Preventing and detecting

Jacques de la Riviere,
Gatewatcher

Abstract:
The heterogeneity of information systems, the migration of data toward the cloud and nomadism make it hard for firms to define their perimeter of protection. Added to this is the professionalization of cybercriminality with more numerous and variegated menaces. In this sinuous environment — a context where the analysts of centers of security operations and the experts in cloud security are scarce resources — it is logical to want to use technology to automate the tasks of detection, evaluation and response. Several solutions are arising, both innovative and complementary but each with its limitations: security orchestration, automation and response (SOAR), cyberthreat intelligence (CTI) and artificial intelligence (AI), specifically machine learning.

The heterogeneity of information systems, the migration of data toward the cloud and roaming users or devices all make it hard for firms to define their perimeter of defense. Meanwhile, cybercriminality is becoming ever more professional; and its menaces, ever more numerous and varied. In this sinuous environment — a context with a lack of analysts for security operations centers (SOCs) and of experts in cloud security — it is logical to try to use technology to automate the tasks of detection, evaluation and response. Solutions, both innovative and complementary, have been designed, each with its own limitations: security orchestration, automation and response (SOAR), cyberthreat intelligence (CTI) and artificial intelligence (AI, specifically machine learning).[1]

Security orchestration, automation and response (SOAR)

At a time when the volume of alerts and data is too big for human beings to handle, it is indispensable to orchestrate and automate the management of incidents on information systems.

SOAR (the acronym for security orchestration, automation and response) comprises the techniques for collecting security alerts and data from multiple sources and, especially, for helping to industrialize the analyzing and sorting of incidents. The idea is to combine the contributions of humans and computers in order to use a standardized process to design, rank and integrate the activities for responding to incidents.

Firms are interested in these tools because they want to try to improve the productivity of their security teams, at a time when budgets are too tight to hire personnel or when the scarcity of skills and qualifications stymies recruiting efforts. Although recourse to third-party service-providers might be helpful, another solution is to use orchestration and automation tools.

[1] This article has been translated from French by Noal Mellott (Omaha Beach, France).

Besides productivity gains, SOAR can help reduce the reaction time for responding to incidents (of confinement and remediation). It can also free analysts from routine — sometimes redundant and often time-consuming — tasks, which inevitably lower a SOC's morale. A good example is the task of initially sorting alerts, many of them false positives. SOAR automatically adds key contextual information to alerts so as to allow for automated sorting — or, at the very least, for speeding up and simplifying manual sorting. Optimizing the sorting process can help improve the system's detection capacity by lowering the risks of overlooking significant alerts. This also holds for all the work of documenting processes, of auditing and monitoring a SOC's performance, or for the initial training of analysts.

Thanks to multiplication of APIs (application programming interfaces), the automation and orchestration of cybersecurity is possible. Only a few tools used to propose APIs, and few of the latter were as standardized and simple as the REST APIs currently in use. The full automation of security processes, assuming that it is feasible, will take longer.

In the meanwhile, SOAR can bring many benefits in several fields, starting with the management of the alerts transferred from the security information and event management (SIEM) system. SOAR can automate tasks such as the search for additional information about the indicators that the information system has been compromised or the submission of suspected samples to third-party analysts. The tickets corresponding to alerts that constitute no real threat can be closed automatically.

Endpoint detection and response tools (EDR) can also be used in a transparent way to collect additional data about the affected host machines; and the search for correlations with network traffic can be launched. In concrete terms, this means handling e-mail messages that are suspected of phishing and obtaining more information about a SIEM-generated alert from a platform of information on cyberthreats. We might also mention the tools for managing the tickets generated by the system for tracking incidents or, more broadly, the ITSM (information technology service management), and, too, team communication tools like Slack or administration systems (like SCCM and SSH).

The principle of security automation (and not of AI) is definitely meant to replace the job of analysts with the work done by a machine. The machine collects the information, aggregates it thanks to a database of well-known suspected causes, and ultimately launches the remediation and communication protocols to users once, of course, the response to this type of alert has been identified.

Cyberthreat intelligence (CTI)

The phrase "cyberthreat intelligence" (CTI) emerged in early 2011, when the media widely reported the first advanced persistent threats (APT). Some organizations and state agencies have been using this approach for a long time.

CTI seeks to collect and organize all information related to cyberthreats so as to draw a portrait of the attacking forces and detect trends (methods and techniques for breaching systems, the infected sectors, etc.). It can help improve knowledge, mount a better defense and anticipate an attack (by detecting its forerunners). Various sorts of information can thus be collected: markers, indicators of compromise (IOCs, such as hash tags, domain names or IP addresses), the history of attacks, the reutilized architecture, and the use of specific services, techniques or methods (signatures, registrations). There are several ways to collect intelligence, among them: open source intelligence (OSINT), social media intelligence (SOCMINT), the data streams from e-commerce, information from the deep web (or darkweb), or even human intelligence (HUMINT), not to mention the capacity for analyzing and correlating the information from these sources of intelligence.

The means used for data-sharing tend to be harmonized as part of a global trend toward standards that can be reused as widely as possible. Several initiatives have been made to improve the communication of CTI information, in particular Open Source, STIX (structured threat information expression), MISP (malware information sharing platform) and TAXII (trusted automated exchange of indicator information). This transmission of information allows for generating rules of detection for supervision procedures (such as IPS tags).

