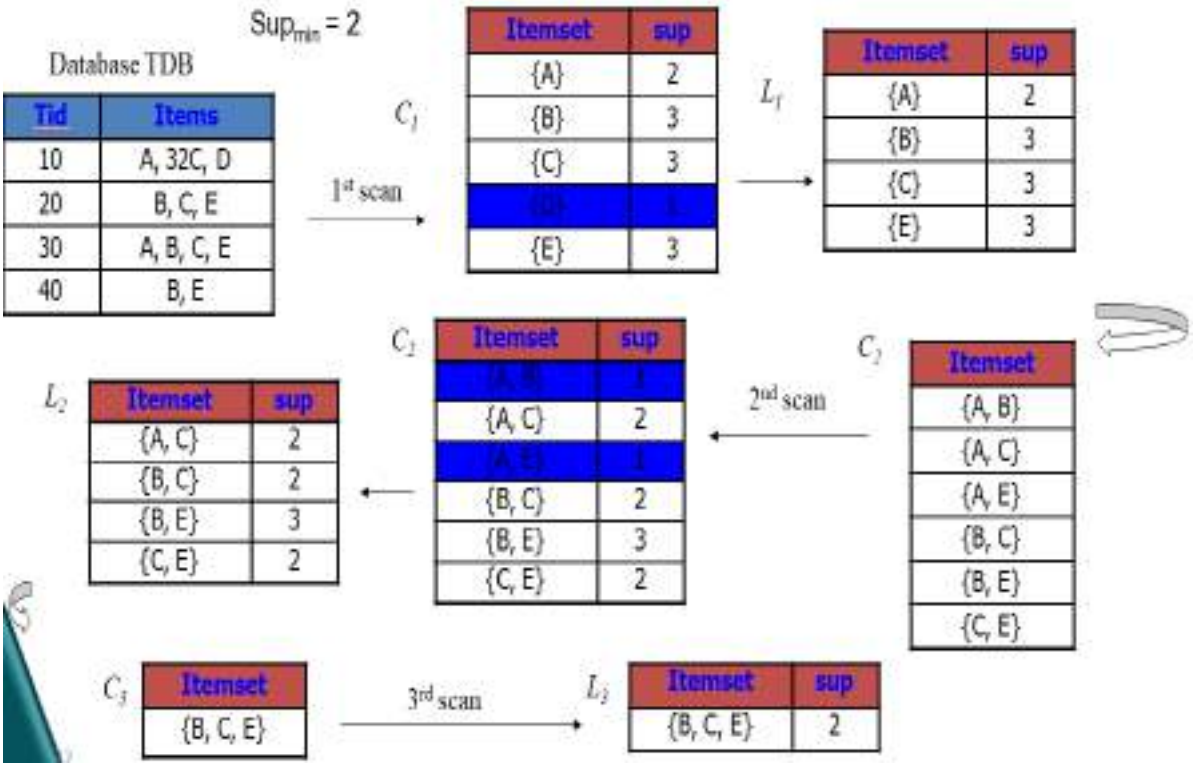
Mining Various Kinds of Association Rules (Constrained Based) ,
https://www.youtube.com/watch?v=3NXVR3opUmI

UNIT IV

MDSS : **Supporting Technology for DS** Session 8

Mining Various Kinds of Association Rules
(Constrained Based)

The Apriori Algorithm—Another Example



Example Generating Association Rules.

Let's try an example based on the transactional data for the table :

*The data contain frequent itemset X = { I1, I2,*I5 }

What are the association rules that can be generated from *X?*

The nonempty subsets of *X are f{I1, I2}, {I1, I5}, {I2, I5}, {I1}, {I2}, and {I5}.*

T100	**I1, I2, I5**
T200	**I2, I4**
T300	**I2, I3**
T400	**I1, I2, I4**
T500	**I1, I3**
T600	**I2, I3**
T700	**I1, I3**
T800	**I1, I2, I3, I5**
T900	**I1, I2, I3**

The resulting association rules are:

$$\{I1, I2\} \Rightarrow I5, \quad confidence = 2/4 = 50\%$$
$$\{I1, I5\} \Rightarrow I2, \quad confidence = 2/2 = 100\%$$
$$\{I2, I5\} \Rightarrow I1, \quad confidence = 2/2 = 100\%$$
$$I1 \Rightarrow \{I2, I5\}, \quad confidence = 2/6 = 33\%$$
$$I2 \Rightarrow \{I1, I5\}, \quad confidence = 2/7 = 29\%$$
$$I5 \Rightarrow \{I1, I2\}, \quad confidence = 2/2 = 100\%$$

If the minimum confidence threshold is, say, 70%, then only the second, third, and last rules are output, because these are the only ones generated that are strong. Note that, unlike conventional classification rules, association rules can contain more than one conjunct in the right side of the rule. ∎

Mining Various Kinds of Association Rules

Now let us consider application requirements by extending our scope to include mining multilevel association rules, multidimensional association rules, and quantitative association rules in transactional and/or relational databases and data warehouses.

Multilevel association rules involve concepts at different levels of abstraction.

Multidimensional association rules involve more than one dimension or predicate

(e.g., rules relating what a customer *buys as well as the customer's age.)*

Quantitative association rules involve numeric attributes that have an implicit ordering among values (e.g., *age).*

Mining Multilevel Association Rules

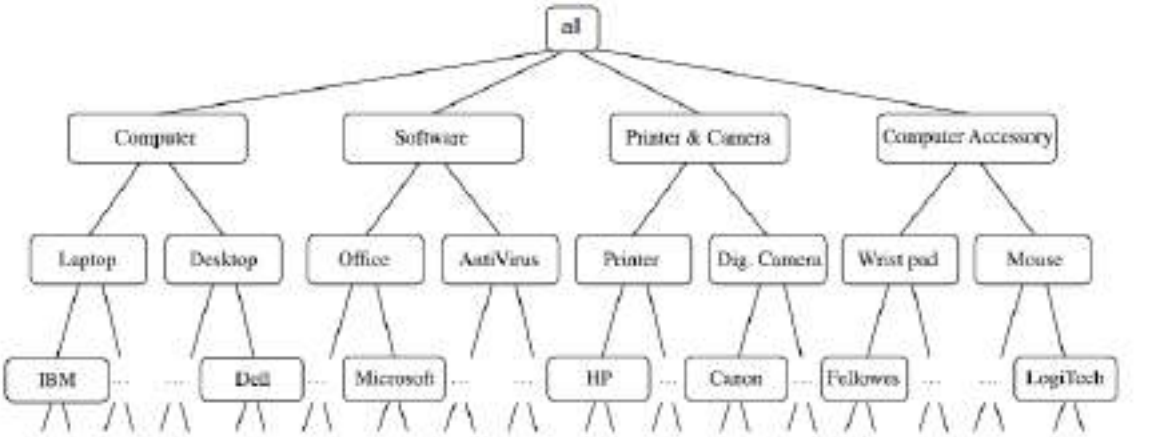
For many applications, it is difficult to find strong associations among data items at low or primitive levels of abstraction due to the sparsity of data at those levels. Strong associations discovered at high levels of abstraction may represent commonsense knowledge.

Moreover, what may represent common sense to one user may seem novel to another.

Therefore, data mining systems should provide capabilities for mining association rules at multiple levels of abstraction, with sufficient flexibility for easy traversal among different abstraction spaces.

Let's examine the following example.

TID	Items Purchased
T100	IBM-ThinkPad-T40/2373, HP-Photosmart-7660
T200	Microsoft-Office-Professional-2003, Microsoft-Plus!-Digital-Media
T300	Logitech-MX700-Cordless-Mouse, Fellowes-Wrist-Rest
T400	Dell-Dimension-XPS, Canon-PowerShot-S400
T500	IBM-ThinkPad-R40/P4M, Symantec-Norton-Antivirus-2003
...	...



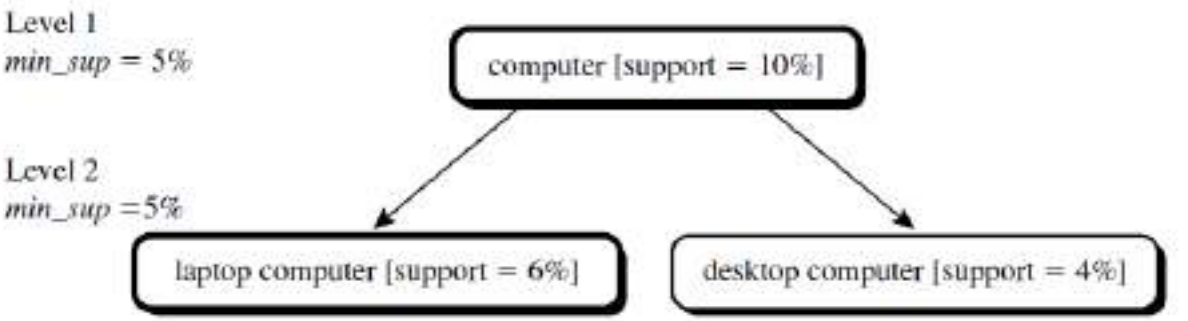
The concept hierarchy of Fig. has five levels, respectively referred to as levels 0 to 4, starting with level 0 at the root node for all (the most general abstraction level).

