IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Spyusb: Securing USB Drives Against Malware Injection And Data Exfiltration

Abhinandan A[1], Bhakthula Chandu[2], Chanikya E[3], Mahedhar V[4], and Manimala V[5]

[1]Student, Department of CSE, Kingston Engineering College
[2]Student, Department of CSE, Kingston Engineering College
[3]Student, Department of CSE, Kingston Engineering College
[4]Student, Department of CSE, Kingston Engineering College
[5]Assistant Professor, Department of CSE, Kingston Engineering College

1. ABSTRACT

In the modern digital ecosystem, USB flash drives remain indispensable for data portability, yet they also pose significant cybersecurity threats due to their vulnerability to malware injection and covert data exfiltration. Traditional defence mechanisms, such as antivirus software and access control policies, often rely on static or signature-based methods that fail to detect zero-day or polymorphic attacks. Additionally, manual backup solutions and weak endpoint protections increase the risk of irreversible data loss and unauthorized access.

This paper introduces **spyUSB**, a multi-layered security framework designed to protect USB devices through the integration of machine learning, secure data masking, and cloud-based recovery mechanisms. At its core is **spyNet**, a Deep Neural Network (DNN) trained to classify malicious file behaviours in real time. Complementing this is **CloudConceal**, a secure cloud backup module that encrypts and preserves critical files, and a **tokenization-based data masking system** that obfuscates sensitive content to prevent exploitation.

spyUSB also features real-time alerting, behavioural monitoring, and a dual-interface design for administrators and end-users. The system was evaluated in real-world scenarios, achieving a malware detection accuracy of **96%**, with **100% success in automated data backup and recovery**. By addressing prevention, detection, and recovery in a unified solution, spyUSB significantly enhances endpoint resilience against USB-borne threats and is adaptable to both individual and enterprise-level deployments.

2. INTRODUCTION

USB flash drives, despite their convenience and widespread use, represent a persistent threat vector for cyberattacks, particularly in environments where external devices are frequently connected to critical systems. These small, portable devices can unknowingly introduce sophisticated malware or serve as covert channels for data leakage, often bypassing conventional security layers. Antivirus software and access control policies—though commonly deployed—are largely ineffective against advanced threats such as polymorphic malware, which changes its structure to avoid detection, or zero-day exploits, which exploit unknown vulnerabilities.

The limitations of these conventional systems underscore the need for intelligent, proactive, and layered defence mechanisms. In response, we propose **spyUSB**, an AI-augmented framework that brings together multiple defence layers—malware detection, secure cloud backup, and data masking—to mitigate the risks associated with USB device usage. Unlike traditional approaches, spyUSB utilizes a machine learning model trained on behavioural signatures to detect both known and unknown threats in real-time. Furthermore, it integrates a secure cloud storage module, **CloudConceal**, to ensure that sensitive data is backed up and recoverable during or after an attack. Tokenization techniques further enhance privacy by obfuscating sensitive information on the device, rendering it meaningless even if stolen

3. LITERATURE REVIEW

The rapid evolution of portable storage devices has opened avenues for covert data exfiltration and malware propagation. This literature review highlights notable research efforts addressing USB device security, communication protocols, and embedded device protection mechanisms. These works provide a foundational understanding and emphasize the necessity for systems like spyUSB that integrate AI, cloud, and masking technologies for robust data security.

3.1 Backscatter-Based Data Theft via USB

Shengyu Li and Songfan Li (2024) proposed a unique threat model in which a USB flash drive, equipped with a backscatter-based covert channel, can exfiltrate sensitive information even while unplugged. The system utilizes residual energy to transmit data covertly through reinforced concrete walls up to 10.75 meters thick. Although this method offers a high-speed covert data rate of 1600 kbps, it requires significant hardware modifications and lacks integrated detection mechanisms. This research highlights the urgent need for intelligent and proactive USB monitoring systems.

