Using Near-Real-Time Threat Intelligence to Detect Email-based Threats

1. Introduction

The cybersecurity landscape continues to evolve at an unprecedented pace, with attackers demonstrating remarkable sophistication in their approach to email-based threats. Modern threat actors operate with a strategic mindset—they systematically analyze their successes and failures, adapting their techniques when defensive measures prove effective, while simultaneously doubling down on attack vectors that continue to yield results. This dynamic creates a perpetual arms race between attackers and defenders.

The complexity of this challenge extends to practically all detection systems. Defenders and defensive security systems must maintain constant vigilance, rapidly identifying emerging attack techniques and implementing countermeasures before they can be widely exploited. Email security systems, serving as the first line of defense for many organizations, face particular pressure to evolve quickly in response to new threats.

Consider a critical scenario that every email security administrator faces: despite best efforts and advanced filtering technologies, a malicious email successfully bypasses all defensive layers and reaches an end user. This represents a false negative from the email security system's perspective—a failure that could potentially compromise the organization more broadly if left unchecked.*

The fundamental question becomes: What can email security systems do to prevent subsequent similar attacks from reaching other end users? The answer lies in implementing a sophisticated feedback loop that transforms false negatives into learning opportunities.

To effectively handle false negatives and strengthen overall security posture, email security systems require three essential capabilities. First, they need a robust mechanism to accurately identify when an email represents a false negative, essentially developing threat intelligence from real-world security events. Second, they must possess the necessary infrastructure to efficiently absorb and process the threat intelligence generated. Third, they need automated mechanisms to act upon this intelligence, preventing future attacks that share similar characteristics.

Cisco® Secure Email Threat Defense (ETD) addresses each of these requirements through an integrated approach that combines human expertise with advanced machine learning techniques. This system demonstrates how near-real-time threat intelligence can be effectively leveraged to enhance email security outcomes while reducing the window of vulnerability that attackers traditionally exploit.

* For customer organizations employing a defense-in-depth approach, the consequences of a false negative are mitigated by other security systems, such as endpoint detection and response (EDR).

2. Determining false negatives

The reality of email security is that no system achieves perfect accuracy. Even the most sophisticated platforms will occasionally allow malicious emails to reach their intended targets, resulting in false negatives that can represent potential security incidents. Accepting this limitation is crucial for developing effective response strategies rather than pursuing the impossible goal of perfect prevention.

False negatives are typically identified through retroactive analysis processes that occur after potentially malicious emails have already been delivered. One method involves dedicated Cisco threat analysts who systematically review email traffic, applying their expertise to manually classify suspicious messages as either malicious or benign. This human-driven analysis remains invaluable because threat analysts can identify subtle indicators and context clues that automated systems might miss.

Additionally, customer security team reporting plays a crucial role in identifying false negatives. End users who receive suspicious emails often serve as the final detection layer, flagging messages that appeared legitimate to automated systems but triggered human intuition. When end users report emails as malicious, the customer security team verifies and forwards the emails to Cisco threat analysts, who investigate these reports and reclassify messages accordingly.

While the percentage of false negatives may appear statistically small compared to the total email volume, their impact should not be underestimated. Each false negative represents a successful breach of defensive measures, and attackers often view initial success as validation to launch additional similar attacks. Without proactive intervention, subsequent attack emails using similar techniques would likely achieve a similar success rate, potentially leading to a broader organizational compromise.

Modern threat landscapes also demand consideration of attack campaign patterns. Sophisticated attackers rarely send a single malicious email; instead, they launch coordinated campaigns targeting multiple recipients across different time frames. Identifying one false negative often reveals broader attack patterns that inform defensive strategies beyond individual message analysis.

3. Storing near-real-time threat intelligence

Identifying false negatives provides valuable intelligence, but this information remains worthless without robust mechanisms to capture, store, and utilize it effectively. Cisco Secure ETD implements a sophisticated context database that serves as the foundation for intelligence-driven email security improvements.

When a false negative is identified through analyst review or customer reporting, the system immediately stores both the email content and its classification label within the context database. However, the storage process involves more than simple archiving. Rather than storing emails in their raw form, ETD converts each message into an embedding vector—a mathematical representation (a fingerprint) consisting of 300+ numerical values that capture the email's essential characteristics and semantic meaning.

This embedding approach offers several advantages over traditional storage methods. Embedding vectors enable efficient similarity comparisons while reducing storage requirements compared to full message retention. They also provide privacy benefits by abstracting away specific content details while preserving the mathematical relationships necessary for threat detection.

