Data analytics for Cyber security

-Introduction-

Vandana P. Janeja

Outline

- What is Cybersecurity?
- Assets Affected
  - Personal assets
  - Public assets
  - Corporate assets at risk
- Motivation, Risks and Security
  - Why do we have security risks?
  - What is the level of damage that can occur?
- Handling Cyber Attacks
  - Sub areas of Cybersecurity
- Data Analytics
  - Why Data Analytics is important for cybersecurity: A case study of understanding the anatomy of an attack

What is Cybersecurity?

Cybersecurity refers to securing valuable electronic assets and physical assets, which have electronic access, against unauthorized access. These assets may include personal devices, networked devices, information assets, and infrastructural assets, among others.

Cybersecurity deals with security against threats also referred to as cyber threats or cyberattacks. Cyberattacks are the mechanism by which security is breached to gain access to assets of value.

Aims of Cybersecurity: prevent, detect, and respond to threats

- Prevention of cyberattacks against critical assets
- Detection of threats
- Respond to threats in the event that they penetrate access to critical assets
- Recover and restore the normal state of the system in the event that an attack is successful

Assets Affected

Personal	Public	Corporate
• Phones (home and mobile), • Tablets , • Personal computers (desktop and laptops), • External physical hard drive, • Cloud drive, • Email accounts, • Fitness trackers, • Smart watches, • Smart glasses, • Media devices (TIVO, apple TV, cable box), • Bank accounts, • Credit cards, • Personal gaming systems	• Smart meters, • Power grid, • Sewage controls, • Nuclear power plant, • Rail lines, • Airplanes and air traffic, • Traffic lights, • Citizen databases, • Websites (county, state and federal), • Space travel programs • Satellites	• Customer database, • Websites, • Business applications, • Business network, • Emails, • Off the shelf software, • Intellectual property

Motivation behind Cyber Threats

1. Stealing intellectual property
2. Gaining access to customer data
3. Making a political statement
4. Performing cyber espionage
5. Damaging reputation, Making a splash/ for fun, Impedeing access to data and applications

Why do we have security risks?

Applications with several dependencies,

Logical errors in software code (such as Heartbleed),

Organizational risks (multiple partners, such as in cyber-attacks at Target and the Pacific Northwest National Laboratories [PNNL]),

Lack of user awareness of cybersecurity risks (such as in social engineering and phishing),

Personality traits of individuals using the systems (phishing), and

Inherent issues in the Internet protocol being used.

Summary of Motivation, Risks and Security

Motivation

- To steal Intellectual property
- To damage reputation
- Gain access to data , which can then be sold
- Gain access to information, which is not generally available
- To make a political statement
- To impede access to critical data and applications
- To make a splash/ for fun

Risks

- Internet protocol which is inherently not secure
- Applications with several dependencies
- Logical errors in software code (ex. Heartbleed)
- Organizational risks (multiple partners ex. Target, PNNL)
- Lack of User awareness of cybersecurity risks (ex. Social engineering, phishing)
- Personality traits of individuals using the systems

