Data analytics for Cyber security

-Understanding sources of Cybersecurity data-

Vandana P. Janeja

Outline

End to End Opportunities for data collection

Sources of cybersecurity data

- **Log Data**
  - Router Connectivity and Log Data
  - Firewall Log Data
- **Raw pay load data**
- **Network Topology data**
- **User system data**
- **Other Datasets**

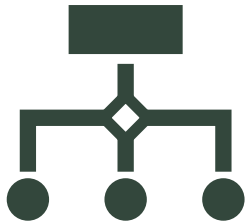
Integrated use of multiple datasets

Summary of sources of Cybersecurity data

Sources of Cybersecurity Data

- Cyber threats often lead to loss of assets
- Multitude of datasets can be harvested and used to track these losses and origins of the attack
- This chapter is not about the data lost during cyberattacks but the data that organizations can scour from their networks to understand threats better so that they can potentially prevent or even predict future attacks

End to End Opportunities for data collection



The information systems used to perform business functions have a well-defined process spanning over connected systems

A typical client server scenario

A user connects to a system via an internet pipeline.

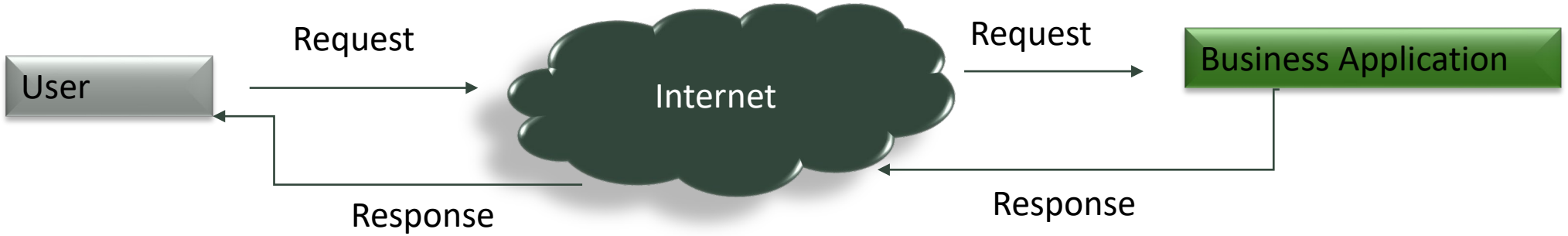
The system has built-in application functionality important to run the business function

A return pipeline sends a response back to the user

The functionality of the system allows the delivery of the information commodity requested by the user

Logical View

Logical View (a)



The logical view of the user requesting access to a business application can appear to be fairly straightforward

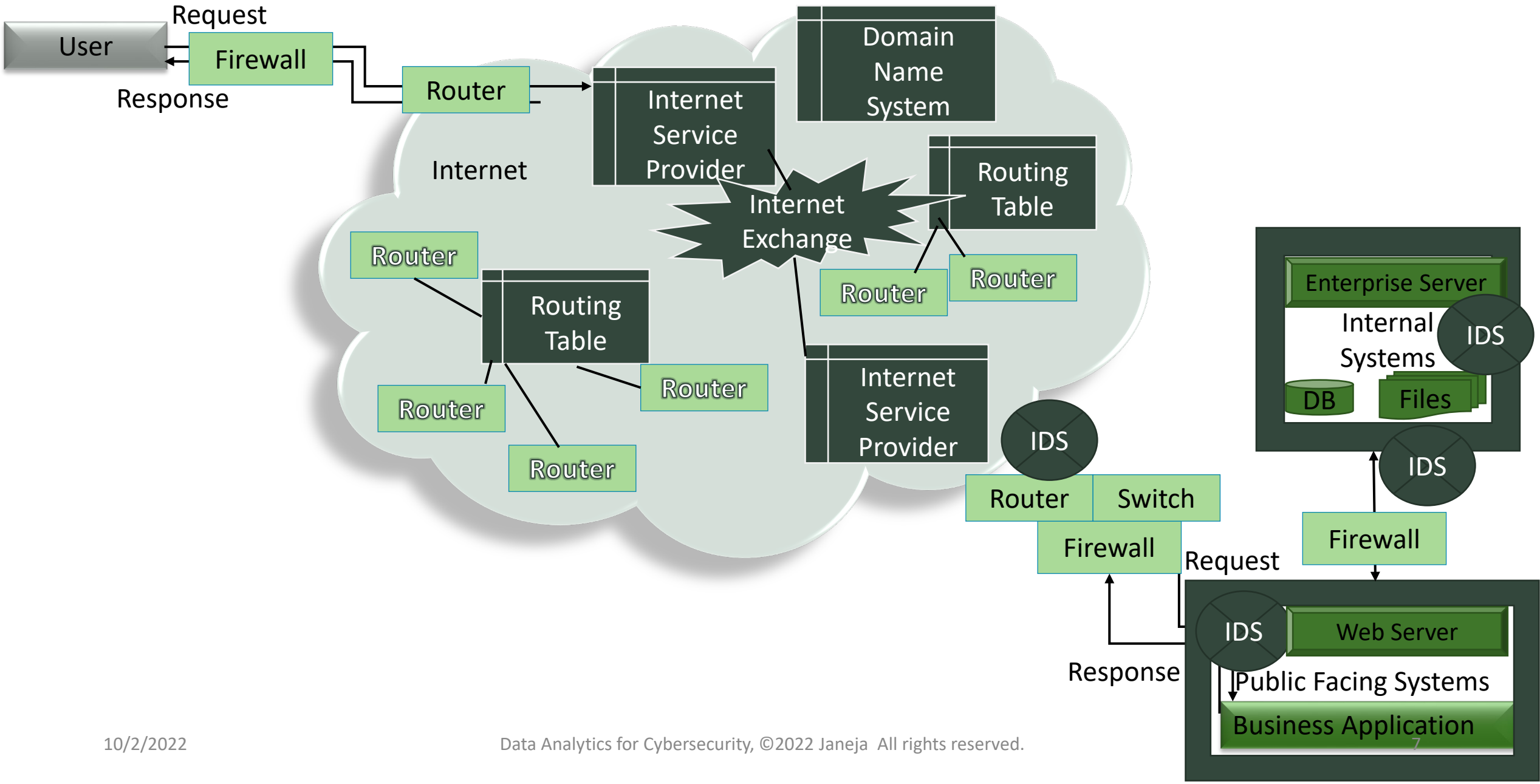
Within this pipeline there could be several points through which the request and response pass

