Mandiant

DTM ALERT SEVERITY DEFINITIONS AND EXAMPLES

This document describes how **Digital Threat Monitoring (DTM)** (https://docs.mandiant.com/home/dtm-digital-threat-monitoring) determines the severity ratings for Alerts.

Alert Scoring Framework

DTM alert scoring (https://cloud.google.com/blog/products/identity-security/alert-scoring-machine-scale/) is governed by two components:

- Confidence, which uses machine learning to score how certain Mandiant is in the given alert actually being malicious
- Severity, which scores what the potential impact of the threat is to determine whether an Alert needs further investigation or is ready for review

Too many alerts, especially false positives, can lead to analyst fatigue. The Confidence score's job is to help initially remove any obvious noise, and any other downstream context is then traversed by the Severity scoring model to further divide alerts into either **High**, **Medium**, or **Low** categories.

Several factors are considered when determining Severity including:

- detected entities within the document
- the document's security classifications
- monitor matching criteria, since the severity of an alert should be considered in the context of your specific environment and risk tolerance

Prioritization of Alerts

Mandiant recommends giving priority to the resolution of High severity Alerts, whereas Medium and Low severity Alerts should be triaged afterwards, time permitting. You can sort Alerts by Severity in order to facilitate this process.

The following table outlines a common understanding of the characteristic features of alerts that give rise to different severity categories.

Alert Severity Definitions		
Alert Severity	**Definitions**	**Examples**
High	Indicates a critical security issue or active exploitation attempt that poses an immediate and significant risk to your organization Could lead to data breaches, system compromise, service disruption, or financial loss Suggested Usage: • Investigate within 24 hours • Remediate within 7 days	• Matches a document showing a darknet security breach notification • Document contains a typosquatted domain matching your monitor criteria • Document contains an infrastructure netloc indicator matching your monitor criteria

Alert Severity Definitions		
Medium	Indicates a security issue or suspicious activity that poses a moderate risk to your organization May lead to unauthorized access, data leakage, or minor service disruption if left unaddressed Suggested Usage: • Investigate within 72 hours • Remediate within 21 days	• Document has a potential leak of confidential matched entity • Contains a credit card bin not matching monitor criteria • Matched document comes from a dark web shop, sells payment cards, accounts, or compromised machines
Low	Indicates a potential security issue or anomaly that poses a low risk to your organization Unlikely to cause immediate harm, but could contribute to a larger security issue if ignored  For common low severity alerts, consider automating responses to free up your security team to focus on more critical issues. Suggested Usage: • Investigate within 1 month • Remediate within 3 months	• Matches monitor criteria, but does not contain any other security-relevant contextual landmarks or derived findings