Attaining Security

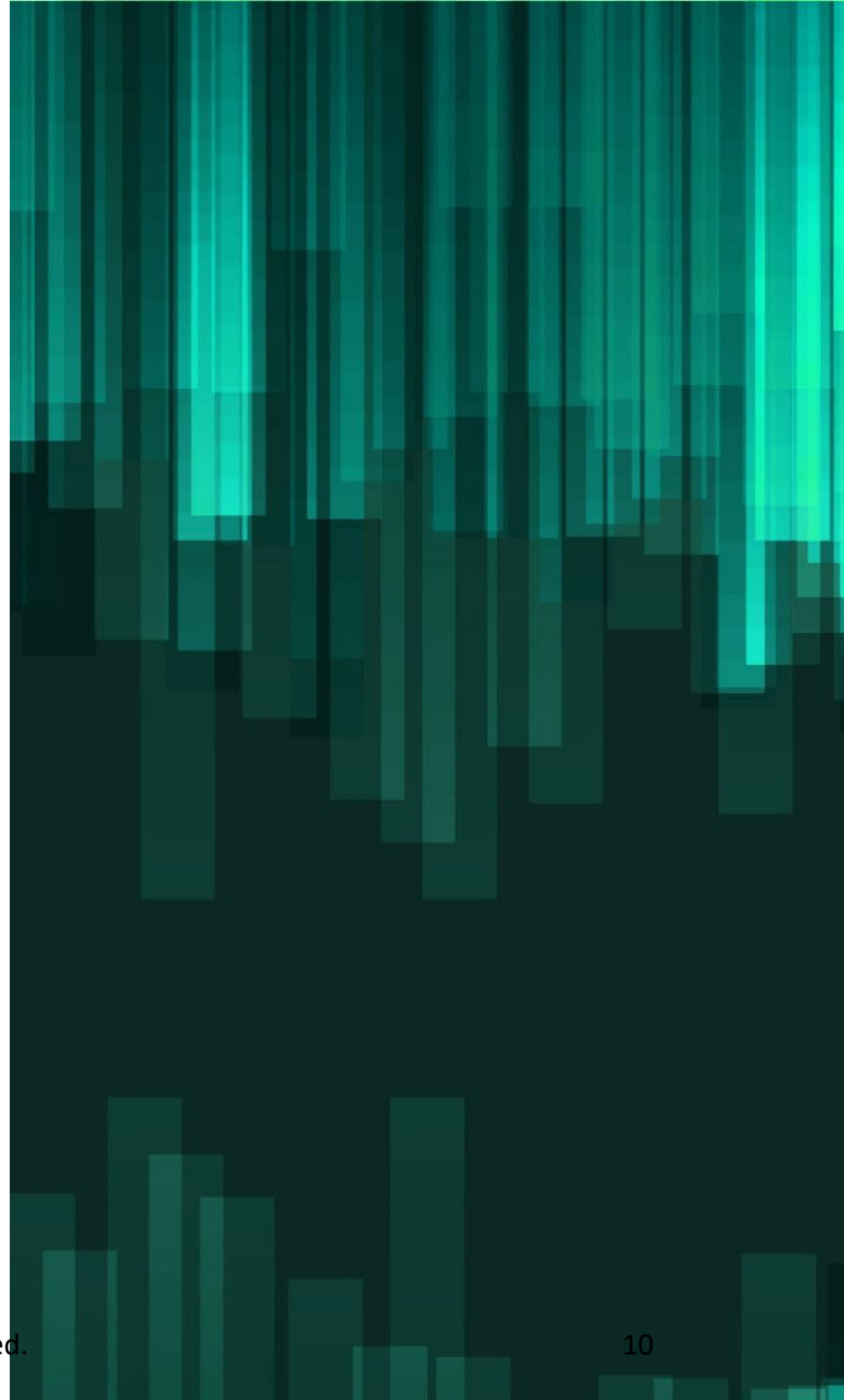
- Protecting resources
- Hardening defenses
- Capturing data logs
- Monitoring systems
- Tracing the attacks
- Predicting risks
- Predicting attacks
- Identifying vulnerabilities

What is the level of damage that can occur?

