The Web graph on page, host and cyber attacks

Sorour Abdul Hussein Dohi
Al-Mustansiriya University/College of Administration and Economics/Information Technology Division
cce3513@gmail. Com

Abstract:

The web graph, composed of interconnected web pages and hosts, is a fundamental structure in the digital realm. Its significance stems from facilitating information dissemination, communication, and online interactions. However, this interconnectedness also brings about vulnerabilities, making the web graph and host systems prime targets for cyber attacks. This abstract provides an overview of the web graph, host systems, and the challenges posed by cyber attacks.

The web graph represents the intricate network of web pages, with each page represented as a node and hyperlinks forming the connections between them. It enables users to navigate through vast amounts of information, creating a dynamic and interconnected web of knowledge. Host systems, on the other hand, provide the infrastructure for web pages, encompassing servers, databases, and network components that facilitate the delivery of web content.

Unfortunately, these components are not immune to cyber threats. Cyber attacks, such as phishing, malware injections, and Distributed Denial of Service (DDoS) attacks, exploit vulnerabilities in the web graph and host systems. These attacks can result in data breaches, service disruptions, and compromise of sensitive information, leading to financial loss, reputational damage, and privacy violations.

To mitigate these risks, comprehensive cybersecurity measures are essential. This includes implementing robust security protocols, regularly updating and patching

software, employing intrusion detection and prevention systems, and conducting security audits. Additionally, user awareness and education play a vital role in preventing cyber attacks through safe browsing practices, strong passwords, and caution when interacting with suspicious content.

Collaboration and information sharing among stakeholders are critical in addressing cyber threats effectively. Governments, organizations, and individuals must work together to share threat intelligence, develop best practices, and establish regulations and standards to enhance the security of the web graph and host systems.

Keywords: (web graph, host systems, cyber attacks, interconnectedness, vulnerabilities).

الرسم البياني للويب على الصفحة والمضيف والهجمات السيبرانية

سرور عبد الحسين دوحي

الجامعة المستنصرية/ كلية الادارة والاقتصاد/ شعبة تكنولوجيا المعلومات

cce3513@gmail. Com

الملخص:

الرابطة الشبكية للويب، المتألفة من صفحات ويب ومضيفين مترابطين، هي هيكل أساسي في المجال الرقمي. ينبع أهميتها من تسهيل نشر المعلومات والاتصال والتفاعل عبر الإنترنت. ومع ذلك، فإن هذه الترابطية تجلب أيضًا ثغرات، مما يجعل الرابطة الشبكية وأنظمة المضيفين هدفًا رئيسيًا للهجمات السيبرانية. يوفر هذا الملخص نظرة عامة على الرابطة الشبكية، وأنظمة المضيفين، والتحديات التي تواجهها الهجمات السيبرانية.

الرابطة الشبكية تمثل الشبكة المعقدة لصفحات الويب، حيث يتم تمثيل كل صفحة كعقدة والروابط الفرعية تشكل الروابط بين تلك العقد. فهي تمكن المستخدمين من التنقل في كميات هائلة من المعلومات، وتخلق شبكة معرفية متغيرة ومترابطة. من ناحية أخرى، توفر أنظمة المضيفين البنية التحتية لصفحات الويب، بما في ذلك الخوادم وقواعد البيانات ومكونات الشبكة التي تسهل توصيل محتوى الويب.

ومع ذلك، فإن هذه العناصر غير محصنة ضد التهديدات السيبرانية. تستغل الهجمات السيبرانية، مثل الصيد الاحتيالي وحقن البرمجيات الخبيثة وهجمات الخدمة المنتشرة (DDoS)، الثغرات في الرابطة الشبكية وأنظمة المضيفين. يمكن أن تؤدي هذه الهجمات إلى اختراقات البيانات وتعطيل الخدمات واختراق المعلومات الحساسة، مما يؤدي إلى خسائر مالية وضرر في السمعة وانتهاكات للخصوصية.

