Technology Stack (Architecture & Stack)

Date	20 October 2023
Team ID	2.5
Project Name	Malware Detection and Classification
Team Members	• Hardik Kankane • Anondita Dutta • Elisabeth Ann Varghese

Technical Architecture:

The system employs Apache Tomcat as the web server and PostgreSQL as the database for storing training data and model parameters. Machine learning is powered by TensorFlow, utilizing a Convolutional Neural Network (CNN) algorithm with a training dataset of 10,000 malware and benign samples, achieving an accuracy of 98.5%. Data privacy is ensured through compliance with GDPR and CCPA, employing SSL/TLS for data in transit and AES-256 encryption for data at rest. IAM roles in AWS enforce access control. The architecture emphasizes scalability with horizontal scaling through Auto Scaling Groups and is integrated with external services like VirusTotal for enhanced threat intelligence. Disaster recovery is facilitated by daily backups to Amazon S3 with versioning enabled, supported by a Multi-AZ deployment for high availability.

Table-1: Components & Technologies

S.No	Component	Description	Technology
1	Web Server	Responsible for handling HTTP requests and responses.	Apache Tomcat 9.0.0-M1
2	Database	Stores training data and model parameters.	PostgreSQL 13.4
3	Machine Learning Library	Utilized for training and deploying the detection model.	TensorFlow 2.7.0

4	Antivirus Software	Provides additional security measures against malware.	McAfee Endpoint Security
5	Operating System	Provides the underlying platform for the system.	Ubuntu Linux 20.04
6	Programming Languages	Utilized for application development and model training.	Python 3.8, Java 11
7	Cloud Provider	Hosts the system infrastructure and resources.	AWS
8	Storage	Provides data storage for various components.	Amazon S3 (100GB), EBS (500GB)

Table-2: Infrastructure Details

S.No	Virtual Machine Type	CPU (vCPUs)	RAM (GB)	Storage Type	Storage Size (GB)
1	Type A	4	16	EBS	100
2	Type B	8	32	EBS	500
3	Type C	16	64	EBS	500

Table-3: Malware Detection Algorithm Details

S.No	Model Type	Training Data Description	Accuracy
1	Convolutional Neural Network (CNN)	10,000 malware samples, 10,000 benign samples	98.5%

Table-4: Data Privacy and Security Measures

S.No	Compliance	Encryption Details	Access Control
1	GDPR, CCPA	SSL/TLS for data in transit, AES-256 for data at rest	IAM Roles for AWS services

Table-5: Compliance and Industry Standards

S.No	Data Privacy Regulations	Industry Standards

1	Compliant with GDPR, CCPA	ISO/IEC 27001, NIST Cybersecurity Framework

Table-6: Integration Details

S.No	Data Sources	APIs and Webhooks
1	API endpoints, File Uploads	Custom APIs for data ingestion, webhook for real-time alerts

Table-7: Scalability and Future Considerations

S.No	Scaling Strategy	Future Enhancements
1	Horizontal Scaling with Auto Scaling Groups	Incorporate advanced heuristics for behavior-based analysis

Table-8: Dependencies and Third-Party Services

S.No	External Services	APIs and SDKs
1	VirusTotal API for additional threat intelligence	AWS SDK for integration with cloud services

Table-9: Cost Estimation

S.No	Cloud Costs Details	Monthly Cost ($)
1	Compute	10,000
2	Storage	5,000
3	Networking	2,000
4	Licensing and Subscription Fees	3,000
5	Operational Costs	4,200

Table-10: Disaster Recovery and Redundancy

S.No	Backup and Restore Details	Redundancy Measures
1	Daily backups to S3 with versioning enabled	Multi-AZ deployment for high availability

Table-11: Glossary and Abbreviations

S.No	Term	Definition
1	API	Application Programming Interface
2	CNN	Convolutional Neural Network

3	GDPR	General Data Protection Regulation
4	CCPA	California Consumer Privacy Act
5	SSL/TLS	Secure Sockets Layer/Transport Layer Security
6	IAM	Identity and Access Management
7	AWS	Amazon Web Services
8	OS	Operating System
9	VM	Virtual Machine
10	CPU	Central Processing Unit
11	RAM	Random Access Memory
12	S3	Simple Storage Service
13	EBS	Elastic Block Store
14	TLS	Transport Layer Security
15	AES	Advanced Encryption Standard

References:

- AWS Documentation: Link
- PostgreSQL Documentation: Link
- TensorFlow Documentation: Link
- NIST Cybersecurity Framework: Link
- ISO/IEC 27001 Standard: Link
- McAfee Endpoint Security: Link