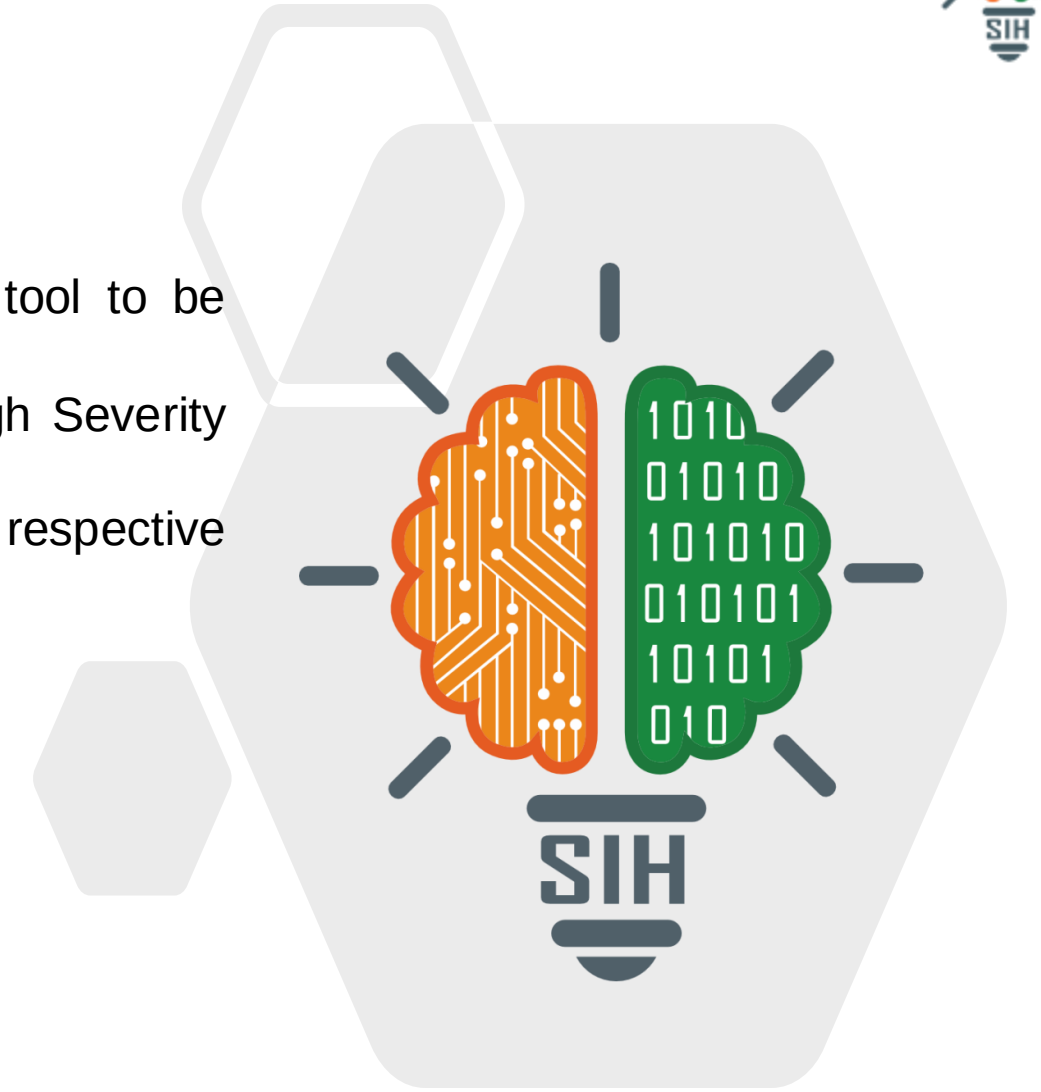
SMART INDIA HACKATHON 2024

- **Problem Statement ID :** SIH1676
- **Problem Statement Title -** Web-scrapping tool to be developed to search and report Critical and High Severity Vulnerabilities of OEM equipment published at respective OEM websites.
- **Theme -** Blockchain & Cybersecurity
- **PS Category -** Software
- **Team ID -** 29961
- **Team Name -** Tech Limit Exceeded



