Unit 4

Data Analysis for Cybersecurity and forensic

Systematic Analysis of cyber security using data, threats and attacks on various layers, threats and attacks on various devices, Ensuring Information Privacy, Anomaly detection, Adversarial Machine Learning, Deep Neural Networks (DNN), Game theoretic approaches

Introduction:Data Analysis for Cybersecurity and forensic

- Leveraging Data for Enhanced Security
- In today's digital age, cybersecurity is paramount.
- Data analysis empowers us to identify threats, investigate incidents, and safeguard our digital assets.

We'll explore various techniques.

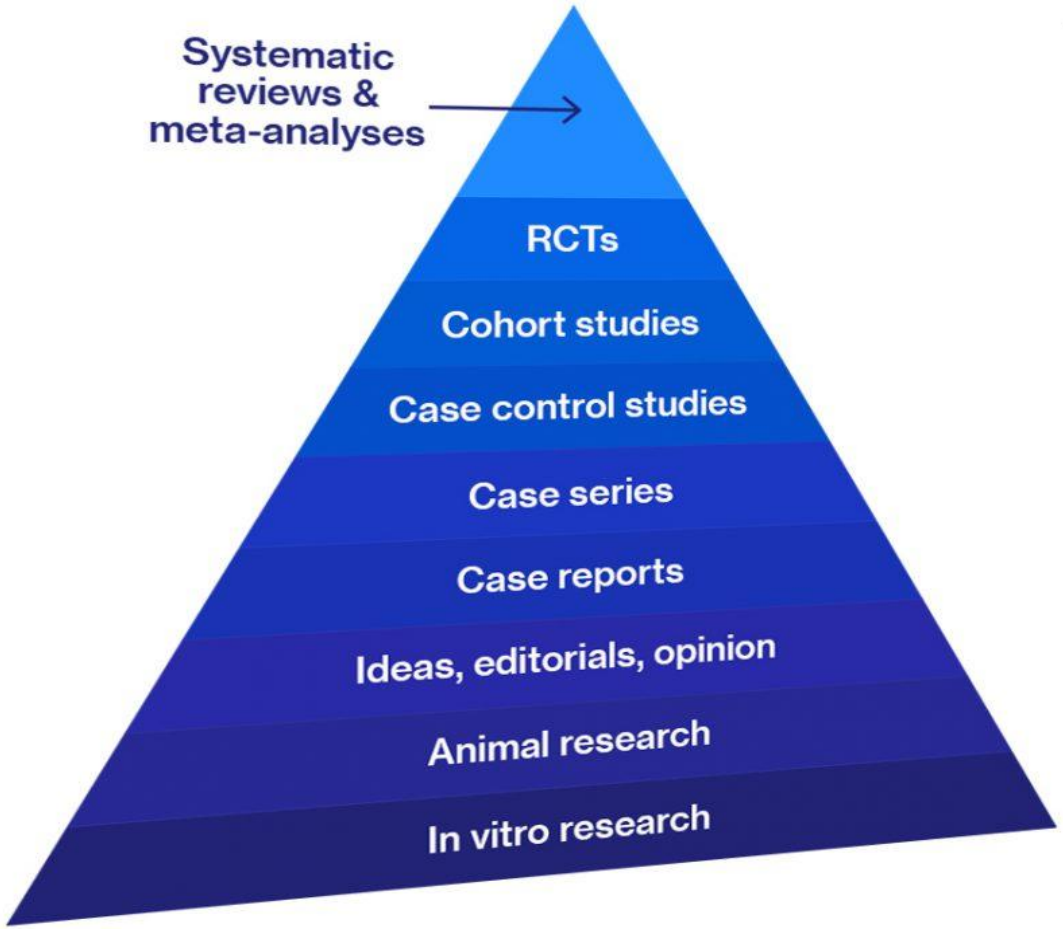

Predictive Modeling
Real-Time Monitoring and Analysis
Identifying Potential Security Risks
Threat Intelligence
Behavioral Analytics
Use Cases for
Data Analytics
in Cyber Security

