Exploring statistical techniques for threat detection and AI security

Raffaele Castagna

Statistics Academic Year 2025-2026

Contents

1 Introduction

Cybersecurity has always been shaped by data analysis, ever since its inception. Even at the start when the *fortress model* was the optimal choice, sysadmins typically relied on the binary classification of traffic, files and behaviours that were either safe or were known malicious patterns, but to actually analyze a file most relied on static analysis of signatures, where the hash of a file was computed and then denoted as malicious, this could also be applied to byte sequences in network traffic or to generalize even more IP ranges could be blocked, either based on suspicious activity or on countries. Under the hood there was a statistical assumption: the distribution of malicious behaviour was disjointed from the distribution of normal behaviour, and therefore there was a boundary between the two that was immutable.

However with the continued expansion of the internet, and the users therein, this assumption became obsolete, the world has changed and with it the amount of data that is ingested, take for example Industrial Control systems or autonomous vehicles navigating public roads[2],in these kinds of environments the amount of data and heterogenity of the devices make signature based maintenance an impossible task. Expanding on this, the convergence of IT and OT (Operational Technology) has exposed critical industrial machine to a wide area of attack, and therefore shifted what can be considered normal behaviour, as well as the attack surface.[4]

With these new surfaces, the entire landscape changed, particularly with the rise of APT (Advanced Persistent Threats), which can even be state-sponsored threats, that are particularly interested in IO, these APTs usually utilize a technique called ”living off the land” which utilizes legitimate software and functions to carry out their attacks, e.g. utilizing group policy changes to gain persistance and access to restricted controls and information, another type of vulnerability that is continuedly exploited are Zero-Day attacks, which due to their nature are impossible to defend against with static hashing, so there's a need to move to a dynamic analysis of behaviour, which is where statistics come into play, so that we see a threat as something that *deviates from a probabilist baseline*, this is what in the modern cyberseucurity field is called **statistical anomaly detection**[1].

Anomaly detection is rooted in the statistical hypothesis that malicious activity is rare and different from normal activity. It does not ask, ”Is this specific packet known to be bad?” but rather, ”What is the probability that this packet belongs to the distribution of normal traffic observed over the last month?” This shift requires security professionals to abandon binary certainty in favor of probabilistic reasoning. It demands a rigorous understanding of distributions, variance, and correlation. The security analyst of the future must be as fluent in covariance matrices and p-values as they are in firewalls and encryption protocols.

1.1 The Dual Role of Artificial Intelligence: Tool and Target

As statistical methods evolved, they gave rise to Machine Learning (ML) and Artificial Intelligence (AI), which are essentially statistical inference engines operating at scale. AI has become a powerful tool for defense, capable of ingesting terabytes of log data to identify subtle correlations that elude human analysts[7].AI-driven statistical anomaly detection has been shown to outperform traditional analytical techniques in mitigating cybersecurity risks in complex environments like wireless networks[5].

However, this reliance on AI introduces a recursive vulnerability: the statistical models themselves are now targets. We are witnessing the rise of Adversarial Machine Learning (AML), where attackers exploit the probabilistic nature of AI. By carefully crafting inputs—adversarial examples—that are statistically indistinguishable from benign data to a human but catastrophic to a model's mathematical logic, attackers can blind surveillance systems or cause autonomous vehicles to misinterpret stop signs[3][6].

Thus, the domain of AI security is not merely about securing the software code but about securing the statistical inference process. It involves using advanced statistical tests—such as Kernel Density Estimation (KDE), Maximum Mean Discrepancy (MMD), and Benford's Law—to detect when a distribution is being manipulated. The defender must now monitor the monitor, applying statistical rigor to ensure that the AI systems protecting the network are not themselves compromised by statistical illusions.

This report provides an exhaustive analysis of the statistical techniques underpinning modern threat detection and AI security. It explores the mathematical foundations of anomaly detection in high-dimensional spaces, the time-series analysis of network traffic, the probabilistic modeling of attack paths using Bayesian networks, and the forensic statistics required to detect adversarial manipulation.

References

[1] VARUN CHANDOLA. “Anomaly Detection: A Survey”. In: (2009).

[2] Meera Sridhar Danial Abshari. “A Survey of Anomaly Detection in Cyber-Physical Systems”. In: *ArXiv* 1.1 (18 Feb 2025).

[3] Nicolas Papernot‡ Michael Backes† Patrick McDaniel‡ Kathrin Grosse† Praveen Manoharan†. “On the Statistical Detection of Adversarial Examples”. In: *CISPA, Saarland Informatics Campus†; Penn State University‡; MPI SWS* (2024).

[4] Karel Kuchar and Radek Fujdiak. “Anomaly Detection in Industrial Networks: Current State, Classification, and Key Challenges”. In: *IEEE SENSORS JOURNAL* 25.3 (2025).

[5] Ali Mohanad Faris Mohammed Q. Mohammed Mohammed G. S. Al-Safi. *Statistical Anomaly Detection for Enhanced Cybersecurity Using AI-Based Wireless Networks.* Tech. rep. Ingénierie des Systèmes d’Information, 2024.

[6] Palo Alto Networks. “What Is an Adversarial AI Attack?” In: (2024). https://www.paloaltonetworks.com/cybe are-adversarial-attacks-on-AI-Machine-Learning.

[7] SITHAMPARANATHAN KANDEEPAN-AKRAM AL-HOURANI KARINA GOMEZ CHAVEZ SONG WANG JUAN FERNANDO BALAREZO and BENJAMIN RUBINSTEIN. *Machine Learning in Network Anomaly Detection: A Survey.* Tech. rep. IEEE, 2021.