Principles of Anomaly-Based Intrusion Detection and Threat Hunting

Numerous challenges of Anomaly-Based Intrusion Detection have been studied in recent years. Most of them are centered on the fact that it is hard to define a baseline to compare against, that this baseline is a moving target, deviations are hard to interpret, or that only limited data are available to train machine learning methods.

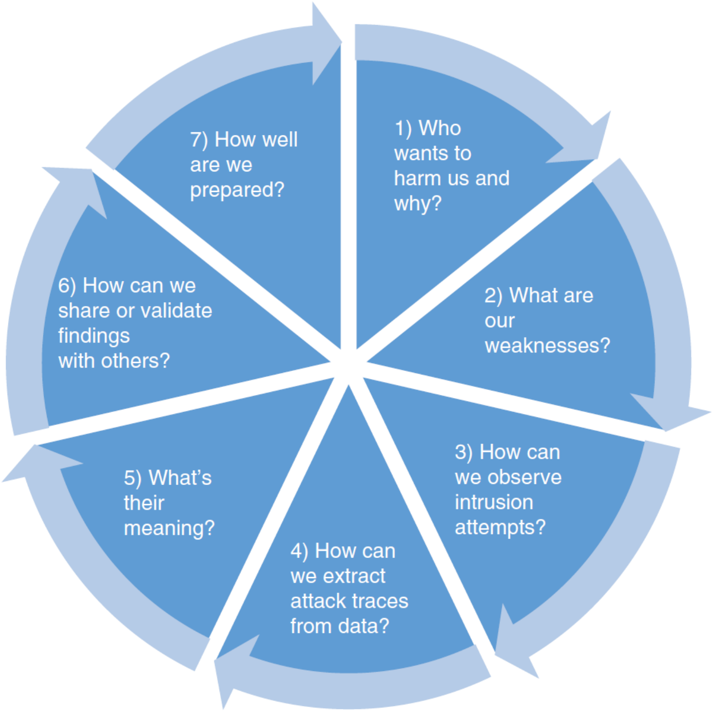
1) Who wants to harm us and why?
2) What are our weaknesses?
3) How can we observe intrusion attempts?
4) How can we extract attack traces from data?
5) What's their meaning?
6) How can we share or validate findings with others?
7) How well are we prepared?

Figure 2. Key questions of effective intrusion detection and threat hunting.

Principle 1: Get to Know Your Enemy

What is true for the real world also applies to the cyber domain. Knowing the motivation and capabilities of adversaries is key to an effective defense. So-called indicators of compromise (IoCs) were important to verify that a system has been penetrated; however, simple IoCs, including file hashes, process names, or certain memory patterns, are easy to circumvent by slightly adapting attacker tools or obfuscating attack techniques. The new hot topic is therefore modeling of more complex tactics, techniques, and procedures (TTPs), which represent the adversaries' modus operandi and are harder to change, and thus are a more sustainable means for detection.

Principle 2: Get to Know Thyself

Equally important is to know the environment, its vulnerabilities, and its weaknesses that constitute its attack surface. This is a mandatory prerequisite for deriving a baseline of good (and predictable) behavior. Unfortunately, it is also one of the most often neglected aspects of intrusion detection.

Principle 3: Be Open to All Sorts of Data

The truth is that in principle there is no right or wrong type of data. The feasibility of data to spot intrusions solely relies on the attack technique applied.

Principle 4: Analyze Smart

Intrusions do not serve up themselves on a silver plate. We need to extract the relevant data points and at the same time reduce the potentially massive amounts of data in a smart way to be able to handle them properly. Feature extraction is the science to do exactly that. For instance, long-tail analysis focuses on identifying rarely occurring events, as does outlier detection based on clustering approaches. Other analysis techniques are time-series approaches, such as the autoregressive integrated moving average (ARIMA) model, used to spot long-term deviations in trends, as well as frequency detection, detection of event correlations, or changes in the value distribution of data fields.

Principle 5: Making Sense Out of Data

Having spot ted an anomaly or, rather, a set of anomalies, the next step is to derive its possible root cause. First, we can derive some useful information simply from knowing the (type of) affected system, the type of anomaly reported, or the specific data source that emitted the anomalous data. Second, another way to gain insights is to make a lookup in historic data if the reported kind of anomaly was observed in the past.

Principle 6: Sharing Means Caring

The fine art to achieve is the sharing of knowledge about concrete attack tactics used by adversaries, combined with the immediate applicability of this knowledge in diverse environments, so-called actionable cyberthreat intelligence (CTI).

Principle 7: Learn and Prepare for the Next Wave

Learning from previous incidents, whether one’s own or those of others, and deriving effective countermeasures to avoid similar problems in the future is a key principle for increasing the own security posture.