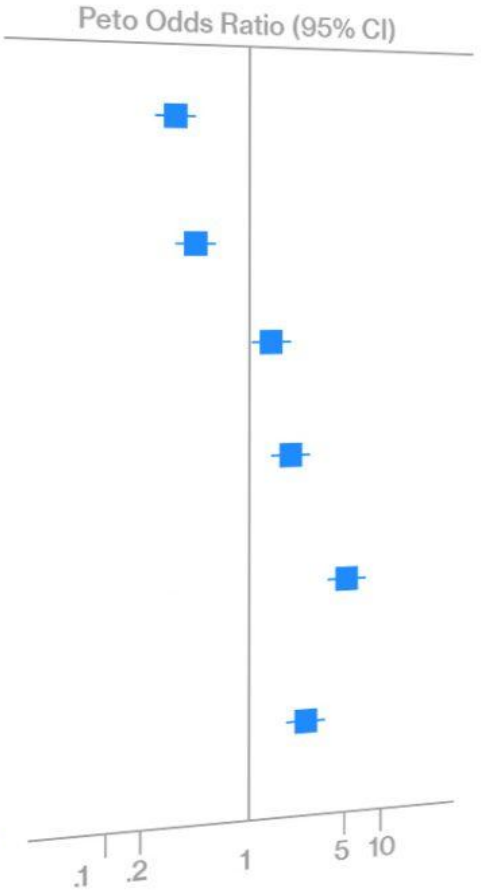
Systematic analysis

- Systematic analysis in research refers to a scientific process of systematically assessing the available evidence from scientific studies to answer a specific research question.
- It involves several key steps, including formulating the research question and assembling a research team , designing and executing a comprehensive search strategy to identify all relevant studies , screening the identified studies for relevance , extracting relevant data from the selected studies , critically appraising each study for potential biases , analyzing and synthesizing the evidence , formulating a report to document the findings , and disseminating the evidence to different stakeholders .
- Systematic analysis aims to provide a *comprehensive* and *reproducible* method for evaluating the available evidence and informing decision-making in various fields, including healthcare and marketing .

- A systematic review is a research method that involves collecting, appraising, and synthesizing evidence to answer a specific question.
- Systematic analysis is a scientific process that involves assessing the available evidence from scientific studies to answer a specific research question.

Meta-analysis is a scientific research methodology consisting of the analysis and synthesis (combination) of data obtained from previously conducted independently from each other studies on the same problem. Most often, it is used to assess the clinical effectiveness of therapeutic interventions. For this purpose, the results of two or more randomised controlled trials are combined.





Systematic review and Systematic analysis are not the same, although they share similarities in their systematic approach to data or literature examination.

1. Systematic Review:
- A systematic review is a research method used to identify, select, evaluate, and synthesize all available evidence relevant to a particular research question or topic.
- It involves a structured and comprehensive search of literature databases, as well as other sources such as gray literature and conference proceedings.
- The inclusion and exclusion criteria are predefined to ensure that only studies meeting specific criteria are included in the review.
- The synthesis of findings often involves statistical analysis or qualitative synthesis techniques such as meta-analysis or thematic analysis.
- Systematic reviews are commonly conducted in academic research to provide evidence-based answers to specific questions and inform decision-making in various fields.

2. Systematic Analysis:
- Systematic analysis, particularly in the context of cybersecurity or data analysis, involves a methodical examination of data or information to identify patterns, trends, anomalies, or insights.
- It may involve the collection, preprocessing, analysis, and interpretation of data from various sources, such as logs, network traffic, or incident reports.
- The objective of systematic analysis is often to uncover insights, detect threats, or improve decision-making related to a specific domain, such as cybersecurity.
- While systematic analysis may utilize structured approaches and techniques, it may not necessarily follow the rigorous methodology of a systematic review, including predefined inclusion/exclusion criteria and formal synthesis methods.

- Both systematic review and systematic analysis involve systematic approaches to examining data or literature, they differ in their objectives, methodologies, and applications. Systematic review is primarily a research method used in academic research to synthesize evidence, while systematic analysis is a broader term that encompasses various analytical techniques used to derive insights from data in fields such as cybersecurity, business intelligence, and data science.