Tech Limit Exceeded

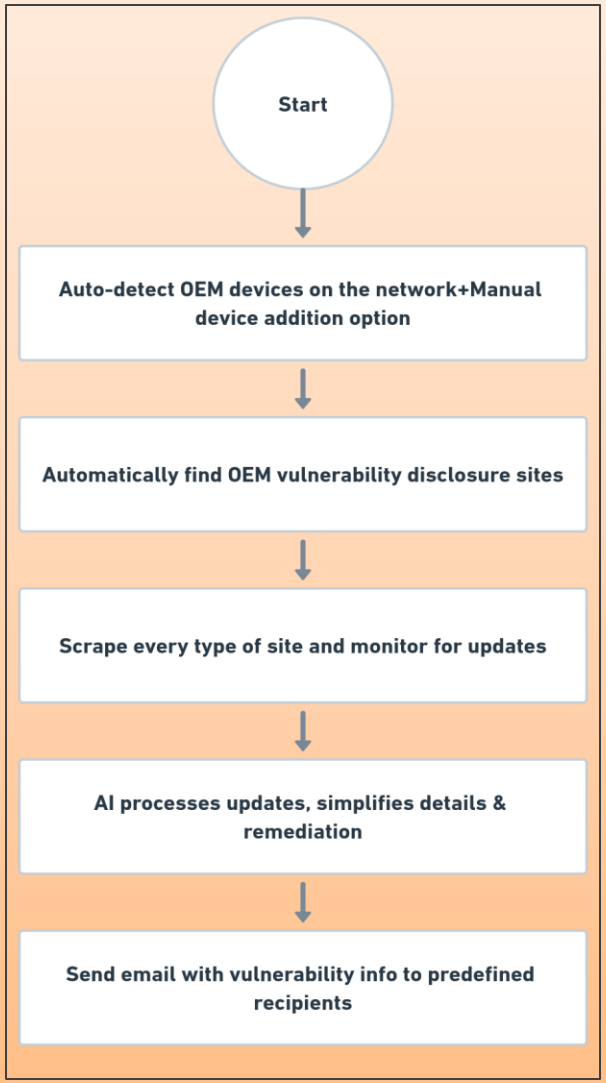
Disclose Tracker: Automated web-scraper

SMART INDIA HACKATHON 2024

1.Idea / Proposed Solution:

The solution automatically scans devices, monitors OEM websites for vulnerability disclosures, and sends real-time email alerts for quick action. It also simplifies remediation.

- ✓ **Automated Device Discovery**: Uses **Nmap** to scan the network and detect OEM equipments without manually typing.
- ✓ **Manual Device Addition**: **Web interface** for manually adding and managing devices with extra information.
- ✓ **Vulnerability Monitoring**: **Playwright** scrapes OEM sites and security sites for **real-time updates**.
- ✓ **Automated OEM Identification**: **Google Dorking with `googlesearch-python`** library find OEM security advisories.
- ✓ **Multi-Source Monitoring**: Scrapes for **(OEM sites+ NVD+ zero-day platforms)** for vulnerabilities.
- ✓ **Email Alerts**: Sends real-time updates to admins using **SMTP.**



2. Problem Resolution:

- ✓ **Immediate Vulnerability Detection:** Provides real-time monitoring and faster updates, ensuring quick responses to critical threats.
- ✓ **Comprehensive Device Coverage:** Protects IT and OT systems by identifying information across all critical devices.
- ✓ **Dual-Mode Operation:** Combines automated scanning with manual input to ensure complete device coverage if some are not on network.
- ✓ **Timely Notifications:** Sends prompt alerts to key stakeholders, enabling swift action to mitigate risks.