للحد من هذه المخاطر، يتعين تنفيذ إجراءات أمان شاملة. يشمل ذلك تنفيذ بروتوكولات أمان قوية، وتحديث وتصحيح البرمجيات بانتظام، واستخدام أنظمة كشف ومنع الاختراق، وإجراء تدقيقات أمان منتظمة. بالإضافة إلى ذلك، فإن إدراك المستخدم والتثقيف يلعبان دورًا حاسمًا في منع الهجمات السيبرانية من خلال ممارسات التصفح الآمنة واستخدام كلمات مرور قوية والحذر عند التفاعل مع محتوى مشبوع

التعاون وتبادل المعلومات بين الجهات المعنية أمر حاسم في التعامل مع التهديدات السيبرانية بشكل فعال. يجب على الحكومات والمؤسسات والأفراد العمل معًا لمشاركة معلومات التهديدات، وتطوير أفضل الممارسات، ووضع التشريعات والمعايير لتعزيز أمان الرابطة الشبكية وأنظمة المضيفين.

من خلال الأولوية الممنوحة للأمان السيبراني وتنفيذ التدابير الاحترازية، يمكننا التقليل من المخاطر المرتبطة بالهجمات السيبرانية وضمان رابطة شبكية وأنظمة مضيفين قوية ومتينة. وسيساعد ذلك في الحفاظ على ثقة المستخدمين وحماية المعلومات الحساسة وتعزيز تجربة آمنة عبر الإنترنت للأفراد والمؤسسات في جميع أنحاء العالم.

الكلمات المفتاحية: (الرابطة الشبكية، أنظمة المضيفين، الهجمات السيبرانية، الترابط، الثغرات).

Introduction

The web graph, consisting of interconnected web pages, is a fundamental structure in the digital landscape. It represents the relationships between web pages, with each page being represented as a node and hyperlinks serving as the connections between nodes. This interconnectedness allows users to navigate through vast amounts of information and facilitates the dissemination of knowledge and communication on the internet.

Host systems play a crucial role in supporting the web graph. They provide the infrastructure for web pages, encompassing servers, databases, and network components. Host systems store and process web content, handle user requests, and ensure the smooth delivery of information across the internet.

However, this interconnected web graph and the host systems are not immune to cyber attacks. Cybersecurity threats pose significant risks to the security and integrity of web pages and host systems. Cyber attacks encompass a wide range of malicious activities, including phishing, malware injections, Distributed Denial of Service (DDoS) attacks, and data breaches.

Phishing attacks attempt to deceive users into revealing sensitive information, such as passwords or financial details, by masquerading as legitimate entities. Malware injections involve the insertion of malicious code into web pages or host systems, compromising their functionality and potentially infecting users' devices. DDoS attacks overwhelm web servers with a flood of traffic, rendering them inaccessible to legitimate users. Data breaches involve unauthorized access to and theft of sensitive information stored within host systems, leading to privacy violations and financial loss.

To mitigate these cyber threats, robust cybersecurity measures are crucial. This includes implementing security protocols such as encryption, regularly updating and patching software to address vulnerabilities, deploying intrusion detection and prevention systems to detect and block malicious

activities, and conducting regular security audits to identify and address potential weaknesses.

User awareness and education are also vital in preventing cyber attacks. Users should be cautious when interacting with suspicious emails, websites, or links, and should follow secure browsing practices, such as using strong passwords and avoiding sharing sensitive information on untrusted platforms.

Collaboration among governments, organizations, and individuals is essential in addressing cyber threats effectively. Sharing threat intelligence, developing best practices, establishing regulations and standards, and fostering a culture of cybersecurity contribute to a safer digital environment.

World Wide Web

The World Wide Web may be seen as a human-managed database for storing and exchanging numerous materials, but it varies from traditional databases in terms of its vastness, very rapid dynamics, and variety. Secondly, the scale of the web cannot be exactly defined or quantified due to its immense size. The number of pages is unrestricted, and their contents are determined by the data submitted by the user. In January 2005, Gulley and Signorini claim that the Web has 11 billion pages. Google stated in 2008 that their technology has processed one trillion URLs on the web [9].

Another difficulty is the quick evolution of website content. In 1999, Chu and Garca-Molina analyzed the pace of change by downloading 720,000

pages over the course of four months [10]. 23% of the collection's page content was amended, and 50% of the collection was modified or eliminated within 50 days [11].