Systematic Analysis of Cybersecurity

-**Data is the new battlefield:** Network traffic logs, system logs, user activity data – all hold valuable insights for security professionals.

- **Data-driven approach:** We'll analyze this data to identify patterns, anomalies, and potential security breaches.
- **Visualization is key:** Charts, graphs, and heatmaps can reveal hidden trends and suspicious activities.

Imagine a vast ocean of data – our weapon against cyber threats. By systematically collecting, analyzing, and visualizing this data, we can gain a comprehensive understanding of the security landscape.

Reference:
https://docs.fortinet.com/document/fortiadc/7.4.2/handbook/254448/data-analytics

- *Image* (search for "cybersecurity data analysis image")

- Example: Utilizing log data to systematically analyze network traffic for potential intrusions.

Bio Metric data for Enhancing Cyber Security in Banking Sector: A Systematic Analysis

The concept from a review paradigm was used.

The review paradigm entails a summary of the research question(s), the topic selection, a search strategy, the examination and obtaining articles, the evaluation of the articles' quality, and data synthesis.

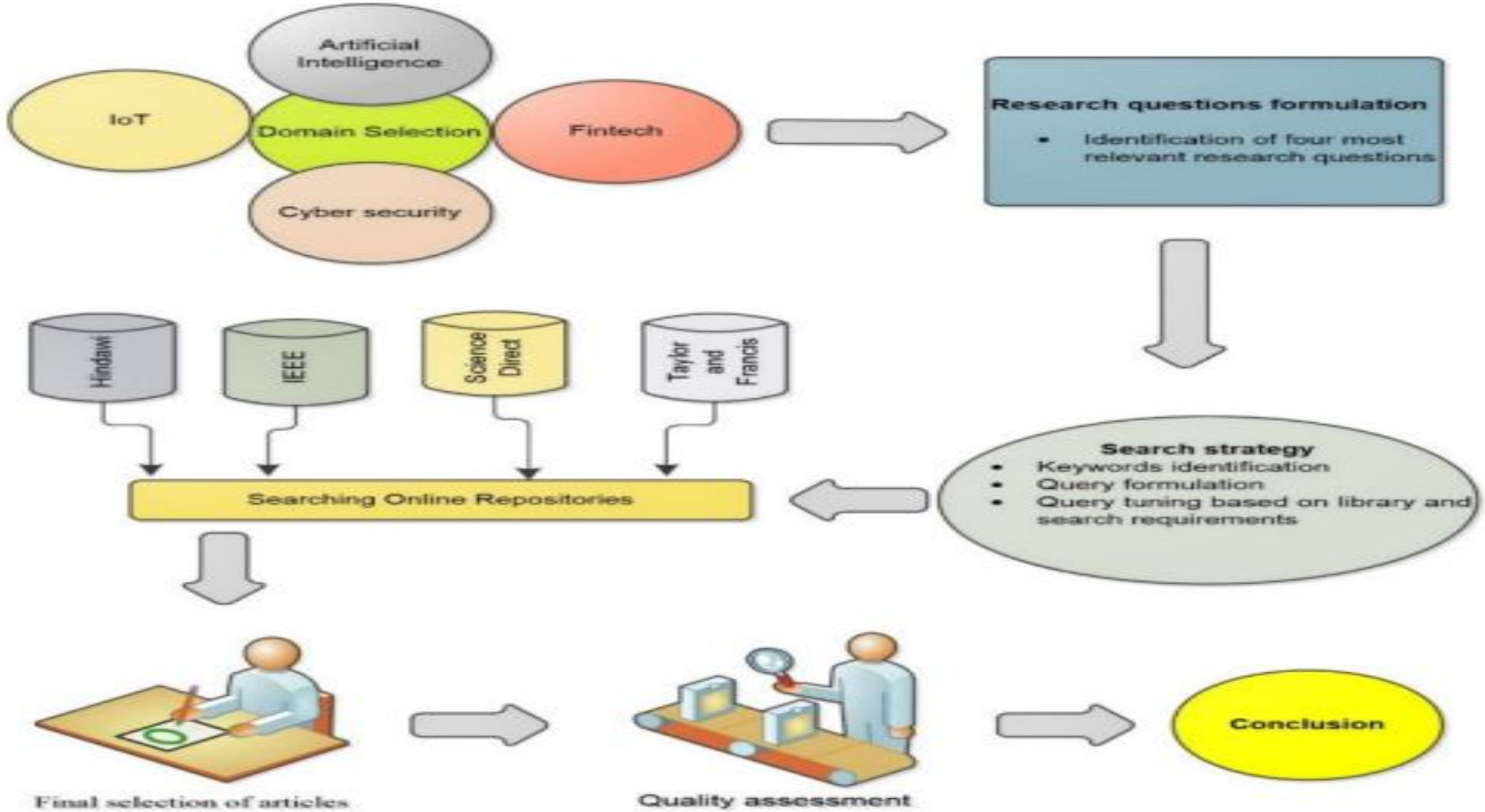
A. SELECTING RESEARCH DOMAIN To have a better understanding of cybersecurity concerns, a thorough investigation of the security topic was carried out. To comprehend the concept of cybersecurity, identify problems in the field, and discover what professionals have done thus far to address the challenges, research papers from a variety of digital information sources were thoroughly analyzed.