3.Solution Uniqueness:

- ✓ **Hybrid Detection Approach:** Combined **automated network scanning** with manual device registration for covering all devices.
- ✓ **Advanced Web Search Tactics:** Leveraged **Google Dorking** to identify hard-to-find OEM vulnerability disclosures.
- ✓ **Dynamic web-scrapping:** Used the playwright library to scrap any type of website regardless of different code-structure(would be able to scrap almost every website.)
- ✓ **Web-interface:** Providing web interface to manage/setup the scraper.
- ✓ **Best Remediation:** Harness the power of gpt4free for expert-level remediation suggestions and detailed vulnerability insights, all at no cost.

Tech Limit Exceeded

TECHNICAL APPROACH

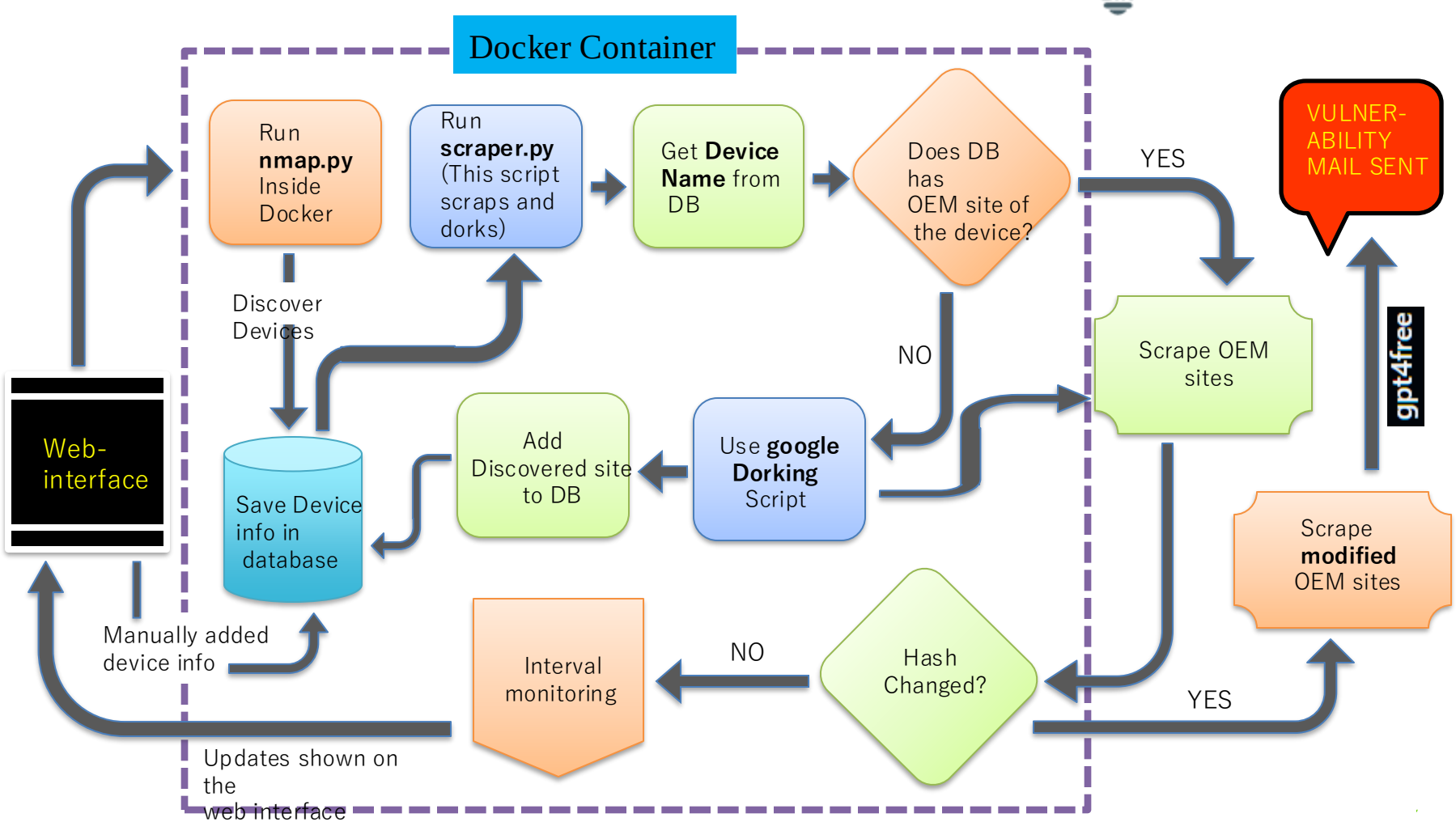
SMART INDIA HACKATHON 2024

1.Technologies/Frameworks

NMAP, docker, Playwright, Streamlit, HTML, CSS, Firebase Realtime Database, JavaScript, GPT4Free, python, SMTP, django

Product Status

- The **web-scrapers are fully ready** and are working nicely when tested on local network having various devices.
- The **web-interface is not fully ready** but it will be ready soon so, **overall 75% product is ready !**



2. Methodology and implementation flow

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2024

Tech Limit Exceeded

Analysis of Feasibility

Technical Feasibility
- Simple Docker Container-ization
- Available free open-source AI models.
- Easy product development using Python scrapers

Financial Feasibility
- Completely free
- no premium tools or costs required
- Zero-budget product with potential for high returns

Operational Feasibility
- Minimal resources needed
- No specialized skills required; user-friendly interface.
- Only a server is needed for monitoring

Potential Challenges and Risks

- Providing fixes without accurate Vulnerability information could lead to issues
- Using one script for all websites was difficult; finding the right libraries was essential.
- Different OEMs post advisories on separate platforms with inconsistent naming, making manual searches necessary
