Virus Scanning in AWS S3 with Hash Checking and VirusTotal Integration

Introduction

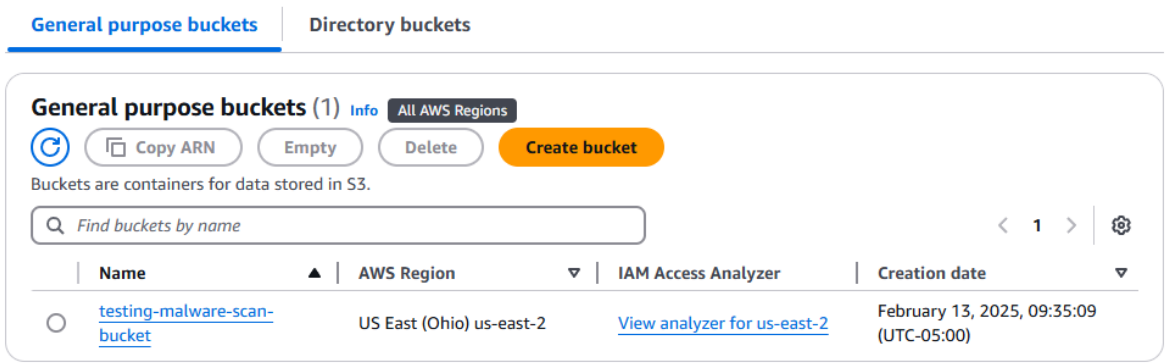
The objective of this project is to implement a script that checks the hash of files uploaded to an S3 bucket against the VirusTotal API. This ensures that potentially malicious files can be identified and handled appropriately. For this, I utilized various AWS functionalities, including Lambda functions, S3 buckets, and GuardDuty.

Implementation

Step 1: Setting Up the S3 Bucket

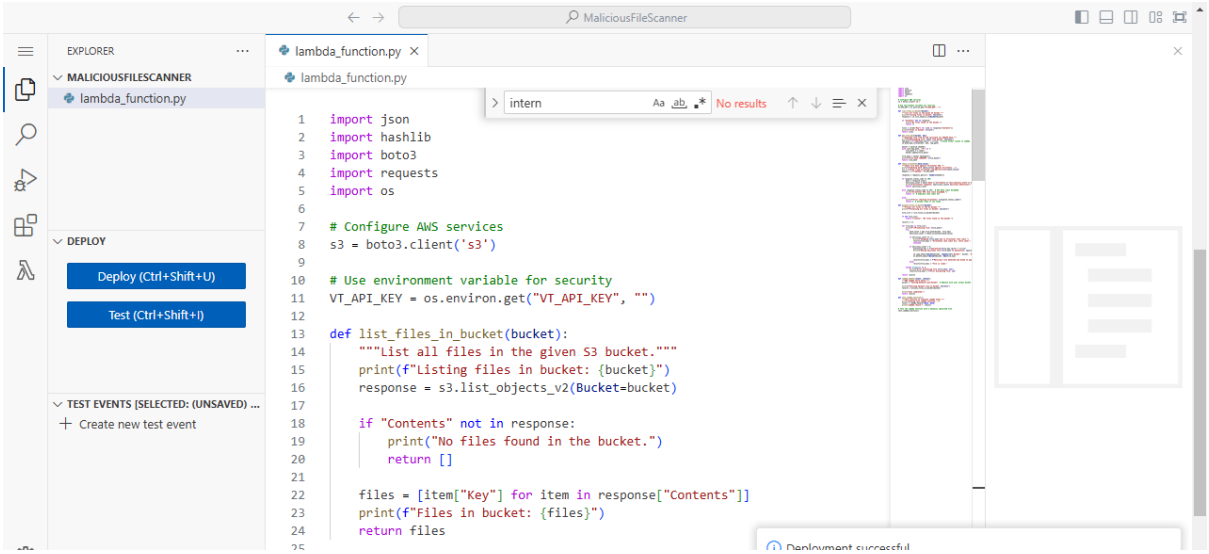
- Created an S3 bucket named testing-malware-scan-bucket.
- Uploaded multiple test files to simulate a real-world scenario.