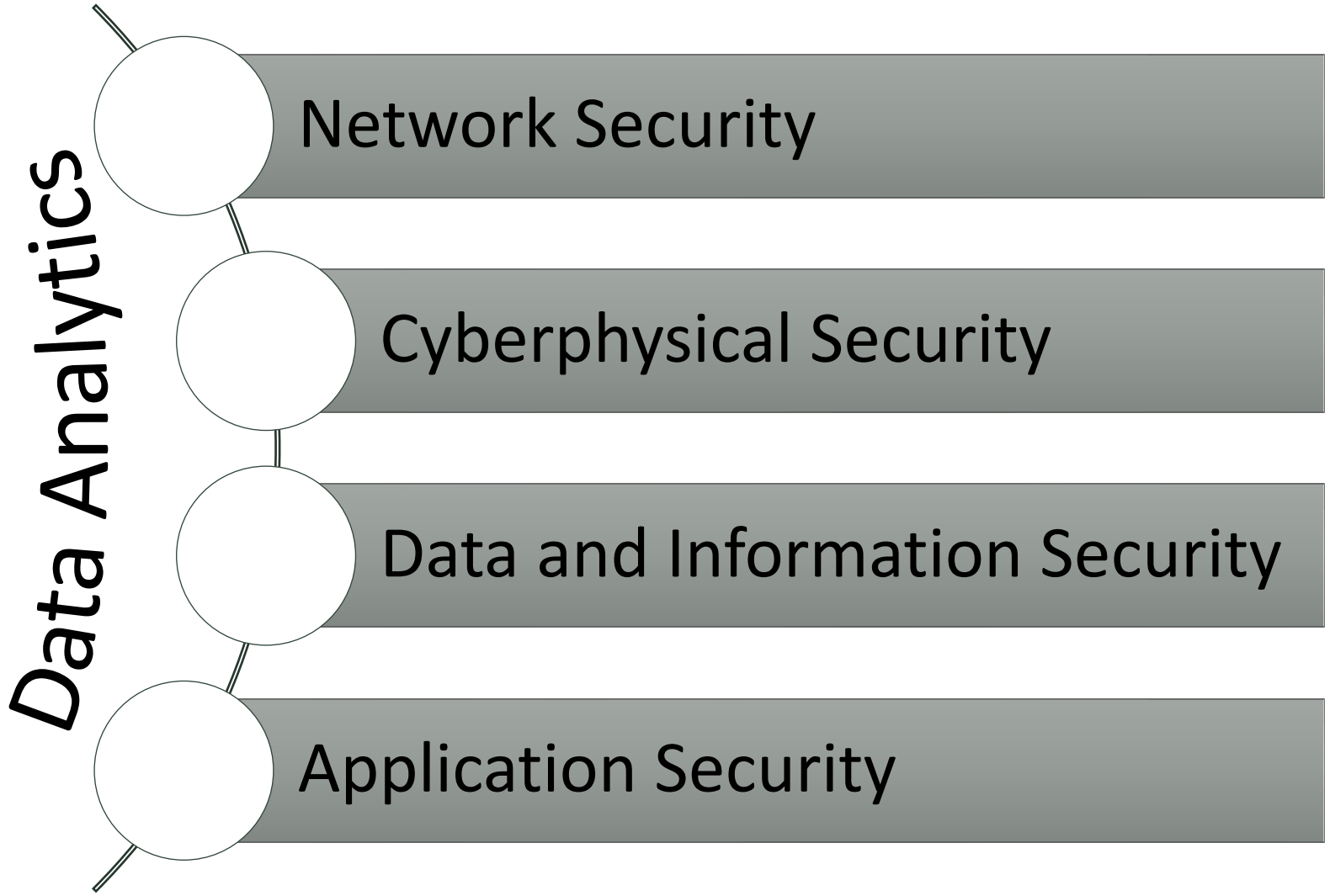
- According to a McAfee report, the monetary loss resulting from cybercrime costs about $600 billion, which that is about 0.8% of the world Gross Domestic Product (GDP) (McAfee––Cybercrime Impact 2018), with malicious actors becoming more and more sophisticated.
- The loss due to cyber-attacks is not simply based on direct financial loss but also based on several indirect factors, which that may lead to a major financial impact.
- Example: Target cyber-attack (RSkariachan and Finkleeuters-Target 2014)
  - Target reported $61 million in expenses related to the cyber-attack out of which $44 million were covered by insurance.
  - The direct financial impact to Target was $17 million.
  - A 46 % drop in net profit in the holiday quarter,
  - 5.5% drop in transactions during the quarter,
  - share price fluctuations led to further losses,
  - cards had to be reissued for to several customers, and
  - Target had to offer identity protection to affected customers.
- All these losses amount to much more than the total $61 million loss. In addition, the trust of the customers was lost, which is not a quantifiable loss and has long-term impacts.

Handling Cyber Attacks

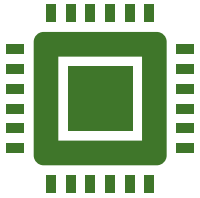
- Protecting resources,
- Hardening defenses,
- Capturing data logs,
- Monitoring systems,
- Tracing the attacks,
- Predicting risks,
- Predicting attacks, and
- Identifying vulnerabilities



Overall Areas of Cybersecurity

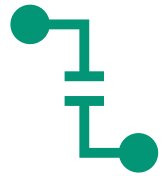


Sub areas of Cybersecurity



Application security: incorporating security in the software development process.

Data and information security: securing data from the risk of unauthorized access and misuse



Network security: securing the traditional computer networks and security measures adopted to secure, prevent unauthorized access and misuse of either the public or the private network.