B. RESEARCH QUESTIONS FORMULATION The objectives of this study are fourfold: to identify the key biometrics that affect banking to prevent cybercrime; to highlight how biometrics system and FinTech has an influence on the financial sector. To enhance the capabilities of the present system to protect the financial sector, it is necessary to study the primary benefits experienced by the banking industry while using biometrics systems. Research was done on pertinent papers, conference proceedings, book chapters, journals, and book chapters that specifically described security problems. Our first findings indicated that cyber security is a significant endeavor. To make this review briefer, it was recommended that the research be conducted with a specific focus on addressing the research questions listed in "Formulation of Research questions (RQs)" table 1 which were developed after the examination of various publications and articles.

C. STRINGS BASED SEARCHING The search approach is made up of The search approach is made up of searching the keywords and search methodology. The descriptions are provided in a sequence of actions. The procedures below are completed for keyword creation. Major important phrases were found in the suggested research topics. The major words' synonyms were filtered out for efficiency.

S.No	RQs	Explanation
Q1.	What are the key features of biometrics system that affect banking and prevent cyberattacks?	The purpose of this question is to describe the various features that can affect banking to prevent cybersecurity by utilising the biometrics system.
Q2.	How does cybercrime influence FinTech in the banking sector?	FinTech is an emerging technology that can influence the banking sector. The aim of the research question is to secure the transaction between the parties by using financial technology and its features.
Q3.	What are the key benefits gained by the banking sector using biometric systems after overcoming cyberattacks?	As the banking sector shifts its business from a traditional to a modern global environment, the risk of cyberattacks and cybercrime is increasing day by day. The aim of this RQ is to provide smooth and secure operations by using biometrics technology to overcome the risk of cyberattack.
Q4.	Using the literature as evidence, how can we enhance the competence of the existing system to secure the banking sector?	By overcoming cyberattacks, what are the major benefits brought to financial organisations? This RQ aims to enhance the competencies of existing available systems to secure the banking sector, enhance customer satisfaction and trustworthiness levels, and build a bridge between banks and clients.

D. THE SEARCHING PROCESSES: To gather information from many researchers' works in cybersecurity for synchronization, the second author of this article did a comprehensive and thorough review of the recommended study on 4digital libraries.



The planned research was completed by looking at the titles, abstracts, and index terms of previously published research publications, including journal articles, conference papers, and book chapters

Phase 1 – To find publications relevant to the proposed topic, four digital libraries are systematically examined. The search's findings were categorized as perspective studies.

Phase 2 - Articles are obtained from these libraries on the bases of keyword string.

Phase 3 – Relevant studies are mined from the online digital libraries.

