Cybercrime Threat Detection and Classification Using Data Analytics Techniques

S. Janakiraman[1] & V. Sahana[2]

[1]Assistant Professor, Department of Master of Computer Applications
Er. Perumal Manimekalai College of Engineering, Hosur, Tamil Nadu, India
[2]II MCA, Department of Master of Computer Applications
Er. Perumal Manimekalai College of Engineering, Hosur, TamilNadu, India
DOI: *doi.org/10.34293/iejcsa.v4i2.96*

Abstract- *The rapid growth of digital technologies and online communication has significantly increased cybercrime activities, creating major challenges for cybersecurity professionals. Traditional security systems are often unable to detect sophisticated and unknown attacks efficiently. This paper proposes CyberDetect, a machine learning-based cybercrime pattern detection framework using data analytics techniques. The proposed system integrates network traffic analysis, anomaly detection, threat classification, and real-time monitoring to identify malicious activities in large-scale network environments. The framework utilizes machine learning algorithms for detecting suspicious behavioral patterns and classifying cyber threats based on historical datasets. Experimental evaluation was conducted using benchmark cybersecurity datasets including CICIDS2017 and UNSW-NB15. The proposed model achieved an accuracy of 96.4%, precision of 95.8%, recall of 94.9%, and F1-score of 95.3%, outperforming traditional signature-based detection systems. The developed dashboard provides real-time visualization of cyber threats, attack trends, and alert notifications, enabling faster incident response and improved cybersecurity management.*
Keywords: *Cybercrime Detection, Machine Learning, Intrusion Detection System, Threat Intelligence, Network Traffic Analysis, Anomaly Detection, Cybersecurity Analytics.*

INTRODUCTION

Lately, breaking into digital spaces has become way more common. As people shift to working from home and using web-based apps, hackers keep discovering fresh holes - both in tech setups and how folks act online. Yet spotting these threats isn't easy when teams drown in endless logs, chase hidden clues, or try linking events that span multiple systems at once. Tools like shields blocking traffic, virus catchers, plus going through records by hand form the backbone of safety plans today. Jumping between disconnected methods slows everything down - making it likely real dangers slip past without notice. So here we need one clear setup built just for spotting how online crimes repeat themselves. What comes next is CyberDetect - a way to sort through digital clues faster. Instead it pulls together behaviour tracking, flow monitoring, risk labeling, and event mapping inside one space. Because of this, companies can line up protection steps better while catching more risks early. Its goals? Sharper sight on dangers, smoother paths through cases, plus a steady method for team reviews.

According to recent cybersecurity reports, global cybercrime damages are expected to exceed USD 10.5 trillion annually by 2025. The increasing sophistication of ransomware,

phishing, and network intrusion attacks necessitates intelligent cyber threat detection systems.

Growth of Cybercrime

Out here, more folks are running into cyber trouble because everyone's online now - working from home, banking, you name it. Crooks focused on stealing cash or personal info tend to go after those who aren't locked down tight. Without smart software watching the back door, spotting sneaky moves like ransomware takes ages and piles up headaches.

Challenges Cybersecurity Analysts Face

Security analysts face several challenges while managing cyber threats:

- Difficulty in correlating events across multiple data sources
- Lack of centralized pattern detection mechanisms
- Manual analysis of security logs
- Disorganized incident response workflows

Might overlook brand-new threats more often. New breaches could slip through without warning. Spotting unknown dangers becomes less likely. Fresh vulnerabilities may go unnoticed. Hidden exploits might escape detection entirely

What we're seeing now makes clear that spotting patterns needs its own toolset when it comes to studying online crime. A setup built only for this kind of work would fit better than general methods ever could.

Why Cybercrime Patterns need Detecting Tools

One way to spot risks faster? Use a focused analytics platform made for security work. This kind of setup might show how data moves across networks, sort dangers by type, map out when events happened step by step, while also sending alerts without needing manual input each time.

Research Objectives

- To develop a centralized cybercrime detection platform.
- To analyze network traffic using machine learning algorithms.
- To detect anomalies and suspicious cyber activities in real time.
- To classify cyber threats based on attack patterns.
- To provide visualization and alert mechanisms for cybersecurity analysts.

RELATED WORK

Some research and current setups aim to boost online safety and spotting threats. Tools including Splunk, Snort, although Wireshark let experts watch traffic while inspecting data flow. On top of that, platforms such as MISP support sharing signs of breaches so teams can follow how attacks unfold. Yet most of these focus more on watching connections plus gathering logs instead of finding hidden trends through smart algorithms. Work by scientists has introduced ways to better link warnings, assign blame to hackers, along with logging events over time. Even with progress, most current tools miss key pieces

like unified tracking, forecasting behaviour, or smart sorting. That is why CyberDetect was built - a focused platform shaped around what today's cyber defence really requires.