The context database extends beyond false negative storage to create a comprehensive repository of email intelligence. It maintains records of emails that were correctly classified by the security system, storing them in the same embedding vector plus label format. This comprehensive approach ensures that the system learns from both its failures and successes, building a more robust understanding of email threat patterns.

Analyst and customer feedback make another significant contribution. When legitimate emails are initially flagged as malicious (false positives), they can be reclassified and stored with appropriate labels.

This feedback mechanism enables the system to understand not only what constitutes a threat, but also what characterizes legitimate business communications.

The combination of professional threat intelligence analysis and customer feedback significantly improves the quality of data feeding ETD's detection engines. As shown in Figure 1, this collaborative approach leverages both the scale of customer insights and the expertise of threat analysts, creating a more comprehensive and accurate intelligence foundation.

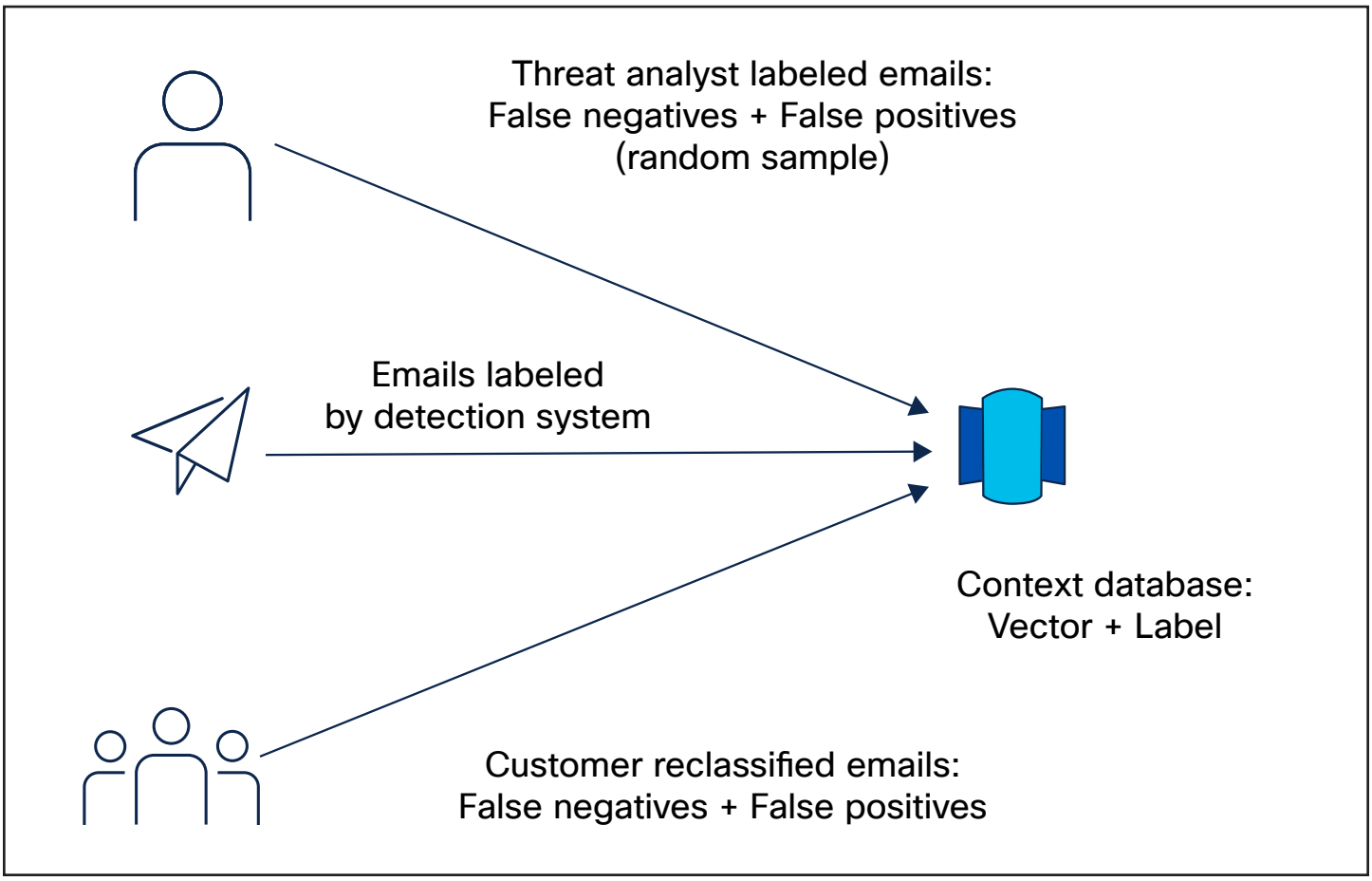

Figure 1. Context database created with input from threat analysts and customers

4. Using the similarity detector to block malicious emails

The stored threat intelligence reaches its full potential through ETD's similarity detector, which operationalizes historical attack data to identify and block new threats. This detector represents a sophisticated application of machine learning principles to cybersecurity challenges.

When ETD processes incoming emails, each message undergoes conversion into its corresponding embedding vector using the same mathematical transformation applied to the stored intelligence. This standardization ensures that new emails can be directly compared against historical threat data using consistent mathematical relationships.