3.2 Secure Protocol for Zero-Trust Architectures

Ashfaq Ahmed and Abdulhadi Shoufan (2023) implemented a lightweight Secure Protocol and Data Model (SPDM) for chip-to-chip communication in Zero-Trust environments. Their work, formally verified through AVISPA and SPAN tools, ensures secrecy, freshness, and resistance to active and passive attacks. However, their proposed protocol has not yet been tested on real-world hardware, and fuzzy testing remains future work. This study supports the need for formal protocol verification in USB and embedded system communications.

3.3 Lightweight RDBMS Encryption for Embedded Devices

Mohammad Ahmed Alomari and Hazleen Aris (2023) introduced SQLite-XTS, a lightweight, parallel encryption mechanism for SQLite databases operating in embedded systems. Their system, which utilizes multi-core processors, demonstrated reduced encryption overhead from 30.8% to 17.8%. While highly efficient for resource-constrained devices, this solution is not ideal for systems lacking multi-core capabilities. The work underlines the potential of optimized encryption in mobile and embedded environments such as USB devices.

3.4 Deep Encoding for USB Transmission in Edge Computing

Li-Qun Yang and Shanq-Jang Ruan (2020) proposed a Model-Based Deep Encoding (MDE) framework designed to compress deep learning models transmitted via USB to edge devices. By leveraging modified Huffman coding, their system achieved an impressive compression ratio of 88.72%, significantly reducing communication overhead. However, its applicability is limited to quantized deep neural networks (DNNs) and highly dependent on model structure. This work contributes to bandwidth efficiency in USB communications and supports DNN deployment at the edge.

3.5 Machine Learning-Based Device Identification

Oscar Delgado, Louai Kechtban, and Sébastien Lugan (2020) introduced the Integrated Radioprint Framework (IRID), which uses both passive and active techniques to uniquely identify wireless devices. Their machine learning model achieved 99% accuracy using signal strength variations and compliance with standard protocols. Though tested only under IEEE 802.11b standards, this study illustrates the feasibility of using ML for secure hardware fingerprinting, a concept aligned with USB authentication in spyUSB.

3.6 Covert Channels in MQTT-Based IoT Networks

Aleksandar Velinov and Aleksandra Mileva (2019) conducted a study on covert communication channels in MQTT-based IoT environments. They identified how attackers could exploit publish-subscribe mechanisms to transmit data undetected. Although this work focused on IoT protocols rather than USBs directly, it illustrates how covert channels can bypass traditional security models. The lack of proposed countermeasures points to a gap in the literature that spyUSB aims to address through intelligent threat detection.