Online documents are diverse for many motives and views. Websites may also include photographs, movies, and audio files in a variety of forms. These papers may range in size from one byte to hundreds of gigabytes. Different versions and pages with erroneous syntax that do not adhere to W3C standards but are nonetheless readable by web browsers may be found among the most prevalent HTML files. Online material is often unstructured, composed of several languages and styles, and of wildly varying quality. While HTML pages include some information, they are often not trustworthy. The primary objective of the Semantic Web is to organise data in a way that facilitates human-machine cooperation [12].

Web as a graph

The web is a rich and helpful source of information despite the fact that it seems to be more unorganized than traditional papers. Web content are connected by a vast network of hyperlinks.

E is (u,v) if page u connects to page v. In this graph, d-(u) represents the input degree of U, such as the number of pages that link to u, and d+(u) represents the output degree, which is the number of pages that u references. As with other human-made networks, the web graph has fascinating characteristics [13].

Web graph on the page and host level

Not all links between pages are identical. In contrast to connections between sites that entail the exchange of human information, leading links between internal pages of a website may not contribute value to the page in question. Davison demonstrated how to target internal links on a website for referral reasons. These linkages may be deleted from the web graph if required [14].

Excluding any connections inside the website and creating a smaller graph where the nodes are the sites and the edges are only our links between the sites makes it simpler to discern anchor links from anchor links.

In this instance, it is not feasible to differentiate between the pages of the same website; hence, when the ranking algorithm distributes scores to the network nodes, all of the pages of that website get the same score [13].

2-2-3 Connection

Except for a few distinct sites, if web linkages are viewed as undirected edges, the web graph forms a single linked component. If the edges are regarded directed, the structure of the web graph changes. A considerable number of network nodes are inaccessible from the remaining nodes, hence the graph nodes may be separated into five sections. Broder et al. found this structure. They determined that 203 gigabytes of online pages often referenced the bow-tie structure. The following structure is specified in web pages, as seen in the picture below[15]. The central core (SCC), the biggest component linked to the graph, is the primary element. Each node in SCC may be reached in a limited number of steps from any other node in SCC. In

Broder's test, the size of the SCC is 50 MB, whereas the second biggest associated component occupies just 150 KB of the page.

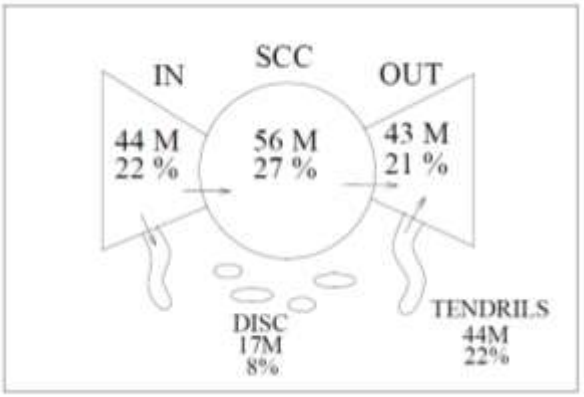

Figure 1: Bowie-tie web structure described in the article [15]

Nodes located outside of the SCC that are accessible through the SCC. 22% of the whole graph consists of both OUT and IN components. The remaining nodes are either reachable from IN and can reach OUT, or they cannot reach any of the previously mentioned nodes (DISK) [15].

Search engines

In the early days of the Internet, there were two primary ways for navigating online sites. Either by typing the URL into the browser's address bar or by clicking on hyperlinks on online sites. Web dictionaries and online search engines have emerged as a result of the expansion in the amount of web pages.

The majority of hyperlinks in online directories go to helpful web sites. These linkages are often structured hierarchically and are typically managed by people. The Yahoo directory at dir.yahoo.com and the dmoz.org open directory project are well-known examples. Regrettably, the creation and maintenance of online directories are labor-intensive processes. Due of the continuous evolution of websites, it is very challenging to manage and update online directories. On the other hand, users must exert a non-negligible amount of effort to choose the correct categorization in the multi-level hierarchy [11].

Search engines are almost automated, need less human work to maintain their databases up-to-date, and are capable of adapting to the constantly changing nature of the web. They provide a service that delivers pages depending on the user's keyword input. Typically, the user inputs keywords or brief sentences into the search box and then clicks the "Search" button. The search engine then provides a number of links that it believes correspond to the user's interests.

