ADVISORY FOR IM SECTORS

Exposed GitHub token leading to data leak

Original report published on: Jun 8, 2024 [1]

Executive Summary

An exposed GitHub token belonging to The New York Times (NYT), a major mass media company, was found by a hacker and used to access 5000 GitHub repositories with 3.6 million files. The data was obtained and posted to 4chan, an online bulletin.

Background

On 6 June 2024, 270 Gigabyte (GB) of data belonging to NYT was posted on 4chan. The post mentioned that data consisting around 5000 GitHub repositories and 3.6 million files had been obtained. The data contained source code for Wordle, a web-based word game owned by NYT, IT documentation, infrastructure tools, WordPress database containing information of around 1500 users and sensitive information like API tokens and secret keys.

The data leak may have implications as it involved source codes which can be reviewed by threat actors to find potential vulnerabilities or gain understanding of internal infrastructure. As the data was accessed likely using a privileged GitHub token, data in the repositories may have been tampered with, to introduce vulnerabilities or backdoors.[2]

The hacker behind the data leak claimed to have accessed them after finding a GitHub token with access to the repositories. NYT was aware of the data leak and mentioned a GitHub token had been accidentally exposed in January 2024 although exact details were not disclosed.

Detection and Mitigation Techniques

- Use access management control to manage access keys and tokens.
- Grant only the required privileges for access keys and tokens to limit potential abuse, in the event if the access keys and tokens are exposed or leaked.
- Practise secure coding and perform source code reviews to ensure access keys and tokens are not hardcoded into the application.
- Deploy API Key, token and secret scanning on code repositories, execution pipelines, configuration files and other data sources and alert for prompt action.
- Implement rotation and revocation policies for access keys and tokens to reduces the amount of time a compromised key or token can be used.
- Proper data segregation is recommended where possible to prevent sensitive and non-sensitive data being stored together.

MITRE ATT&CK Tactics and Techniques

Tactic	Technique ID and name	Details
Initial Access	T1078.004 - Valid Accounts: Cloud Accounts	Cloud accounts are those created and configured by an organization for use by users, remote support, services, or for administration of resources within a cloud service provider or SaaS application.
Credential Access	T1528 - Steal Application Access Token	Adversaries can steal application access tokens as a means of acquiring credentials to access remote systems and resources.
Exfiltration	TA0010 - Exfiltration	Exfiltration consists of techniques that adversaries may use to steal data from your network.

References

1. https://www.bleepingcomputer.com/news/security/new-york-times-source-code-stolen-using-exposed-github-token/
2. https://www.darkreading.com/cloud-security/new-york-times-internal-data-nabbed-from-github