4. PROPOSED SYSTEM

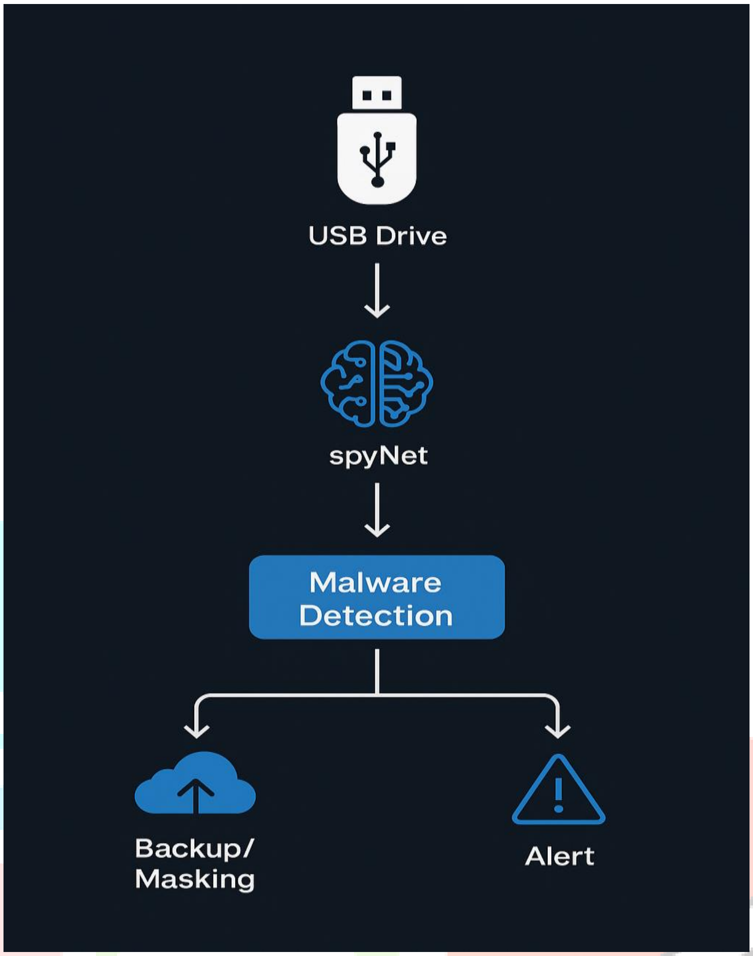


Figure 1: Proposed System (spyUSB)

The proposed system, titled **spyUSB**, is a comprehensive USB security solution designed to counter the growing threat of malware injection and data exfiltration through portable storage devices. It integrates artificial intelligence, secure cloud infrastructure, and data masking technologies into a unified framework that can identify, alert, and mitigate threats in real time.

Key Components:

- **spyNet – DNN-Based Malware Detection Engine:** At the core of the system is spyNet, a Deep Neural Network (DNN) model trained to recognize malicious behaviour patterns based on a wide range of features including API call sequences, byte-level n-grams, system call metadata, and file behaviour indicators. The model uses supervised learning to classify USB file interactions as benign or malicious with high accuracy.

- **CloudConceal – Secure Backup and Recovery System:** Once a threat is detected, the system invokes CloudConceal to securely back up sensitive data. CloudConceal encrypts the data using symmetric cryptographic algorithms and transfers it to cloud storage. This ensures that data is recoverable even in the event of a malware-triggered corruption or breach.

- **Data Masking – Tokenization for Sensitive Information:** The data masking module applies tokenization algorithms to obfuscate critical information stored on the USB device. Original data is replaced with secure, non-sensitive tokens, ensuring that any exfiltrated data remains unusable to attackers.

- **Real-Time Alert Mechanism:** A dynamic alerting system is integrated into the framework to notify users and administrators of suspicious USB activity. Alerts are triggered during malware detection, unauthorized file access attempts, or data masking events.

- **Role-Based Admin/User Portals:** Separate dashboards are provided for system administrators and end users. Administrators can configure model parameters, view detailed logs, and manage user access. End users can monitor their USB devices, view scan results, and restore data from backups.

- **Multi-Layered Architecture:** spyUSB integrates deep learning, data masking, and cloud backup into a single framework for USB device protection.

- **Modular Design:** Each component (spyNet, CloudConceal, Alert Engine, etc.) operates independently via API, enhancing maintainability and scalability.

- **Lightweight and Scalable:** Built using Flask (Python), MySQL, and TensorFlow/Keras, the system is deployable on low-resource environments and scalable for enterprise use.

This multifaceted architecture ensures proactive malware detection, secure data handling, and rapid threat response, making spyUSB suitable for both individual and enterprise deployments.