Here, level 1 includes *computer, software, printer&camera, and computer accessory, level* 2 includes *laptop computer, desktop computer, office software, antivirus software, . . . , and* level 3 includes *IBM desktop computer, . . . , Microsoft office software, and so on.* Level 4 is the most specific abstraction level of this hierarchy. It consists of the raw data values.

Association rules generated from mining data at multiple levels of abstraction are called multiple-level or multilevel association rules. Multilevel association rules can be mined efficiently using concept hierarchies under a support-confidence framework.

In general, a top-down strategy is employed, where counts are accumulated for the calculation of frequent itemsets at each concept level, starting at the concept level 1 and working downward in the hierarchy toward the more specific concept levels, until no more frequent itemsets can be found. For each level, any algorithm for discovering frequent itemsets may be used, such as Apriori or its variations. A number of variations to this approach are described below, where each variation involves "playing" with the support threshold in a slightly different way. The variations are illustrated in Figures 5.11 and 5.12, where nodes indicate an item or itemset that has been examined, and nodes with thick borders indicate that an examined item or itemset is frequent.

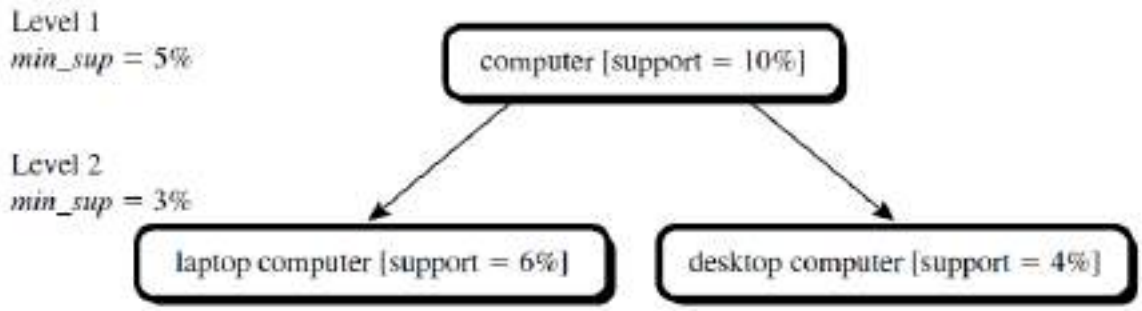
Using uniform minimum support for all levels (referred to as uniform support): The same minimum support threshold is used when mining at each level of abstraction. For example, in Fig., a minimum support threshold of 5% is used throughout (e.g., for mining from *"computer" down to "laptop computer").* Both *"computer" and "laptop computer" are found to be frequent, while "desktop computer" is not.*



Multilevel mining with uniform support.

Using reduced minimum support at lower levels (referred to as reduced support): Each level of abstraction has its own minimum support threshold. The deeper the level of abstraction, the smaller the corresponding threshold is. For example,

in Fig., the minimum support thresholds for levels 1 and 2 are 5% and 3%, respectively. In this way, *"computer," "laptop computer," and "desktop computer" are* all considered frequent.



2 Multilevel mining with reduced support.

Constraint-Based Association Mining

A data mining process may uncover thousands of rules from a given set of data, most of which end up being unrelated or uninteresting to the users. Often, users have a good sense of which "direction" of mining may lead to interesting patterns and the "form" of the patterns or rules they would like to find. Thus, a good heuristic is to have the users specify such intuition or expectations as constraints to confine the search space. This strategy is known as constraint-based mining.

The constraints can include the following:

Knowledge type constraints: These specify the type of knowledge to be mined, such as association or correlation.

Data constraints: These specify the set of task-relevant data.

Dimension/level constraints: These specify the desired dimensions (or attributes) of the data, or levels of the concept hierarchies, to be used in mining.

Interestingness constraints: These specify thresholds on statistical measures of rule interestingness, such as support, confidence, and correlation.

Rule constraints: These specify the form of rules to be mined. Such constraints may be expressed as metarules (rule templates), as the maximum or minimum number of predicates that can occur in the rule antecedent or consequent, or as relationships among attributes, attribute values, and/or aggregates.

End of session 8

End of UNIT IV

Data Science for Cyber Security ,
https://www.youtube.com/watch?v=xv56JH6hxGM

UNIT V

MDSS : **Data Science for (in) Cyber Security** Session 1

The volume of data generated every day is increasing at a surprising rate. Nearly 5 quintillion bytes of data are being created daily. With the rise in data, there has also been a surge in data breaches. ***Hacking and penetrating a system using various tools has become a significant cause of concern for organizations and individuals worldwide.*** Sophisticated data science techniques are now widely used by attackers to break into a system. The question is if data science can be used to take charge of the system, can it be used to prevent it from hacking?

The answer is yes; with the use of data science in cyber security, it has become easy to predict vulnerability in a system, which in turn prevents the potential risk of breach by taking appropriate measures.

According to the Identity theft Resource Centre (ITRC)'s 2022 report, there have been 817 publicly-reported data compromises in the United States alone - loosing $ 4.35 Million Dollars due to data breaches.

Cyber-attackers have drilled their way into various domains such as Healthcare, Finance, retail, Insurance , etc., thus increasing the target spectrum. Data science and Cybersecurity serve as powerful weapons to dilute these losses.

What (Why) is Data Science in Cyber security?

Data Science for cyber security has been a game changer in resisting fraudulent activities. Data Science uses Machine Learning tools on past data to predict the likelihood of an intrusion or attack. It involves developing algorithms to deduce patterns from previous attacks and beforehand warning about the reliability of the system in use.

Example: Detecting unauthorized access in an institution. The AI model would grant access to only pre-registered users based on their credentials and analyze the activity of these users so that there is no activity beyond authorization. All these steps are used to prevent any sort of data breach or misuse of information.

What Do Data Science Cyber Security Professionals Do?

Data Science professionals analyze large amounts of data using statistical and programmable skills. They develop solutions to cater to an organization's needs. It involves interpreting raw data and extracting valuable information from it. This information is further used to interpret the underlying trend and derive a solution using machine learning algorithms.

Data science cyber security professionals are exposed to a large amount of data provided by institutions that thrive on collecting more and more data to leverage data science solutions. Data to be used must be managed. Handling large amounts of data without the help of data scientists is a very big challenge. Taking the predictive way tightens not only the security of the sensitive data but also blocks any sort of penetration.

Cyber Security Before Data Science

Cyber security in the initial times was associated with Fear and uncertainty. This fear arose from the fact that all the security strategies made by the companies were purely based on assumptions. How the attack will take place, and which area is more prone to attack all these parameters were assumption based.

With data science coming into the picture, it changed the face of the entire Cyber industry. Since cyber security is mainly about technology decisions, the predictions from data science have helped a lot in minimizing the chances of making wrong decisions, as most of the judgments are facts based. These data-driven tools have made the jobs of cyber security analysts and experts a lot better by increasing their scope of resources, which in a way, helps them to carve out better security improvement plans.

It becomes extremely important that the security team is actively involved with the data science team right from the beginning. This collaboration from the beginning can facilitate both teams in numerous ways, the data science will become aware of the cybersecurity controls while the cyber security team will become well versed with the possible loopholes.

Importance of Data Science in Cybersecurity Risk

The pandemic has changed our lives in numerous ways. Our lives have moved to online platforms, be it anything purchasing, transferring money, or the shift of companies to online models. A system can be attacked through different mediums, as our usage is just not limited to one aspect. We use a

variety of appliances every day, hence furthermore increasing the bandwidth for the attackers to cause data breaches.

Considering all of the above arguments, it is obvious why data science plays an important role in managing cybersecurity risk. This approach helps to reduce the percentage of attacks, it can't stop the attack, but it helps in notifying the concerned stakeholder about the estimated risk involved. The security team then takes the necessary steps to stop the attack or minimize the damage due to the same. All this can be possible only if we have a risk assessment report by the data science team. In the wake of cybersecurity, data science is of extreme value.

Data Science in Cyber Security to Protect the Digital Footprint

Today, everyone is under the threat of an attack, and these attacks are not limited to just large organizations or governments. Hackers are always looking for the minutest opportunity to get sensitive information. These include personal information, bank account details, etc. Different frauds can be conducted using this information.

Anything which is put on digital platforms gets immortal. There is no way one can wash away their digital activities. With every round of surfing, we are leaving a huge amount of information that helps businesses to grow their trades by making user-oriented choices. Data science becomes fundamental in protecting our digital traces as they can be misused.

For example, My personal information can be used for identity theft. A person can claim my identity and thus create a lot of chaos by accessing private and confidential accounts, thus creating a lot of loss.

How Data Science and Machine Learning Work Together to Improve Cyber Security

Technology is enhancing day by day. Thus, the potential risk of cybercrimes is also increasing.

If you are still wondering about data science or cyber security which is better?

The best possible answer to this question is data science for cyber security.

The amount of sensitive data within an organization is increasing day by day, it becomes increasingly important for each one of them to include data science in their risk analysis plans.

There are various ways in which data science help to alleviate the threats, below are some mentioned evidence:

1. Protection of Data

Data is extremely vital to any organization and it is extremely crucial that it is been protected at any cost, data science helps to create impermeable data channels for transferring the data using machine learning algorithms.

2. Enhanced Intrusion Detection

With improvements in technology, hackers do not use just one pathway to hack a system. Refined techniques have increased challenges for companies to recognize the paths for penetrating the system. Machine Learning models developed on current and past attack information provide a wholesome understanding to model different attacks. These models then predict the type of attack and the probability of breaking the system.