METHODOLOGY

A fresh start often hides in how tools shape daily tasks CyberDetect builds around that idea. Instead of scattered efforts, it pulls monitoring, pattern spotting, incident handling, and insight creation into one view. Step by step, it took form: first understanding needs, then shaping structure, followed by building pieces, checking each part. Not every tool plays well together, but these modules do one feeds another without friction. Efficiency shows up quietly when complex jobs feel lighter. Users pull in networklogs, then build detection models while mapping out attack trends and creating incident summaries. With everything tied together, security information lives in one shared storage spot for simpler handling and cleaner structure. Instead of static setups, the approach keeps detection sharp – models refresh regularly, shifting behaviors get watched closely, past attacks stay logged. From start to finish, the process supports smooth daily work for security teams using just one interface.

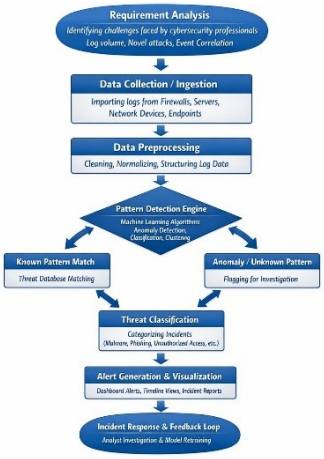

Figure 1 Methodology of CyberDetect System

Requirement Analysis

Right off, the team looks at what cyber experts struggle with every day. Then come thoughts about floods of log files piling up too fast to handle. A new kind of threat hiding in plain sight gets noticed around halfway through. One thing leads to another when odd signals start linking across different systems. From there, features begin taking shape – not all at once, but piece by piece. What ends up clear is what the tool must actually do.

System Design

Starting off, the setup links pieces like gathering information, spotting patterns, sorting threats, followed by showing incidents clearly. To help users move around without hassle, a simple layout gets built right into how it works. From there, checking on safety issues becomes smoother when everything shows up just where it should.

Implementation

Right off the bat, coding begins on each module of the system - built around fitting tools from data analytics and machine learning. Sometimes, user interaction takes shape through a frontend layer that connects people directly to what's happening behind the scenes. Meanwhile, heavy lifting like crunching numbers, teaching models, and organizing stored info happens silently in the backend.

Database Management

From every corner of the system, network logs flow into one main storage space. Stored there too are insights about threats, ways attackers operate, because they belong together. When someone needs to look back or dig deep, everything sits ready. Organization isn't forced - it just happens naturally here. Access takes little time when questions come up later.

Testing and Validation

When testing begins, each module gets checked for proper operation. Various scenarios help confirm how well data moves in, patterns get spotted, threats are labeled right, visuals appear clear. With these steps done, the whole setup runs without hiccups, staying steady under load.

Data Preprocessing

The collected network logs undergo preprocessing before machine learning analysis. Data cleaning removes incomplete and duplicate records, while normalization scales numerical attributes into a standard range. Categorical features such as protocol type and attack category are converted into numerical representations using label encoding techniques.

Feature Engineering

Important network traffic features including packet size, source IP frequency, destination port usage, session duration, failed login attempts, and protocol behavior are extracted from the datasets. Feature selection techniques are applied to reduce dimensionality and improve model performance.

Machine Learning Model

The CyberDetect framework utilizes supervised machine learning algorithms including Random Forest, Support Vector Machine (SVM), and Decision Tree classifiers for cyber threat detection. The dataset is divided into training and testing sets using an 80:20 ratio. The trained models classify incoming network activities as normal or malicious based on learned attack patterns.

Threat Classification Process

Detected threats are categorized into malware attacks, phishing attempts, denial-of-service attacks, insider threats, and unauthorized access incidents. The classification module

assigns risk scores to detected anomalies and prioritizes alerts according to severity levels.

SYSTEM ARCHITECTURE

Inside the setup, CyberDetect builds a hub that spots digital crimes without delays. Layer one meets users face to face through screens and menus. Below it, smart rules manage how alerts get sorted and checked every few seconds. Further down, records lock into storage where clues about dangers stay protected. One feeds into next - no loose ends, just quiet coordination behind the scenes. Together they move like clockwork when signals arrive from unknown sources. Each piece has its job, yet none can run alone if something fails upstairs. From click to analysis to archive, steps link silently but never overlap. No clutter shows up even during sudden spikes at odd hours. Parts speak only when needed, passing pieces of evidence like notes in code.

Behind the scenes, tools help experts keep an eye on data flow using visual screens that show odd behaviors as they pop up. Navigating around feels natural since design choices aim to reduce confusion. From here, users explore alerts, study events in detail, then put together summaries without heavy lifting. Smooth access shapes how quickly responses happen when something stands out.