Leading to several opportunities in the end-to-end process for data collection to help understand when a cyber threat may occur in this process

Physical View

- User request on a network– follow a complex networking pipeline
  - The user may have a firewall on their own system and the router through which they send out the request
  - This request can be filtered through the internet service provider
  - lookups can be performed in the domain name system (DNS) and
  - the data can be routed through multiple paths of routers, which are linked through the routing table
- The request on the other side may again have to pass through the routers and firewalls at multiple points in the system being accessed by the user
  - There may be multiple intrusion detection systems (IDS) posted throughout the systems to monitor the network flow for malicious activity
- This is just one example scenario; different network layouts will result in different types of intermediate steps in this process of request and response, particularly based on
  - the type of response
  - the type of network being used
  - the type of organization of business applications
  - the cloud infrastructure being used
- However, certain key components are always present that allow for multiple opportunities to glean and scour for data related to potential cyber threats

Physical View



Opportunities for Data Collection - to understand potential threats.

Data collection begins at

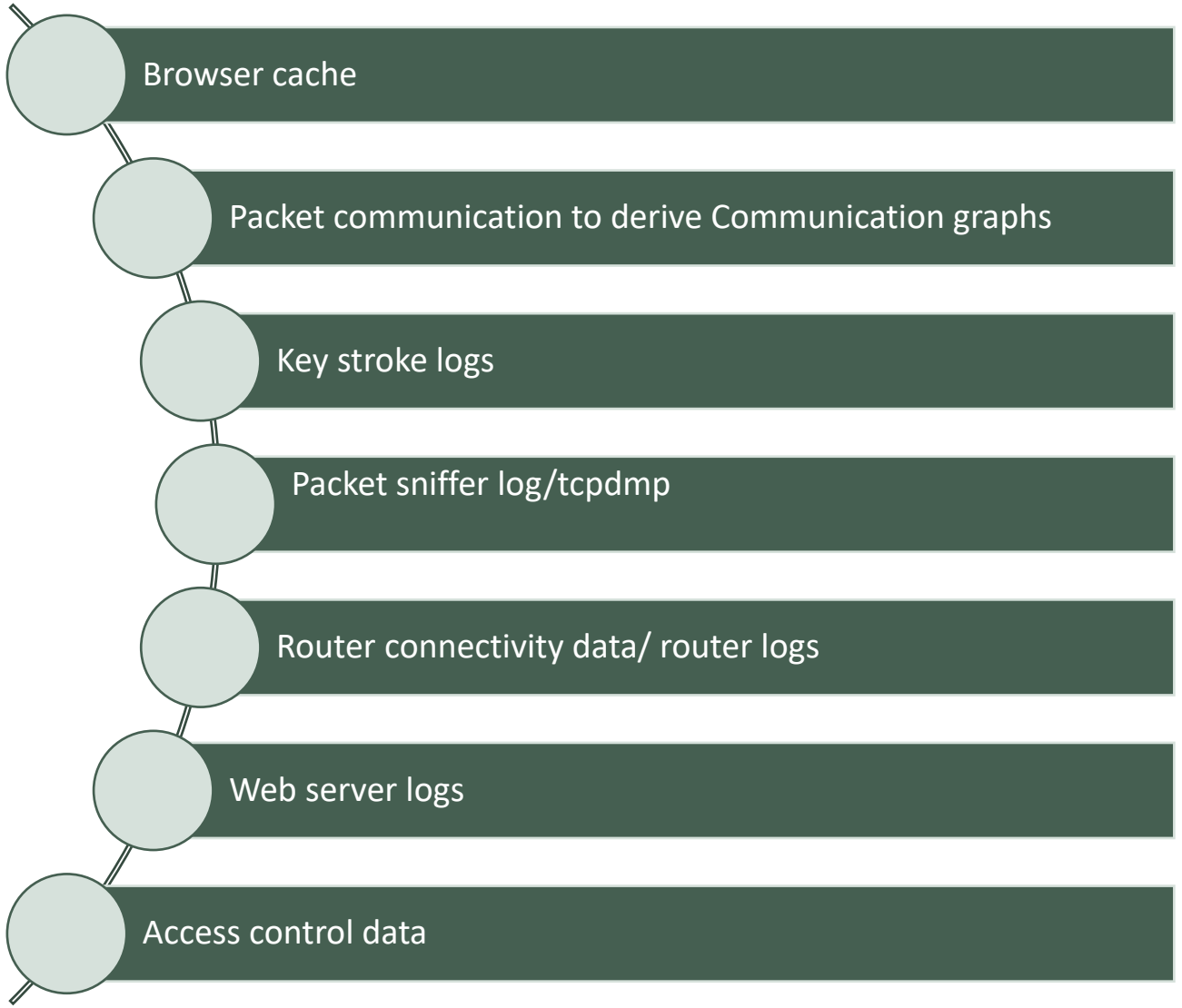
- a user access point
- system functionality level, and
- commodity level (particularly if the data is being delivered)

Example – Questions to consider at the user level

- Who is the user?
- The psychology of the user, personality types, etc.
- What type of interface is being used by the user?
- Is there clear information about what is acceptable or not acceptable in the interface? (c) What type of access system is being used? Is there access control for users?
- What data are available about the access pipeline, such as the type of network or cloud being used?