The similarity detector operates by querying the context database to identify emails with embedding vectors that closely match the incoming message's embedding vector. Rather than requiring exact matches, the system identifies a small number of similar emails based on mathematical calculations of vector distances. This approach enables the detection of attack variations and evolutionary techniques that might evade signature-based systems.

The detector extracts both vectors and their associated labels from similar historical emails and then calculates a maliciousness grade for the new message based on two primary factors: the labels of similar emails and the mathematical closeness of their vectors. A message similar to previously identified threats receives a higher maliciousness grade, while one resembling legitimate communications gets a lower grade.

This maliciousness grade feeds into ETD's broader detection architecture, where multiple detector outputs are aggregated to produce a cumulative maliciousness assessment. As shown in Figure 2, the similarity detector provides one input among several, ensuring that detection decisions consider multiple analytical perspectives rather than relying on any single technique.

ETD also integrates large language models to further refine classifications for emails that fall into gray areas—messages that aren't clearly malicious or benign based on traditional analysis (see reference [1]). The similarity detector contributes to this enhanced analysis by providing label information on similar historical messages, enabling a more nuanced decision-making process.

The detector's effectiveness depends heavily on the quality and diversity of the underlying intelligence database. Rich historical data enables more accurate similarity assessments, while sparse data may lead to false conclusions. Continuous intelligence gathering and database maintenance, therefore, directly impact detector performance.

Note that the similarity detector is also used to validate legitimate business communication (true negatives) and correct messages incorrectly identified as threats (false positives). We omit a more detailed discussion in the interest of brevity.

