Precognitive DLP Orchestration

Degree programme: MAS Data Science

Even without proper statistics, the collective memories of data-related incidents from the last decade show that the number of events is constantly rising. Regarding reasons & sources, reports and surveys not only point at external attackers or malware, but also hint at malicious insiders regularly. This master thesis work provides a thorough analysis of the malicious insider phenomenon and introduces a concept for automating DLP with UEBA, using artificial neural networks...

The Malicious Insider Problem

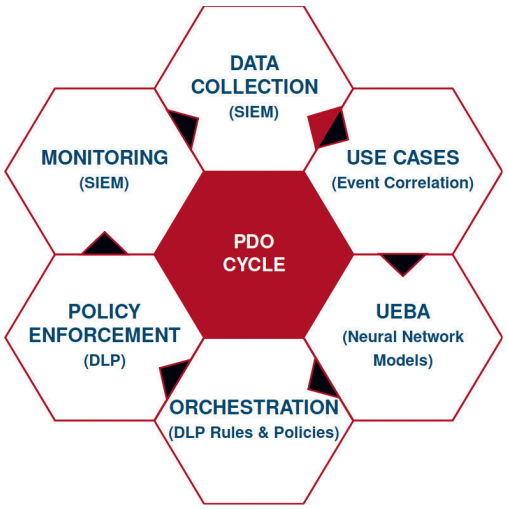
The urge to secure and protect one's own properties and assets from undesirable exposure is a basic desire in most civilizations and cultures. The latest by the shift from hunter-gatherers to farmers, we started engineering tools and concepts to protect our belongings against theft and destruction. And to this date, all our understanding of information security and data protection methods are adaptations of the fortress principle, where people inside the wall are considered friendly and those outside regarded as enemies. Even our vocabulary remained the same; we are using **firewalls** to protect computer systems and **proxies** to handle the traffic between internal and external networks; you have to know a **password** to get access to hosts & data, and need to be granted certain **privileges** to execute **administrator**-tier commands on a computer.

However, if we take globalization at society-level and distributed computing principles, like multi-site networks, mobile devices, the Internet or Cloud with vanishing perimeters into consideration, then it becomes hard to distinguishing between friendly- & malicious entities. The fortress principle is not only failing at providing protection against malicious insiders but also complicates the balancing act between key security properties (confidentiality, integrity, availability) on one hand and smooth business operations using laptops, smartphones, tablets, apps and other modern tools on the other. This is in particular true for protecting sensitive data from being leaked or breached. The state-of-the-art for preventing data thefts in corporate networks are so called **Data Loss Protection** (DLP) systems, which perform deep-content and -context analyses based on manually set rules and policies. The problem with DLP systems and any other rule-based strategy is that you are either missing hits, because of slack rules, or will risk too much false-positives with exaggeratedly strict rules. Both situations are of course undesirable – but how would it be possible to protect valuable information assets from malicious insiders and keep hindrance-free workflows for the others? And how to deal with relevant incidents, false-negatives or false-positives?

Proposed Solution

One answer is **data-driven security automation**, using data mining, machine learning and similar Data Science techniques to extend and automate methods for the mentioned key security properties. This master thesis work introduces a concept for automating Data Loss Protection using artificial neural networks for user- & entity behavior analysis (**UEBA**) with malicious insiders as the main threat source. The concept is called **Precognitive DLP Orchestration** (PDO). The PDO makes use of **SIEM** log aggregations to create continuous risk ratings for users and entities. These risk ratings again are being used for dynamic adaptations of DLP context rules and policies – see illustration for the full PDO cycle. The concept proposes an autonomous system which could significantly increase protection against malicious insiders and reduce costs & resources at the same time.

Engin Akdeniz
engin@enginakdeniz.com



Suggested cycle for the proposed Precognitive DLP Orchestration concept