Sub areas of Cybersecurity

Cyber physical security

- Emerging challenges due to the coupling of the cyber systems with the physical systems.
  - The power plants being controlled by a cyber system,
  - risk of disruption of the cyber component or
  - risk of unauthorized control of the cyber system,
  - gaining unauthorized control of the physical systems.

Sub areas of Cybersecurity

Data analytics

- Cross cutting across areas to learn from existing threats and develop solutions for novel and unknown threats towards networks, infrastructure, data, and information
  - Example: Threat hunting proactively looks for malicious players across the myriad data sources in an organization
  - Does not necessarily have to be a completely machine-driven process and should account for user behaviors
  - Must look at the operational context.
  - Provide security analysts a much focused field of vision to security analysts to zero in on solutions for potential threats

Hardware and Network Landscape

- Multiple types of networks and devices
  - computer networks, Cyber Physical Systems (CPS), Internet of Things (IoT), sensor networks, smart grids, and wired or wireless networks.
- Computer networks - Traditional type of networks
  - Groups of computers are connected in pre-specified configurations. These configurations can be designed using security policy deciding who has access to what areas of networks. Another way networks form is by determining patterns of use over a period of time. In both cases, zones can be created for access and connectivity where each computer in the network and sub-networks can be monitored.
- Cyber Physical Systems - an amalgamation of two interacting sub-systems, cyber and physical
  - used to monitor and perform the day- to- day functions of the many automated systems that we rely on, including power stations, chemical factories, and nuclear power plants, to name a few.
- Ubiquitous connected technology - “smart” things - Internet of Things

Data Analytics

- Data analytics deals with analyzing large amounts of data from disparate sources to discover actionable information leading to gains for an organization.
  - Includes techniques from data mining, statistics, and business management, among other fields.
- Big data
  - Massive datasets (volume)
  - Generated at a rapid rate (velocity)
  - Heterogeneous nature (variety)
  - Can provide valid findings or patterns in this complex environment (veracity)
  - Changing by location (venue)
- Every device, action, transaction, and event generates data. Cyber threats leave a series of such data pieces in different environments and domains. Sifting through these data can lead to novel insight not why a certain event occurred and potentially allow the identification of the responsible parties and lead to knowledge for preventing such attacks in the future.

Anatomy of an attack

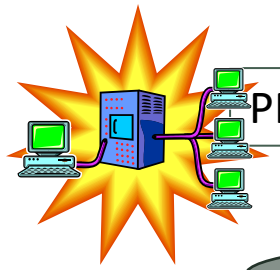
vulnerability in one of the lab's public-facing web servers

Drive by attack

PCs of site visitors (lab employees)

