]
65. Ring, T. Threat intelligence: Why people do not share. *Comput. Fraud. Secur.* **2014**, *2014*, 5–9. [[CrossRef](#)]
66. Wagner, T.D.; Mahbub, K.; Palomar, E.; Abdallah, A.E. Cyber threat intelligence sharing: Survey and research directions. *Comput. Secur.* **2019**, *87*, 101589. [[CrossRef](#)]
67. Peng, K.; Li, M.; Huang, H.; Wang, C.; Wan, S.; Choo, K.K.R. Security Challenges and Opportunities for Smart Contracts in Internet of Things: A Survey. *IEEE Internet Things J.* **2021**, *8*, 12004–12020. [[CrossRef](#)]
68. Elmaghraby, A.S.; Losavio, M.M. Cyber security challenges in Smart Cities: Safety, security and privacy. *J. Adv. Res.* **2014**, *5*, 491–497. [[CrossRef](#)] [[PubMed](#)]
69. Malhotra, P.; Singh, Y.; Anand, P.; Bangotra, D.K.; Singh, P.K.; Hong, W.C. Internet of Things: Evolution, Concerns and Security Challenges. *Sensors* **2021**, *21*, 1809. [[CrossRef](#)] [[PubMed](#)]
70. Shaukat, K.; Alam, T.M.; Hameed, I.A.; Khan, W.A.; Abbas, N.; Luo, S. A Review on Security Challenges in Internet of Things (IoT). In Proceedings of the 2021 26th International Conference on Automation and Computing (ICAC), Portsmouth, UK, 2–4 September 2021; pp. 1–6.
71. Latif, S.; Idrees, Z.; e Huma, Z.; Ahmad, J. Blockchain technology for the industrial Internet of Things: A comprehensive survey on security challenges, architectures, applications, and future research directions. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e4337. [[CrossRef](#)]
72. Sullivan, S.; Brighente, A.; Kumar, S.; Conti, M. 5G Security Challenges and Solutions: A Review by OSI Layers. *IEEE Access* **2021**, *9*, 116294–116314. [[CrossRef](#)]
73. Navamani, T. A Review on Cryptocurrencies Security. *J. Appl. Secur. Res.* **2021**, 1–21. [[CrossRef](#)]