Common types of cybersecurity data

- Several common types of datasets can be collected and evaluated, including various types of log data such as key stroke logs, web server logs, and intrusion detection logs, to name a few



Sources of Cybersecurity Data and Variations

- Cybersecurity-related data collection will vary across the type of networks, including computer networks, sensor networks, or cyberphysical systems
- The method and level of data collection will also vary based on the application domains for which the networks are being used and the important assets being protected
    - Social media businesses, such Facebook, are primarily user data driven, where the revenue is based on providing access to user data and monitoring usage data
    - E-commerce businesses, such as Amazon, are usage and product delivery based
    - Portals, such as Yahoo, are again user data driven but more heavily reliant on advertisements, which can target users based on what they see and use most often
    - Cyberphysical systems, such as systems for monitoring and managing power grids, are based on accurate functioning of physical systems and delivery of services to users over these physical infrastructural elements
- The level of monitoring and management of data will vary with the level of prevention, detection, or recovery expected in the domain
- Some domains have a high emphasis on prevention; others may have a high level of emphasis on detection or recovery
- In all such cases, multiple types of datasets can be collected to provide intelligence on the cyber threats, and user behaviors can be evaluated to prevent future threats or even identify an insider propagating the threats

Log Data

Intrusion detection system (IDS) logs including alarms raised by IDS;

- Alerts are raised by matching any known signatures of malicious activities in the header and payload data
- Logs analyze the packets based on malicious signatures and provide information on time stamp, service used, protocol, source, and destination
- IDS can be placed at various points in a network, and multiple such datasets can be collected and correlated

Key stroke logs

- To capture every key being pressed on a keyboard
- Capture actions such as copying materials to the clipboard or other interactions with the user system

Log Data

Router connectivity data/ router logs;

- Routers not only provide route information but also all the raw IP addresses that pass through it
- IP addresses can be mapped to identify possible malware activity when data is sent to suspicious geolocations in an unauthorized manner
- Router data can also be utilized to study and possibly identify traffic hijacking and bogus routes by looking at historic route data stored in a knowledge base

Firewall logs

- Firewalls are typically designed to look at the header information in the data packets to match against pre specified rule sets
- Firewalls differ from IDS since they are generally limited to header information screening whereas IDS can look at the payload data as well and block connections with malicious signatures

Raw Pay Load Data

Data over the network contains

The header information, which stores data about source and destination among other things and

the actual content being transmitted, referred to as payload

There are several privacy concerns in accessing this payload data since this data is the actual content that is being sent which may be under strict access controls

Payload data can be accessed only where legally allowed and users have provided permissions to access this data

This data may be encrypted, so its usefulness as raw data to be mined is limited

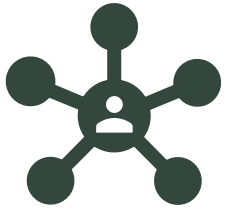
Payload data is accessible through packet sniffers such as Wireshark, where the data dump of the traffic can be retrieved

Payload data can be massive even for a few minutes of data capture

Raw Pay Load Data - Uses

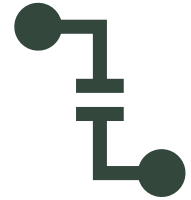
- To discover individual user's behavior
- To detect presence of malwares in the payloads
- To detect other security threats based on the actual content of the payload
- To identify threats based on signatures of malwares that may be present in the payload
  - For example, if a virus is embedded in a packet and this virus has a known signature then this can be captured by traditional intrusion detection system rules
  - Packets with malware embedded in them can be detected using multiple mechanisms such as simple keyword searches or complex regular expression matches and flagged
  - The traffic can be blocked or marked for further analysis, such as using Snort alarms or Wireshark coloring rules

Network Topology Data



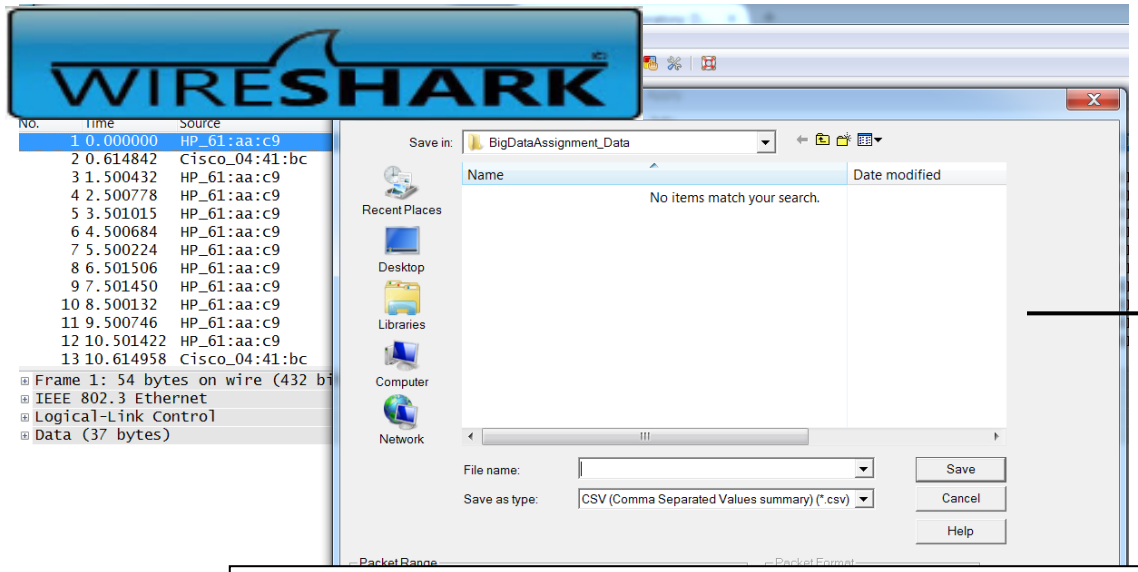
A computer network can be represented as a graph in terms of the structure of the network and in terms of the communication taking place over the networks