Compromised Workstations

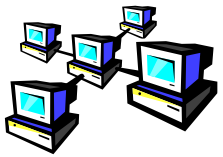
PNNL's network



scouting from the compromised workstations for weeks

Shared Network resources

Spear Phishing attack



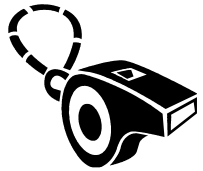
Business Partners

root domain controller compromised

Obtained a privileged account

recreate and elevate account privileges

Raise alert

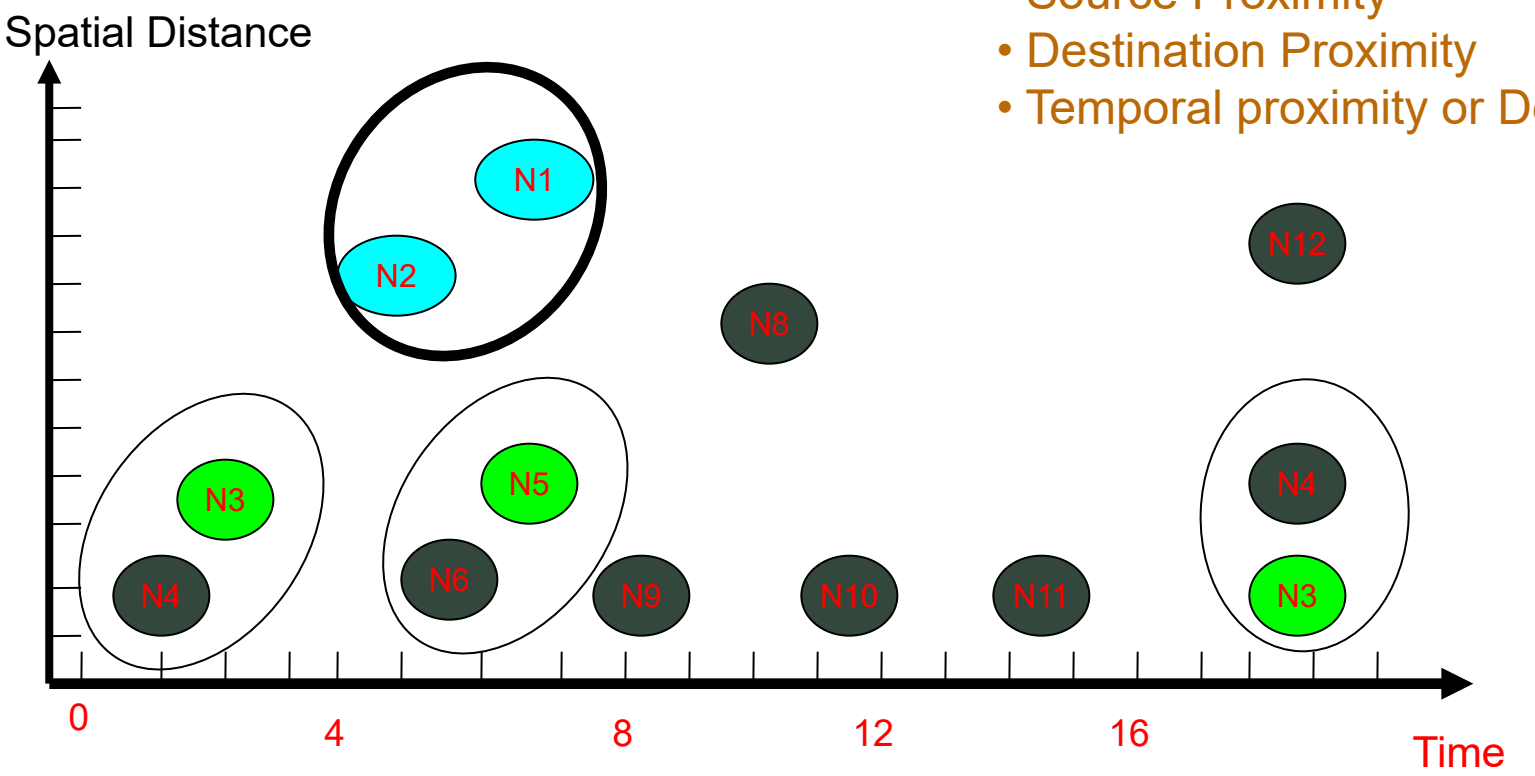


Why Data Analytics is important for cybersecurity: A case study of understanding the anatomy of an attack

- The three aspects are temporal, spatial, and data -driven understanding of human behavioral aspects (particularly of attackers)
- Firstly, computer networks evolve over time, and communication patterns change over time. Can we identify these key changes, which deviateare deviant from the normal changes in a communication pattern, and associate them with anomalies in the network traffic?
- Secondly, attacks may have a spatial pattern. Sources and destinations in certain key geo locations are more important for monitoring and preventing an attack. Can key geo locations, which are sources or destinations of attacks, be identified?
- Thirdly, any type of an attack has common underpinnings of how it is carried out; this has not changed from physical security breaches to computer security breaches. Can this knowledge be leveraged to identify anomalies in the data where we can see certain patterns of misuse?
- Utilizing the temporal, spatial, and human behavioral aspects of learning new knowledge from the vast amount of cyber data can lead to new insights of understanding the challenges faced in this important domain of cybersecurity

