



Smart India Hackathon

Team: Lemon

Institute Name: Indian Institute of Technology Bhilai

Problem Statement ID: 1452

Problem Statement Title: Develop Ransomware Readiness Assessment tool

Organization: National Technical Research Organisation,(NTRO)

20.12.2023

Key Aspects

Our solution is divided into two steps:

- Dynamic Questionnaire
- Automated multi-step process

We've created a comprehensive system that contains both client and server architecture. Additionally, we've developed standalone applications that operate seamlessly without the need for an internet connection.

Types of users

- Our product has three different types of users with varying access levels.
 - National Technical Research Organisation,(NTRO)
 - Organization
 - User

Dynamic Questionnaire

● Dataset:

- RSAT [\[LINK\]](#)
- Cisagov cset dataset [\[LINK\]](#): specifically from sql queries [\[LINK\]](#)
- ISO/IEC [\[LINK\]](#): Ransomware protection provision
- Cis controls from cisecurity v8 guide[\[LINK\]](#)
- [Veeam](#), [Prevalent](#) and [Cybersecop](#) ransomware risk assessment tool for predefined questions.
- Final dataset contained 500+ data points.
- Dataset schema:
 - Id: Unique key for each record in dataset
 - Category: Questionnaire category for classification. Our data includes 10+ categories ranging to *DATA PROTECTION (DP)*, *NETWORK SECURITY*, *SECURITY AWARENESS AND TRAINING (SAT)*, *SYSTEM AND SOFTWARE MANAGEMENT (SSM)*, *ACCESS CONTROL AND IDENTITY MANAGEMENT (ACIM)*, *INCIDENT MANAGEMENT* and many more.
 - Question: Query for each organization. This includes Yes/No and multiple choice questions.
 - Response: Answer given by the user.
 - Description: Additional context related to particular queries.
 - Confidence: Numerical value quantifying the confidence level of the user for various questions.
 - Predetermined / Dynamic: Users are initially presented with a set of predetermined questions, averaging two questions from each category. Subsequently, based on the user's confidence level, the system dynamically generates additional questions. The emphasis is on providing the user with more questions from categories they express confidence in, while still including a subset of questions from categories where confidence may be lower.

● Information Retrieval

- Users begin with 20 predetermined questions, 2 from each category.
- Each questionnaire page consists of 5 questions.
- Users express confidence for each question.
- If confident in a category:
 - Receive 4 questions from confident areas.
 - Get 1 question from a less confident area.
- This approach ensures fairness, accounting for potential knowledge in less confident categories.
- Results are obtained with just 30-40 questions, approximately 10% of the current dataset size.
- This streamlined approach maximizes efficiency and reduces the total time taken by the user.
- BM25 IR technique is used to get matching questions.

Automated multi-step process

1. Mail Attachments auto check:

- Download single file from drive.
- Download all the attachments from all the emails.

2. User's search history: The program comprises two phases: data cleaning for Logistic Regression and machine training to identify malicious URLs. The data was structured with URL and label columns for model training. Tf-idf feature extraction from sklearn was employed.

3. Ransomware detection crawler using Machine Learning:

- Classification of various files present in the user's system using Machine learning model trained on yara signature and heuristic.
- However it does not work on polymorphic malware, which can alter its signatures, as well as new and unknown malware for which signatures have not been established yet.

- Dataset: [LINK](#)
- 4. **Open port checking:**
 - Network Mapper(Nmap) : Open source linux command line tool. We are using it to scan all the devices connected in the network. Not just websites, it scans all the devices, mobile phones present in the network.
 - Devices that are directly accessible by the open internet have the highest risk of ransomware attack.
 - We have implemented it by iterating through the whole subnet and finding all the vulnerable devices.
 - Dataset: [LINK](#)
- 5. **Remote Desktop Protocol:** Ports opened due to the RDP connection:
- 6. Checking whether our system is allowing for download of ransomware-like files.
- 7. **Access new domains:** Attackers open up new domains that are active for very less time. We are making a connection to a new domain that appears on the world wide web.
- 8. **Uploading encrypted file:** Blocks upload of password protected/encrypted files to obscure cloud storages.
- 9. **Financial information:** Upload of files containing financial information, such as credit card details, to unfamiliar or unknown cloud servers poses potential security risks. Giving warning before upload.

Technical Stack

- Frontend: ReactJS
- Backend: NodeJS
- Database: MongoDB
- External API:

Cost Analysis

We are doing cost analysis by following factors:


- **Deployment type:** On-premises or cloud-based deployments
- **Company size:** Small / Medium / Large
- **Specific features:** Basic / Standard / Premium
- Number of **devices** in the company

Category:

- **DATA PROTECTION (DP):**
 - Veeam Backup
 - Acronis Backup
 - Bitlocker
 - IBM Guardium
 - Veritas Enterprise vault
- **NETWORK SECURITY:**
 - IBM QRadar Darktrace
 - Zscaler Internet Access
 - iboss Cloud Platform
 - Cisco Umbrella
 - Palo Alto Networks Firewall
- **SECURITY AWARENESS AND TRAINING (SAT)**
 - nowBe4
 - Proofpoint Security Awareness Training
 - SANS Security Awareness
 - Hoxhunt
 - Infosec IQ
- **ACCESS CONTROL AND IDENTITY MANAGEMENT (ACIM)**
 - Okta
 - Azure Active Directory
 - IBM Security Verify
 - SailPoint IdentityIQ
 - CyberArk Identity

Market Readiness / Impact

- **Efficiency Overload:** Existing solutions flood users with an extensive number of questions, causing fatigue and inefficiency.
- **Strategic Optimization:** Our solution revolutionizes the user experience by minimizing the number of questions required for a thorough assessment—only 10% (40-50 questions) needed from a pool of 500+.
- **Automated Testing Synergy:** Uniquely, we combine a user-friendly automated testing tool with our optimized questionnaire, eliminating the time-consuming manual data entry prevalent in the market.

- 
- **Cost Analysis Edge:** Introducing a groundbreaking feature, our tool provides users with approximate budgets, ensuring not just security efficacy but also financial prudence.
 - **Holistic Solution:** The integration of automation, optimized questioning, and cost analysis positions our tool as a comprehensive and efficient choice, filling a gap in the market.

Future Scope

- **Global Benchmark:** Inspired by the USA and UK, our initiative aims to create India's dedicated ransomware assessment tool in collaboration with NTRO.
- **Integrating with the security Ecosystem**
- **Expanding dataset**
- **User Training and Awareness**