Network traffic data dump can be used to generate the communication graphs



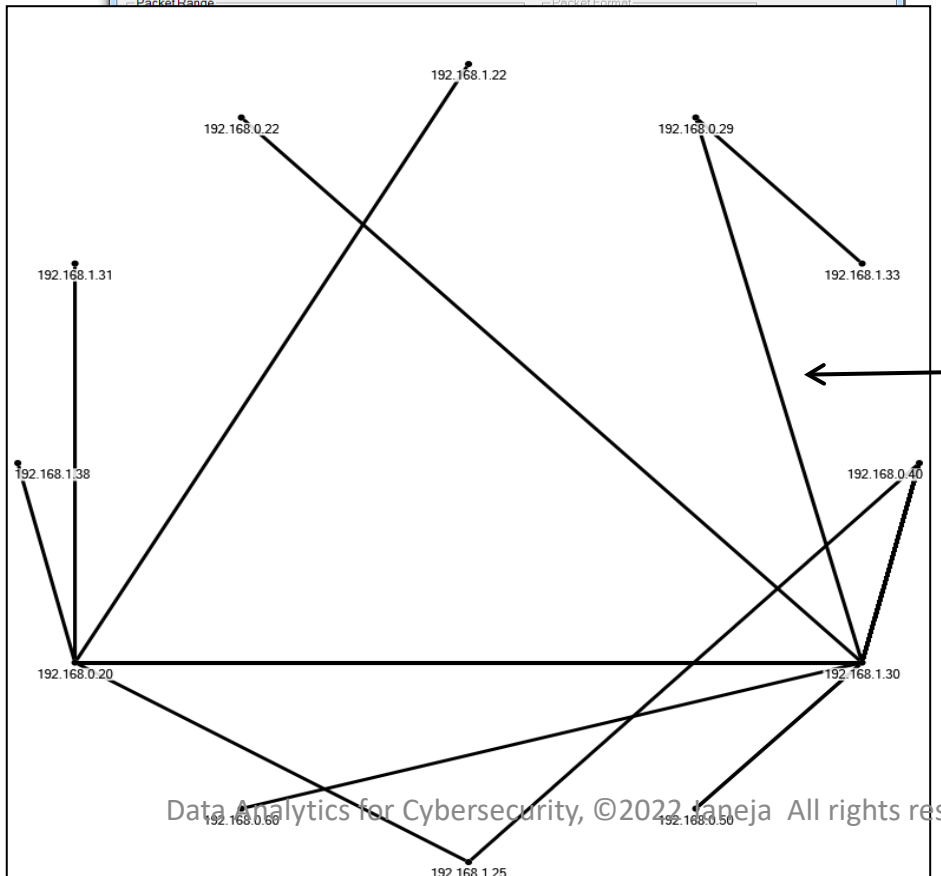
Header data collected from a traffic dump file through Wireshark can be utilized to plot the communication between the source and destination IP addresses, which become the vertices of each edge in the graph

Network Topology



Data Preprocessing

Vertex 1	Vertex 2	Color
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.1.2	192.168.0.20	
192.168.1.3	192.168.0.22	
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.0.2	192.168.1.30	
192.168.0.2	192.168.1.30	
192.168.1.3	192.168.0.20	
192.168.1.2	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.20	
192.168.0.2	192.168.1.30	
192.168.1.3	192.168.0.20	
192.168.0.2	192.168.1.30	
192.168.0.2	192.168.1.30	
192.168.0.2	192.168.1.30	
192.168.1.3	192.168.0.20	
192.168.1.3	192.168.0.29	
192.168.1.3	192.168.0.29	



Example extraction of communication graph from network traffic

Header data collected from a traffic dump file through Wireshark can be utilized to plot the communication between the source and destination IP addresses, which become the vertices of each edge in the graph.

Communications to Graphs

Communication data in the graph form,

- graph metrics can be computed such as node level metrics including centrality, page rank and network level metrics such as diameter, density
- Based on the network properties future predictions can also be made about the network evolution

The example illustrates one such task, in a sample traffic dump data.

- Data from network traffic can be collected through packet sniffers such as Wireshark
- The data from the network traffic is preprocessed and this preprocessing will change with the task being performed

Example - Network traffic to Graph Analysis

- Analysis by day of the week.
  - The traffic data is sorted by the day of the week to get patterns by day, such as all Mondays, all Tuesdays
- Compute the degree of each node by day of the week.
  - Can be performed by specific date also
- Can sort the IP addresses by their degrees across days of the week and the top ones appear to be consistently present in the traffic
- Nodes with low degrees can also be identified
  - In such a scenario it would be interesting to find a node which is highly consistent as a high degree node to appear in the lower degree list indicating a shift in the traffic pattern
- Consider the bar chart of the degrees for each IP address across each day of the week. We can observe that some IP addresses are consistently higher across all days of the week, which is further illustrated by the plot for IP1, IP2, and IP3 across all days of the week
- Degree of IP9 and IP7 seem to be higher on some days but lower on other days
- Clarified by the plot for IP9, which shows Wednesday as a day where IP9 has inconsistent behavior

WIRESHARK

Data Preprocessing

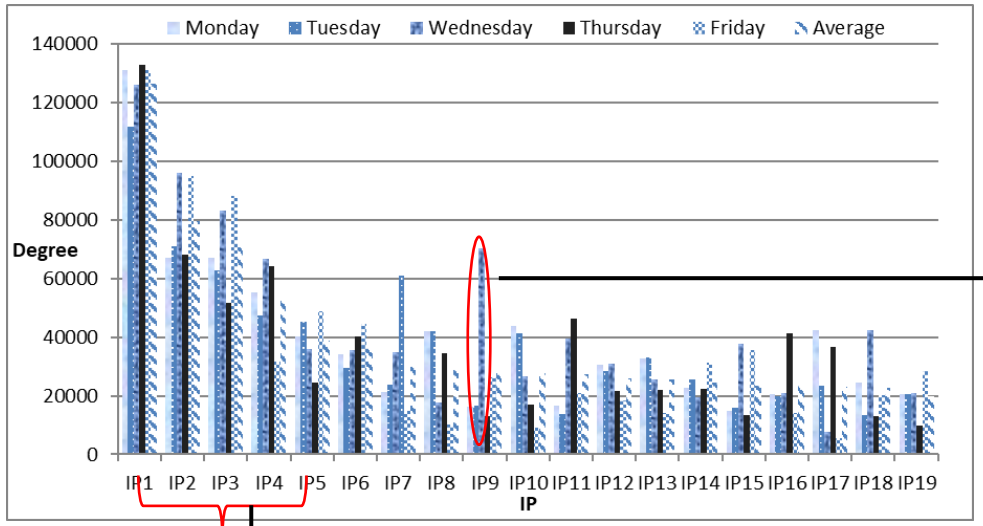
ID	Time	Source	Destination	Protocol	Length	Info	DayOfWeek
1	0	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
2	0.614842	Cisco_04:41:bc	Cisco_04:41:bc	LOOP	118	Reply	Monday
3	1.500432	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
4	2.500778	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
5	3.501015	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
6	4.500684	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
7	5.500224	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
8	6.501506	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
9	7.50145	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
10	8.500132	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
11	9.500746	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
12	10.50142	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
13	10.61496	Cisco_04:41:bc	Cisco_04:41:bc	LOOP	118	Reply	Monday
14	12.00152	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday
15	13.0004	HP_61:aa:c9	HP_61:aa:c9	LLC	54	U P, func=TEST; DSAP NULL LSAP Individual, SSAP NetBIOS Command	Monday

