Novel attack resilience by fusing events related to objectives

Mark Vella

mark.vella@um.edu.mt

Research in intrusion detection systems (IDS) is mainly restricted to the misuse and anomaly detection dichotomy, and therefore to their limitations. Web attack detectors are a case in point, where ones that perform misuse detection are prone to miss novel attacks, whilst those performing anomaly detection produce impractical amounts of daily false alerts. Detectors inspired from the workings of the human immune system (HIS) have proposed new effective detection approaches, however without tackling the issue of novel attack resilience separately from anomaly detection.

Danger Theory-inspired detection. This paper attempts to leverage inspiration from the Danger Theory (DT), which is a recent model of the HIS [1, 2]. Specifically, the focus is on the process that enables the HIS to respond to previously unseen infections without attacking host cells or symbiotic organisms. At a high level of abstraction, the process consists of the sensing of generic signs of an ongoing infection which triggers a process that identifies the make-up of the germ responsible, resulting in a response that is highly specific. Seen from a computational point of view, immune responses as suggested by DT follow a *generic-to-specific information fusion process.* Generic signals of infection that germs cannot avoid producing, enable the immune system to respond to novel infections. Their fusion identifies the make-up of the responsible germ, thereby preventing the production of anti-bodies that do harm. The aim of this work is to explore how this process can be used to develop effective detectors, meaning that they are resilient to novel attacks and suppress false positives. This is achieved by translating the process from the HIS to the web attack domain, and then subsequently realized as concrete detectors.

Approach. There are two types of generic signals of ongoing infection: Pathogenic Associated Molecular Patterns (PAMP) and danger signals, that originate externally and internally to the human body respectively. PAMPs constitute molecular patterns associated with evolutionary distant organisms. PAMPs are not specific to germs, but are rather associated with entire germ classes and so any germ is expected to feature such patterns. Danger signals on the other hand are molecules that originally form part of the internals of the body's own cells. Successful infections cause cells to explode, causing cell internals to be released and picked up by the HIS as danger signals. Like PAMPs, danger signals do not identify the germs causing them, and are expected to be provoked by any successful infection.

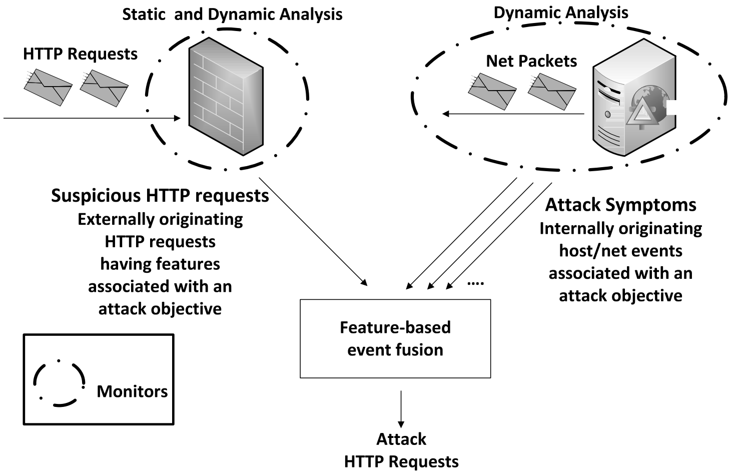
Static and Dynamic Analysis
HTTP Requests
Dynamic Analysis
Net Packets
Suspicious HTTP requests
Externally originating HTTP requests having features associated with an attack objective
Attack Symptoms
Internally originating host/net events associated with an attack objective
....
Feature-based event fusion
Monitors
Attack HTTP Requests

Fig. 1. An attack objective-centric translation of the generic-to-specific information fusion process

Figure 1 summarizes the proposed translation, with suspicious HTTP requests modeled on PAMPs and attack symptoms modeled on danger signals, both defined as attack objective-related events that attacks cannot avoid producing. The fusion process however does not follow the workings of the HIS since it was found to be counterproductive in earlier work [3]. Rather, it leverages feature-based correlation of events in order to identify attack HTTP requests. Since suspicious HTTP requests and attack symptoms are not exclusive to attack behavior, this component is required to distinguish those related to attacks. This can be achieved through feature similarity links that capture both the causal relation between suspects and symptoms as well as their association with an ongoing attack, in a similar manner to existing security event correlation systems.

Results and conclusions. The approach has been evaluated through three detectors that cover both high impact and popular web attacks within their scope. Effectiveness results are encouraging, with novel attack resilience demonstrated in terms of attacks aiming for a specific objective but modify the exploited vulnerability, payload or use obfuscation. False positive suppression is achieved until requests that do not contain the same attack content as per coincident and successful attack requests. However, their implementation is rendered difficult by requiring extensive knowledge of the deployment platform and secure coding. A performance study showed that the efficiency challenges tied with the stateful detection of events can be mitigated. These results merit a follow-up through further detector development and a formalization of the method.

References

1. Aickelin, U., Cayzer, S.: The danger theory and its application to artificial immune systems. CoRR abs/0801.3549 (2008)
2. Matzinger, P.: Friendly and dangerous signals: Is the tissue in control? Nat Immunol 8(1), 11–13 (2007)
3. Vella, M., Roper, M., Terzis, S.: Danger theory and intrusion detection: Possibilities and limitations of the analogy. In: Artificial Immune Systems, pp. 276–289. Lecture Notes in Computer Science, Springer Berlin / Heidelberg (2010)