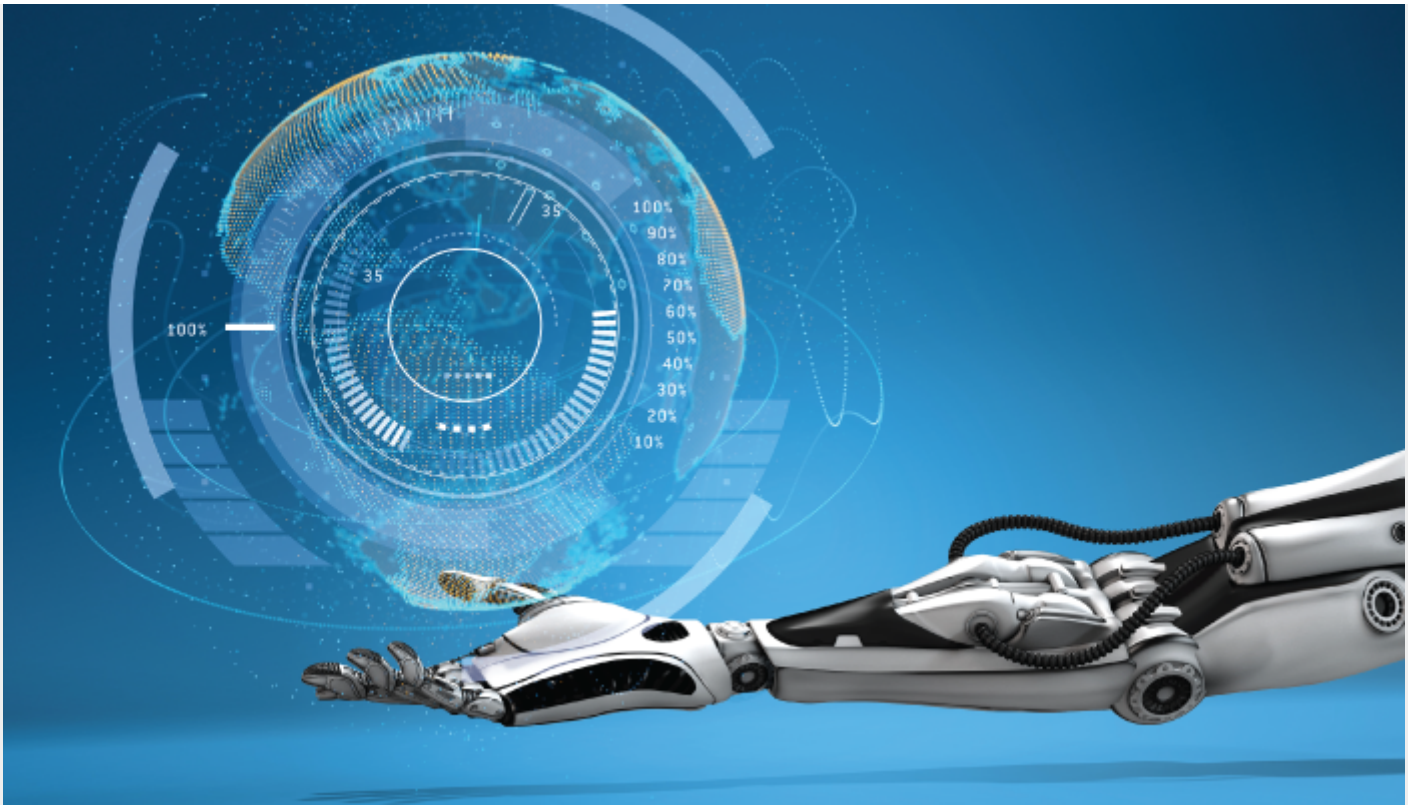
THE ROLE OF MACHINE LEARNING IN DATACENTER NETWORK SECURITY



Modern datacenters store large volumes of data that is critical to reliable online services such as financial transactions, healthcare services, national security and defense, as well as our online activities and communications. This data must be available on demand and accessible only by the authorized entities. Vendors offering datacenter services must therefore fight ongoing threats that risk the security, integrity and availability of sensitive data. As the threat evolves exponentially and cybercriminals leverage sophisticated tools to compromise datacenter network security, vendors rely on technologies that could outpace any human's capability to identify and proactively respond to network security threats. At the heart of these technologies is Machine Learning that automates the mechanism involved in defending against cyber-attacks, protecting sensitive business information of customers and streamlining end-user experience.

