Models Used in Cyber Attack Detection

Cyber attack detection is a critical component of cybersecurity, aiming to identify malicious activities such as intrusions, malware, phishing, denial-of-service (DoS) attacks, insider threats, and advanced persistent threats (APTs). Machine Learning (ML) and Deep Learning (DL) models are widely used to analyze network traffic, system logs, user behavior, and application data to detect anomalies and known attack patterns.

Below is a detailed research-oriented explanation of **10+ widely used models** in Cyber attack detection, covering their **purpose, working principle, accuracy, use cases, contribution to threat detection, and real-world applications**.

1. Logistic Regression (LR)

Purpose

Logistic Regression is a supervised learning algorithm primarily used for **binary classification**, such as distinguishing between normal and malicious traffic.

How It Works

It models the probability of an event using a logistic (sigmoid) function. Features such as packet size, protocol type, connection duration, and failed login attempts are used as inputs.

Accuracy

- Typical accuracy: **80–90%** on benchmark datasets like NSL-KDD
- High precision for well-defined attack patterns

How It Helps in Cyber Threat Detection

- Identifies known attack signatures
- Works well for linear relationships

Uses

- Intrusion Detection Systems (IDS)
- Spam and phishing detection

Real-World Applications

- Used in **SIEM tools** for baseline threat classification
- Email spam filtering systems

2. Decision Tree (DT)

Purpose - Decision Trees classify traffic by learning decision rules from data.

How It Works - The model splits data based on feature thresholds (e.g., number of packets, port number) to classify attacks.

Accuracy - Accuracy range: **85–92%**

How It Helps

- Easy to interpret attack rules
- Detects misuse-based attacks

Uses

- Network intrusion detection
- Policy-based security systems

Real-World Applications

- Rule engines in firewalls
- Security analytics platforms

3. Random Forest (RF)

Purpose - Random Forest is an ensemble model that improves detection accuracy by combining multiple decision trees.

How It Works - Each tree is trained on a random subset of features and data; final prediction is based on majority voting.

Accuracy - Accuracy: **92–97%** on datasets like CICIDS2017

How It Helps

- Reduces false positives
- Handles high-dimensional cybersecurity data

Uses

- Malware detection
- Intrusion detection

Real-World Applications

- IBM QRadar
- Splunk security analytics

4. Support Vector Machine (SVM)

Purpose - SVM is used for both linear and non-linear attack classification.

How It Works - It finds an optimal hyperplane separating benign and malicious activities using kernel functions.

Accuracy - Accuracy: **88–95%**

How It Helps

- Effective for small-to-medium datasets
- Robust to overfitting

Uses

- DDoS detection
- Network traffic classification

Real-World Applications

- Academic IDS prototypes
- Network security appliances

5. k-Nearest Neighbors (KNN)

Purpose - KNN is a distance-based algorithm for classifying unknown traffic based on similarity.

How It Works - It compares a data point with its nearest neighbors and assigns the most common class.

Accuracy - Accuracy: **85–93%**

How It Helps

- Detects anomalous behavior
- Useful for unknown attack patterns

Uses - Anomaly-based IDS

Real-World Applications

- Experimental IDS systems
- Research-based cyber labs

6. Naïve Bayes (NB)

Purpose - Naïve Bayes is a probabilistic classifier used in threat detection.

How It Works - Uses Bayes' theorem assuming independence between features.

Accuracy - Accuracy: **75–88%**

How It Helps

- Fast detection
- Works well with text-based threats

Uses - Phishing and spam detection

Real-World Applications- Email security gateways

7. Artificial Neural Network (ANN)
Purpose - ANN models complex non-linear attack patterns.
How It Works - Uses interconnected neurons to learn representations of malicious behavior.
Accuracy - Accuracy: **90–96%**
How It Helps - Learns complex attack signatures
Uses

- Intrusion detection
- Malware classification

Real-World Applications - Enterprise security platforms

8. Convolutional Neural Network (CNN)
Purpose - CNNs are used for spatial feature extraction in cybersecurity data.
How It Works - Network traffic is transformed into matrices/images for convolution operations.
Accuracy - Accuracy: **94–98%**
How It Helps - Detects sophisticated attack patterns
Uses

- Malware image classification
- Network intrusion detection

Real-World Applications - Advanced malware detection tools.

9. Recurrent Neural Network (RNN) / LSTM
Purpose - Designed to analyze **sequential attack behavior**.
How It Works - Processes time-series data such as login attempts and traffic flows.
Accuracy - Accuracy: **93–98%**
How It Helps - Detects slow and stealthy attacks (APTs)
Uses - Insider threat detection
Real-World Applications - User Behavior Analytics (UBA)

10. Autoencoders
Purpose - Used for unsupervised anomaly detection.
How It Works - Learns normal behavior and flags deviations as attacks.
Accuracy - Detection rate: **90–97%**
How It Helps - Detects zero-day attacks

Uses - Network anomaly detection
Real-World Applications - Cloud security monitoring

11. Gradient Boosting (XGBoost / LightGBM)

Purpose - Boosted models improve detection performance.
How It Works - Sequentially trains weak learners to correct errors.
Accuracy - Accuracy: **95–99%**
How It Helps – High precision and recall
Uses - IDS and threat intelligence systems
Real-World Applications - Financial and enterprise SOCs

Conclusion

Cyber attack detection relies on a combination of **traditional ML models, deep learning architectures, and ensemble methods**. While classical models offer interpretability and speed, deep learning models provide superior accuracy and adaptability to modern cyber threats. In real-world deployments, **hybrid models** are often used to balance accuracy, explainability, and scalability.