ARN: arn:aws:lambda:us-east-2:575108960751:function:MaliciousFileScanner



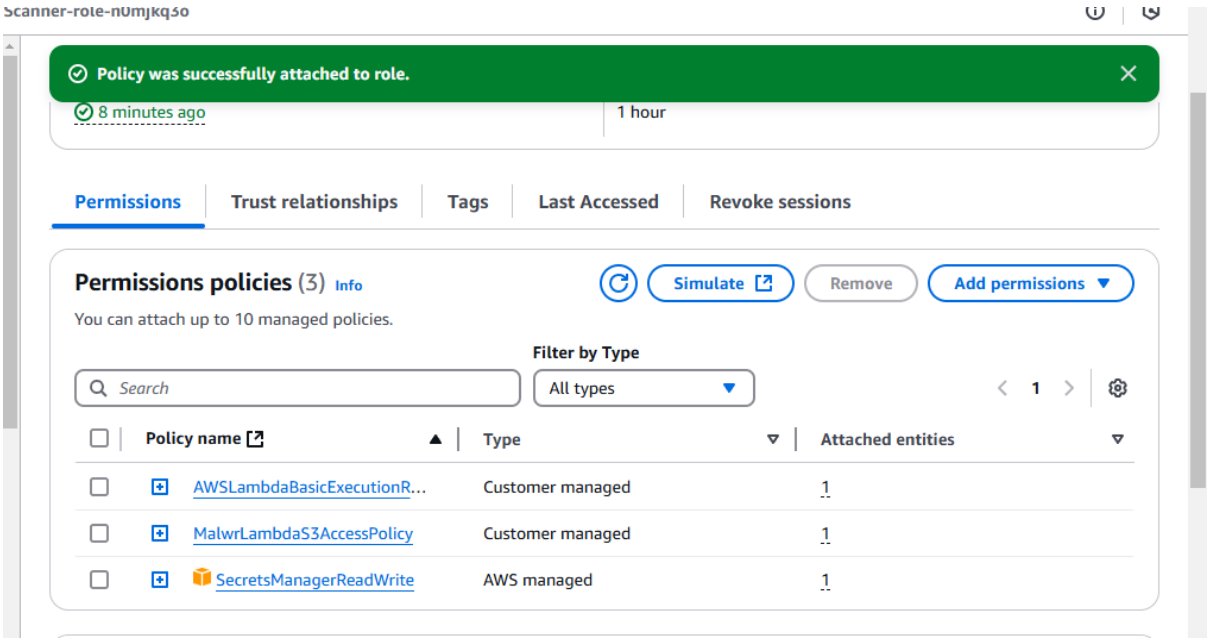
Step 2: Writing the Lambda Function

- Developed a Python script to:
    - List all files in the S3 bucket.
    - Compute the SHA-256 hash of each file.
    - Query the VirusTotal API to check if the file is flagged as malicious.
- Handled API rate limits and logging mechanisms for better tracking.



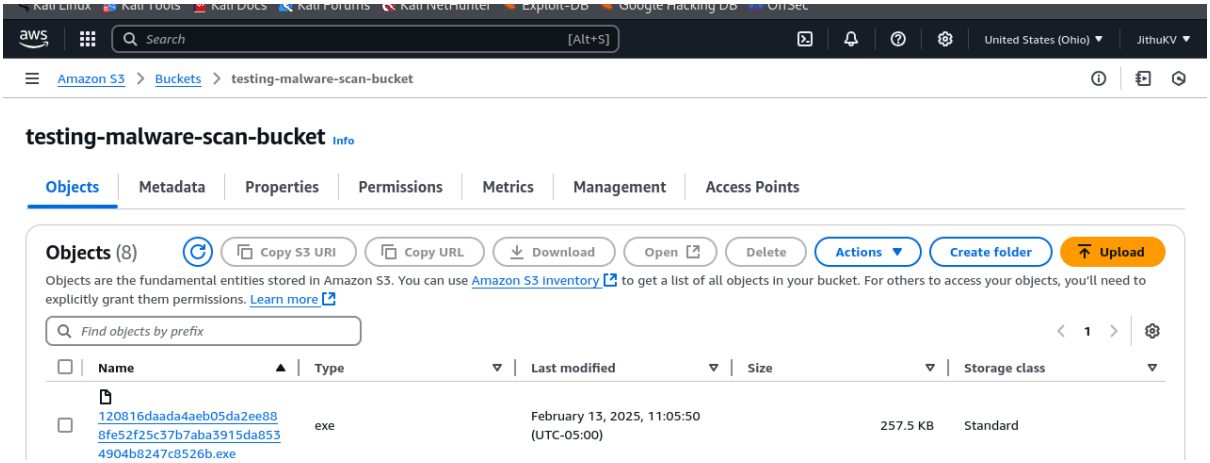
Step 3: Configuring IAM Permissions

- Ensured that the Lambda function had proper permissions to:
  - Read objects from the S3 bucket.
  - Write logs to CloudWatch.
- Adjusted IAM roles to avoid 403 Forbidden errors when accessing S3 objects.