3. Efficient Prediction

Prediction doesn't just only mean detecting True positives. A data science cyber security model should also generate very few False positives, this will help to combat the problem of spam calls. These techniques help to create real-world hypotheses rather than old-school assumptions related to threats and Cyber risk.

4. Behavioral Analysis

Just understanding the type of attack or knowing the probability of it affecting the system is not enough, one must understand a hacker's behavioral pattern. This can serve great advantages as we will be in a position to predict his/her next move or next attack. This **behavioral analysis** is done by combining different datasets, studying the network logs, and finding correlations between systems help to draw a hacker's behavioral pattern and take preventive measures accordingly.

Considering the need of the hour, there has been a rise in data science cyber security jobs. A person confused about which is best, data science or cyber security, to pursue as a career must explore their options as a data scientist in cyber security. In order to gain skills in this sector, one can take up any cyber security data science course and become an expert. Pursuing small cyber security data science projects provides practical understanding in addition to theoretical knowledge. Today in the field of cyber security data science salary spectrum varies from $100k- $150k.

Future of Data Science in the Realm of Cybersecurity

Data Science has one of the most promising futures. Hackers are constantly trying to find ways and loopholes to break into a system, with the advancement in approaches, more and more sophisticated attacks are surfacing, in order to prevent this, data science seems like a long-term solution. As already mentioned, the generation of data is not going to stop anytime soon, instead in

the coming decades, we will see an exponential rise in the data this will, in turn, result in better-performing data science models as they will have more and more information to connect the dots.

Data Science is not just limited to developing models or algorithms. Analyzing and maintaining the existing data science model is also one of the important aspects involved in this branch. Analysis helps distinguish between what behavior is normal and what can be considered an anomaly. Large enterprises are facing huge losses due to data breaches. They are in dire need of finding ways to reduce these losses. Protecting data using data seems like an encouraging method.

Some Questions (FAQs)

1. Is data science used in cyber security?

Yes, data science is used in cyber security, it's a modern-day technology that works by predicting the potential risk to a system. Data science uses the past history of the attacks based on which they alert the system about forthcoming attacks.

2. Which is better, data science or cyber security?

Both fields are in high demand. Considering that almost all companies are moving to digital platforms, professionals skilled in data science are required to manage the ever-increasing data, and people skilled in cyber security are needed to protect this data.

3. Who earns more in cybersecurity or data science?

Data science jobs are in high demand. Hence, a data scientist has a high salary as compared to Cyber security professional. The average salary of a data scientist is $100,000 USD, while that of a Cyber security professional is around $85,000 USD.

4. Does cyber security require coding?

Little to Not much as compared to other coding , but it is a great skill to have. It facilitates in gaining deeper knowledge of the system one is trying to protect, which in turn helps in deducing ways to prevent the system from going under attack.

End of session 1

Malicious Executables ,
https://www.youtube.com/watch?v=tI7daAlH7Ec&feature=youtu.be

UNIT V

MDSS : **Data Science for (in) Cyber Security** Session 2

Malicious Executables

A malicious executable is defined to be a program that performs a malicious function, such as compromising a system's security, damaging a system or obtaining sensitive information without the user's permission.

Examples of a malicious code:

Taking advantage of common system vulnerabilities, malicious code examples include **computer viruses**, **worms**, Trojan horses, **logic bombs**, **spyware**, **adware**, and **backdoor programs**.

Visiting infected websites or clicking on a bad email link or attachment are ways for malicious code to sneak its way into a system.

What is malicious software and examples?

Malware, short for malicious software, refers to any **intrusive software** developed by cybercriminals

(often called hackers) to steal data and damage or destroy computers and computer systems.

(Examples of common malware include viruses, worms, Trojan viruses, spyware, adware, and ransomware).

What is an example of a malicious website?

The most classic example of a malicious site is one built to enable a phishing attack or scam. Generally, cyber criminals build a clone site of a known site, for example, their own bank's site, or Amazon, or even a courier company's site.

How to identify a malicious website?

Here are the most prevalent tell-tale signs of a threatening website

and some ways that you can protect yourself:

- Never click on a link embedded in an email. ...
- Use your common sense. ...
- Look for signs of legitimacy. ...
- Read the URL carefully. ...
- If it looks too good to be true, it probably is. ...
- Check the properties of any links.

How to Identify and Protect Yourself from an Unsafe Website

With more people storing personal information on their computers, it has never been more important to protect yourself from internet predators looking to gain access to your files. One of the many ways they can do this is by attacking your computer or trying to gather your information from an infected or malicious website you may visit, even if only once. The best thing you can do is to avoid malicious websites altogether.

Here are the most prevalent tell-tale signs of a threatening website and some ways that you can protect yourself:

- Never click on a link embedded in an email. Even if sent from someone you trust, always type the link into your browser.

Use your common sense. Does a website look strange to you? Is it asking for sensitive personal information? If it looks unsafe, don’t take the risk.

Look for signs of legitimacy. Does the website list contact information or some signs of a real-world presence. If doubtful, contact them by phone or email to establish their legitimacy.

Read the URL carefully. If this is a website you frequent, is the URL spelled correctly? Often times, phishers will set up websites almost identical to the spelling of the site you are trying to visit. An accidental mistype may lead you to a fraudulent version of the site.

If it looks too good to be true, it probably is. Is the website offering you a product or service at an unheard of price? Or maybe they are promising you a huge return on investment?

If the offer looks too good to be true, trust your instincts. Do some research to find reviews or warnings from other users.

Check the properties of any links. Right-clicking a hyperlink and selecting "Properties" will reveal the true destination of the link. Does it look different from what it claimed to lead you to?

You should also always be on the lookout for the clues and telltale hints that you are on a malicious website. After all, it is by smart people noticing something wrong and reporting it that the above tools can do their job.

Things to look for in a secure website

When visiting a website that asks for sensitive information such as credit card numbers or your social security number, the first step you can take to securing your privacy is creating a strong password. Equally important is verifying that any information you enter on this site is transmitted and stored properly. Once your information is entered online, it is transmitted as plain text for anyone to intercept. To avoid this, make sure that the website is encrypted over a secure connection.

HTTPS

One such sign to look for is in the URL of the website. A secure website's URL should begin with "https" rather than "http". The "s" at the end of "http" stands for secure and is using an SSL (Secure Sockets Layer) connection. Your information will be encrypted before being sent to a server.

THE LOCK ICON

Another sign to look for is the “Lock” icon that is displayed somewhere in the window of your web browser. Different browsers may position the lock in different places, but a few examples of what it may look like can be found here:

Be sure to click on the "lock" icon to verify that a website is trustworthy. Do not simply look for the icon and assume a website is secure! Your web browser will have detailed information on the website's authenticity if you click on the icon, so be sure to read this carefully before entering any of your information on the site.

End of session 1

Data Mining tools for Cyber Security ,
https://www.youtube.com/watch?v=wu9QfOqfjT4

UNIT V

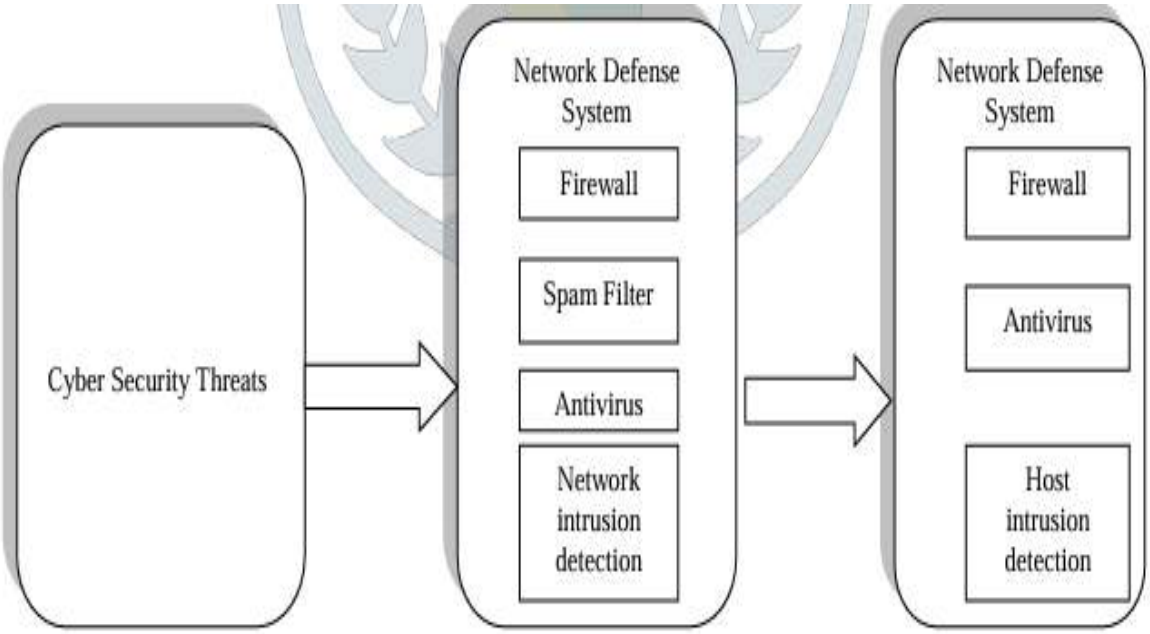
MDSS : **Data Science for (in) Cyber Security** Session 3

Data Mining Tools in Cyber Security : Applications

Introduction

Cyber security is concerned with protecting computer and network systems from corruption due to malicious software including Trojan horses and viruses. Data mining for cyber security applications For example, anomaly detection techniques could be used to detect unusual patterns and behaviors. Data mining is the process of identifying patterns in large datasets. Data mining techniques are heavily used in scientific research as well as in business, mostly to gather statistics and valuable information to enhance customer relations and marketing strategies. Data mining has also proven a useful tool in cyber security solutions for discovering vulnerabilities and gathering indicators for baseline.