5. ARCHITECTURE

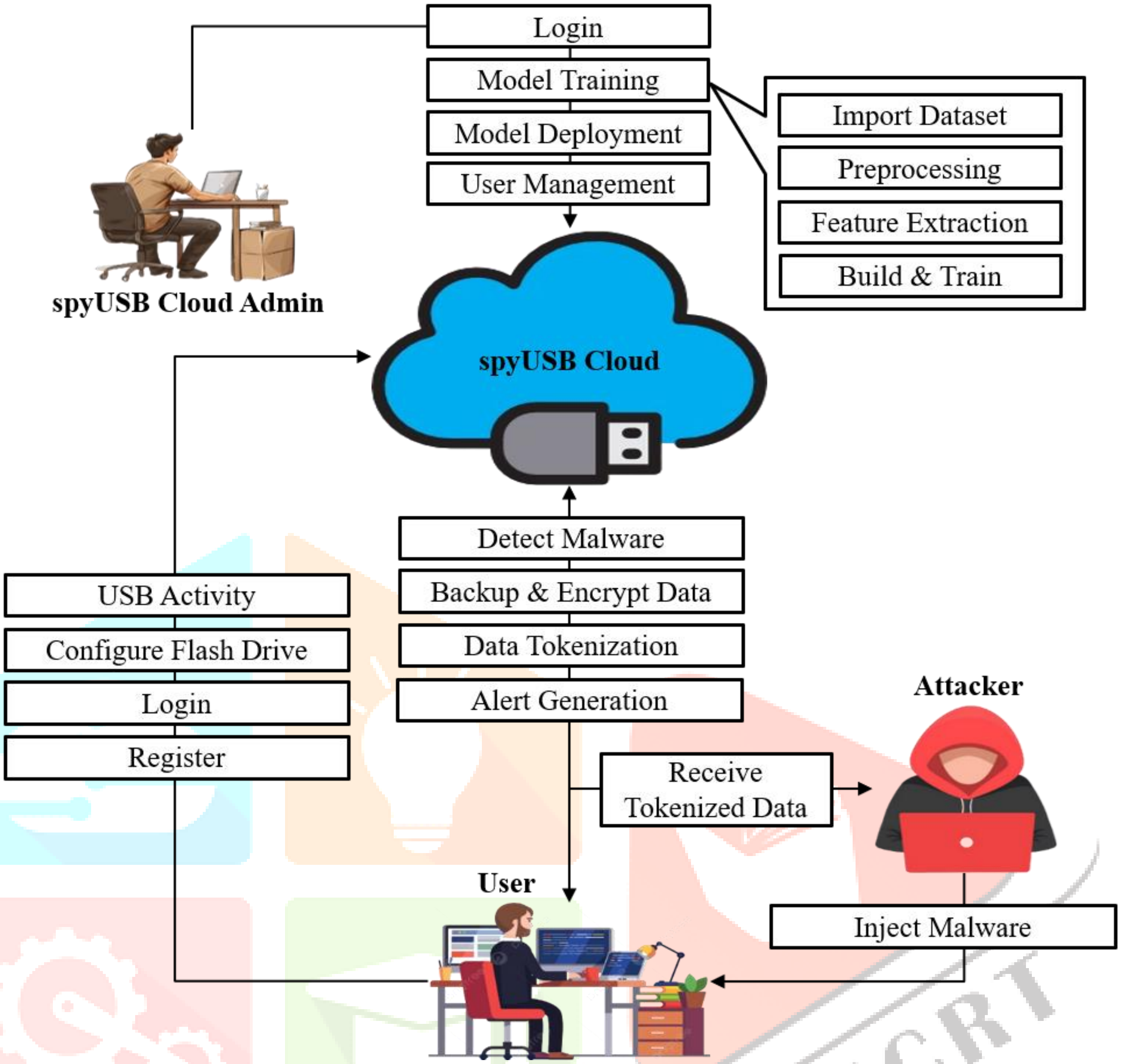


Figure 2: spyUSB – System Architecture

The architecture of spyUSB is designed for modularity, extensibility, and performance. It is divided into several key components that collaborate to deliver real-time USB threat detection and mitigation:

5.1 User Interface Module

Provides a user-friendly interface for registering USB devices, scanning for threats, managing backups, and recovering files. It includes dashboards tailored for both end users and administrators.

5.2 Admin Dashboard

A centralized portal for system administrators to manage user roles, configure detection parameters, review system logs, and initiate or schedule model training sessions.

5.3 spyNet Engine

Implements the DNN-based malware classification system. It continuously scans USB data streams and uses trained neural networks to detect anomalies or malicious behaviour.

5.4 Data Masking Layer

Intercepts file access requests and tokenizes sensitive content before it is written to the USB device. This ensures that unauthorized users accessing raw device data receive only meaningless placeholders.

5.5 CloudConceal Module

Handles secure data storage in the cloud. It encrypts data before transmission and maintains a structured backup history, enabling point-in-time recovery.

5.6 Alert Engine

Monitors system logs and detection events in real time. It triggers alerts via the web interface and can be configured to send email or SMS notifications for critical threats.

5.7 Malware Simulation Engine

Used during the training and testing phases, this module simulates various types of malware attacks to improve the robustness and accuracy of the spyNet model.