E. SCRUTINIZATION AND RETRIEVAL OF RELEVANT ARTICLES Four digital databases were examined as part of the search procedure, yielding 1283 research papers. The proportion of studies can be check via the number of studies found. Metadata can found in the initial search phase (keyword, title, abstract and the contents).

F. QUALITY OF ASSESSMENT AND DATA SYNTHESIS Quality evaluation criteria were applied based on the filtered studies found after carefully examining and retrieving pertinent publications, i.e., each article was carefully reviewed to see if the research addressed at least two specified questions. The three writers of this study each made an equal contribution throughout this stage.

Then

RESULTS AND DISCUSSIONS :

A Each research question is explored in the following subsections, which also categorize the pertinent publications according to the research questions posed. Based on the research question, a summary of each research article is given

Give Result and your finding

Conclusion:

Investigate the role of Deep Neural Networks (DNN) in cybersecurity and forensic analysis.

- Artificial neural networks (ANNs) represent a directed graph in which a set of artificial neuron generally called as units in mathematical model that are connected together with edges. This influenced by the characteristics of biological neural networks, where nodes represent biological neurons and edges represent synapses. A feed forward network is a type of ANNs. A feed forward network (FFN) consists of a set of units that are connected together with edges in a single direction without formation of a cycle.
- They are simple and most commonly used algorithm. Multi-layer perceptron (MLP) is a subset of FFN that consist of 3 or more layers with a number of artificial neurons, termed as units. The 3 layers are input layer, a hidden layer and output layer. There is a possibility to increase the number of hidden layers when the data is complex in nature.
- So, the number of hidden layer is parameterized and relies on the complexity of the data. These units together form an acyclic graph that passes information or signals in forward direction from layer to layer without the dependence of past input

Description of Data sets

Task 1 (Android Malware Classification) data set includes 37,107 unique API information from 61,730 APK files. These APK (application package) files were collected from the Opera Mobile Store over the period of January to September of 2014. When a user runs an application, a set of APIs will be called. Each API is related to a particular permission. The execution of the API may solely achieve success within the case that the permission is granted by the user. These permissions are grouped into Normal, Dangerous, Signature and SignatureOrSystem in Android. These permissions are explicitly mentioned in the AndroidManifest.xml file of APK by application developers.

Task 2 (Incident Detection) dataset contains operational log file that was captured from Unified Threat Management (UTM) of UniteCloud . UniteCloud uses resilient private cloud infrastructure to supply e-learning and e-research services for tertiary students and staffs in New Zealand. Unified Threat Management is a rule based real-time running system for UniteCloud server. Each sample of a log file contains nine features. These features are operational measurements of 9 different sensors in UTM system. Each sample is labeled based on the knowledge related to the incident status of the log samples.

Task 3 (Fraud Detection) dataset is anonymised data that was unified using the highly correlated rule based uniformly distributed synthetic data (HCRUD) approach by considering similar distribution of features . The detailed statistics of Task 1, Task 2 and Task 3 data sets are reported in Table 1.

Task name	Total APK's	Unique APIs	Classes	Training Samples	Testing Samples
Task 1	61,730	37,107	2	30,897	30,000
	Total Samples	**Features**	**Classes**	**Training Samples**	**Testing Samples**
Task 2	100,000	9	2	70,000	30,000
Task 3	100,000	12	3	70,000	30,000

For more details: you can study
https://arxiv.org/pdf/1812.03519

Threats and Attacks on Various Layers

- Discussion of common threats and attacks targeting different layers of the network (e.g., network layer, application layer, data layer).
- Examples: DDoS attacks targeting the network layer, SQL injection attacks targeting the application layer, and data exfiltration attacks targeting the data layer.
- Reference: "Network Security Essentials" by William Stallings.

Threats and Attacks on Various Devices

- Exploration of threats and attacks directed at various devices, including computers, mobile devices, and IoT devices.
- Example: Malware targeting IoT devices to create botnets for launching large-scale attacks.
- Reference: "Practical Internet of Things Security" by Brian Russell, Drew Van Duren, and John Whitman.

