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

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. Threat intelligence 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. For this reason, we must focus on the effective detection and sharing of behavioral IOCs to face advanced threats, as these indicators are harder for a hostile actor to modify.

In Cyber Threat Intelligence (CTI), an indicator of compromise is defined 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. The three main categories of Indicators of Compromise are as follows:

- 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 indicators are those which are derived from data involved in an incident. Examples of computed indicators include hash values and regular expressions.
- Behavioral indicators are collections of computed and atomic indicators, often subject to qualification by quantity and possibly combinatorial logic. Such indicators are captured as tactics, techniques and procedures, representing the modus operandi of the attacker

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.

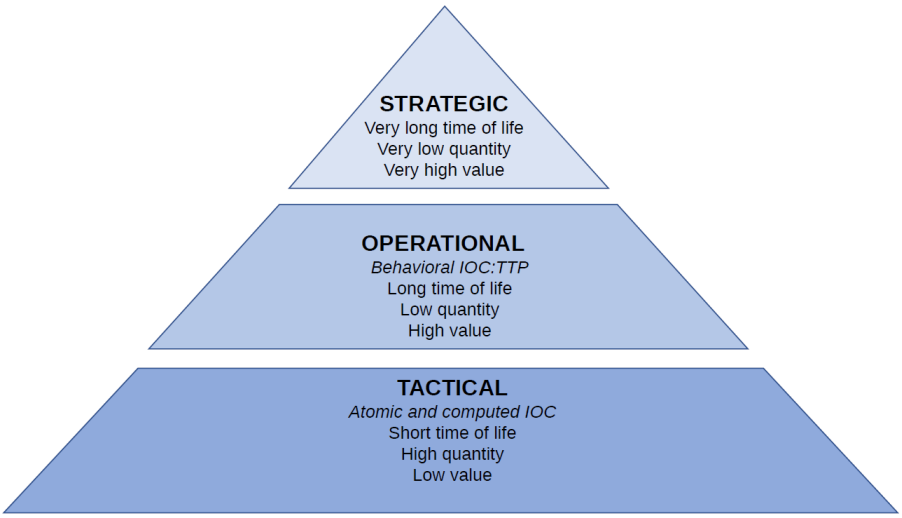

Figure 1. Indicators of compromise and intelligence levels.

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. Any hostile actor who wants to evade detection will defeat, at least, these three types of IOC, as they are the most used.

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.

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.

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.

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

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, 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

- 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.

Processing

When dealing with behavioral IOC 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.

Analysis

The first identified key requirement for the analysis is to be able to specify the behavioral IOC in a technology-agnostic way. 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.

Most techniques cannot be identified by analyzing events from a single data source; in fact, only about ten particular MITRE ATT&CK techniques (out of 185) can be detected using a single data source. Thus, the ability to establish relationships between events from different sources is a must.

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 reference. 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.

Dissemination

Threat intelligence sharing between defensive centers is a must: without 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.