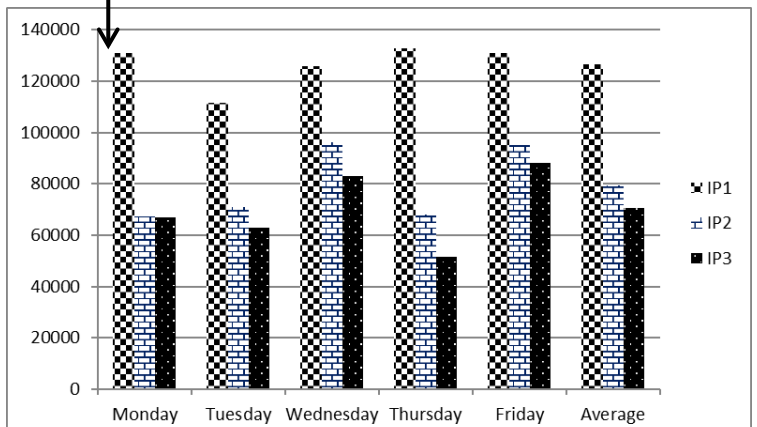
Consistently Central Nodes

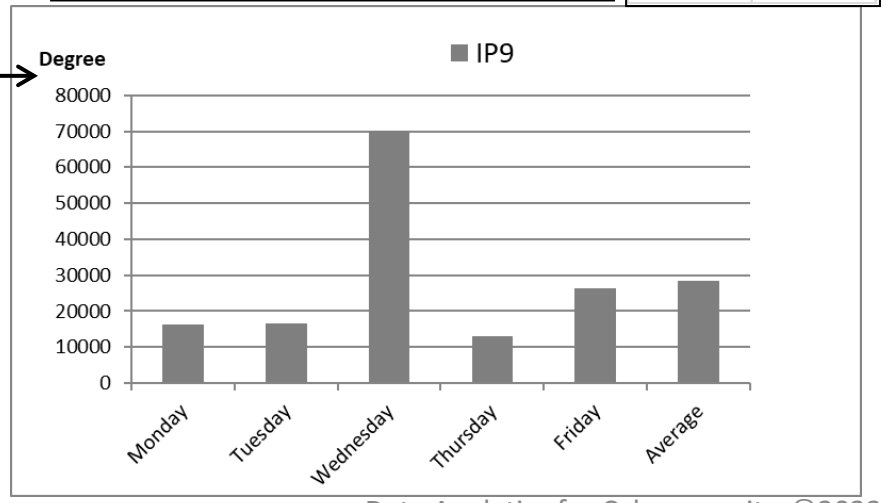
	Monday	Tuesday	Wednesday	Thursday	Friday	Average
IP1	131102	111662	125942	132846	130788	126468
IP2	67138	71083	96073	68084	95038	79483.2
IP3	66826	62722	83025	51658	88219	70490
IP4	55023	47465	66643	64093	31606	52966
IP5	40120	45089	35835	24626	48834	38900.8
IP6	34296	29614	35447	40116	44573	36809.2
IP7	21365	23757	34677	60977	14980	31151.2
IP8	42043	42046	17620	34496	10305	29302
IP9	16360	16455	70181	13043	26243	28456.4
IP10	43743	41197	26674	16867	9186	27533.4
IP11	16642	13627	39616	46150	21056	27418.2
IP12	30477	28288	30733	21490	18483	25894.2
IP13	32864	33012	25680	21880	14237	25534.6
IP14	22739	25507	19570	22308	31388	24302.4
IP15	14875	15960	37694	13526	35597	23530.4
IP16	20709	19761	20987	41284	14005	23349.2
IP17	42364	23275	7512	36506	5389	23009.2
IP18	24420	13220	42303	13150	20214	22661.4
IP19	20462	20539	20749	9778	28250	19955.6

High centrality

	Monday
IP1	131102
IP2	67138
IP3	66826
IP4	55023
IP5	40120
IP6	34296
IP7	21365
IP8	42043
IP9	16360
IP10	43743
IP11	16642
IP12	30477
IP13	32864
IP14	22739
IP15	14875
IP16	20709
IP17	42364
IP18	24420
IP19	20462







User System Data

- Key features can be extracted to monitor unusual activities at the individual system level
  - Examples: active process resident memory usage, which is available for all operating systems (OS) and allows for building a profile on the normal memory usage of a process over time
  - An abnormal spike in memory usage can be attributed to processing a large volume of data
  - Useful in detecting a potential insider threat, especially when integrated with other user behavioral data from sensors monitoring user stress levels or integrating with other log datasets
  - CPU time utilization can be used for measuring system usage
  - Several OS-specific features, such as kernel modules and changes in registry values
- It is important to use multiple signatures over time from several of the features to eliminate the regular spikes of day-to-day operations
- Key differentiator for a robust analysis where we do not simply rely on one or two features but multiple features and their stable signatures (as compared to historical data) to distinguish alerts
- Tools such as OSQuery and Snare can facilitate capture of these features

Key Features To Monitor Unusual Activities at the Individual System Level

Feature Name	OS Specific
Active process name, Active process filesystem path, Active process ports and sockets, Active process file access, Active process resident memory usage, Active process CPU time utilization, Active process system calls, Active process priority value, Active process owner and group information, Loaded peripherals drivers, Key-store access Patterns	All OS
Loaded kernel modules	Linux/Unix
Loaded Kernel Extensions	Mac OS X
Change in registry values	Windows
File System Journaling (meta-data) information	All Major File System (NTFS, ext4, HFS+)
Network Routing Tables	All Major OS
Network Firewall rules	All Major Firewall implementations
System level sensors (current, voltage in different bus inside PC, CPU/GPU fan speed etc)	Almost all peripherals