There are genuine reasons why a firm or organization would want to install CTI. The starting point is to ask the right questions. Since not all data collected are significant, a framework for the program has to be set up upstream in the project. First, define the uses of CTI and the consequences of a breach of security for the firm. What should be protected? Which information might hackers want to obtain or destroy? Once these questions have been answered, set the objectives for installing CTI in the organization. These objectives should be clear, achievable and, too, easily measured thanks to predetermined indicators and criteria. Once the needs have been expressed and the objectives set, it is time to collect data by, for example, working the open sources accessible via the Internet (OSINT).

It is then necessary to pass to the phase of data processing so as to make it simpler for analysts to put the collected data to use. All the information collected (IOCs, malware, geopolitical context, methods used by group of hackers) has to be contextualized and enriched — to be turned into genuine intelligence. Thus starts the phase of analysis, which still very much depends on the expertise and skills of analysts. Their work will lead to a report that should be communicated to the right persons, at the right time and in the right format.

CTI could interest the following teams:

● at headquarters: top executives. The head of information systems and the chief information officer have to be provided intelligence about cyberthreats in their sectors of activity.

● risk managers and their teams, who have to have a global view of actual cyberthreats to jobs and processes in the organization.

● the security operation center (SOC), which needs intelligence to be operational and contextualized: the most recent cyberthreats in the form of IOCs, rules of detection and correlation, and even strategies for adapted responses/remediation (block such and such an IP, investigate certain gateways, block a vulnerable protocol, etc.).

● computer emergency response teams (CERT) and computer security incident response teams (CSIRT). Whether in the organization or outside, they need both strategic and operational intelligence (that is contextualized with information about the groups attacking the organization or its customers) in order to speed up investigations and responses to security incidents.

Artificial intelligence (AI): The future of detection?

Let us start by defining what is hidden underneath the phrase “artificial intelligence”. AI is a set of concepts, theories and techniques for solving highly complex logical or algorithmic problems. It combines several disciplines, among them: computational neurobiology, mathematical logic and computer science. All of this might seem complicated; but as with any technology, we should pay attention to needs and the proper uses of AI.

Let us focus on the potential of machine learning for security purposes and the detection of menaces. The differences between deep learning and machine learning have been the focus of many a discussion. Machine learning apparently provides a good solution to problems that traditional algorithms were unable to solve.

Machine learning helps us build a model of normality: what is called supervised AI. Algorithms describe a given situation and determine its distance from a normal distribution. If the difference is too great, there is an anomaly, which might generate an alert (a security incident) or signal a fraud or false positive. The advantage of machine learning comes from its handling of false positives. Once the machine has learned what is a “normal” case, this model of normality will be improved to avoid committing the same mistake twice.

AI, specifically machine learning, has several advantages.

AI can be used to update security databases. By analyzing logs from several sources, it can detect imminent threats. In other words, AI can collect exhaustive data from several logs and records, and make comparisons for identifying new threats of piracy.

With regard to malware and spyware, AI can identify trends by analyzing data from several channels. Thanks to a faster detection of new malware, damages are kept from becoming enormous. Engineers thus have more time to look for methods of prevention and to patch the security loopholes that malware or a virus might exploit.

Besides detecting large-scale malware transfers, AI can also analyze the information system to detect abnormal activities. Since the system is continuously monitored, sufficient data can be collected to draw conclusions about abnormal activities and build a model of normality. Furthermore, users can be monitored continuously to detect unauthorized accesses. If the system detects an anomaly, AI can use certain parameters to determine upstream in the process whether or not the threat is for real and should be reported.

Machine learning will enrich AI so that it more accurately distinguishes normal from abnormal activities. The more it is enriched, the more AI will be efficient and better able to detect slight anomalies that might signal a problem. As pointed out, what is important is to make the necessary comparisons. Some minor anomalies seem insignificant as such; but when taken together, they help us form a clear idea of the underlying causes. AI is capable of continuously monitoring a system, analyzing different activities, comparing them and sounding alerts. It seeks to identify potential points of vulnerability, bugs and security flaws. Machine learning, for example, can help detect when the data generated by an application are not reliable. SQL injection is among the vulnerabilities most frequently exploited by malware and viruses to steal data and breach systems. AI can also help detect another anomaly, namely buffer overflow resulting from the transfer of an unusually huge volume of data. AI can also be useful for limiting human errors. After all, mistakes by employees are a major cause of data protection violations. Here too, AI can help limit the damages.

In general, AI can determine an information system’s eventual points of vulnerability by monitoring current threats, in particular from malware. As it evolves, it will detect not just the defects of a system or of updates, but will also automatically keep the defects from being exploited.

AI is an excellent way to prevent problems, whether by reinforcing firewalls or correcting the coding errors that create points of vulnerability.

Although this might seem like prevention, it is a phase that takes place later, once malware has already been installed on the system. As mentioned, AI can be used to detect abnormal patterns and make correlations so as to determine the profile of a virus or malware. The appropriate response can then be adopted: control damages, eliminate the virus from the system, patch security flaws and set up supplementary protection to keep the virus from reinfecting the system.

Although AI carries several advantages for cybersecurity, it still has to make progress. The detection of anomalies, which blocks points of unauthorized access to an account or detects malware in the early stages of an attack, also generates false positives. AI can be improved to better detect truly abnormal activities. After all, a connection from a new location might simply mean that the user is roaming.

Cybersecurity companies and software editors will continue relying on machine learning to reduce detection time, increase detection rates, prevent malware from spreading, protect information systems and improve security for customers. Although AI still has to make headway, its impact is starting to be felt in cybersecurity.