Cyber security is set of rules and technologies meant to protect our systems, network, and data from unauthorized access, attacks, and unwanted interrupts. They are aimed to maintain the confidentiality, integrity, and availability of information and information management systems through various cyber defense systems. To secure cyber infrastructure against potentially malicious threats, a growing collaborative effort between cyber security professionals and researchers from institutions, private industries, academia, and government agencies has engaged in exploiting and designing a variety of cyber defense systems.



Intrusion Detection System (IDS)

An IDS maintains network traffic looks for unusual activity and sends alerts when it occurs. The main duties of an IDS are anomaly detection and reporting, however, certain IDS can take action when malicious activity or unusual traffic is discovered.

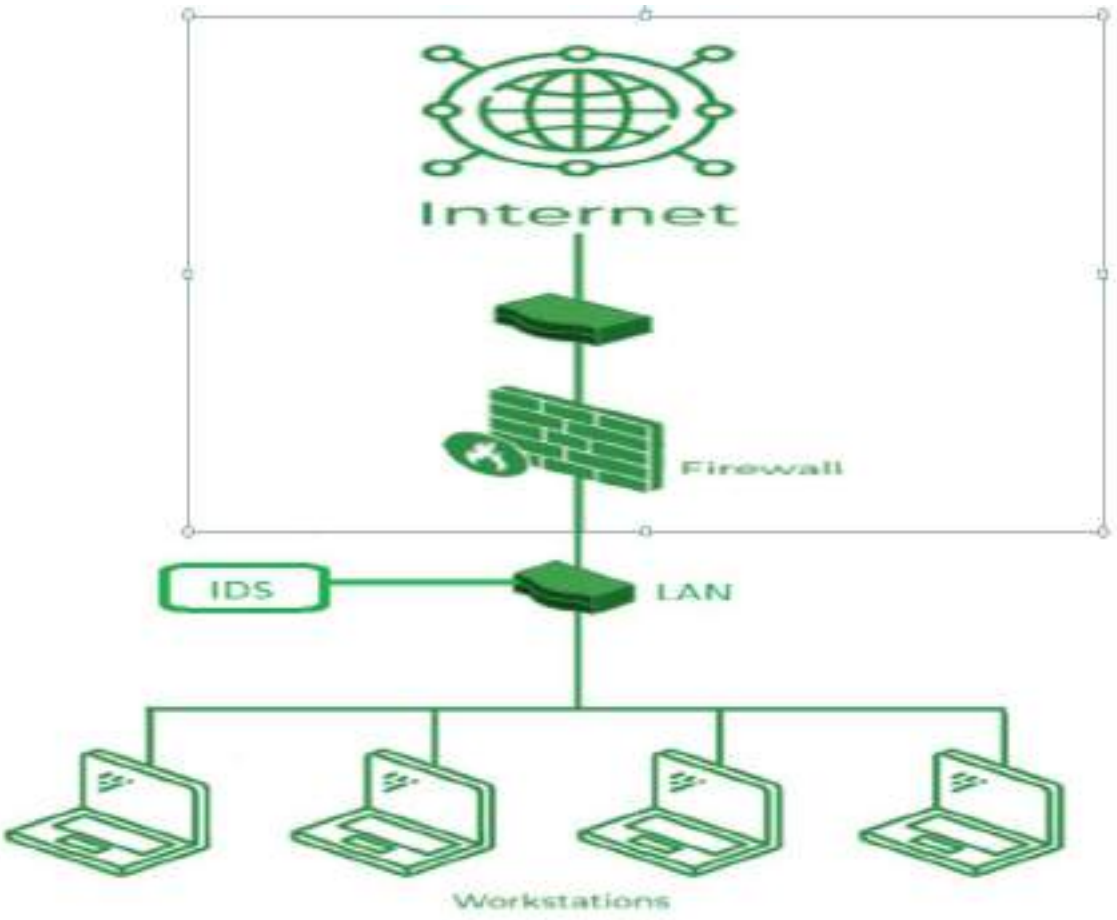
What is an Intrusion Detection System?

A system called an IDS observes network traffic for malicious transactions and sends immediate alerts when it is observed. It is software that checks a network or system for malicious activities or policy violations. Each illegal activity or violation is often recorded either centrally using an SIEM system or notified to an administration. IDS monitor a network or system for malicious activity and protect a computer network from unauthorized access from users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between 'bad connections' (intrusion/attacks) and 'good (normal) connections'.

Working of Intrusion Detection System(IDS)

- An IDS monitors the traffic on a computer network to detect any suspicious activity.

- It analyzes the data flowing through the network to look for patterns and signs of abnormal behavior.
- The IDS compares the network activity to a set of predefined rules and patterns to identify any activity that might indicate an attack or intrusion.
- If the IDS detects something that matches one of these rules or patterns, it sends an alert to the system administrator.
- The system administrator can then investigate the alert and take action to prevent any damage or further intrusion.



Classification of IDS:

Network Intrusion Detection System (NIDS): NIDS are set up at a planned point within the network to examine traffic from all devices on the network. It performs an observation of passing traffic on the entire subnet and matches the traffic that is passed on the subnets to the collection of known

attacks. Once an attack is identified or abnormal behavior is observed, the alert can be sent to the administrator. An example of a NIDS is installing it on the subnet where firewalls are located in order to see if someone is trying to crack the firewall.

Host Intrusion Detection System (HIDS): HIDS run on independent hosts or devices on the network. A **HIDS** monitors the incoming and outgoing packets from the device only and will alert the administrator if suspicious or malicious activity is detected. It takes a snapshot of existing system files and compares it with the previous snapshot.

If the analytical system files were edited or deleted, an alert is sent to the administrator to investigate. An example of HIDS usage can be seen on mission-critical machines, which are not expected to change their layout.

Protocol-based Intrusion Detection System :PIDS comprises a system or agent that would consistently reside at the front end of a server, controlling and interpreting the protocol between a user/device and the server. It is trying to secure the web server by regularly monitoring the HTTPS protocol stream and accepting the related HTTP protocol. As HTTPS is unencrypted and before instantly entering its web presentation layer then this system would need to reside in this interface, between to use the HTTPS.

Application Protocol-based Intrusion Detection System (APIDS):

An APIDS is a system or agent that generally resides within a group of servers. It identifies the intrusions by monitoring and interpreting the communication on application-specific protocols. For example, this would monitor the SQL protocol explicitly to the middleware as it transacts with the database in the web server.

Hybrid Intrusion Detection System:

Hybrid intrusion detection system is made by the combination of two or more approaches to the intrusion detection system. In the hybrid intrusion detection system, the host agent or system data is combined with network information to develop a complete view of the network system. The hybrid intrusion detection system is more effective in comparison to the other intrusion detection system. Prelude is an example of Hybrid IDS.

Intrusion Detection System Evasion Techniques

- **Fragmentation:** Dividing the packet into smaller packet called fragment and the process is known as fragmentation. This makes it impossible to identify an intrusion because there can't be a malware signature.
- **Packet Encoding:** Encoding packets using methods like Base64 or hexadecimal can hide malicious content from signature-based IDS.
- **Traffic Obfuscation:** By making message more complicated to interpret, obfuscation can be utilized to hide an attack and avoid detection.
- **Encryption:** Several security features, such as data integrity, confidentiality, and data privacy, are provided by encryption. Unfortunately, security features are used by malware developers to hide attacks and avoid detection.

Benefits of IDS

- **Detects malicious activity:** IDS can detect any suspicious activities and alert the system administrator before any significant damage is done.
- **Improves network performance:** IDS can identify any performance issues on the network, which can be addressed to improve network performance.
- **Compliance requirements:** IDS can help in meeting compliance requirements by monitoring network activity and generating reports.
- **Provides insights:** IDS generates valuable insights into network traffic, which can be used to identify any weaknesses and improve network security.

Detection Method of IDS

- **Signature-based Method:** Signature-based IDS detects the attacks on the basis of the specific patterns such as the number of bytes or a number of 1s or the number of 0s in the network traffic. It also detects on the basis of the already known malicious instruction sequence that is used by the malware. The detected patterns in the IDS are known as signatures. Signature-based IDS can easily detect the attacks whose pattern (signature) already exists in the system but it is quite difficult to detect new malware attacks as their pattern (signature) is not known.

Detection Method of IDS

- **Anomaly-based Method:** Anomaly-based IDS was introduced to detect unknown malware attacks as new malware is developed rapidly. In anomaly-based IDS there is the use of machine learning to create a trustful activity model and anything coming is compared with that model and it is declared suspicious if it is not found in the model. The machine learning-based method has a better-generalized property in comparison to signature-based IDS as these models can be trained according to the applications and hardware configurations.

Placement of IDS

- The most optimal and common position for an IDS to be placed is behind the firewall. Although this position varies considering the network. The 'behind-the-firewall' placement allows the IDS with high visibility of incoming network traffic and will not receive traffic between users and network. The edge of the network point provides the network the possibility of connecting to the extranet.