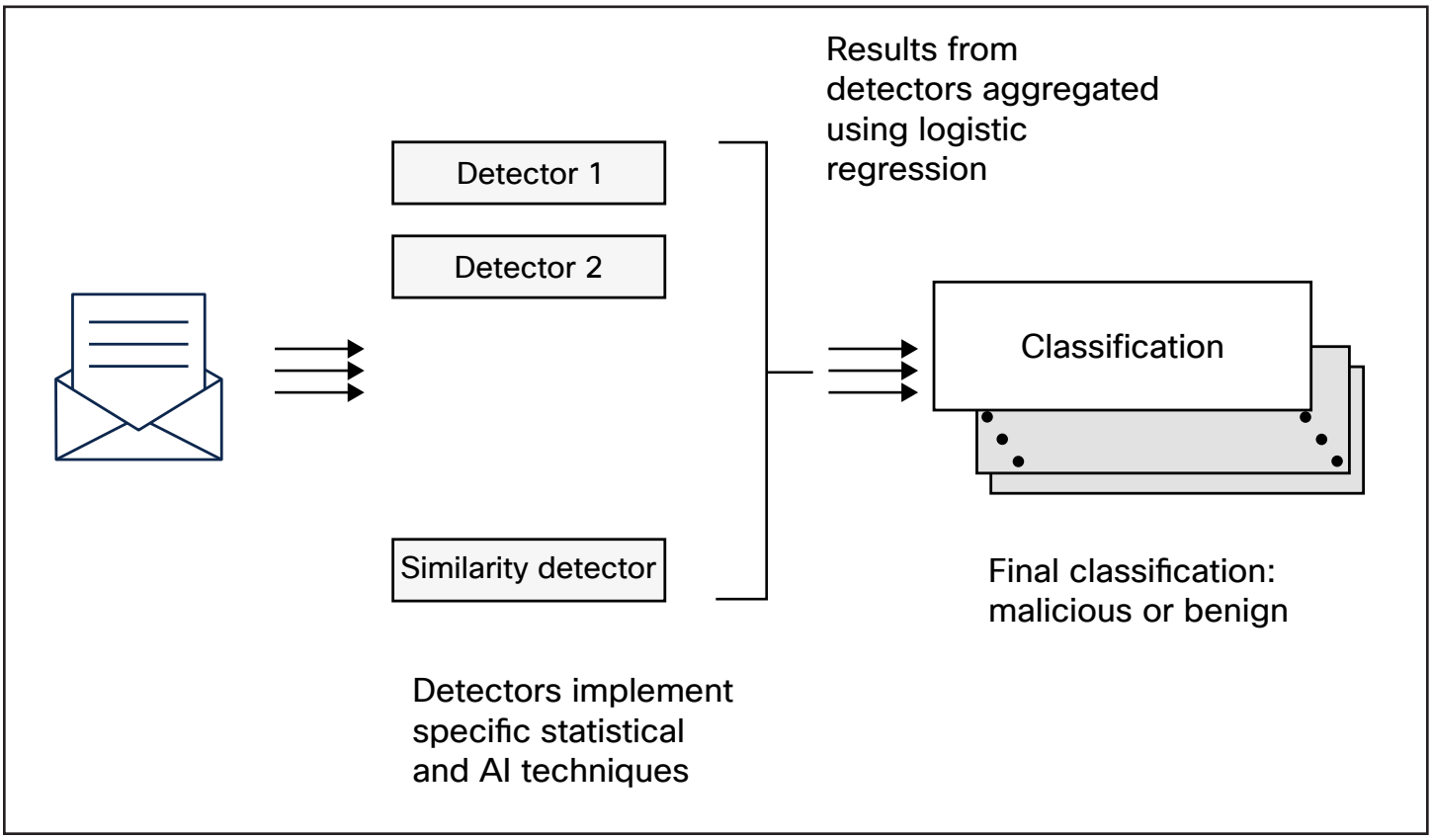

Figure 2. Similarity detector as part of the ETD detection system

5. Putting it all together

The practical application of near-real-time threat intelligence becomes clear when examining a typical attack scenario and the system's response to it. Consider an attacker who successfully delivers a malicious email to an end user within an organization. Despite multiple layers of defense, the email bypasses all automated detection systems and reaches its intended target.

The end user, recognizing suspicious characteristics or experiencing negative consequences from the email, reports it to their security team. The security team analyzes the email and classifies it as a threat. Alternatively, routine threat analysis might identify the email during systematic review processes. Regardless of the identification method, the email is reclassified as malicious and added to the context database with the appropriate labeling.

Subsequently, the same attacker attempts to expand their campaign by sending similar malicious emails to different users within the organization. For example, the attacker may change the subject line from "Urgent update" to "Important update." Further, these follow-up emails may originate from previously unknown sender addresses, potentially evading reputation-based filtering systems that rely on known bad actors.

However, the email security system now possesses threat intelligence from the initial attack. When processing the new malicious email, the similarity detector identifies mathematical similarity between the new message and the previously classified threat stored in the context database. Based on this similarity, the detector assigns a higher maliciousness grade than would have been possible without the historical intelligence.

The elevated maliciousness grade increases the probability that the cumulative detection system will classify the new email as malicious. If the context database contains multiple similar emails with malicious classifications, the cumulative score may exceed blocking thresholds, preventing delivery to the intended recipient. Figure 3 illustrates emails blocked via threat intelligence generated by ETD.

This process demonstrates how human-generated intelligence directly influences automated decision-making systems. The initial manual classification by analysts or customer reports creates a learning opportunity that improves future detection capabilities. Each false negative becomes a stepping-stone toward better protection, rather than simply representing a security failure.

The system's effectiveness compounds over time as the intelligence database grows. Early attack attempts might succeed due to limited historical data, but subsequent similar attacks face increasingly sophisticated detection as the system learns from each encounter. This evolutionary approach mirrors the adaptive strategies employed by attackers themselves!