- Optimizing scraping schedules for various sources was challenging

Strategies for Overcoming Challenges

- Used **gpt4free** for detailed vulnerability explanations
- Leveraged **Playwright** along with **crawl4ai** to scrape most websites.
- Applied **Google Dorking** via the **googlesearch** library to find OEM sites
- Scheduled tasks using **cron jobs** or **apscheduler** in Python

Tech Limit Exceeded

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2024

IMPACTS

- ✓ Automated real-time monitoring systems could reduce the financial impact of data breaches by up to 30%.
- ✓ Early detection of vulnerabilities, especially in OEM systems, can **reduce overall security incidents by up to 25%.**
- ✓ Clients and partners feel more secure with well-maintained systems.

BENEFITS

- ✓ Reduced expenses on incident response and recovery.
- ✓ Avoiding system downtime and customer loss due to security breaches.
- ✓ Prevent data breaches and unauthorized access to sensitive information.
- ✓ Automated monitoring reduces the need for extensive manual labour.

Tech Limit Exceeded

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2024

- **Nmap** - https://nmap.org/book/man-os-detection.html
- **Vulnerability disclosure portals –**
  https://nvd.nist.gov/
  https://www.zero-day.cz/
  https://portswigger.net/daily-swig/
  https://threatpost.com/
  https://www.exploit-db.com/
- **Google Dorking techniques** - https://en.wikipedia.org/wiki/Google_hacking
- **Understanding IT and OT devices –**
  https://www.cisco.com/c/en/us/solutions/internet-of-things/what-is-ot-vs-it.html
  https://www.shiprocket.in/blog/original-equipment-manufacturer-oem/
- **Research on Scrapping tools –**
  https://medium.com/geekculture/web-scraping-101-tools-techniques-and-best-practices-417e377fbeaf
  https://oxylabs.io/blog/playwright-web-scraping
- **Scheduling Scripts** -
  https://www.reportserver.net/de/dokumentation/scripting-guide-4-6-1/4-monitoring-multithreading-scheduling-and-other-advanced-techniques