- In cases, where the IDS is positioned beyond a network's firewall, it would be to defend against noise from internet or defend against attacks such as port scans and network mapper.An IDS in this position would monitor layers 4 through 7 of the OSI model and would use Signature-based detection method. Showing the number of attempted breaches instead of actual breaches that made it through the firewall is better as it reduces the amount of false positives. It also takes less time to discover successful attacks against network.

- An advanced IDS incorporated with a firewall can be used to intercept complex attacks entering the network. Features of advanced IDS include multiple security contexts in the routing level and bridging mode. All of this in turn potentially reduces cost and operational complexity.
- Another choice for IDS placement is within the network. This choice reveals attacks or suspicious activity within the network. Not acknowledging security inside a network is detrimental as it may allow users to bring about security risk, or allow an attacker who has broken into the system to roam around freely.

End of session 3

Cyber Crimes,
https://www.youtube.com/watch?v=sFiVnx9OP9g

UNIT V

MDSS : **Data Science for (in) Cyber Security** Session 4

Cyber Crimes

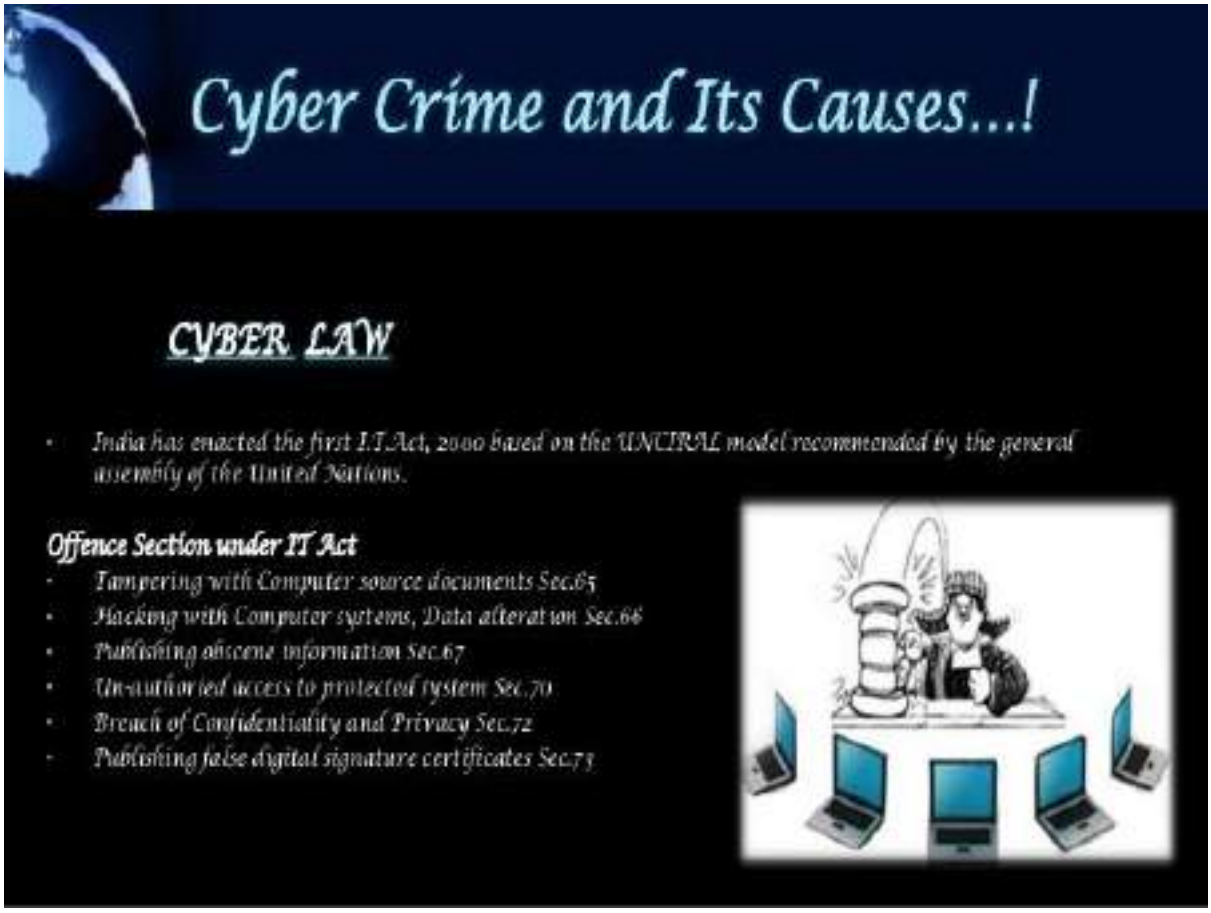
Introduction

- What is cybercrime?
- Running a cybercrime syndicate
- Cybercrime attacks
- Countermeasures

Cyber-offences: new type of crime (illegal access, illegal interference with data and system,…).

Cybercrime at a Glace

Who, Where, When, Why ?



Confusion and misconceptions

Cybersecurity-related terms: “cybercrime”, “cyberwar”, “cyberattack”, “cyberterrorism”

- absence of a clear consensus

• Terms are used interchangeably, sometimes with

little regard for what they actually mean

• Sensationalization and exaggeration

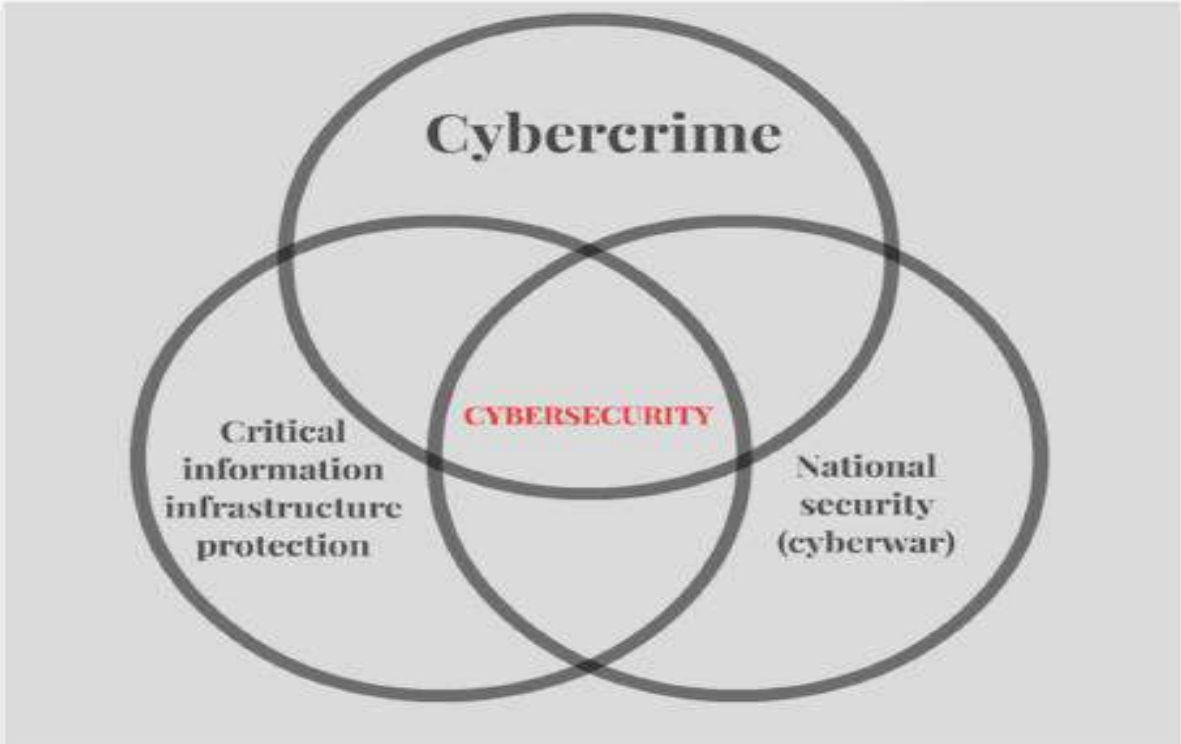
• Overuse of such terms as ‘cyberwar’ and

‘cyber-weapons’

- tendency to view the situation in catastrophic terms

• Legal and regulatory responses: confusion and

misunderstanding



DIFFERENT TYPES OF CYBER CRIMES

Cyber Crimes can be categorized in two ways:

1. The crimes in which the computer is the target. Examples of such crimes are hacking, virus attacks, DOS attack etc.

2. The crimes in

which the computer is used as a weapon. These types of crimes include cyber terrorism, IPR violations, credit card frauds, EFT frauds, pornography etc.

Unauthorized Access and Hacking:

Unauthorized access means any kind of access without the permission of either of the rightful or person in charge of the computer, computer system or computer network. Hacking means an illegal intrusion into a computer system and/or network. Every act committed towards breaking into a computer and/or network is hacking. Hackers write or use ready-made computer programs to attack the target computer.

They possess the desire to destruct and they get the kick out of such destruction. Some hackers hack for personal monetary gains, such as to stealing the credit card information, transferring money from various bank accounts to their own account followed by withdrawal of money. Government websites are the most targeted sites for the hackers.

