CYBER ANALYTICS:
FROM DATA TO DEFENSE

FROM PERIMETER TO INTELLIGENCE-LED SECURITY

DATA INGESTION & NORMALIZATION

Collecting & standardizing logs from Cloud, Network, Endpoints & Applications

ANALYTICS & MACHINE LEARNING

Detecting Anomalies & Patterns

UEBA BASELINING

Learning Normal Behavior Profiles

THREAT INTELLIGENCE ENRICHMENT

Concextualizing Alerts with Global Threat Data

LATERAL MOVEMENT TRACKING

Tracing the Attacker's Path Through the Network

LOW & SLOW DETECTION

Finding Stealthy Attacks

RISK-BASED PRIORITIZATION

Focusing on Critical Threats

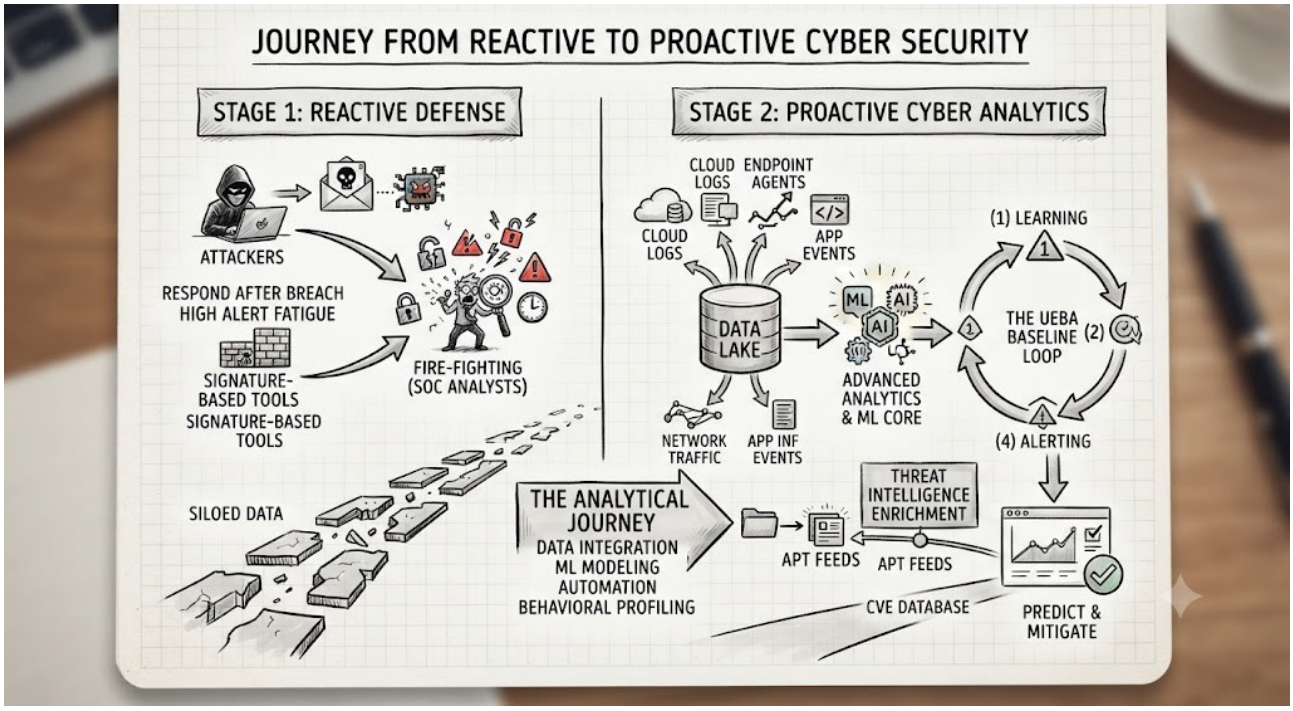
WHITE PAPER 8

JAYANTA CHAKRABORTY
CEO, ASPIRATION PUBLISHING TRUST

White Paper on Cyber Analytics

The Architecture of Vigilance: Expanding Cyber Analytics

Cybersecurity is undergoing a paradigm shift. In an era where attackers deploy automated, polymorphic threats and exploit zero-day vulnerabilities, static perimeter defenses—like traditional firewalls and signature-based antivirus—are increasingly insufficient. Cyber Analytics represents the evolution of security from a "gatekeeper" model to an "intelligence-led" model. It transforms raw, chaotic machine data into actionable, high-fidelity security intelligence.