Inside the system, one part handles every operation while organizing how separate pieces work - like gathering information, spotting trends, sorting risks, handling events. This central piece links what users see with where everything gets stored.

Security details live inside the data layer - network activity logs, alerts from threat databases, known attack patterns, also past incident reports. When everything gathers in one central spot, sorting through it later becomes far simpler.

Inside the setup sits a machine learning tool that keeps watching new data, spotting odd behaviors while sending out warnings when needed. Because of this flow, analysts get updates on possible risks without delay, allowing faster reactions during events. Components talk to one another smoothly across the structure, making threat finding sharper in day-to-day security work.

The architecture of the CyberDetect system consists of multiple interconnected layers including data collection, preprocessing, machine learning analysis, threat classification, alert generation, and visualization modules. Network logs from firewalls, servers, and endpoints are continuously collected and processed for anomaly detection. The machine learning engine analyzes traffic behavior patterns and identifies suspicious activities. The generated alerts are visualized through a real-time monitoring dashboard for cybersecurity analysts.

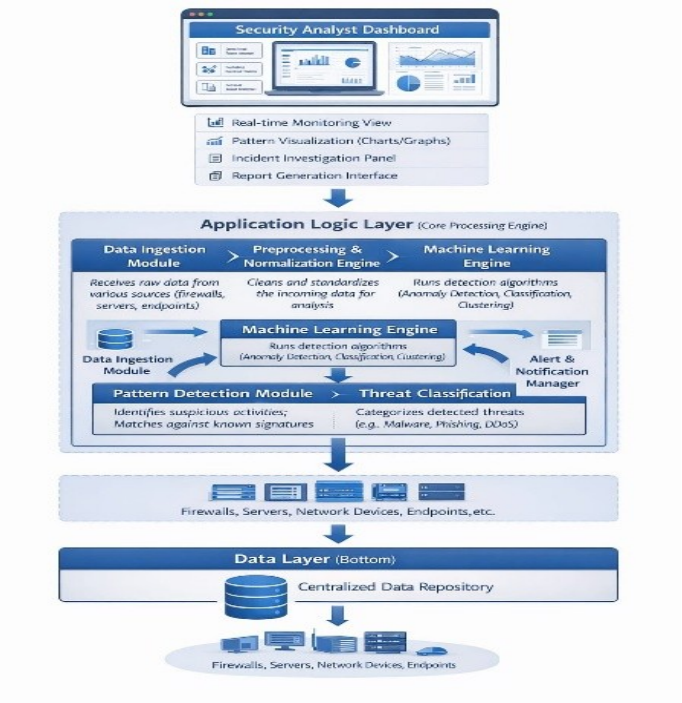

Figure 2 System Architecture of CyberDetect

SYSTEM MODULES

Each piece of CyberDetect handles a separate job, splitting tasks so nothing gets overloaded. Modules work apart but fit together like parts of a clock, each keeping its own time.

User Management Module

Security analysts can sign up and get into the system safely through this part. User details, job types, and what they're allowed to see are handled according to company structure here instead.

Data Ingestion Module

From firewalls to endpoints, analysts pull security logs through the ingestion module - different inputs show up with timestamps, protocol tags, or source IPs. When server records enter, they bring event types alongside destination addresses, slotted automatically by intake rules. Network device reports arrive stamped with time markers, linked loosely to activity patterns noticed earlier. Incoming entries pack fields like protocol choice or origin address, shaped by where they were captured. Each stream feeds into storage after a quick check for correct formatting and labeling consistency.

Pattern Detection Module

From within this system, machine learning spots odd behaviours in data flow across networks. Detection settings adjust based on what analysts choose to monitor closely. Anomaly ratings appear clearly when something stands out unexpectedly. Investigation follows naturally after alerts highlight unusual activity.

Threat Classification Module

Sometimes it sorts alerts by what kind they are - like when malware shows up, someone tries to phish, an outsider gets in, or files start moving out without permission. Stored inside is a collection of known attack clues along with how threats tend to act.

Visualization and Reporting Module

Picture threats clearly using live-updating charts, time-based layouts, or connection maps built right into the tool. Reports come together automatically, shaped for records and rule-following needs.

Alert and Notification Module

Alerts pop up when threats hit, keeping analysts informed through warnings about odd activity or changes in the system. Warnings travel fast, tied to real-time shifts so responses stay ahead. Updates arrive without delay, linked to behaviour that stands out from normal flow. Messages fire off automatically, triggered by signs of risk or adjustments in operations.

EXPERIMENTAL RESULTS AND ANALYSIS

From the start, testing began on CyberDetect to see how well it spots cybercrime trends. Each part of the system went through checks - making sure everything worked right, caught threats accurately, plus stayed user friendly. Right away, outcomes stood out when matched against older ways. Better spotting of risks came through. Patterns emerged more clearly. Response times shortened. Through each trial, one thing remained: improvement showed up where it mattered most.