Cyber Stalking:

In general terms, stalking can be termed as the repeated acts of harassment targeting the victim such as following the victim, making harassing phone calls, killing the victims pet, vandalizing victims property, leaving written messages or objects. Stalking may be followed by serious violent acts such as physical harm to the victim. Cyber Stalking means repeated acts of harassment or threatening behavior of the cyber criminal towards the victim by using internet services. Both kind of Stalkers i.e., Online & Offline – have desire to control the victims life.

Denial of service Attack:

This is an attack in which the criminal floods the bandwidth of the victim's network or fills his e-mail box with spam mail depriving him of the services he is entitled to access or provide. This kind of attack is designed to bring the network to crash by flooding it with useless traffic. Another variation to a typical denial of service attack is known as a Distributed Denial of Service (DDoS) attack wherein the perpetrators are many and are geographically widespread. Many DoS attacks, such as the Ping of Death and Teardrop attacks, exploit limitations in the TCP/IP protocols. For all known DoS attacks, there are

software fixes that system administrators can install to limit the damage caused by the attacks. But, like Virus, new DoS attacks are constantly being dreamed up by Hacker.

Virus attacks:

Viruses are the programs that have the capability to infect other programs and make copies of itself and spread into other program. Programs that multiply like viruses but spread from computer to computer are called as worms. These are malicious software that attach themselves to other software. Virus, worms, Trojan Horse, Time bomb, Logic Bomb, Rabbit and Bacterium are the malicious. Viruses usually affect the data on a computer, either by altering or deleting it. On the other hand worms merely make functional copies of themselves and do this repeatedly till they eat up all the available.

Trojan horse is a program that acts like something useful but do the things that are quiet damping. Trojans come in two parts, a Client part and a Server part. When the victim (unknowingly) runs the server on its machine, the attacker will then use the Client to connect to the Server and start using the Trojan. TCP/IP protocol is the usual protocol type used for communications, but some functions of the Trojans use the UDP protocol as well.

Software Piracy:

Software piracy refers to the illegal copying of genuine programs or the counterfeiting and distribution of products intended to pass for the original. These kind of crimes also include copyright infringement, trademarks violations, theft of computer source code, patent violations etc.

Domain names are also trademarks and protected by ICANN's domain dispute resolution policy and also under trademark laws. Cyber squatters register domain name identical to popular service provider's name so as to attract their users and get benefit from them .

Salami attacks :

These attacks are used for the commission of financial crimes. The key here is to make the

alteration so insignificant that in a single case it would go completely unnoticed. E.g. a bank employee inserts a program, into the bank's servers, that deducts a small amount of money (say Rs. 5 a month) from the account of every customer. No account holder will probably notice this unauthorized debit, but the bank employee will make a sizable amount of money every month.

Phishing:

Phishing is the act of sending an e-mail to a user falsely claiming to be an established legitimate enterprise in an attempt to scam the user into surrendering private information that will be used for identity theft. The e-mail directs the user to visit a web site where they are asked to update personal information, such as passwords and credit card, social security, and bank account numbers that the legitimate organization already has. The Web site, however, is bogus and set up only to steal the user's information. By spamming large groups of people, the phisher counted on the e-mail being read by a percentage of people who actually had listed credit card numbers with legitimately.

Sale of illegal articles:

This category of cyber crimes includes sale of narcotics, weapons and wildlife etc., by posting information on websites, auction websites, and bulletin boards or simply by using email communication.

Online gambling :

There are millions of websites; all hosted on servers abroad, that offer online gambling. In fact, it is believed that many of these websites are actually fronts for money laundering. Cases of hawala transactions and money laundering over the Internet have been reported.

Email spoofing :

Email spoofing refers to email that appears to originate from one source but actually has been sent from another source. Email spoofing can also cause monetary damage.

Cyber Defamation:

When a person publishes defamatory matter about someone on a website or sends e-mails containing defamatory information to all of that person's friends, it is termed as cyber defamation.

Forgery:

Computers, printers and scanners are used to forge counterfeit currency notes, postage and revenue stamps, mark sheets etc. These are made using computers, and high quality scanners and printers.

Theft of information contained in electronic form :

This includes theft of information stored in computer hard disks, removable storage media etc.

Email bombing :

Email bombing refers to sending a large number of emails to the victim resulting in the victim's email account (in case of an individual) or mail servers (in case of a company or an email service provider) crashing.

Data diddling :

This kind of an attack involves altering raw data just before it is processed by a computer and then changing it back after the processing is completed.

Internet time theft :

Internet time refers to usage by an unauthorized person of the Internet hours paid for by another person.

Theft of computer system :

This type of offence involves the theft of a computer, some part(s) of a computer or a peripheral attached to the computer.

Physically damaging a computer system :

This crime is committed by physically damaging a computer or its peripherals.

E-commerce/ Investment Frauds:

An offering that uses false or fraudulent claims to solicit investments or loans, or that provides for the purchase, use, or trade of forged or counterfeit securities. Merchandise or services that were purchased or contracted by individuals online are never delivered. The fraud attributable to the misrepresentation of a product advertised for sale through an Internet auction site or the non-delivery of products purchased through an Internet auction site. Investors are enticed to invest in this fraudulent scheme by the promises of abnormally high profits.

Cyber Terrorism:

Targeted attacks on military installations, power plants, air traffic control, banks, trail traffic control, telecommunication networks are the most likely targets. Others like police, medical, fire and rescue systems etc.

Cyber terrorism is an attractive option for modern terrorists for several reasons.

It is cheaper than traditional terrorist methods.

Cyber terrorism is more anonymous than traditional terrorist methods.

The variety and number of targets are enormous.

Cyber terrorism can be conducted remotely, a feature that is especially appealing to terrorists.

Cyber terrorism has the potential to affect directly a larger number of people.

CYBER LAWS

Cyber crimes are anew class of crimes which are increasing day by day due to extensive use of internet these days. To combat the crimes related to internet The Information Technology Act, 2000 was enacted with prime objective to create an enabling environment for commercial use of I.T. The IT Act specifies the acts which have been made punishable. The Indian Penal Code, 1860 has also been amended to take into its purview cyber crimes.

The various offenses related to internet which have been made punishable under the IT Act and the IPC are enumerated below:

The list of offenses given above is not exhaustive and would also include any other types of offenses that would be committed through a computer or against a computer in the future.

1. Cyber crimes under the IT Act :

Tampering with Computer source documents - Sec.65

Hacking with Computer systems, Data alteration - Sec.66

Publishing obscene information - Sec.67

Un-authorised access to protected system Sec.70 Breach of Confidentiality and Privacy - Sec.72

Publishing false digital signature certificates - Sec.7

2. Cyber Crimes under IPC and Special Laws :

Sending threatening messages by email - Sec 503 IPC

Sending defamatory messages by email - Sec 499 IPC

Forgery of electronic records - Sec 463 IPC

Bogus websites, cyber frauds - Sec 420 IPC

Email spoofing - Sec 463 IPC

Web-Jacking - Sec. 383 IPC

E-Mail Abuse - Sec.500 IPC

3. Cyber Crimes under the Special Acts:

Online sale of Drugs under Narcotic Drugs and Psychotropic Substances Act

Online sale of Arms Arms Act

Web site:

http://www.cyberlawsindia.net/internet-crime.html

https://www.digit.in/technology-guides/fasttrack-to-cyber-crime/the-12-types-of-cyber-crime.html

End of session 4

Streaming Mining,
https://www.youtube.com/watch?v=PGaRwXMGSOM

UNIT V

MDSS : **Data Science for (in) Cyber Security** Session 5

Streaming Mining

Streaming data analysis in real time is becoming the standard to obtain useful knowledge from what is happening right now, allowing organizations to react quickly when problems appear, or to detect new trends, helping them to improve their performance.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Big Data

Motivation

- The world's technological per-capita capacity to **store information doubled** every 40 months
  - As of 2012, 2.5 exabytes (2.5×1018) of data/day
    - Relational database management systems and desktop statistics and visualization packages often **have difficulty** handling big data.
  - Big Data: new driver for digital economy&society
    - Gartner: hundreds of billions of GDP by 2020.
    - Intangible factor after labor and capital
    - Data Science: The fourth paradigm

The Power of Big Data

- Big Data can bring "**big values**" to our life in almost every aspects.
- Technologically, Big Data is bringing about changes in our lives because it allows **diverse and heterogeneous data to be fully integrated and analyzed to help us make decisions**.
- Today, with the Big Data technology, **thousands of data from seemingly unrelated areas can help support important decisions**. This is the power of Big Data.
- Areas of Applications
  - Health and Well being
  - Policy making and public opinions
  - Smart cities and more efficient society
  - New online educational models: MOOC and Student-Teacher modeling
  - Robotics and human-robot interaction
- Much of this power hinges on Research on Analytics

Big Data Stream Mining

The conventional ML setting(batch setting), operates assuming the training data is available as a whole set—any example can be retrieved as needed for little cost.

An alternative is to treat the training data as a **stream**, a potentially endless flow of data that arrives in an order that cannot be controlled.

An algorithm capable of learning from a stream is, by definition, a data mining algorithm.

Placing classification in a data stream setting offers several advantages.

An example of such an application is the monitoring of high-speed network traffic, where the unending flow of data is too overwhelming to consider storing and revisiting.

A stream of items, also called instances or examples, that are continuously arriving.

The main algorithms in data stream mining are:

- Classification
- Regression
- Clustering
- Frequent pattern mining.
-

- Classification and Regression need a set of properly labelled examples to learn a model, so that we can use this model to predict the labels of unseen examples (Supervised Learning).