Other Datasets

- **Access control data:** These data can help better understand usage of the assets that need to be protected. Role mining from access control data can help shape and create better and more robust roles
- **Eye tracker data:** A user's behavior can be judged by the interactions of the user with the system being used. One such mode of input is the screen. Data collected from the user's eye gaze, captured through an eye tracker, can help analyze the user's level of engagement with a system and user preferences or positioning important items on the screen
- **Vulnerability data:** Software vulnerability is a defect in the system (such as a software bug) that allows an attacker to exploit the system and potentially pose a security threat. Vulnerabilities can be investigated, and trends can be discovered in various operating systems to determine levels of strength or defense against cyberattacks

Example: NVD Datasets

National Vulnerability Database from the National Institute of Standards and Technology (NIST)

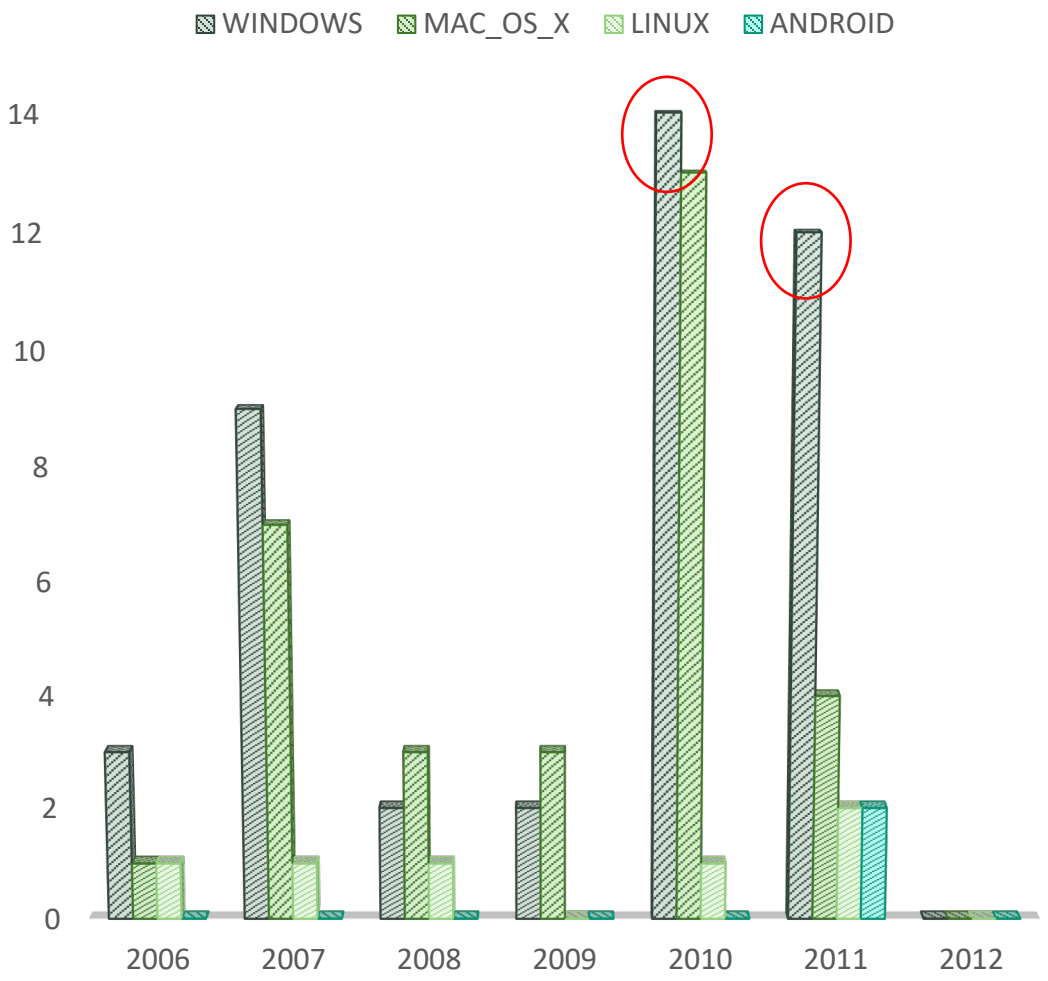
Trends can be analyzed for several years and across major releases for operating systems to reinforce knowledge of choices for critical infrastructural or network projects

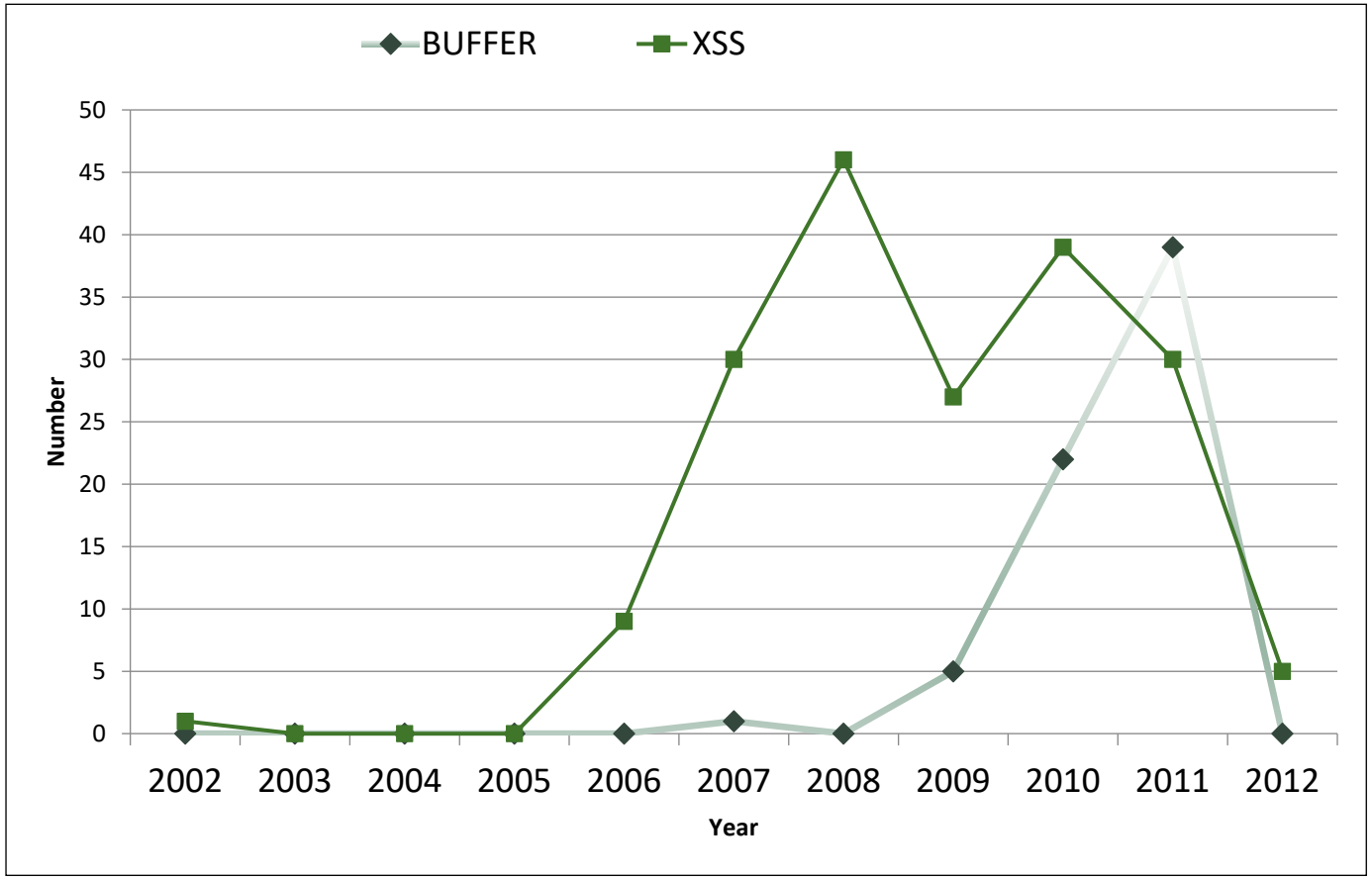
NVD is built on the concept of Common Vulnerabilities and Exposures (CVE), which is a dictionary of publicly known vulnerabilities and exposures