Web directories are unable to deliver more targeted results than search engines. Unfortunately, for ambiguous and poorly-formed inquiries, the resulting information is often irrelevant and misleading. The greatest problem for search engines is to comprehend the query and provide the most relevant results. There are frequently millions of online pages that match a user's search terms, yet consumers only need to visit a small number of

relevant websites. Study of online query log data revealed that consumers view the top 10 search engine results [16].

Ranking is primarily concerned with information retrieval. A interdisciplinary branch of study concerned with the selection and rating of materials that match keywords. The Internet differs from traditional literature. Hyperlinks connect pages, which may be used in relational computations [11].

Search engine query server

The primary function of the query engine is to process user inquiries. If the query comprises more than one phrase, search engines will typically consider an AND between them; hence, the user will obtain a page including all the sought-after terms. The query engine evaluates the list of query terms in the index, assesses their appropriateness, applies the ranking, and displays the results to the user in order to build this list of pages. Delivering results to the user may include further computations and database access, such as the generation of a snippet, a brief extract of a document containing the query terms. All of these procedures must be completed in a fraction of a second, which necessitates fast algorithms and supports the use of pre-calculations during indexing [11].

Cyber and security

Cyber is a prefix in English and a suffix in Farsi that is added to new and contemporary phrases to give them meanings relating to the computer environment or the Internet. Cyber is derived from the term Cyberspace,

which refers to the study of processes used to govern and regulate complex human or machine systems [18]. In 1984, "William Gibson" used the term "cyber space" for the first time in his science fiction book "Nuromancer." In the absence of today's worldwide computer networks and systems, Gibson characterized cyberspace as follows: "Cyberspace is a common delusion experienced everyday by billions of operators and children who are taught mathematics [19]." [Others have characterized cyberspace as follows: "Cyberspace is a collection of internal human conversations using computers and telecommunications equipment, independent of physical location." [20] Note that cyberspace is a more advanced notion than the Internet. Cyberspace is thus, in a general sense, a virtual and unreal realm in which computers exchange electronic data. Cyberspace is thus an ethereal, virtual, and unreal place generated by the connection and communication of computers and internet technologies. The three fundamental components of cyberspace are a computer system, an internet network, and the users. In cyberspace, the breadth of a user's activity is not restricted by the physical boundaries of a home or office or even the borders of a nation, and any user may meet individuals anywhere in the globe for a very cheap cost at any time and location. communicate any amount and quality of information without knowing his presence and identity [21].

Definition of cyber attacks

With the inception of the first communication on the World Wide Web, which was subsequently renamed the Internet, more than three decades have

passed. Few individuals expected the Internet's extensive impact on people's personal and professional life at that time. This decentralized network was portrayed in grand terms as a tool for democracy and empowerment. While the assertion of the establishment of a global village and the dominance of a certain culture over the whole globe, which was referred to as globalization, was an exaggeration, the revolution of communication and its fast growth in a short period of time had good results. And its negative impact on societies and the attempts of some forces to manipulate global opinion in this manner cannot be concealed. Several governments have sought centralized control over such a chaotic network as a result of this evident danger posed by this dispersed power. Via the unsuccessful Communications Decency Act and the Key Recovery Initiative, the United States attempted to prohibit the use of cryptographic technology and control verbal communication, respectively. Several nations have established many stringent rules. The Internet and its stockholders have vehemently fought the implementation of such rules, resulting in several conflicts. [22] There are several definitions of cyber assaults, some of which are included here. Cyber assaults may be characterized as unlawful access to computers and data, or in a simpler sense, as the malicious destruction of information systems, including hardware- and software-affecting events. [23] Includes spam, denial of service assaults, and malware. [24]

Pushing a government or group to behave in order to threaten, halt, or seize control [25].

Using the situational awareness system in cyber command and control, it will be possible to identify attacks", "discover relationships between attacks", "track attacks", "perform a situational assessment of cyberspace," and "predict the effects of attacks" in the massive volume of data collected from various sources. To develop a cyber situational awareness system, as seen in figure (2-1), three basic subsystems of cyber assault comprehension, comprehension, and visualization are required. The subsystem for comprehending cyber assaults collects signs and comprehends ideas; without this subsystem to extract accurate information, the likelihood of developing an inaccurate picture of the cyber environment under command and control grows significantly. This subsystem will address the question, "What are the present realities of cyberspace?" The understanding subsystem integrates and determines the connection between various bits of information. This subsystem makes it feasible to comprehend what is occurring in the cyber command and control area.