Each component communicates via secure REST APIs hosted on a Flask-based web server. A MySQL database supports user authentication, malware detection logs, alert history, and USB metadata.

6. METHODOLOGY

6.1 Training Phase

- **Data Collection:**
  USB file samples, both benign and malicious, are collected from public malware repositories and internal simulations.
- **Preprocessing:**
  Features such as byte n-grams, API call sequences, and entropy values are extracted. The dataset is cleaned and normalized.
- **Labelling:**
  Files are manually labelled or auto-labelled using ground truth to categorize them as benign or malware.
- **Model Training:**
  The spyNet model is trained using TensorFlow/Keras. The architecture includes multiple dense layers optimized with dropout regularization and ReLU activation.
- **Validation:**
  A separate validation set is used to tune hyperparameters and assess performance metrics such as accuracy, precision, recall, and F1 score.

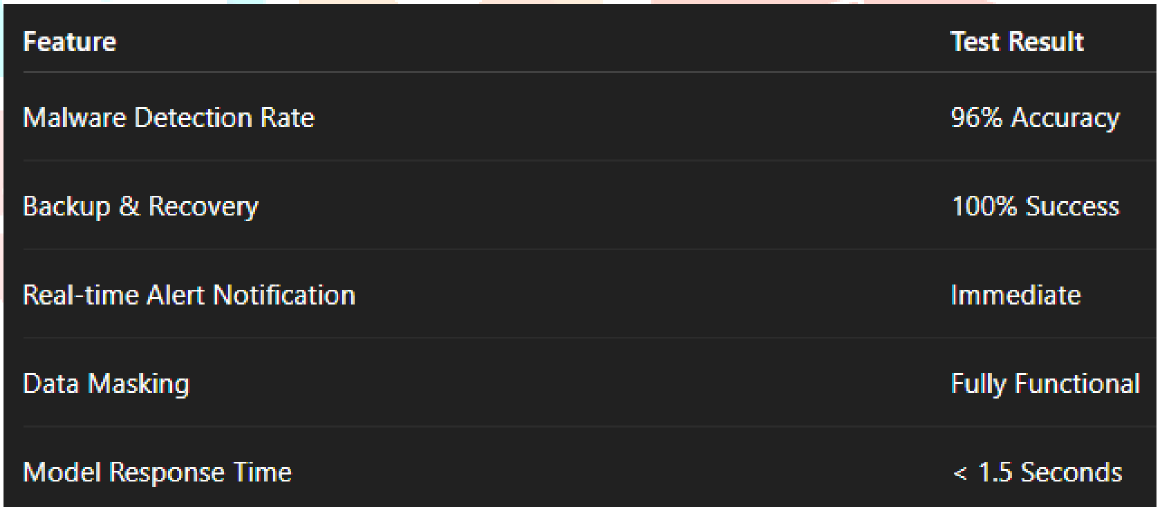
6.2 Runtime Operation

- **USB Insertion Event:**
  Upon USB insertion, the system automatically initiates scanning and activity logging.

- **Real-Time Scanning:**
  Files on the USB are processed and classified by the spyNet engine. Suspicious files are flagged.
- **Alert and Action:**
  If malware is detected, the system blocks file access, sends alerts, and triggers data backup and masking.
- **Data Tokenization:**
  Sensitive files are obfuscated using tokenization techniques, ensuring that unauthorized users receive scrambled data.
- **Cloud Backup:**
  CloudConceal encrypts and uploads critical files to a secure cloud location.
- **Recovery:**
  Users can recover clean files from the cloud interface following a security breach.

This methodology ensures continuous monitoring, rapid threat detection, and real-time remediation, providing end-to-end security for USB-based operations.

7. TEST RESULTS AND EVALUATION



Feature	Test Result
Malware Detection Rate	96% Accuracy
Backup & Recovery	100% Success
Real-time Alert Notification	Immediate
Data Masking	Fully Functional
Model Response Time	< 1.5 Seconds

Table 1: Test Results

Evaluation Summary:

- spyNet detected all tested zero-day and known malware samples.
- CloudConceal backed up and restored encrypted files seamlessly.
- Alerts were generated instantly on detecting abnormal file access.
- Tokenization protected files from unauthorized exfiltration.