CVEs allow the standardization of vulnerabilities across products around the world. NVD scores every vulnerability using the Common Vulnerability Scoring System (CVSS)

CVSS is comprised of several submetrics, including (a) base, (b) temporal, and (c) environmental metrics. Each of these metrics quantifies some type of feature of a vulnerability

Cross-site Scripting vulnerability:
Data regarding the number of vulnerabilities pulled from NVD across 2006 to 2012
Comparing the occurrences of different types of vulnerabilities such as cross-site scripting and buffer overflow





Integrated Use of Multiple Datasets

- If multiple datasets result in similar types of anomalies, then the credibility of labeling an anomaly is higher
  - Example: to discover anomalies in network traffic data with a temporal, spatial, and human behavioral perspective
  - Studying how network traffic changes over time, which locations are the sources, where is it headed, and how are people generating this traffic – all these aspects become very critical in distinguishing the normal from the abnormal in the domain of cybersecurity
  - This requires shifting gears to view cybersecurity as a holistic people problem rather than a hardened defense problem

Integrated Use of Multiple Datasets: Key Questions to Consider

Computer networks evolve over time, and communication patterns change over time

- Can we identify these key changes that are deviant from the normal changes in a communication pattern and associate them with anomalies in the network traffic?

As attacks may have a spatial pattern, sources and destinations in certain geolocations can be more important for monitoring and preventing an attack

- Can key geolocations that are sources of attacks, or key geolocations that are destinations of attacks, be identified?
- Can IP spoofing be mitigated by looking at multiple data sources to supplement the knowledge of a geospatial traffic pattern?

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 behavioral models of anomalies where we can see patterns of misuse?

Summary of Sources of Cybersecurity Data

Source of cybersecurity data	Literature study examples	Type of detection it can be used for
Keystroke logging	Simon 2007 , Gupta et al. 2016, Cai and Hao 2011, Muzzamil et al. 2016	User behavior, malicious use to detect user credentials
IDS log data	Deokar and Hazarnis 2012, Vaarandi and Podiņš 2010, Quader and Janeja 2015, Quader and Janeja 2014, Chen et al. 2014, Janeja et al. 2014, Abad et al. 2003, Koike and Ohno 2004	Association rule mining, human behavior modeling, log visualization, temporal analysis, anomaly detection
Router connectivity and log data	Sklower 1991, Tsuchiya 1988, Geocoding Infosec 2013, Kim Zetter Security 2013 , Jian 2007	Suspicious rerouting, traffic hijacking, bogus routes
Firewall log data	Golnabi et al. 2006, Abedin et al. 2010	Generate efficient rule sets, anomaly detection in policy rules
Raw payload data	Wang and Stolfo 2004, Kim et al. 2014, Limmer and Dressler 2010, Parekh et al. 2006, Roy 2014	Malware detection, embedded malware, user behavior
Network topology	Massicotte et al. 2003, Nicosia 2013, Namayanja and Janeja 2015 and 2017,	Consistent and inconsistent nodes, time points corresponding to anomalous activity
User system data	Stephens and Maloof 2014, Meigham 2016	User profiles, user behavior data, insider threats
Access control Data	Vaidya et al. 2007, Mitra et al. 2016	Generate efficient access control roles
Eye tracker data	Darwish and Bataineh 2012	Browser security indicators, security cues, user behavior
Vulnerability data	Frei et al. 2006	Vulnerability trend discovery