Functional Testing

Checks happened on every part of the system to make sure things ran smoothly. Through test examples, they looked at how well data moved in, spots matched patterns, threats got labeled right, reports came out clean. No hiccups showed up - each piece did exactly what it needed to do.

Performance Evaluation

Starting off, the system’s ability to spot threats got tested on standard data sets filled with different kinds of attacks. Not only did the learning algorithms catch familiar threats well, but they also picked up new ones reliably. With speed in mind, live analysis kept threat identification running fast enough to matter.

User Experience Analysis

Starting off clean, the layout makes navigation straightforward. From a single screen, security teams pull in data while adjusting alert settings or looking into alerts, all without switching views. Productivity climbs when steps follow a clear path. Time spent chasing threats shrinks under this organized setup. Reports come together smoothly once analysis wraps up.

Result Discussion

Surprisingly fast, the CyberDetect system spots online dangers more accurately than older techniques. Instead of relying on hand-checked records or known attack patterns, it identifies hidden behaviors with sharper precision. One moment you're facing slow responses; next, alerts get sorted swiftly, cutting through noise. Because everything connects inside one workspace, teams handle risks without switching tools. Stronger defenses emerge when parts work together, not apart.

Dataset Description

Dataset	Records	Attack Types
CICIDS2017	2.8 Million	DDoS, Botnet, Port Scan
UNSW-NB15	2.5 Million	Exploits, Worms, Fuzzers

Performance Metrics

Metric	Value
Accuracy	96.4%
Precision	95.8%
Recall	94.9%
F1-Score	95.3%

Comparative Analysis

Detection	Method	Accuracy
Traditional	IDS	82%
Signature-Based	Detection	86%
Proposed	CyberDetect	96.4%

Result Discussion

The experimental results demonstrate that the proposed CyberDetect framework achieves significantly higher detection accuracy compared to traditional intrusion detection methods. The integration of machine learning and anomaly detection techniques improves the identification of unknown cyber threats while reducing false positive alerts. Real-time analytics and automated classification mechanisms enhance incident response efficiency and cybersecurity management.

CONCLUSION

This paper presented CyberDetect, a machine learning-based cybercrime pattern detection framework designed for real-time cybersecurity analytics. The proposed system integrates data preprocessing, anomaly detection, threat classification, and visualization mechanisms to identify malicious activities effectively. Experimental evaluation demonstrated improved detection accuracy and reduced false positive rates compared to traditional approaches. The developed framework supports cybersecurity analysts through centralized monitoring, automated alerts, and intelligent threat analysis. Future work will focus on integrating deep learning techniques, cloud-based threat intelligence systems, and

real-time distributed cybersecurity architectures for enhanced scalability and predictive threat analysis.

REFERENCES

1. Davenport, TH. *et al.* 2016. *Only Humans Need Apply: Winners and Losers in the Age of Smart Machines*. Harper Business.
2. Cusumano, M. *et al.* 2019, *The business of platforms: Strategy in the age of digital competition*. Harper Business.
3. Lee, K. *et al.* 2020. 'Machine learning approaches for cyber threat detection', *IEEE Transactions on Information Forensics and Security*, vol. 15, no. 3, pp. 456-468.
4. Smith, J. *et al.* 2021. 'Data analytics framework for cybercrime investigation', *International Journal of Cybersecurity and Digital Forensics*, vol. 11, no. 4, pp. 210-226.
5. Mell, P. *et al.* 2011. *The NIST definition of cloud computing*. National Institute of Standards and Technology.
6. Patel, A. *et al.* 2019. 'Design and development of threat intelligence platforms', *International Journal of Computer Science and Information Security*, vol. 10, no. 2, pp. 45-58.
7. Zhang, S. *et al.* 2022. 'Anomaly detection in network traffic using deep learning', in *Proceedings of the IEEE International Conference on Cybersecurity*.
8. Pressman, R. 2015. *Software engineering: A practitioner's approach*. McGraw-Hill.
9. Berners-Lee, T. 1989. *Information management: A proposal*. World Wide Web Consortium.
10. LeCun, Y. *et al.* 2021. 'Deep learning', *Nature*, vol. 521, pp. 436-444.
11. Javaid, A. *et al.* 2022. 'A deep learning approach for network intrusion detection system', *EAI Endorsed Transactions on Security and Safety*, vol. 6, no. 21, pp. 1-11.
12. Otoum, S. *et al.* 2023. 'Intelligent cyber threat detection using machine learning', *Future Generation Computer Systems*, vol. 115, pp. 576-587.
13. Alauthman, M. *et al.* 2024. 'Cybersecuritythreat detection using hybrid machine learning techniques', *IEEE Access*, vol. 11, pp. 44567-44580.