Under the cyber attack visualization system, which will result in the maximum degree of situational awareness, the issue of foresight will be solved. This capability to foresee future occurrences based on present dynamic events would enable cyber battlefield commanders to make timely judgments [26].

While the words "visualization" and "prediction" are thought to be synonymous in most research publications, it should be highlighted that visualization and prediction have distinct meanings. Theoretically,

forecasting involves determining what will occur in the future by making educated assumptions, while visualizing only determines what will occur in the future. Forecasting is the act of predicting uncertain conditions, or, to put it more simply, forecasting is the estimation of future events with varying probability. Actually, prediction gives a forecast of future occurrences, while visualization is a subset of the set of forecasts that will most likely occur in the future. This difference is essential because identifying A Due to random events or a lack of understanding, the future is not always definite. [27]

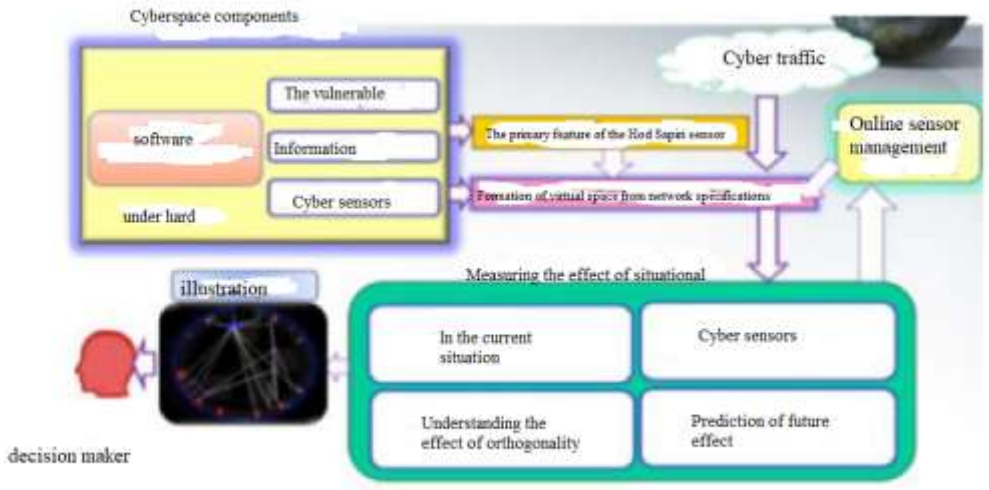

Figure 2-3: cyber situational awareness system [27]

Visualization enables the analyst to examine potential future situations prior to referring to the probable future in order to prepare for various outcomes. Hence, by visualizing the cyber combat scene and forecasting cyber assaults, it is able to anticipate the unique patterns of each attack in the system and assess the network weaknesses that are appealing to attackers. In addition,

this technology enables the tracking of multi-stage, coordinated cyber assaults and the display of the future status of these sorts of operations. As depicted in Figure (2-4), attacks can be conducted utilizing a cyber battle scene system equipped to evaluate the effects of cyber attacks, and by depicting the current state of the cyber battle scene and displaying the scores of the effects of various threats for hosts, services, users, and the entire network, enabling analysts to better comprehend and monitor the current and future situation. [28]

