Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 05/10/2026 **Time:** 11:45 AM - 01:15 PM

Course Code: CBC3407 **Course Name:** Machine Learning for Cyber Threat Detection

Semester: Seventh Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	44	46

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Cybersecurity is the practice of protecting computers, networks, applications, and data from unauthorized access, cyberattacks, damage, or theft. It aims to ensure confidentiality, integrity, and availability of informatio	2	CO1	BT1
2	A cyber threat is a potential event or activity that can exploit a vulnerability and harm a computer system or its data. Example: A phishing attack that steals user credentials.	2	CO1	BT1
3	Machine learning-based threat modeling is the use of ML techniques to analyze system, network, or security data to identify potential threats, suspicious patterns, and attack risks.	2	CO2	BT2
4	Supervised learning trains a model using labeled examples. Example: Training a classifier using labeled network traffic to identify normal and malicious connections.	2	CO2	BT1
5	Classification: Predicts a category or class, such as normal or attack.Regression: Predicts a continuous numerical value, such as expected network traffic volume.	2	CO2	BT2

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	The cyber threat landscape refers to the different types of cyberattacks, attackers, vulnerabilities, and techniques that threaten computer systems and networks. ThreatDescriptionExamplePotential impactPhishingDeceives users into revealing sensitive informationFake bank login emailCredential theftMalwareMalicious software that harms or compromises systemsTrojan horseData theft or system damageDoSOverwhelms a service with excessive requestsFlooding a web serverService unavailabilityRansomwareRestricts access to data and demands paymentEncrypting company filesData loss and operational disruption Analysis: Phishing primarily exploits human trust and may provide attackers with valid credentials.Malware can compromise the confidentiality, integrity, or availability of a system.DoS attacks mainly target availability by exhausting system or network resources.Ransomware may affect availability and integrity and can expose confidential data if combined with data theft.Conclusion: These threats require a combination of user awareness, secure configurations, access controls, monitoring, backups, and appropriate detection systems.	10	CO1	BT2

ANSWER KEY

PTO (Please Turn Over)

7	Machine learning (ML) is a branch of artificial intelligence that enables computers to learn patterns from data and use those patterns to make predictions or decisions. In cybersecurity, ML analyzes large volumes of network traffic, system logs, emails, and user activities to identify suspicious behavior. Applications of ML in cybersecurity: Intrusion detection: Identifies unusual network traffic and potential attacks.Malware detection: Learns characteristics of malicious files and processes.Phishing detection: Analyzes email content, URLs, and sender information.Anomaly detection: Identifies activities that differ from normal user or network behavior.Fraud detection: Identifies unusual transaction patterns.User behavior analysis: Detects abnormal login locations, times, or access patterns.Example: An ML model is trained using network traffic labeled as normal or malicious. When new traffic arrives, the model analyzes its features and predicts whether it is suspicious. Conclusion: ML supports automated and data-driven threat detection. However, its effectiveness depends on data quality, model design, and continuous evaluation.	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
8	Explain Indicators of Compromise (IoCs), their types, and their role in identifying cybersecurity incidents. Illustrate how IoCs can be used to investigate a suspected cyberattack.	10	CO1	BT2
9	An Intrusion Detection System (IDS) monitors network traffic or host activity to identify suspicious behavior. FeatureSignature-based IDSAnomaly-based IDSWorking principleMatches activity against known attack signaturesDetects deviations from normal behaviorDetectionKnown attacksUnusual or previously unseen activityUpdatesRequires signature updatesRequires baseline or model maintenanceFalse positivesOften lower for well-matched signaturesMay be higher when normal behavior changesLimitationMay miss new attack patternsMay classify legitimate unusual activity as suspicious Working principles: Signature-based IDS compares observed traffic with a database of known attack patterns.Anomaly-based IDS establishes a baseline of normal behavior and flags significant deviations.Example: A signature-based IDS can identify a known malware communication pattern. An anomaly-based IDS may flag a sudden increase in outbound traffic that differs from the normal network profile. Conclusion: Both approaches can complement each other. Combining known-pattern detection with anomaly detection can help broaden coverage, though neither guarantees detection of every attack.	10	CO1	BT2