Multi-dimensional view of Threats

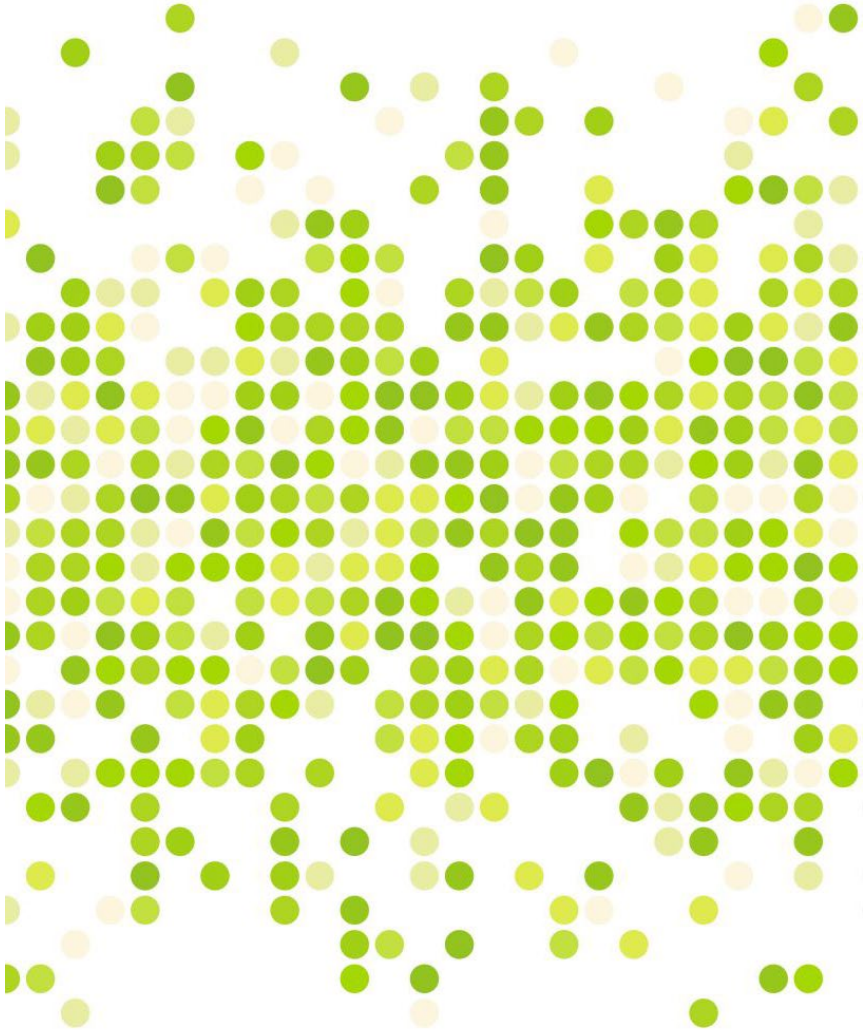
- Events become relevant when they occur *together*
- These events become relevant with proximities rather than causation
- The two items are in close Proximity, based on
  - Source Proximity
  - Destination Proximity
  - Temporal proximity or Delay



Goal : to identify potential "collusions" among the entities responsible for these two events

Why Data Analytics is important for cybersecurity: A case study of understanding the anatomy of an attack

- Looking at one dimension of the data is not enough in such prolonged attack scenarios.
- For such a multipronged attacks, we need a multilevel framework
  - Brings together data from several different databases.
  - Events of interest can be identified using a combination of factors such as proximity of events in time, in terms of series of communications and even in terms of the geographic origin or destination of the communication.



Understanding the Anatomy of an attack: Clustering based on feature combinations

- Intruder Detection System (IDS) logs such as SNORT
- A keyword matrix and a word frequency matrix
- Perform alarm clustering and alarm data fusion
- Identify critical alerts (a combination of log entries)
- Perform clustering based on a combination of features

Understanding the Anatomy of an attack: Collusions and associations

- Extract associations to identify potentially repeated or targeted communications
- Utilize network mapping
- Determine attacks consistently targeted to specific types of machines or individuals

Understanding the Anatomy of an attack: Time proximity and network evolution

- Time intervals accounts for time proximity
- Allows mining the data in proximity of time
- Evaluating how the networks evolve over time
- Identify which time interval may be critical: for example, Identify repeated events of interest in certain time periods
- Clustering in different segments of time
- Mining for possible attack paths based on variations in cluster content and cluster cohesion

How Can Data Analytics Help?

Data from multiple sources can be used to glean novel information

Supports the defense of cyber systems

Tracing Attacks,

Predicting risks

Identifying critical systems in a network flow,

predicting attacks based on prior or similar attacks,

Identifying vulnerabilities by mining software code,

Understanding user behavior by mining network logs, and

Creating robust access control rules by evaluating prior usage and security policies.

Focus of this Course

What this course is not about: This course does not address the traditional views of security configurations and shoring up the defenses, including, setting up computer networks, setting up firewalls, web server management, and patching of vulnerabilities.