- When examples are not labelled, one interesting task is to group them in homogeneous clusters. Clustering can be used to obtain user profiles in a website (**Unsupervised Learning**).

Frequent pattern mining looks for the most relevant patterns within the

examples. For instance, in a sales supermarket dataset, it is possible to know what items are bought together and obtain association rules, as for example: ***Most times customers buy cheese, they also buy bread and sauce.***

The most significant requirements for a stream mining algorithm are the same for predictors, clusterers, and frequent pattern miners:

Requirement 1: Process an instance at a time, and inspect it (at most) once.

Requirement 2: Use a limited amount of time to process each instance.

Requirement 3: Use a limited amount of memory.

Requirement 4: Be ready to give an answer (prediction, clustering, patterns) at any time.

Requirement 5: Adapt to temporal changes.

What are the use cases for streaming data?

A stream processing system is beneficial in most scenarios where new and dynamic data is generated continually. It applies to most of the industry segments and big data use cases.

Companies generally begin with simple applications, such as collecting system logs and rudimentary processing like rolling min-max computations. Then, these applications evolve to more sophisticated near real-time processing.

Here are some more examples of streaming data.

Data analysis

Applications process data streams to produce reports and perform actions in response, such as emitting alarms when key measures exceed certain thresholds. More sophisticated stream processing applications extract deeper insights by applying machine learning algorithms to business and customer activity data.

IoT applications

Internet of Things (IoT) devices are another use case for streaming data. Sensors in vehicles, industrial equipment, and farm machinery send data to a streaming application. The

application monitors performance, detects potential defects in advance, and automatically places a spare part order, preventing equipment downtime.

Financial analysis

Financial institutions use stream data to track real-time changes in the stock market, compute value at risk, and automatically rebalance portfolios based on stock price movements. Another financial use case is fraud detection of credit card transactions using real-time inferencing against streaming transaction data.

Real-time recommendations

Real estate applications track geolocation data from consumers' mobile devices and make real-time recommendations of properties to visit. Similarly, advertising, food, retail, and consumer applications can integrate real-time recommendations to give more value to customers.

Service guarantees

You can implement data stream processing to track and maintain service levels in applications and equipment. For example, a solar power company has to maintain power throughput for its customers or pay penalties. It implements a streaming data application that monitors all panels in the field and schedules service in real time. Thus, it can minimize each panel's periods of low throughput and the associated penalty payouts.

Media and gaming

Media publishers stream billions of clickstream records from their online properties, aggregate and enrich the data with user demographic information, and optimize the content placement. This helps publishers deliver a better, more relevant experience to audiences. Similarly, online gaming companies use event stream processing to analyze player-game interactions and offer dynamic experiences to engage players.

Risk control

Live streaming and social platforms capture user behavior data in real time for risk control over users' financial activity, such as recharge, refund, and rewards. They view real-time dashboards to flexibly adjust risk strategies.

End of session 5

Cloud Based DS for Malware Detection,
https://www.youtube.com/watch?v=GfkhuixV1jw

UNIT V

MDSS : **Data Science for (in) Cyber Security** Session 6

Cloud Based DS for Malware Detection

What is cloud-based malware detection?

The cloud-based malware protection feature helps protect endpoints from high risk file types from external sources such as the Internet or network drives by querying FortiGuard to determine whether files are malicious.

CASBs can identify sensitive data, control access to it, and block malicious content and files. They provide real-time threat protection against both known and unknown malware by using advanced threat protection techniques such as machine learning and behavior analytics.

What is FortiGuard used for?

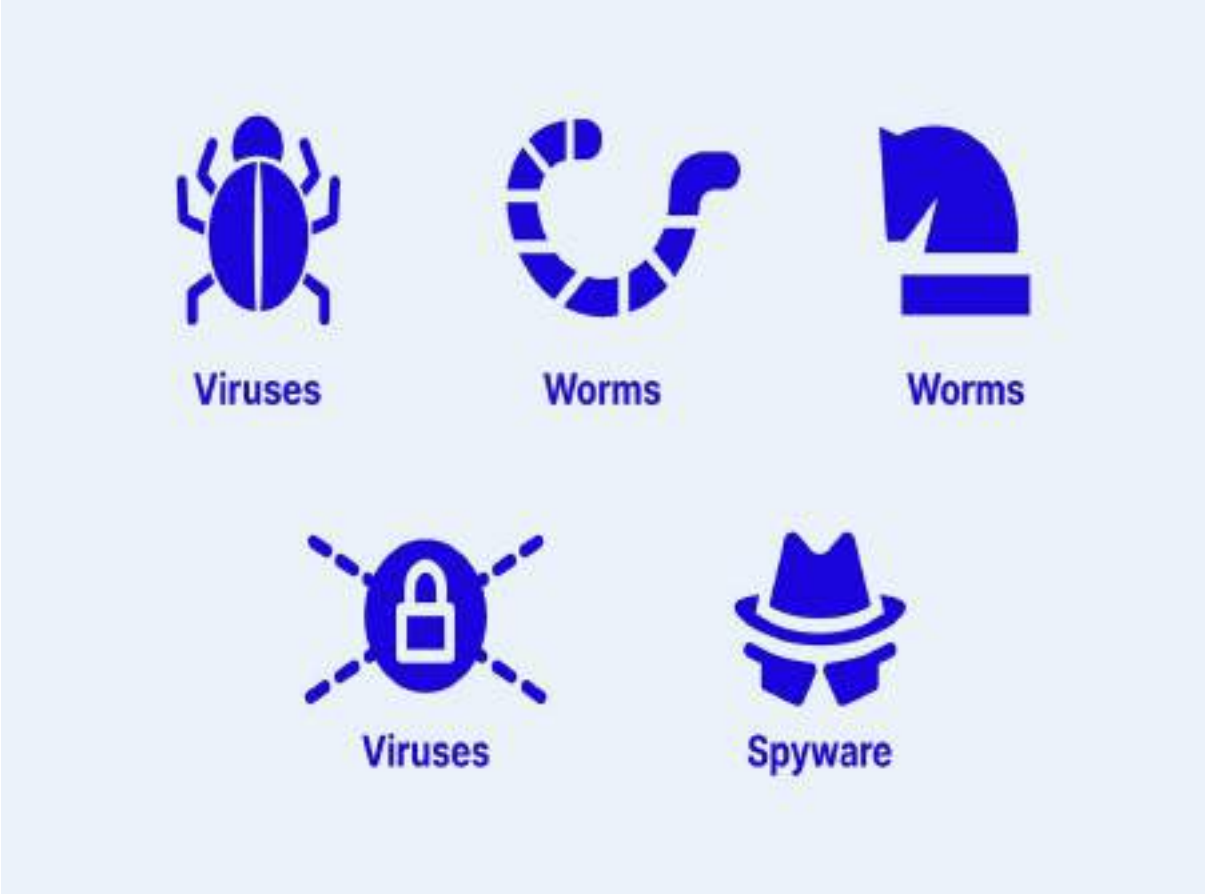
FortiGuard device security services offer advanced capabilities to monitor and protect IT, IoT, and OT devices and applications. FortiGuard NOC/SOC security services enable faster identification, containment, and response to attacks

---

Malware Detection in the Cloud Computing Era

Malware, short for malicious software, refers to any software designed to cause harm to a computer, server, client, or network. The counter to this digital menace is malware detection. Malware detection is a set of practices aimed at identifying the presence and nature of malware in a system. It's a gatekeeper, tasked with identifying, isolating, and eliminating malicious software that could compromise the integrity of your environment.

Modern malware detection is complex, involving multiple techniques and tools designed to address the wide array of malicious software that exists. It's a continual process of vigilance and adaptation, evolving as new types of malware emerge and old ones become more sophisticated.



Viruses

The term “virus” in the context of computers was inspired by biological viruses. Much like their biological counterparts, computer viruses attach themselves to clean files and infect other clean files. They can spread uncontrollably, damaging a system’s core functionality and deleting or corrupting files. They usually appear as an executable file and require human action to propagate.

Examples:

Zeus Gameover: A sophisticated variant of the Zeus Trojan that steals banking credentials using peer-to-peer botnet capabilities.

DarkHotel: A cyber-espionage campaign targeting hotel Wi-Fi networks to spy on high-profile guests.

Triada: Mobile malware targeting Android devices, primarily for financial fraud and data theft.

Worms

Worms are a type of malware that can replicate themselves to spread to other computers over a network. Unlike viruses, they can propagate independently without any human action. They exploit vulnerabilities in operating systems, causing widespread damage by consuming bandwidth or overloading servers. Some worms carry more dangerous payloads, such as ransomware (see below).

Examples:

Stuxnet: A highly sophisticated worm designed to sabotage Iran's nuclear program by targeting industrial control systems.

Raspberry Robin worm: Initially a low-profile worm spread via USB drives since September 2021, it has evolved to be a precursor for severe ransomware attacks, impacting nearly 1,000 organizations.

HermeticWizard: Also known as Foxblade, it's a worm distributed through ".DLL" files, primarily designed to distribute the HermeticWiper malware, with the ability to identify local network IPs and spread the "wiper" via WMI or SMB protocols.

Trojans

Trojans, named after the ancient Greek story of the Trojan horse, are deceptive types of malware that disguise themselves as legitimate software. Once downloaded and installed, they create a backdoor in a system's security, allowing cybercriminals to gain control and access sensitive information.