References

- Heron, Simon. "The rise and rise of the keyloggers." *Network Security* 2007.6 (2007): 4-6.
- Gupta, Haritabh, et al. "Deciphering Text from Touchscreen Key Taps." *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer International Publishing, 2016.
- Cai, Liang, and Hao Chen. "TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion." *HotSec* 11 (2011): 9-9.
- Hussain, Muzammil, et al. "The rise of keyloggers on smartphones: A survey and insight into motion-based tap inference attacks." *Pervasive and Mobile Computing* 25 (2016): 1-25.
- Deokar, Bhagyashree, and Ambarish Hazarnis. "Intrusion Detection System using log files and reinforcement learning." *International Journal of Computer Applications* 45.19 (2012): 28-35.
- Vaarandi, Risto, and Kārlis Podiņš. "Network ids alert classification with frequent itemset mining and data clustering." *2010 International Conference on Network and Service Management*. IEEE, 2010.
- Quader, Faisal, Vandana Janeja, and Justin Stauffer. "Persistent threat pattern discovery." *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*. IEEE, 2015.
- Chen Song, Janeja V., Human Perspective to Anomaly Detection for Cybersecurity, Journal of Intelligent Information Systems, Journal of Intelligent Information Systems, February 2014 (Accepted 2013) , Volume 42, Issue 1, pp 133-153
- Quader, Faisal; Janeja, Vandana, Computational Models to Capture Human Behavior in Cybersecurity Attacks Academy of Science and Engineering (ASE), USA, ©ASE 2014, 2014-06-16
-
- Janeja, Vandana P., et al. "B-dids: Mining anomalies in a Big-distributed Intrusion Detection System." *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 2014.
- Abad, Cristina, et al. "Log correlation for intrusion detection: A proof of concept." *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*. IEEE, 2003.
- Koike, Hideki, and Kazuhiro Ohno. "SnortView: visualization system of snort logs." *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*. ACM, 2004.
- Sklower, Keith. "A tree-based packet routing table for Berkeley unix." *USENIX Winter*. Vol. 1991. 1991.
- Tsuchiya, Paul F. "The Landmark Hierarchy: A new hierarchy for routing in very large networks." *ACM SIGCOMM Computer Communication Review*. Vol. 18. No. 4. ACM, 1988.
- Qiu, Jian, et al. "Detecting bogus BGP route information: Going beyond prefix hijacking." *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on*. IEEE, 2007.

References

- Kim Zetter Security, WIRED, Someone's Been Siphoning Data Through a Huge Security Hole in the Internet, https://www.wired.com/2013/12/bgp-hijacking-belarus-iceland/ 2013, Last accessed 12/26/16
- Geocoding-Infosec: Robert Barnes, Infosec Institute, http://resources.infosecinstitute.com/geocoding-router-log-data/#gref, Geocoding Router Log Data, Aug, 8 2013
- Korosh Golnabi, Richard K. Min, Latifur Khan, Ehab Al-Shaer. Analysis of Firewall Policy Rules Using Data Mining Techniques[C]. Network Operations and Management Symposium,2006. NOMS 2006. 10th IEEE/IFIP
- Abedin, Muhammad, et al. "Analysis of firewall policy rules using traffic mining techniques." International Journal of Internet Protocol Technology 5.1-2 (2010): 3-22.
- Wireshark, https://www.wireshark.org
- Wang K, Stolfo S. J. 2004. Anomalous Payload-based Network Intrusion Detection. In: Symposium on Recent Advances in Intrusion Detection, Sophia Antipolis, France.
- Sun-il Kim, William Edmonds, and Nnamdi Nwanze. 2014. On GPU accelerated tuning for a payload anomaly-based network intrusion detection scheme. In *Proceedings of the 9th Annual Cyber and Information Security Research Conference* (CISR '14),
- Robert K. Abercrombie and J. Todd McDonald (Eds.). ACM, New York, NY, USA, 1-4. DOI=http://dx.doi.org/10.1145/2602087.2602093
- Tobias Limmer and Falko Dressler. 2010. Dialog-based payload aggregation for intrusion detection. In *Proceedings of the 17th ACM conference on Computer and communications security* (CCS '10). ACM, New York, NY, USA, 708-710. DOI=http://dx.doi.org/10.1145/1866307.1866405
- Janak J. Parekh, Ke Wang, and Salvatore J. Stolfo. 2006. Privacy-preserving payload-based correlation for accurate malicious traffic detection. In *Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense* (LSAD '06). ACM, New York, NY, USA, 99-106. DOI=http://dx.doi.org/10.1145/1162666.1162667
- SNORT Rules Infographic https://snort-org-site.s3.amazonaws.com/production/document_files/files/000/000/116/original/Snort_rule_infographic.pdf?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIXACIED2SPMSC7GA%2F20210316%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20210316T191343Z&X-Amz-Expires=172800&X-Amz-SignedHeaders=host&X-Amz-Signature=bcfc7d75d223ab40badd8bd9e89ded29cc98ed3896f0140f55320ee9bcdf1383, last accessed March 2020
- Cheok, Roy. "Wire shark: A Guide to Color My Packets Detecting Network Reconnaissance to Host Exploitation." *GIAC certification paper* (2014), SANS Institute Reading Room
- NodeXL: https://www.smrfoundation.org/nodexl/
- Nicosia, Vincenzo, et al. "Graph metrics for temporal networks." Temporal networks. Springer Berlin Heidelberg, 2013. 15-40.
- Namayanja, Josephine M., and Vandana P. Janeja. "Change detection in evolving computer networks: Changes in densification and diameter over time." *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*.
- Namayanja, Josephine M., and Vandana P. Janeja. "Characterization of Evolving Networks for Cybersecurity." Information Fusion for Cyber-Security Analytics. Springer International Publishing, 2017. 111-127.
- Massicotte, Frédéric, Tara Whalen, and Claude Bilodeau. "Network Mapping Tool for Real-Time Security Analysis." *Real Time Intrusion Detection* (2003).
- OSQuery. OSquery, last accessed, March 2020,. WWW: https://osquery.io/, 2016.
- Snare. Snare, last accessed, March 2020, WWW: https://www.snaresolutions.com/central-83/.
- Stephens, Gregory D., and Marcus A. Maloof. "Insider threat detection." U.S. Patent No. 8,707,431. 22 Apr. 2014.
- Van Mieghem, Vincent. Masters Thesis Delft University, "Detecting malicious behaviour using system calls.", 2016
- Vaidya, Jaideep, Vijayalakshmi Atluri, and Qi Guo. "The role mining problem: finding a minimal descriptive set of roles." Proceedings of the 12th ACM symposium on Access control models and technologies. ACM, 2007.
- Mitra, Barsha, et al. "A Survey of Role Mining." ACM Computing Surveys (CSUR) 48.4 (2016): 50.
- Darwish, A.; Bataineh, E., "Eye tracking analysis of browser security indicators," Computer Systems and Industrial Informatics (ICCSII), 2012 International Conference on , vol., no., pp.1,6, 18-20 Dec. 2012 doi: 10.1109/ICCSII.2012.6454330
- Frei, S., May, M., Fiedler, U. and Plattner, B., 2006, September. Large-scale vulnerability analysis. In Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense (pp. 131-138). ACM.
- NIST. National institute of standards and technology: National vulnerability database. Accessed Sept, 2017, http://nvd.nist.gov/.