What this course is about: This course addresses the challenges in cybersecurity that data analytics can help address, including analytics for threat hunting or threat detection, discovering knowledge for attack prevention or mitigation, discovering knowledge about vulnerabilities, and performing retrospective and prospective analysis for understanding the mechanics of attacks to help prevent for preventing them in the future.

References

- Digital Attack Map: A Global Threat Visualization https://www.netscout.com/global-threat-intelligence Last accessed Nov, 2020
- Check Point - Threat Cloud, https://threatmap.checkpoint.com/ Last accessed March, 2020
- Cyberthreat Real-Time Map, https://cybermap.kaspersky.com/ , Last accessed Nov, 2020
- Target data flood stolen-card market, https://www.washingtonpost.com/business/economy/target-cyberattack-by-overseas-hackers-may-have-compromised-up-to-40-million-cards/2013/12/20/2c2943cc-69b5-11e3-a0b9-249bbb34602c_story.html?utm_term=.42a8cd8b6c0e , Last accessed Nov, 2016
- Alexandra Whitney Samuel, Hactivism and future of Political Participation, http://www.alexandrasamuel.com/dissertation/pdfs/Samuel-Hacktivism-entire.pdf , Sept 2004
- CSMonitor-Estonia: Arthur Bright, Estonia accuses Russia of 'cyberattack', 2007, http://www.csmonitor.com/2007/0517/p99s01-duts.html , last accessed March 2020
- SANS: Maxwell Chi, Cyberspace: America's New Battleground, https://www.sans.org/reading-room/whitepapers/warfare/cyberspace-americas-battleground-35612, 2014
- James. A. Lewis, Computer Espionage, Titan Rain and China, 2005, http://csis.org/files/media/csis/pubs/051214_china_titan_rain.pdf, Last accessed March 2020
- Reuters-Solarwinds: Jack Stubbs, Raphael Satter, Joseph Menn, U.S. Homeland Security, thousands of businesses scramble after suspected Russian hack, 2020, https://www.reuters.com/article/global-cyber/global-security-teams-assess-impact-of-suspected-russian-cyber-attack-idUKKBN28O1KN
- FireEye: Pascal Geenens, FireEye Hack Turns into a Global Supply Chain Attack, 2020, https://securityboulevard.com/2020/12/fireeye-hack-turns-into-a-global-supply-chain-attack/
- Bloomberg-Target: Matt Townsend, Lindsey Rupp and Jeff Green, Target CEO Ouster, http://www.bloomberg.com/news/2014-05-05/target-ceo-ouster-shows-new-board-focus-on-cyber-attacks.html, 2014
- Google Dorking, Amy Gesenhues, http://searchengineland.com/google-dorking-fun-games-hackers-show-202191, 2014
- Risk Based Security-Sony, A Breakdown and Analysis of the December, 2014 Sony Hack
- , https://www.riskbasedsecurity.com/2014/12/05/a-breakdown-and-analysis-of-the-december-2014-sony-hack/ , 2014
- McAfee, The Economic Impact of Cybercrime—No Slowing Down, McAfee, Center for Strategic and International Studies (CSIS), 2018, https://www.mcafee.com/enterprise/en-us/solutions/lp/economics-cybercrime.html
- Reuters-Target: Target shares recover after reassurance on data breach impact , 2014, https://www.reuters.com/article/us-target-results/target-shares-recover-after-reassurance-on-data-breach-impact-idUSBREA1P0WC20140226 , Last accessed March 2020
- Hiren Sadhwani, Introduction to Threat Hunting, 2020, https://medium.com/@hirensadhwani2619/introduction-to-threat-hunting-8dff62ba52ca
- M. Shashanka, M. Shen and J. Wang, "User and entity behavior analytics for enterprise security," *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 1867-1874, doi: 10.1109/BigData.2016.7840805.
- Manyika, James, et al. "Big data: The next frontier for innovation, competition, and productivity." (2011).
- Chen, Min, Shiwen Mao, and Yunhao Liu. "Big data: a survey." Mobile Networks and Applications 19.2 (2014): 171-209.
- PNNL Attack: 7 Lessons: Surviving A Zero-Day Attack, 2011 http://www.darkreading.com/attacks-and-breaches/7-lessons-surviving-a-zero-day-attack/d/d-id/1100226 , Last accessed March 2020