Unauthorized Intrusion Detection: Essential Data Requirements, Security Triggers, and Prevention

İsmail Şimşek

Data Analytics and Risk Management

Professor Mario Spremić

December 9, 2025

Abstract

Detecting unauthorized intrusions, like malware attacks, requires gathering and analyzing extensive datasets from various sources. These sources include host systems, network infrastructure, identity management platforms, and cloud environments. This paper explores the key datasets needed for intrusion detection. It identifies their sources and access methods. It also proposes security triggers for early threat identification. Additionally, the paper presents practical detection methodologies and prevention strategies aligned with modern cybersecurity frameworks and data analytics practices.

Data Requirements for Unauthorized Intrusion Detection

For intrusion detection to succeed, integrating multiple data sources is crucial. These sources offer different views of system activity. When correlated properly, these datasets help security professionals identify malicious behavior and unusual patterns that suggest compromise.

Host and System Data

Telemetry from hosts provides critical insights into endpoint activity and system state. Important datasets include system event logs, process execution records, outputs from file integrity monitoring, registry modification logs, and telemetry from endpoint detection and response (EDR) systems. These sources reveal suspicious behaviors such as unusual process spawning, unauthorized modifications to protected system files, and deployment of persistence mechanisms. Organizations

typically obtain this information through security agents like Sysmon, OSQuery, or commercial EDR platforms installed at the operating system level.

Network Traffic Data

Network telemetry offers visibility into communication patterns and traffic flows within organizational infrastructure. Essential datasets include NetFlow and IPFIX logs, packet capture (PCAP) data, Domain Name System (DNS) query logs, and firewall traffic records. Analyzing these datasets helps detect command and control communications, lateral movement activities, port scanning operations, and data exfiltration attempts. Network infrastructure components like firewalls, routers, intrusion detection systems (IDS), and dedicated network sensors capture and forward this information to centralized analysis platforms.

Identity and Access Management Data

Compromised authentication systems are a frequent vector for unauthorized intrusions. Critical datasets include Active Directory logs, multi factor authentication (MFA) event records, and user behavior analytics (UBA) data. These sources enable detection of suspicious authentication attempts, privilege escalation activities, and unauthorized creation of privileged accounts. Organizations aggregate this information from domain controllers, cloud identity platforms such as Microsoft Entra, and single sign on (SSO) providers.

Cloud Infrastructure and Application Data

Modern organizational environments generate substantial volumes of cloud audit logs. Representative examples include AWS CloudTrail, Azure Activity Logs, and Google Cloud Platform (GCP) Audit Logs. These datasets reveal unauthorized API key generation, unexpected permission modifications, and alterations to security controls. Access to cloud telemetry is established through log export mechanisms configured via cloud native monitoring services.

Recommended Security Triggers

Security triggers are predefined alert conditions that notify security personnel when observed behavior deviates from established baselines or expected patterns.

Triggers at the Host Level

Execution of unsigned or previously unknown executable files

Process initiation from atypical directories or paths

Registry modifications or configuration changes affecting persistence mechanisms

Memory injection techniques or anomalous behavior between processes

Triggers at the Network Level

Outbound connections to known malicious IP addresses or domains

Multiple failed authentication attempts across network services

Lateral movement indicators including Server Message Block (SMB) brute force attempts

Periodic beaconing patterns to external hosts

Triggers for Identity and Access

Impossible travel scenarios indicating compromised credentials

Unexpected privilege escalation outside authorized change windows

Creation of administrative accounts during non maintenance periods

Authentication patterns showing multiple failures followed by successful login

Triggers for Cloud Infrastructure

Disabling of logging or monitoring capabilities

Creation of API keys without corresponding approved change requests

Modifications to Identity and Access Management (IAM) roles granting elevated privileges

Dataset Access and Utilization

Organizations usually consolidate these diverse datasets within Security Information and Event Management (SIEM) platforms such as Splunk, Elastic Stack, IBM QRadar, or Microsoft Sentinel. Endpoint agents forward host level logs, network infrastructure components export flow records and traffic logs, and cloud service providers offer connectors or APIs enabling automated log ingestion. After collection, datasets undergo correlation using rule based detection systems, statistical anomaly detection algorithms, or machine learning models. This correlation enables identification of abnormal behaviors that individual datasets cannot reveal independently. Integration of threat intelligence feeds provides additional context through identification of known malicious indicators including IP addresses, domains, and file hashes.

Detection and Prevention Methodologies

Effective intrusion management requires implementing multiple complementary approaches that address both detection and prevention.

Detection Approaches

Detection Based on Signatures: This method identifies known malware patterns by comparing against databases of previously identified threats. While effective against documented threats, it cannot detect novel or modified attacks.

Analysis of Behavior: Examining abnormal process execution patterns and network activity enables detection of threats lacking known signatures. This approach identifies deviations from typical operational behavior.

Detection of Anomalies: Statistical and machine learning techniques identify deviations from established baseline behaviors for users, systems, and networks. This method excels at detecting insider threats and advanced persistent threats.

Integration of Threat Intelligence: External reputation data and indicators of compromise enhance alert context and enable rapid identification of known malicious infrastructure.

Prevention Strategies

Protection at Endpoints: Deploying EDR solutions with prevention capabilities provides real time blocking of malicious activities at the host level.

Hardening Authentication: Mandatory implementation of multi factor authentication significantly reduces the risk of compromise based on credentials.

Network Architecture: Application of network segmentation and least privilege principles limits the potential impact of successful intrusions.

Management of Vulnerabilities: Continuous patch management and system updates reduce the attack surface available to threat actors.

Filtering Egress Traffic: Blocking outbound communications to unknown or suspicious destinations disrupts command and control channels and prevents data exfiltration.

Management of Cloud Security Posture: Monitoring and restricting administrative actions within cloud environments prevents privilege abuse and unauthorized configuration changes.

Conclusion

Detecting unauthorized intrusions demands comprehensive integration of diverse datasets from endpoint systems, network infrastructure, identity management platforms, and cloud environments. Implementing appropriate security triggers enables early identification of malicious activity, while advanced analytics techniques facilitate recognition of abnormal patterns indicative of system compromise. Organizations that effectively consolidate these datasets within integrated SIEM and EDR ecosystems achieve superior intrusion detection capabilities and can implement proactive preventive measures to reduce organizational risk. A security posture that is driven by data and proactive remains fundamental to effective cybersecurity operations in contemporary threat environments.