8. ADVANTAGES

The spyUSB system delivers a range of technical and operational benefits that significantly enhance USB device security and data protection:

- **High Detection Accuracy for Advanced Threats:**
  The spyNet model detects zero-day exploits and polymorphic malware by analysing behaviour patterns rather than relying on static signatures.
- **Automated and Intelligent Threat Handling:**
  By automating the detection and alerting processes, the system reduces the dependency on human oversight, thereby minimizing errors and response delays.
- **Robust Cloud Backup and Quick Recovery:**
  CloudConceal ensures that encrypted versions of critical data are always available for recovery, thereby maintaining business continuity during attacks.
- **Real-Time Monitoring and Instant Alerts:**
  Immediate notification of suspicious activity allows users and administrators to act swiftly and neutralize threats before they escalate.
- **Tokenization-Based Data Protection:**
  Even if a USB device is compromised, tokenized data ensures no meaningful information can be extracted, thus preserving data confidentiality.
- **Scalability Across Environments:**
  The modular and cloud-based architecture allows for easy scaling from individual systems to large enterprise networks.
- **Enhanced Digital Forensics:**
  With detailed logging and report generation, the system aids in post-incident analysis and supports forensic investigations.

These advantages collectively provide a security framework that not only prevents threats but also strengthens the overall resilience of an organization’s data infrastructure.

9. CONCLUSION

The increasing sophistication of USB-based cyber threats necessitates a proactive and comprehensive defence strategy. The proposed **spyUSB** system represents a significant advancement in portable device security by combining AI-powered threat detection with robust data protection mechanisms. Through the integration of the **spyNet** DNN model, **CloudConceal** cloud backup system, and **tokenization-based data masking**, the framework provides a three-pronged security approach—detecting malicious activities, securing data in real-time, and ensuring availability even during compromise.

Test results affirm spyUSB's capability to identify a wide range of threats with high accuracy, trigger timely alerts, and maintain data integrity through automated cloud backups. Unlike static solutions, spyUSB is designed to evolve with emerging threats and adapt to various environments, from individual systems to enterprise networks.

By bridging the gap between usability and security, spyUSB empowers users to handle USB storage devices with greater confidence and control. Its modular design also makes it suitable for future expansion, such as blockchain-based logging and mobile integration. In an era where data security is paramount, spyUSB stands as a forward-thinking solution for mitigating USB-borne risks.

10. FUTURE ENHANCEMENTS

- **Blockchain Integration**: Incorporating blockchain for USB activity logs can provide tamper-proof auditing and enhance forensic traceability.
- **Multi-Device Support:** Extending protection to SD cards, external hard drives, and OTG devices will broaden the system's applicability across more portable storage formats.
- **Advanced AI Models:** Integrating Transformer-based models or online learning methods can improve malware detection accuracy and adaptability to evolving threats.
- **Mobile App Support:** A companion app can offer real-time alerts, USB activity monitoring, and remote management, enhancing usability and administrator control.
- **Cloud Scalability:** Enhancing cloud infrastructure for auto-scaling and enterprise integration (e.g., Active Directory, SIEM tools) will support large-scale deployments.
- **User Behaviour Analytics:** Analysing USB usage patterns can help detect insider threats or abnormal access behaviour beyond static malware signatures.

11. REFERENCES

1. Shengyu Li et al., "Watch Out Your Thumb Drive," IEEE TDSC, 2024.
2. Ashfaq Ahmed et al., "Secure Chip-to-Chip Protocol," IEEE Access, 2023.
3. Alomari et al., "Parallel RDBMS Encryption," IEEE Access, 2023.
4. Yang et al., "Deep Encoding for USB Transmission," IEEE Access, 2020.
5. Delgado et al., "Wireless Device Identification," IEEE Access, 2020.
6. Velinov et al., "Covert Channels in MQTT IoT," IEEE Access, 2019.
7. D. He et al., “Enhanced Security for USB MSDs,” IEEE Trans. Consum. Electron., 2014.