A Research Proposal: Using Empirical MTTD Benchmarking for Cloud Security Services to Go Beyond Vendor Claims

Arnav Khinvasara
University of California, San Diego
akhinvasara@ucsd.edu

April 17, 2025

Abstract

Organizations face challenges in objectively evaluating and comparing the threat detection capabilities of commercial cloud security services. This paper proposes a standardized methodology and open-source framework for measuring Mean Time to Detect (MTTD) across different cloud security offerings. Our approach uses realistic attack simulations based on the MITRE ATT&CK framework to provide quantitative metrics for direct comparison, enabling evidence-based security decisions. We discuss the limitations of current evaluation methods, the novelty of our approach, its potential security applications, and the importance of establishing MTTD benchmarks in cloud security.

1 Introduction

As organizations increasingly migrate their critical infrastructure to cloud environments, the selection of effective security services becomes paramount. A key metric for evaluating these services is Mean Time to Detect (MTTD), which measures how quickly a security service identifies a threat. A shorter MTTD allows for faster response and mitigation, reducing the potential impact of security incidents.

Currently, organizations lack objective methods to evaluate and compare detection capabilities across different security services. They often rely on vendor claims, qualitative assessments, or ad-hoc testing, making informed security decisions difficult. This paper proposes a standardized methodology and framework to address this gap, providing a repeatable, platform-agnostic approach to security service evaluation.

2 Problem Statement

Today, organizations rely on vendor claims, qualitative assessments in analyst reports, or ad-hoc testing that lacks methodological consistency [2]. Industry reports from Gartner and Forrester evaluate cloud security services but rarely include quantitative MTTD metrics based on standardized testing [3]. When MTTD is reported, methodologies vary significantly, making direct comparisons problematic.

Existing security testing tools like Atomic Red Team and Caldera implement MITRE ATT&CK techniques but focus primarily on on-premises environments and don't provide standardized metrics for comparison [4]. Cloud-specific attack simulation is less developed, with most approaches using proprietary methodologies that lack transparency [6].

The few security benchmarking efforts, such as those by NIST and the Center for Internet Security (CIS), focus on configuration compliance rather than detection performance [1, 5]. This creates several challenges for organizations:

- Difficulty in making evidence-based security service selections
- Inability to accurately assess security investments
- Limited understanding of detection blind spots and gaps
- Lack of industry benchmarks for reasonable detection times

3 Proposed Solution

Our approach is novel in providing:

- A standardized, platform-agnostic methodology for measuring MTTD across cloud providers
- Realistic attack simulations based on the MITRE ATT&CK framework that generate consistent indicators
- Objective metrics (MTTD, detection rates, false positive rates) for direct comparison
- An open-source framework that enables reproducible benchmarking

We've designed a modular architecture with six core components:

- **Threat Simulation Engine**: Executes controlled attack scenarios across platforms.
- **Detection Monitoring System**: Collects security events from various services.
- **Metric Collection & Analysis**: Processes event data to calculate MTTD and other metrics.
- **Reporting & Visualization**: Generates comparative reports and visualizations.
- **Test Scenario Manager**: Orchestrates test execution and manages environments.
- **Service Integration Layer**: Provides standardized interfaces to various security services.

We are developing working implementations for AWS GuardDuty, AWS Security Hub, Azure Sentinel, and Microsoft Defender for Cloud. Our initial testing has demonstrated that the framework can successfully execute attack scenarios, monitor for alerts, and calculate MTTD metrics across these services. The modular design allows us to extend support to additional services and cloud platforms incrementally.

4 Security Applications

The MTTD Benchmarking Framework has several immediate security applications:

- **Evidence-based security decisions**: Organizations can select cloud security services based on quantitative performance data.
- **Detection gap identification**: Security teams can identify blind spots where certain attack techniques go undetected.
- **Configuration optimization**: Organizations can measure how different security configurations impact detection performance.
- **Cost-benefit analysis**: Security leaders can evaluate the detection performance relative to the cost of services.
- **Compliance verification**: Organizations can verify that their security controls meet detection time requirements.
- **Vendor accountability**: Security vendors can be held accountable for their detection performance claims.
- **Industry benchmarking**: Establishes baseline expectations for detection performance across different attack techniques.

Beyond cloud security, our methodology could be adapted for evaluating detection capabilities in other security domains like endpoint protection, network monitoring, and SIEM systems.

5 Importance of the Research

As organizations migrate critical infrastructure to cloud environments, selecting effective security services becomes paramount. Without objective performance metrics, organizations may invest in suboptimal security solutions, leaving them vulnerable to attacks. By establishing standardized MTTD benchmarks, our research will:

- Improve security decision-making based on empirical evidence
- Enable vendors to quantify and improve their detection capabilities
- Establish baseline performance expectations for the industry
- Promote transparency in security service effectiveness
- Reduce security incidents by helping organizations select more effective detection tools
- Lower costs by enabling more informed security investments

The emergence of regulatory frameworks requiring rapid detection and response to security incidents (like GDPR's 72-hour breach notification requirement) further emphasizes the importance of understanding and optimizing MTTD. Our framework provides a systematic way to measure and improve this critical metric.

6 Risks and Mitigation Strategies

We've identified several risks to our research:

- **Methodology limitations**: Our focus on non-destructive attack techniques may not fully represent sophisticated threats.
- **Platform changes**: Cloud service updates could affect measurement consistency over time.
- **Vendor resistance**: Security providers might be reluctant to participate in objective benchmarking.
- **Ethical considerations**: Attack simulations must be carefully designed to avoid unintended consequences.
- **Resource constraints**: As Masters students, we have limited time and computing resources.
- **Detection correlation challenges**: Accurately correlating attack indicators with security alerts can be complex.
- **API limitations**: Reliance on vendor APIs for alert collection may not capture all detection channels

To mitigate these risks, we are:

- Designing our framework to be adaptable to changing cloud environments.
- Focusing on a core set of well-documented attack techniques.
- Maintaining transparent communication with security service providers.
- Implementing comprehensive logging and audit trails.
- Using isolated testing environments.
- Developing robust correlation algorithms that can handle alert variability.
- Documenting API limitations and their impact on measurements

7 References

References

[1] Center for Internet Security. CIS Benchmarks for Cloud Security, 2022. URL `https://www.cisecurity.org/cis-benchmarks/`. Accessed: [Insert Date of Access].

[2] Pedro Garcia-Teodoro, Jesus Diaz-Verdejo, Hugo Maciá-Fernández, and Félix E. Vázquez. Intrusion detection systems for cloud computing environments: A systematic review. *ACM Computing Surveys (CSUR)*, 52(1):1–35, 2019.

[3] Gartner. Magic quadrant for cloud security service providers, 2023.

[4] MITRE ATT&CK. MITRE ATT&CK Framework for Cloud Environments, 2023. URL `https://attack.mitre.org/`. Accessed: [Insert Date of Access].

[5] National Institute of Standards and Technology. Cloud Security Technical Reference Architecture, 2023. URL `https://www.nist.gov/`. Accessed: [Insert Date of Access].

[6] Ruoyu Shu, Suman Jha, Shiqing Wang, Qi Duan, and Hao Zhang. Cloudstrike: A framework for cloud attack simulation and testing. 2022. URL `https://www.usenix.org/conference/usenixsecurity22/presentation/shu`.