Examples:

TOITOIN: A Windows-based banking trojan active since 2023, employing advanced techniques for infection and evasion.

PlugX: A remote access tool (RAT) associated with several cyber-espionage campaigns, often targeting entities in Asia.

Carberp: A banking Trojan known for stealing financial information from infected hosts.

Ransomware

Ransomware is a particularly malicious type of malware that encrypts the victim's files, then demands a ransom to restore access. It holds the user's data hostage, often threatening to delete or publish it if the ransom is not paid within a certain timeframe.

Examples

WannaCry: A ransomware worm that exploited the Windows SMB protocol and affected systems worldwide in 2017.

NotPetya: A destructive malware initially masquerading as ransomware, designed primarily to damage and disrupt systems.

Hive: Advanced malware associated with the APT group named Chafer, primarily used for espionage activities.

Spyware

Spyware is a type of malware designed to spy on the user's activity without their knowledge. It collects information such as keystrokes, browsing habits, and personal information, which can then be used for identity theft or other illicit activities.

Examples

PhoneSpy: A malicious software targeting mobile devices to siphon off sensitive user data.

Android/SpyC23.A: A mobile malware strain targeting Android devices, often used for espionage purposes.

Pegasus: A sophisticated spyware developed by NSO Group, capable of infiltrating smartphones and extracting user data without detection.

Techniques for Malware Detection

Signature-Based Detection

Signature-based detection is the most common method for detecting malware. This technique involves identifying unique patterns or "signatures" in the code of known malware and then scanning files to look for these signatures. When a match is found, the file is flagged as potentially malicious.

However, signature-based detection has its limitations. First, it can only identify known malware. This means that if a new or altered version of malware is released, the system won't be able to recognize it until its signature has been added to the database. Secondly, malware developers are always finding ways to evade signature-based detection by obfuscation or encryption, rendering their malware unrecognizable to the system.

Static File Analysis

Static file analysis is another method used for malware detection. This technique involves analyzing a file without executing it to understand its behavior and identify any potential threats. It examines the internal structure, the code, and other metadata of the file.

The advantage of static file analysis is that it can detect malware without having to run the malicious code. Another advantage is that it can detect unknown malware using techniques like heuristic analysis, which involves making an educated guess about the maliciousness of a file based on its characteristics.

However, like signature-based detection, static file analysis is not foolproof. Some malware can disguise itself so that it appears benign during static analysis. Others can thwart static analysis by encrypting their code or by using other obfuscation techniques.

Dynamic Malware Analysis

Dynamic malware analysis involves running the suspicious file in a controlled environment (like a sandbox) and observing its behavior. This method can reveal a lot about the malware's functionality and the potential damage it can cause.

The advantage of dynamic analysis is that it can discover malware that static file analysis and signature-based detection might miss, especially if the malware uses advanced evasion techniques. It can also provide a more complete picture of the malware's behavior, such as what changes it makes to the system, what network activity it initiates, and so on.

However, dynamic malware analysis can be time-consuming and requires a more sophisticated setup than the other methods. Moreover, some advanced malware can detect when they are being run in a sandbox and behave differently or refuse to run at all, thereby evading detection.

Checksumming

Checksumming is a method used to verify the integrity of a file. By calculating a checksum (a unique value derived from the file's content) and comparing it with a known good checksum, you can determine if the file has been altered in any way – for instance, by malware.

Checksumming is a simple and effective way to detect malware, especially if you have a baseline checksum for comparison. It's also very fast and doesn't require much computational power. However, checksumming can only detect changes in a file, not the presence of malware per se. Therefore, it's usually used in conjunction with other malware detection techniques.

Application Allow listing

This strategy involves creating a list of trusted software applications that are permitted to run on your computer system. If an application isn't on this list, it's denied execution, regardless of whether it's malware or not.

Application allowlisting is a proactive approach to malware detection. By only allowing approved applications to run, you significantly reduce the risk of malware infection. However, this technique requires careful management and regular updating of the allowlist to accommodate new, safe applications.

Block listing

The next malware detection technique is blocklisting, which is essentially the opposite of allowlisting. Instead of creating a list of trusted applications, you create a list of known malicious or potentially harmful applications. Any application on this list is denied execution, protecting your computer system from known threats.

Blocklisting is reactive, as it depends on identifying and listing malware after it's been discovered. It's effective against known threats, but it can't protect your system from new, unknown malware. Therefore, like allowlisting, blocklisting should be used as part of a multi-layered security strategy.

Machine Learning Behavioral Analysis

Machine learning behavioral analysis is a way to automatically analyze the behavior of applications and identify any abnormal or malicious activities.

Machine learning behavioral analysis differs from traditional detection techniques in that it doesn't rely on predefined malware signatures. Instead, it learns from the data it's fed and improves its detection capabilities over time. This makes it highly effective at detecting new, unknown malware.

Detecting Malware in the Cloud: Cloud Native Security Services

Cloud computing environments are more complex, and more dynamic, than traditional on-premise data centers. This makes it more difficult to deploy anti-malware technology and ensure systems are safe, raising the need for dedicated cloud native security solutions.

Cloud native security services are specifically designed to protect and secure applications running in cloud environments. These services are integrated within the cloud platform and offer a layer of security spanning across cloud resources, such as virtual machines, containers, and serverless functions. Some cloud native security solutions provide real-time malware detection.

These services provide a view of the security posture across multiple cloud resources, enabling you to monitor and respond to threats effectively. Many cloud native security services also perform automated security operations, such as automatically scanning resources for malware and cleaning systems when malware is detected.

Malware Detection Tools and Services for Cloud Environments

Cloud Access Security Brokers (CASBs)

Cloud Access Security Brokers (CASBs) act as a gatekeeper between your on-premise infrastructure and the cloud service provider. They provide visibility into your cloud applications and enforce security policies to protect your data. CASBs play a crucial role in malware detection by monitoring data traffic and files passing between your network and the cloud.

CASBs can identify sensitive data, control access to it, and block malicious content and files. They provide real-time threat protection against both known and unknown malware by using advanced threat protection techniques such as machine learning and behavior analytics. Also, CASBs provide detailed logs and alerts, helping you respond promptly to any malware threats.

Cloud Workload Protection Platforms (CWPPs)

Cloud Workload Protection Platforms (CWPPs) provide security for workloads running in the cloud. They offer a suite of security capabilities, including malware detection, vulnerability management, and network segmentation. CWPPs monitor the behavior of workloads to detect any unusual activities that could indicate a malware attack.

CWPPs provide real-time protection against malware attacks, reducing the dwell time of threats and minimizing the potential damage. They offer a centralized view of the security posture across all workloads, enabling you to better manage the security of your cloud environment. Moreover, CWPPs provide automated response capabilities, helping you mitigate the impact of malware attacks more effectively.

Cloud Security Posture Management (CSPM) Tools

Cloud Security Posture Management (CSPM) tools help you maintain an optimal security posture in the cloud. They provide continuous visibility into your cloud configurations and assess them against security best practices, helping you identify and rectify any misconfigurations that could expose you to malware attacks.

CSPM tools provide real-time monitoring of your cloud environment, ensuring that malware detection and protection systems are enabled on every cloud resource. They offer actionable insights into your security posture, helping you make informed decisions about your security strategy. Moreover, CSPM tools automate the process of managing your security posture, reducing the manual effort required and ensuring a consistent level of protection against malware.

Best Practices for Malware Protection in the Cloud

Regular Monitoring and Auditing of Cloud Resources

It is important to keep a constant watch over your cloud environment and analyze your security logs to identify any unusual activities that could indicate a malware attack.

Regular monitoring provides real-time visibility into your cloud resources, enabling you to detect and respond to malware threats promptly. Auditing, on the other hand, helps you understand the security events that have occurred in your cloud environment, providing you with insights into your threat landscape.

Implementing a Strong IAM Policy

Implementing a strong Identity and Access Management (IAM) policy is another essential practice for effective malware detection. It involves managing who has access to your cloud resources and what they can do with that access. A strong IAM policy can help you prevent unauthorized access to your cloud resources, reducing the risk of malware attacks.

A strong IAM policy includes principles like least privilege, where users are granted only the permissions they need to perform their tasks, and segregation of duties, where critical tasks are divided among multiple users. Implementing such a policy not only helps you protect your cloud resources from malware but also provides you with a clear view of who has access to your resources, enabling you to detect any unauthorized activities promptly.

Encrypting Data at Rest and in Transit

Encrypting data at rest and in transit is a crucial practice for protecting your data against malware attacks. It involves converting your data into a format that can only be read by those who have the decryption key. By encrypting your data, you can ensure that even if a malware attack occurs, your data remains safe.

Encrypting data at rest protects your stored data from malware attacks, while encrypting data in transit protects your data as it moves across networks. Both types of encryption are essential for a robust defense against malware.

Keeping Cloud Software and Services Updated

Keeping your cloud software and services updated is a vital practice for effective malware detection. It involves regularly updating your cloud platforms, applications, and security tools to ensure that you have the latest security patches and features. By keeping your

software and services updated, you can protect your cloud environment from known malware threats and enhance your ability to detect new ones.

Regular updates not only provide you with the latest security patches but also introduce new security features and improvements. This not only enhances your defense against malware but also improves your overall security posture. Moreover, regular updates to security software provide you with the latest threat intelligence, helping you stay ahead of the evolving malware landscape.

End of Session 6

End of session 6