1. The Deep Workflow of Cyber Analytics

The efficacy of cyber analytics is predicated on the quality of its data pipeline. This is not merely a collection exercise; it is an engineering challenge requiring high-throughput data processing and sophisticated correlation.

Pillar 1: Data Collection & Telemetry (The Foundation)

Data must be gathered from every corner of the IT estate. This includes:

- **Network Layer:** NetFlow, DNS queries, and full packet captures (PCAP).
- **Endpoint Layer:** EDR logs, process execution history, and PowerShell command-line arguments.
- **Identity Layer:** Authentication logs (Active Directory, Okta, OAuth tokens).
- **Cloud Infrastructure:** Control plane logs (AWS CloudTrail, Azure Monitor) and serverless function telemetry.
- **Application Layer:** API call logs and database transaction logs.

Pillar 2: Normalization and Contextualization

Data arriving from these sources is inherently messy and inconsistent. Normalization involves mapping disparate fields into a common schema (like the Open Cybersecurity Schema Framework - OCSF). Contextualization adds "business logic"—identifying that an IP address belongs to a critical database server or that a specific user is a domain administrator. Without this, alerts are just data points; with it, they become security events.

Pillar 3: Advanced Analysis & Machine Learning

This is the "brain" of the operation. It typically employs:

- **Supervised Learning:** Training models on known malicious patterns (e.g., classifying phishing emails).
- **Unsupervised Learning:** Clustering algorithms (like K-Means) that group "normal" user behavior and flag outliers (e.g., a developer suddenly accessing HR records at 3:00 AM).
- **Graph Analytics:** Analyzing relationships between entities to find lateral movement paths that individual logs would miss.

---

2. Strategic Advantages: Beyond Reactive Defence

The "Low and Slow" Detection

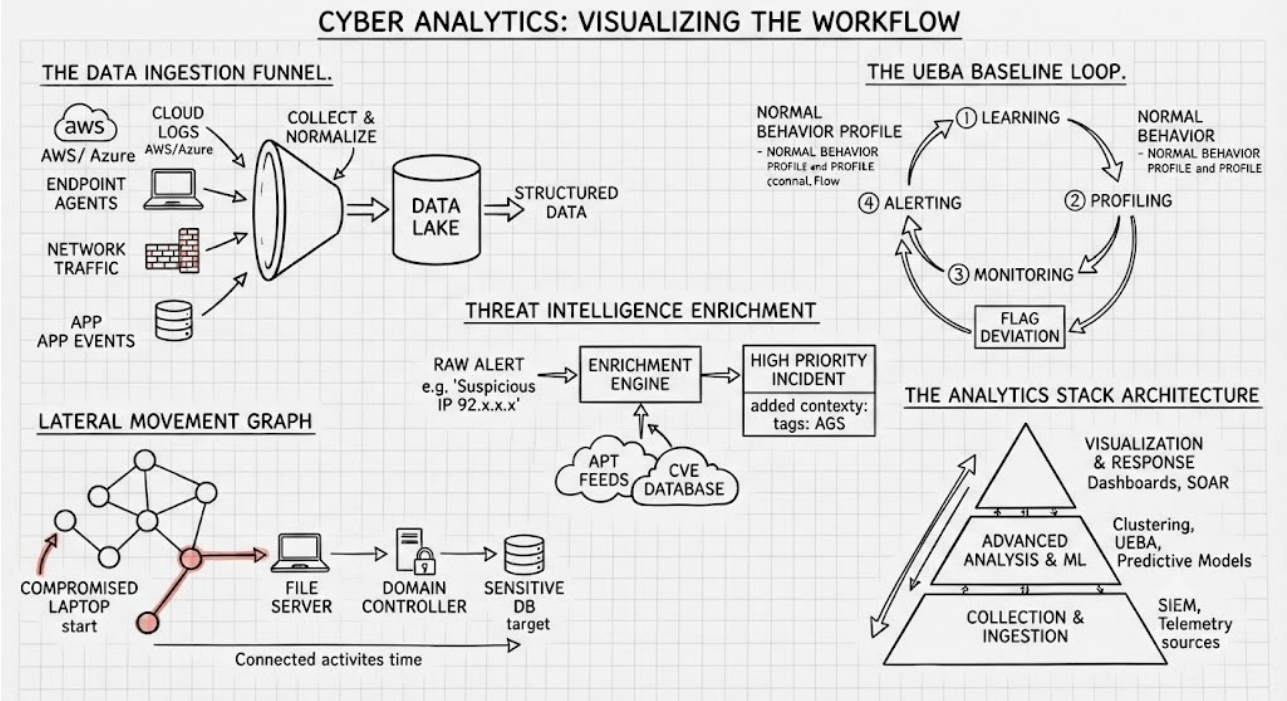
Sophisticated adversaries often employ "living-off-the-land" (LotL) techniques, using built-in administrative tools to avoid triggering signature-based alarms. Behavioral analytics detects these by identifying deviations from a user’s historical baseline, such as an unusual spike in file read operations or unconventional cross-server remote execution.