- When you think of networks as being structured in the seven layers of the ISO-OSI model, it makes sense that cybersecurity threats can happen at any layer. We can think of these layers as the “links” in our metaphorical chain. Moving outward from the user, data is entered into the network through software running on the Application layer. Through the Session, Transport, Network, and Data-Link layers and arriving at the other end, the Physical layer, the data travels back up the seven layers to arrive at its intended destination. Each layer has its own protocols and other communication standards that govern its efficient operation. So, you may be asking, where is the Security layer? Where does security fit in? The answer is “Yes.”
- Your most comprehensive cybersecurity plan - built layer by layer. Download the eBook >
- Imagine a building with seven doors providing entry. If all seven doors are locked, the building can be considered secure. If one is left unlocked, the entire building is insecure. It really is just that simple. Unless every layer of the network is secured, penetration can occur. Data can be compromised. And compromised data creates an existential danger. According to Inc. Magazine, 60% of businesses whose data is significantly compromised go out of business and don’t return.
- Many providers of data and network security products emphasize the importance of “multi-layer” security, but here is the reality; if security is not efficiently and effectively embedded into every layer of the ISO-OSI model, every step along the path data takes from origin to destination, it is vulnerable and ineffective. Only as secure as its weakest link.

Where do Cybersecurity threats happen?

- Cybersecurity threats exist at all OSI-ISO model layers beginning at Layer 7 – the Application Layer because that's the place where users begin by interfacing to the network. For the purposes of creating the most comprehensive cybersecurity plan we must actually start BEFORE the Application Layer and address perhaps the biggest vulnerability in the entire network – **the user**. Users are human and far more subject to making costly errors than are computers and other digital devices which will perform the same function the same way every time.
- The best example is found in one of the top malware attacks or threats in the cyber landscape – ransomware. Fraudsters send out a "phishing" email that looks very authentic, very much as if it actually comes from where it says it does. But somewhere in that email is a link for the user to click or an attachment for the user to open. The text provides powerful inducements to get the user to do so. Once they do their data is either encrypted, corrupted, or stolen. The only way to get it back is to pay a ransom, thus ransomware.
- The attackers know the user is their best place to gain access.
- Threats at each layer of the ISO-OSI model include:

Ensuring Information Privacy

- Overview of techniques for ensuring information privacy, including encryption, access control, and data masking.
- Example: Implementation of end-to-end encryption in messaging applications to protect user privacy.
- Reference: "Cryptography and Network Security: Principles and Practice" by William Stallings.

Anomaly Detection
- Explanation of anomaly detection techniques in cybersecurity.
- Discussion of statistical methods and machine learning algorithms for detecting anomalous behavior.
- Example: Using machine learning algorithms to detect unusual patterns in user activity indicating potential security breaches.
- Reference: "Machine Learning for Cybersecurity Cookbook" by Pedro Serrano, Andrea De Capua, and Giuseppe Bonaccorso.

Adversarial Machine Learning

- Introduction to adversarial machine learning and its relevance in cybersecurity.
- Explanation of how attackers can manipulate machine learning models to evade detection.
- Example: Adversarial attacks on image recognition systems to deceive automated security measures.
- Reference: "Adversarial Machine Learning" by Anthony D. Joseph, Blaine Nelson, and Benjamin I. P. Rubinstein.

Deep Neural Networks (DNN)

- Exploration of the applications of deep neural networks (DNNs) in cybersecurity.
- Discussion of DNNs for intrusion detection, malware analysis, and security analytics.
- Example: Using deep learning models to classify malware based on behavioral features.
- Reference: "Deep Learning and Cyber Security" by Soman Kp, Amiya Nayak, and N. C. Mahanti.

Game Theoretic Approaches

- Overview of game theoretic approaches to cybersecurity.
- Explanation of how game theory can model interactions between attackers and defenders.
- Example: Using game theory to analyze optimal strategies for defending against cyber attacks in a networked environment.
- Reference: "Game Theory for Security and Risk Management" by Stefan Rass.

Conclusion

- Summary of key points covered in the lecture.
- Reinforcement of the importance of data analysis in cybersecurity and forensic investigations.
- Encouragement for further exploration of the topics discussed.
- Invitation for questions and discussion.

References

-