A Case-based Reasoning Model for Intrusion Detection

N. Nobelis, K. Boudaoud, M. Riveill

Laboratoire I3S, University of Nice Sophia Antipolis, France

{nobelis, karima, riveill}@esi.fr

In improving security for the networks of today and tomorrow there are two important features that must be considered. The first is the dynamic of today’s enterprise networks, not just in terms of size, but in terms of changing services offered and the mobility of users that make them more vulnerable to various kinds of complex attacks. And the second is the continuous evolution of these attacks. Of particular importance, then, is the ability to rapidly apply new security policies in a network in order to detect and react as quickly as possible to attacks occurring in this network. In this context, an intrusion detection system (IDS) driven by security policies and based on a multi-agents system has been proposed. In this IDS, intelligent agents cooperate and communicate to efficiently detect attacks according to schema attacks registered in their knowledge base. The key characteristics of Intelligent Agents (IA) used in this system are delegation, cooperation and communication. However, an important property of IA, which is their ability of learning, hasn’t been used. This characteristic seems to us very useful to detect new attacks. Hence, we propose to extend this agent-based IDS by adding the learning functionality for new attack schemas.

Generally, in existing IDSs, the learning notion is used to learn normal behaviours of the system to secure. Thus, in a dedicated learning phase, normal profiles are built. These profiles are then compared with the current activity. In our case, we think that it will be more interesting to use this notion to learn abnormal behaviours that correspond to attacks. To learn a new attack, we need first to detect it then to update the attack schema base.

For the detection phase, it is necessary to select the appropriate detection approach. Data Mining (DM) seems today one of the most popular technique aiming to uncover significant non-intuitive correlations and patterns within a large data set. It is applied to both misuse and anomaly detection models. In the first one, each data record is classified and labeled as normal or abnormal activity, which permits to generate rules used to sort new data easily. In the second one, a data set is stored during a normal usage of the system to be compared to the current activity and detect potential attacks. In our case, we don't want to build a normal profile before the detection phase, this is why we didn't choose the DM. Another model that could fulfill our requirement is the Case-Based Reasoning (CBR) model. As defined by Aamodt&Plaza, “CBR is able to utilize the specific knowledge of previously experienced, concrete problem situation (cases). A new problem is solved by finding a similar past case, and reusing it in the new problem situation.” This definition is based on a problem/solution paradigm, which is slightly different from our approach. Actually, the key idea of our model is: when a suspicious event sequence is detected, the CBR engine retrieves in its case base, containing attacks, the case that matches as much as possible the suspicious sequence. This research is done by using a similarity function that weighs all the attributes of the reference cases to elect the best case with a certain similarity degree. If this degree satisfies a minimal given trust rate, the suspicious event sequence is stored as a new case in the CBR’s case base, the administrator is notified and the updating phase starts. Otherwise, if the matching fails, the agents could ignore the sequence or propose to the administrator to analyze it manually.

The updating phase of the attack schema base starts when a suspicious event sequence is identified as a new attack. At this step, it is required to generate automatically the attack schema related to the suspicious event sequence. To do so, the attack schema generator must identify which events characterize the attack, which attributes (for instance the source of the attack) may be omitted, if the time between two events is significant, etc. When the schema is created, and after a confirmation of the administrator, it is stored in the attack schema base. We consider that the generation of the attack schema is too critical a point to be achieved automatically. Thus, human intervention is necessary before updating the attack schema base.

In our approach, we have proposed a detection model that doesn't require to learn normal behaviors before the detection phase. We found in Case-Based Reasoning an appropriate technique to permit agents to discover similarity between past attacks and new ones. After having designed the learning (detection-updating) model, we plan to implement it soon.