Quantifying Risk & Prioritization

Security Operations Centers (SOCs) are often overwhelmed by thousands of alerts daily. Cyber analytics provides **Risk Scoring**. Instead of treating every failed login as an alert, the system aggregates signals: *[Failed Login] + [New Process Started] + [Unexpected Outbound Connection]*. The analytics platform correlates these into a single, high-severity incident, drastically reducing "Alert Fatigue."

---

3. Diagrams: Visualizing the Security Ecosystem



To understand how these components work together in a modern security operations center, we can view these five sketches as a sequential, integrated pipeline. Here is the step-by-step breakdown of how data travels from raw logs to actionable incident response.

Step 1: The Data Ingestion Funnel

The Objective: Establishing Visibility

Before you can analyze data, you must gather it. The Ingestion Funnel acts as the entry point for the entire system.

- **The Process:** Raw logs are pulled from disparate sources like Cloud platforms (AWS/Azure), physical Endpoints, Network traffic (firewalls/routers), and Application events.
- **The Key Transformation:** These raw logs are messy and use different formats. The "Normalization Filter" acts as a translator, converting all these varied inputs into a single, standardized format that the system can understand, ultimately storing them as **Structured Security Data** in a Data Lake.
-

Step 2: The Analytics Stack Architecture

The Objective: Building the Foundation

While Sketch 1 shows the flow, Sketch 5 explains the structural hierarchy of the tools required to process that data.

- **The Bottom (Data Collection):** This is the ingestion layer from Sketch 1.
- **The Middle (Processing & ML):** This is the intelligence layer where statistical models, Machine Learning (ML), and User and Entity Behavior Analytics (UEBA) run. It processes the structured data to find patterns.
- **The Top (Visualization & Response):** The final layer that transforms technical findings into human-readable dashboards and automated security orchestration (SOAR) playbooks.

Step 3: The UEBA Baseline Loop

The Objective: Defining "Normal"

Once the stack is ready, you need a baseline to know what "abnormal" looks like.

1. **The Process:** The system enters a perpetual cycle:
2. **Learning:** It watches user and device behavior over time.
3. **Profiling:** It creates a "Normal Behavior Profile" for every user.
4. **Monitoring:** It compares real-time actions against that profile.
5. **Alerting:** If someone acts outside their profile (e.g., a standard employee suddenly running high-level administrative scripts), the system flags the deviation.

Step 4: The Threat Intelligence Enrichment Path

The Objective: Providing Context

Not every deviation is a threat; you need external context to know what is dangerous.

- **The Process:** You take a raw alert (e.g., a connection to an unknown IP address). On its own, it's just a data point. The system passes this through an "Enrichment Engine" that cross-references the IP against external Threat Intelligence feeds (such as lists of known Command & Control servers or CVE databases).
- **The Outcome:** The engine adds context to the alert, transforming a low-level "unknown connection" into a "High Priority Incident."
-

Step 5: The Lateral Movement Graph

The Objective: Mapping the Attack Chain

Finally, when a real attack occurs, it rarely happens in one place. Advanced attackers move laterally.

- **The Process:** Using graph analytics, the system maps the relationship between different nodes (devices/servers) in your network.
- **The Visualization:** It visualizes the "chain" of an attack: an attacker starts at a compromised laptop, jumps to a file server, moves to a domain controller, and finally reaches a sensitive database. By showing the *connections* between these steps, security teams can see the full path of the breach rather than just isolated, disconnected events.

Summary Flow:

You **collect** the data , **process** it through your analytical **stack** , establish a **baseline** of behavior , **enrich** any suspicious findings with global threat intelligence , and finally, **visualize** the attacker's path through the network .

Conclusion: The Surveillance Mandate

Cyber analytics is the transition from "knowing what we are told" (threat signatures) to "knowing what is happening" (behavioral reality). By treating the IT environment as a living system, organizations can detect unauthorized behavior in real-time, regardless of the tools or techniques the adversary uses.

Ultimately, cyber analytics does not just secure the environment; it provides the visibility required to govern it. In an increasingly complex digital world, this surveillance capability is not just an advantage—it is the baseline requirement for operational survival.