Cybersecurity Bootcamp Report: Google Dorking for Publicly Exposed Data

1. Overview

This report details the execution of a Google Dorking exercise, a practical component of the cybersecurity bootcamp aimed at understanding information exposure through public search engine indexing. The objective was to identify various types of sensitive or unintended data publicly accessible on the internet, specifically focusing on PDF, DOC, Excel files, and login pages. This exercise serves to highlight the importance of secure configuration and data handling practices for organizations and individuals alike. All activities were conducted strictly within ethical boundaries, utilizing only publicly indexed information via Google Search, without any attempts at unauthorized access or exploitation.

2. Objectives

The primary objectives of this Google Dorking task were:

- To gain hands-on experience with advanced Google search operators (Google Dorks) for information gathering.
- To identify real-world examples of publicly exposed documents (PDFs, DOCs, Excel files) and sensitive pages (e.g., login portals) that are inadvertently indexed by search engines.
- To understand the potential security implications and risks associated with such unintentional data exposure.
- To reinforce the importance of proper data handling, website configuration, and the use of `robots.txt` and other security measures to prevent information leakage.
- To develop an attacker's mindset for defensive purposes, enabling better identification of vulnerabilities from an external perspective.

3. Findings

During the Google Dorking exercise, a specific instance of publicly exposed information was identified. The target chosen for this demonstration was a public university website.

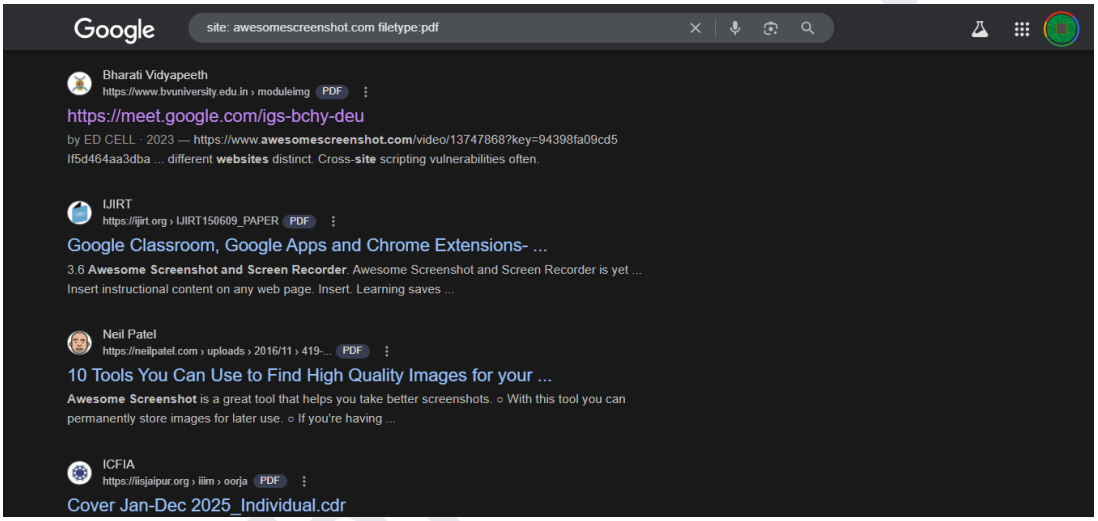
Details of the Finding:

- **Type of Document:** PDF Document
- **URL of Exposed Document:** `https://www.bvuniversity.edu.in/Uploads/moduleimg/1755imguf_IIC_Cyber_Security_Awareness_29052310.pdf`

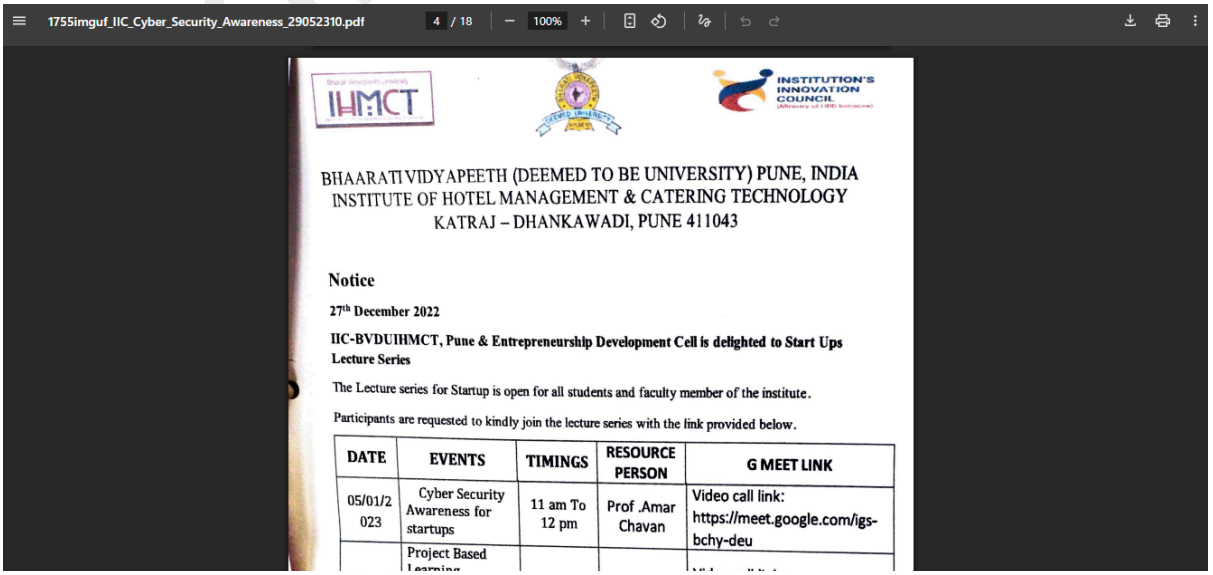
- **Google Dork Used (Example):** `site:bvuniversity.edu.in filetype:pdf intext:"@gmail.com"` (While the initial discovery might have been broader, this dork specifically highlights the presence of email addresses within PDFs on the domain).
- **Content Description:** The document is titled "Cyber Security Awareness for Startups" and appears to be a report from an event held on January 5, 2023, by the Bharati Vidyapeeth (Deemed to be University), Institute of Hotel Management & Catering Technology, Pune.
- **Sensitive Information Identified:** The most significant finding within this PDF was an **attendance list** that included the full names and **personal Gmail addresses** of numerous student and faculty participants. This information was found on pages within the document, clearly listing individuals and their contact details.