PART D: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
10	Machine learning supports threat modeling by identifying patterns in security data that may indicate vulnerabilities, suspicious behavior, or potential attacks. Major stages: Data collection: Gather network traffic, system logs, and security events.Data preprocessing: Clean missing, duplicate, or inconsistent values.Feature selection: Identify relevant variables, such as packet rate or failed login count.Model selection: Choose a suitable algorithm, such as Decision Tree or Random Forest.Training: Train the model using historical data.Testing: Evaluate performance using unseen data.Deployment: Use the trained model to analyze incoming traffic.Monitoring: Track false alarms and model performance over time.Example: A university collects network traffic data and trains a classifier to identify normal and malicious connections. The model analyzes new traffic and flags connections that resemble known attack patterns. Conclusion: ML can assist threat detection by learning from data and supporting automated analysis, but model outputs should be validated and interpreted within the security context.	10	CO2	BT2

ANSWER KEY

11	Supervised learning uses labeled training data, where each example has a known output. Examples of algorithms: Decision Tree, Random Forest, SVM, and KNN. Unsupervised learning uses unlabeled data to discover hidden patterns, clusters, or unusual observations. Examples of algorithms: K-means clustering and DBSCAN. FeatureSupervised LearningUnsupervised LearningTraining dataLabeledUnlabeledGoalPredict known classes or valuesDiscover hidden patternsOutputClass or numerical predictionClusters, structure, or anomaliesCybersecurity useKnown attack classificationUnusual traffic discoveryLimitationNeeds labeled dataDiscovered patterns may be difficult to interpret Example: Supervised: Train a model using normal and attack traffic labels.Unsupervised: Cluster network flows and investigate unusual clusters.Conclusion: Supervised learning is useful when reliable labels are available. Unsupervised learning can help explore data where labels are missing or incomplete. Both methods may be combined in an intrusion detection system.	10	CO2	BT2

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	A DoS attack attempts to disrupt a service by overwhelming its resources. Network features can help identify traffic patterns associated with such attacks. FeatureRelevance to DoS detectionSource IPHelps identify traffic sourcesDestination IPIdentifies the targeted systemProtocolHelps identify the traffic typePacket sizeMay reveal unusual packet patternsFlow durationMay indicate abnormal connection behaviorPacket rateHigh packet rates may indicate flooding Proposed feature selection strategy: Collect labeled normal and DoS traffic.Clean missing and inconsistent values.Calculate derived features such as packets per second and bytes per second.Remove duplicate or redundant features.Use statistical tests or feature importance methods to identify useful features.Train a model using the selected features.Evaluate the model on separate test data.Conclusion: Packet rate, flow duration, protocol, and traffic volume may be useful for DoS detection. Source IP and destination IP should be handled carefully to avoid overfitting to particular network addresses.	10	CO2	BT4
13	Data preprocessing transforms raw data into a clean and consistent format suitable for machine learning. Major techniques: Handling missing values: Remove records when appropriate or replace missing numerical values using the training-set median or mean.Removing duplicates: Identify and remove duplicate records when they represent repeated copies rather than legitimate repeated events.Handling inconsistent values: Standardize units, correct data types, and normalize category names.Feature scaling: Transform numerical features to comparable ranges.Encoding: Convert categorical variables into numerical representations.Data splitting: Divide data into training, validation, and test sets.Example: A network traffic dataset contains: Missing packet sizes.Duplicate flow records.Protocol values written as TCP, tcp, and Tcp.Different numerical scales for packet size and flow duration.The preprocessing steps are: Impute or remove missing values appropriately.Remove accidental duplicate records.Standardize protocol names.Apply standardization or min-max scaling where needed.Split the data before fitting preprocessing transformations to avoid data leakage.Conclusion: Preprocessing improves data consistency and helps ensure that model evaluation reflects performance on unseen data.	10	CO2	BT4

ANSWER KEY