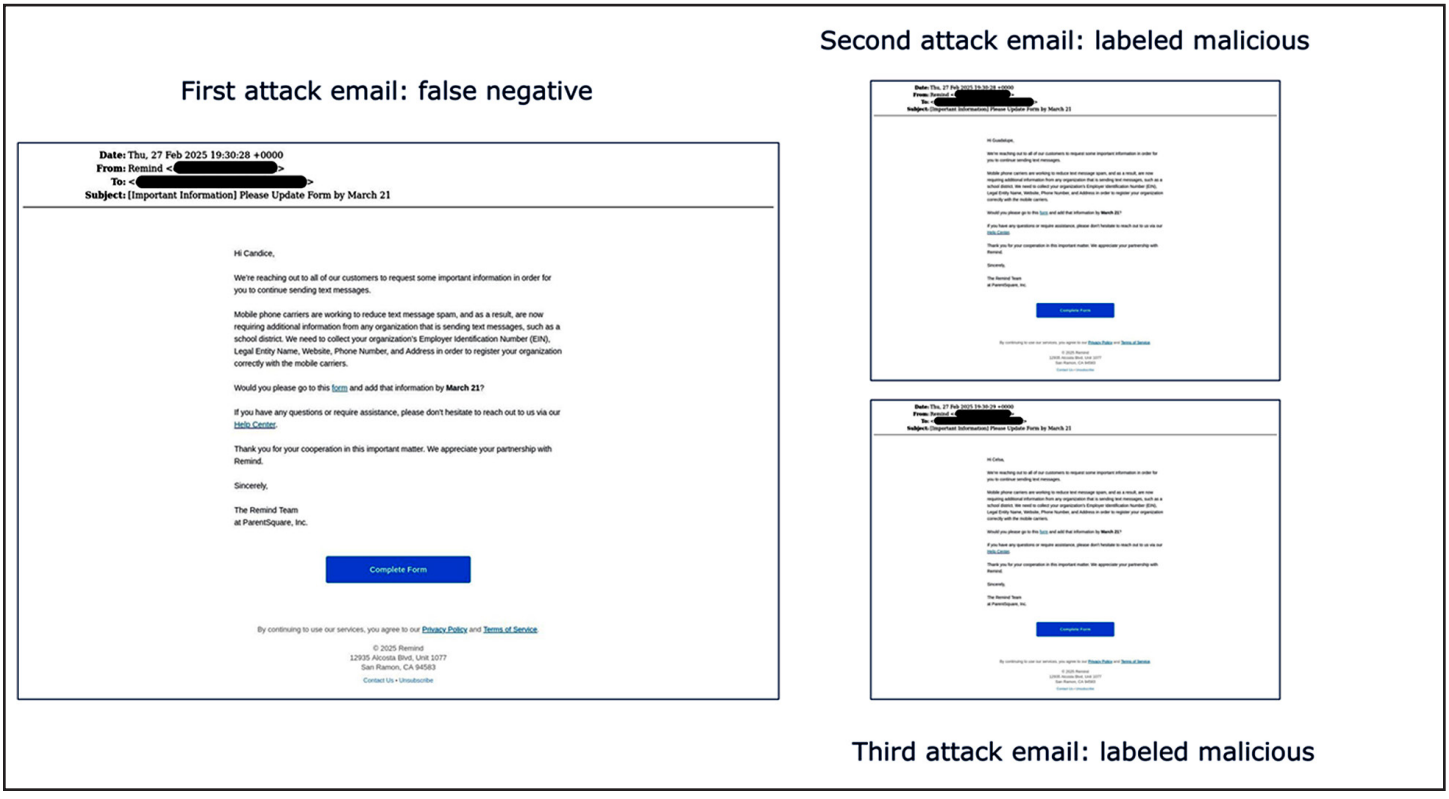

Figure 3. Impact of the similarity detector and the context database

6. Conclusion

Cisco has been working on email security for two decades, continuously investing in security innovations to address evolving threats. The development of a near-real-time email threat intelligence ecosystem and the use of an email similarity detector exemplify this ongoing commitment to advancing cybersecurity capabilities. Additionally, the mechanisms described above have proven their effectiveness at scale in real-world scenarios.

For more detailed information on ETD's technical architecture, interested readers are encouraged to consult the references below, which provide in-depth explanations of the system's design and capabilities.

To experience ETD's capabilities first-hand, start a free trial today.

7. References

[1] How large language models enhance Cisco Secure Email Threat Defense, May 2024, Cisco Systems, Inc.

[2] Using Relationship Graphs to Mitigate Email-based Threats, May 2025, Cisco Systems, Inc.

[3] Brabec et al., A Modular and Adaptive System for Business Email Compromise Detection, August 2023, arxiv.org.