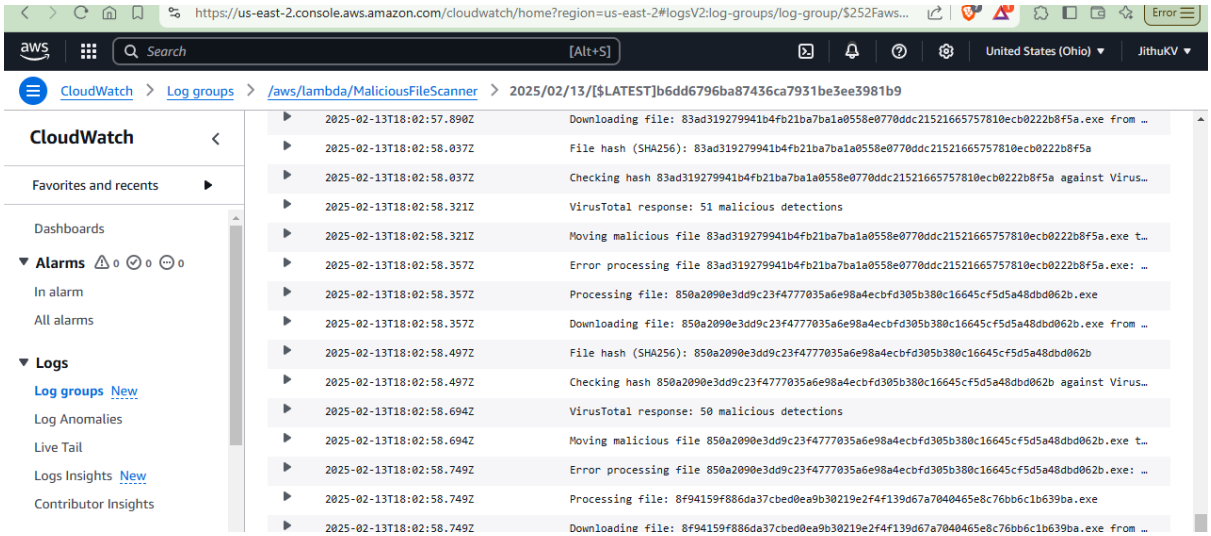
Step 4: Testing the Function

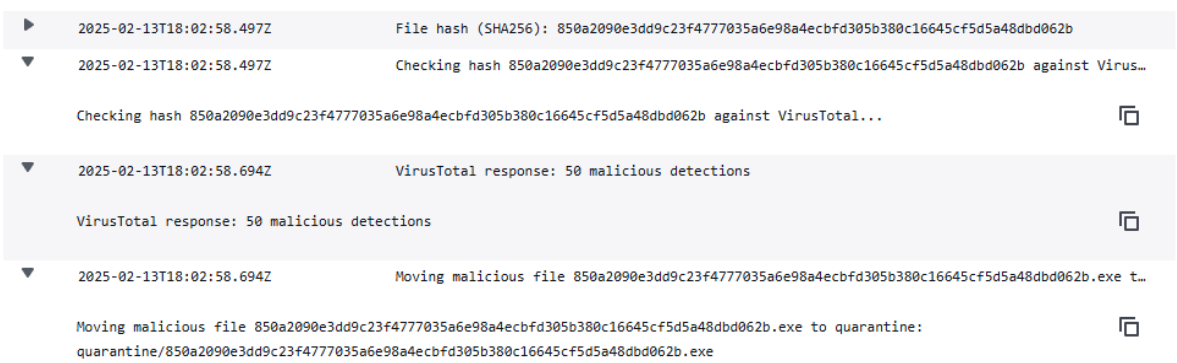
- Deployed the Lambda function and executed test cases.
- Verified logs in CloudWatch to confirm file hashes and VirusTotal responses.
- Addressed issues such as missing file permissions and API request errors.