Machine Learning is a subset of Artificial Intelligence (AI) and refers to the study of algorithms and mathematical models that can be used in applications involving compute processing. The mathematical models derive the behavior of a system by learning the behavior pattern of output in response to input data. Using these models, computers can predict the response behavior of a system based on the parameters associated with it. For instance, if a datacenter network is compromised and the data traffic is routed to unauthorized channels, machine learning can identify the traffic behavior as anomalous and trigger appropriate security actions. This is effectively the primary role of machine learning in the security of modern datacenter networks.

The prominent use cases in this role include:

Intelligent Intrusion Detection

Traditional intrusion detection systems were designed to identify attempts to access the network based on known threats. However, modern datacenter infrastructure face a range of new and variable threats in attempts to deceive intrusion detection systems that lack intelligence. While cyber-attackers continue to evolve the technologies and practices used in intruding a datacenter network, implementing a security system that covers all present and future threats may be virtually impossible. To address this issue, machine learning based intrusion detection systems evaluate each attempt to access the network against a range of predefined metrics. These metrics can be static and dynamic – the machine learning system doesn't rely on a fixed set of static policies but incorporates the dynamic behavior of the network. As a result, the intrusion detection system accurately classifies security threats against legitimate intrusion attempts as the network behavior, traffic flows and other parameters change. The machine learning system learns from the false negatives and false positive alerts to continuously improve its accuracy in detecting unauthorized attempts to access the network in the future.

Security Forensics and Root Cause Analysis

Cloud datacenters generate a deluge of log big data that contain unprecedented information about network incidents that could potentially lead to downtime, security breach or performance issues. The vast information requires technology capable of evaluating the nature of each incident, classifying the alerts and triggering appropriate actions based on accurate results. Machine learning aids the forensic analysis of datacenter infrastructure and network logs and the data traffic. Machine learning models can be applied to collect the right piece of information faster than humans and target its search in close proximity of the root cause. Machine learning algorithms can be trained to improve their outcome of data extraction and facilitate accurate analysis accordingly.

Considering the complex architecture of large-scale IT infrastructure, service degradation can escalate exponentially across the network when the root cause is not addressed proactively. Machine learning tools facilitate classification, visualization and analysis of the datasets to glean insightful information more intuitively. A manual approach to perform root cause analysis of large-scale datacenters may require expertise and skills in data analysis acquired over decades of experience in the industry.

Response Automation for Proactive Security

Identifying a security issue constitutes one part of the risk mitigation process. The other key element is related to the response to potential attacks and the ability to ensure business continuity during possible security disasters. To achieve these goals, the systems should be capable of accurately identifying the root cause and performing appropriate risk mitigation actions without causing service disruption. This means that the security system must evaluate real-time performance of the datacenter and evaluate a range of parameters in reducing the impact of risk mitigation actions to end-users. If certain datacenter resources have to be isolated, the IT workloads should be transferred to the available infrastructure instances. This transfer should also balance the load on the available infrastructure to limit performance degradation for other users. With machine learning capabilities, the algorithms can be developed and tuned to align security response actions with the

organizational goals of service availability, performance and security as promised to end-users as part of their SLA agreements.

Preventive Maintenance and Security Upgrades

Cloud datacenter companies are well positioned to fight against cybercrime with their ability to invest in the latest and greatest datacenter resources. The maintenance and upgrades often incur downtime and costly business opportunities for both the cloud vendor and customers. Failure to upgrade the systems on time can open the network to zero-day or unpatched security attacks. With machine learning technologies, organizations can evaluate the performance of their infrastructure and security resources against ongoing threats, understand which systems are underperforming and need upgrades or replacement.

Machine learning allows datacenter companies to monitor the equipment for potential failures, schedule maintenance only when its needed and prolong the life of its hardware. A well-functioning hardware that doesn't collapse upon extensive usage in enterprise-grade infrastructure environment is less likely to cause networking issues that may leave security loopholes in the system.

Effectively, machine learning enables an automated and autonomous security system that is designed to incorporate the needs of end-users while addressing security challenges. Especially when cloud vendors target mass market with large volumes of storage and sever resources, the security challenges grow exponentially complicated for traditional or manual security systems. Machine learning fills in this gap and allows organizations to maximize the value potential of their datacenter investments.