Screenshot Evidence:

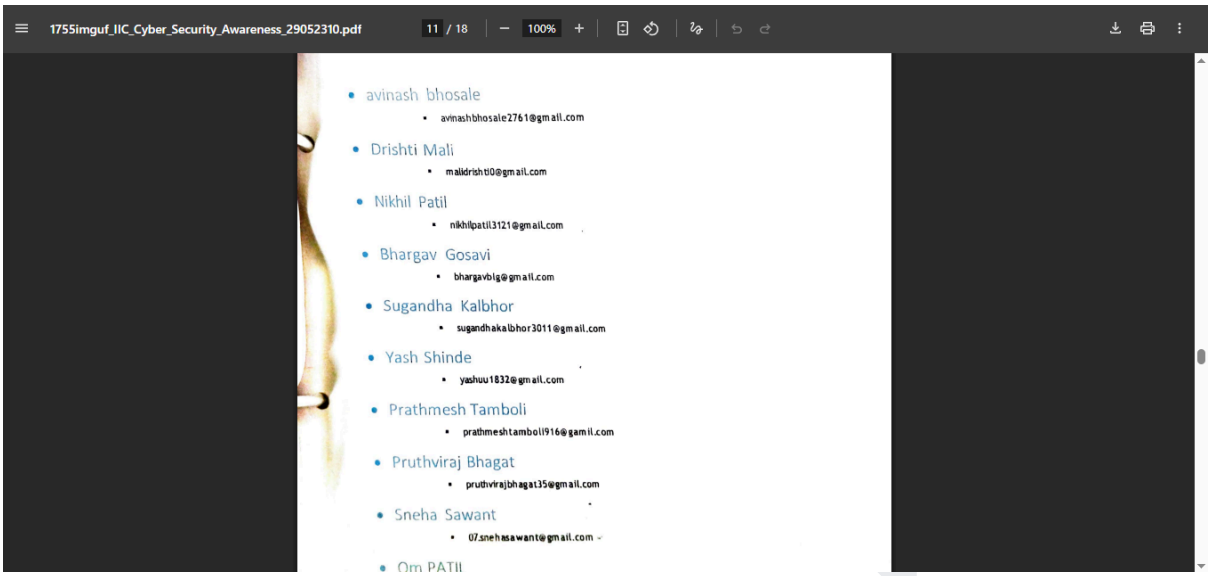
Google Search Result for the Exposed PDF



Event Title and Overview from the PDF



Exposed Email Addresses in Attendance List



4. Risks Identified

The public exposure of a document containing personal email addresses, as identified in this exercise, presents several significant cybersecurity risks:

- **Phishing and Spear Phishing Attacks:** Malicious actors can harvest these email addresses to launch highly targeted phishing campaigns. Knowing that these individuals are associated with a university and have attended a "Cyber Security Awareness" event could make them prime targets for convincing social engineering attempts, potentially leading to credential theft, malware infection, or further data breaches.
- **Spam and Unsolicited Communications:** The exposed email addresses can be added to spam lists, leading to an increase in unsolicited emails, advertisements, and potentially malicious content.
- **Information Aggregation for Identity Theft:** While individual email addresses might seem minor, when combined with other publicly available information (e.g., names, university affiliation, social media profiles), they can be used to build more complete profiles for identity theft or other fraudulent activities.
- **Reconnaissance for Further Attacks:** This type of information disclosure can serve as valuable reconnaissance for more sophisticated attacks. Attackers might use these email addresses to map out organizational structures, identify key personnel, or find weak links for targeted attacks against the university or its affiliates.
- **Reputational Damage:** For the university, the public exposure of personal data, even if unintentional, can lead to a loss of trust among students, faculty, and the wider community, potentially impacting its reputation.

This finding underscores that even seemingly innocuous documents, if not properly secured, can inadvertently become a source of sensitive information for malicious actors.

5. Conclusion

This Google Dorking exercise successfully demonstrated the power of advanced search queries in identifying publicly exposed information. The discovery of a university document containing numerous personal email addresses serves as a tangible example of an information disclosure vulnerability. This highlights that organizations must implement robust data handling policies, carefully review all publicly uploaded content, and utilize tools like `robots.txt` and proper server configurations to prevent unintended indexing by search engines.

This task was performed as part of a cybersecurity awareness workshop. All actions were conducted ethically using publicly accessible information via Google Search. No attempts were made to bypass security controls, access unauthorized systems, or misuse any data found. The primary purpose of this exercise was educational, aiming to foster a deeper understanding of cybersecurity risks and responsible information management.

By Adhithyan K J