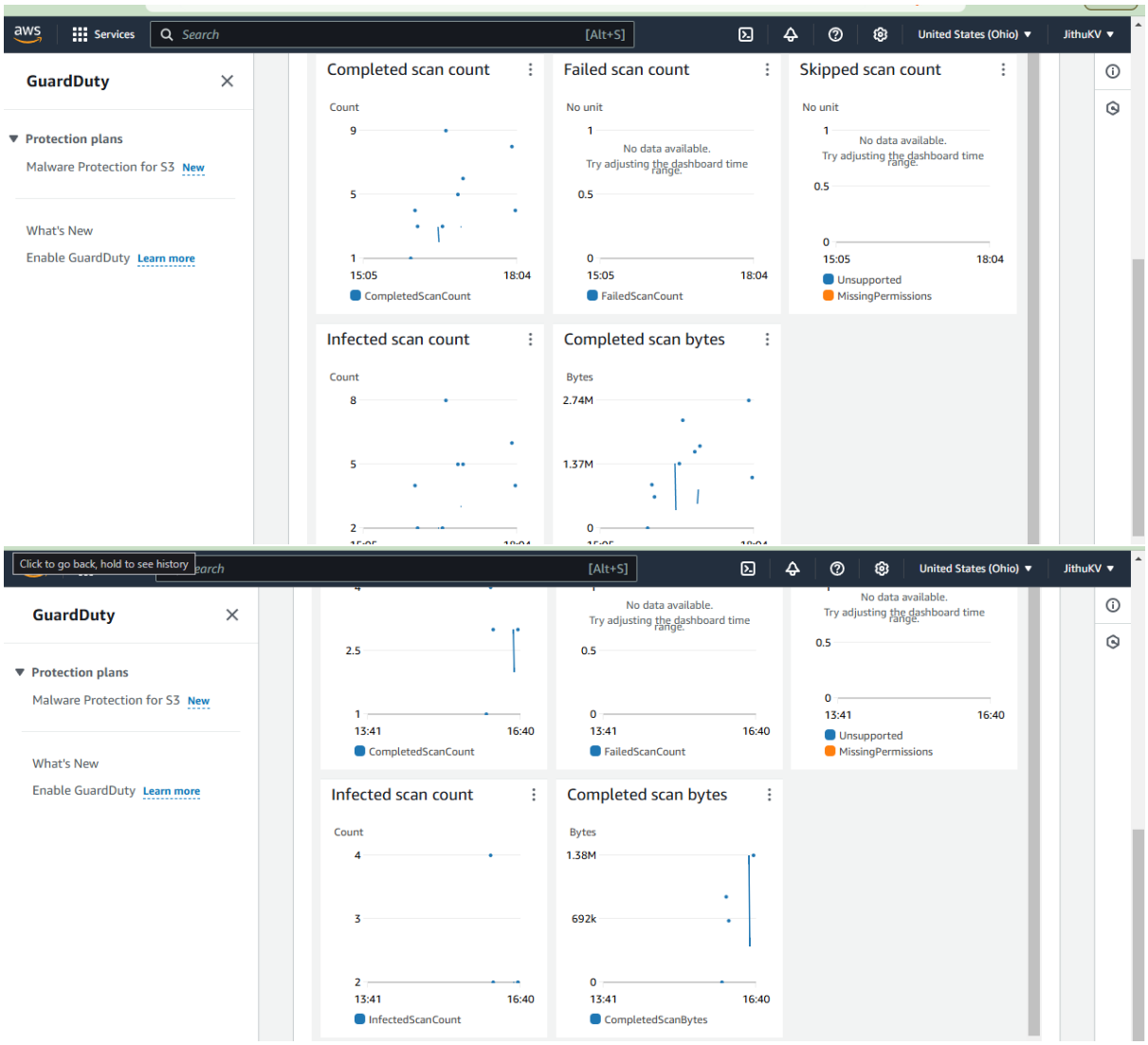
Results

- Successfully retrieved file hashes and checked them against VirusTotal.
- Encountered and handled API rate limits appropriately.
- Detected potential malicious files and logged them for further action.





Guard duty malware agent is able to detect the malicious files too



Conclusion

This project demonstrates an automated approach to scanning files in an S3 bucket for potential malware. Using AWS services like Lambda and GuardDuty, along with VirusTotal integration, ensures a proactive security posture. Future improvements could include integrating SNS for real-time alerts and optimizing API requests for large-scale file scanning.