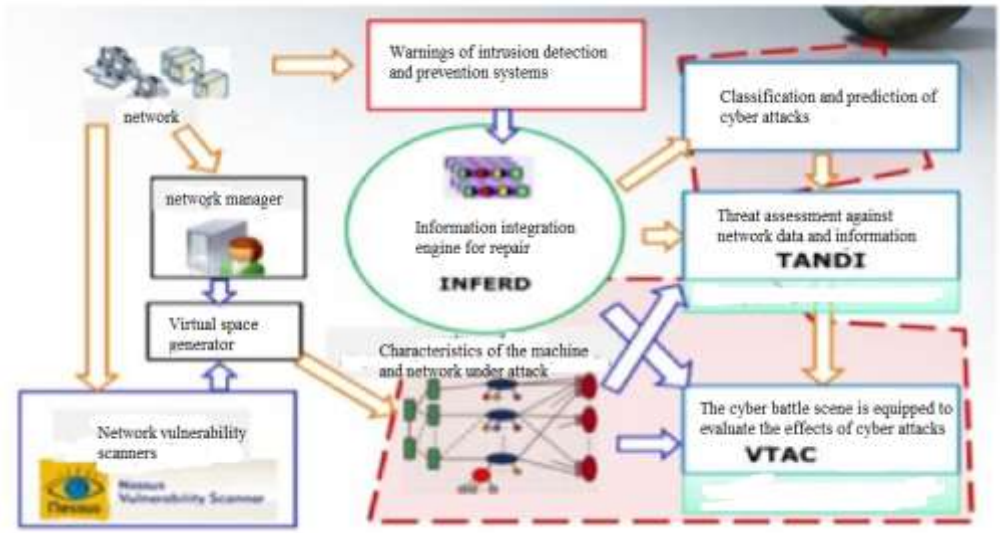

Figure 2: the process of achieving visualization of cyber attacks

Based on the above summary, it can be said that the attack visualization system analyzes associated alerts connected to multi-stage assaults and forecasts the prospective targets in a network that are at danger. The extraction of a model for the progression of cyber assaults is one of the system's primary obstacles. Due to the complexity, unpredictability, and

nature of networks and systems' distributability, there is a chance of disruption in the aforementioned models. Components like purpose, capacity, opportunity, and behavior are modeled to overcome this challenge in the depiction of cyber assaults. Whereas past modeling efforts have rendered the previously listed components independent [28].

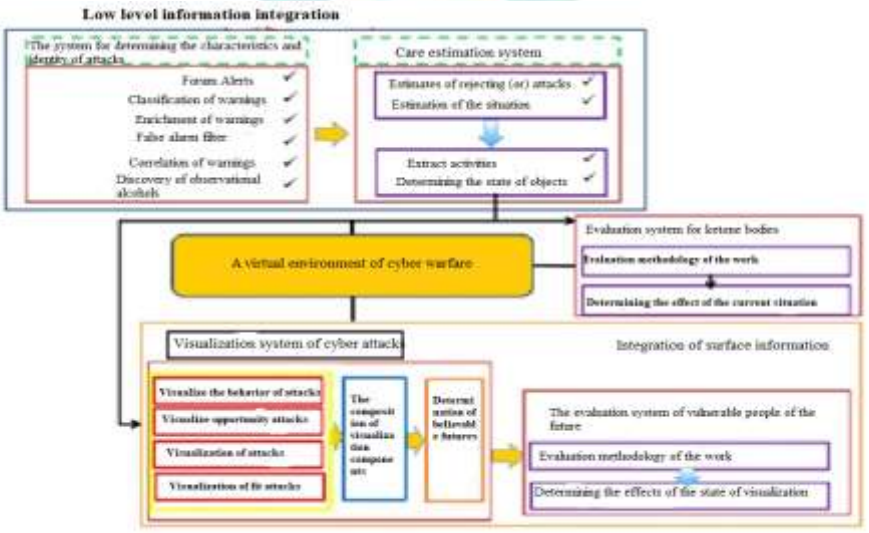

Figure 3: proposed architecture to visualize cyber attacks

Proposed architecture to visualize cyber attacks

Figure (6-2) depicts the architecture of awareness of the future situation and evaluation of its effects. It consists of three parts of the virtual environment of the cyber battle scene, the integration of high-level and low-level information, and four systems of identity meter, estimator, visualization, and vulnerability evaluation. In the architecture described above, attack rejection is generated based on a collection of correlated environment observations. Information integration engine for real-time decision-making in low-level

information integration is the heart of its construction. In this design, the estimator system describes the system's state as collections of actions and objects, while the visualization system detects potential futures. In addition to calculating the degree of vulnerability using impact assessment algorithms, the evaluation system analyzes the consequences of each prospective circumstance. This architecture produces an assessment of the present state, its impacts, and potential future circumstances. The future impacts are derived through the computation of the consequences of potential future scenarios; feasible futures are also assessed by effect assessment algorithms, resulting in this outcome. Algorithms predict future impacts. One of the responsibilities of this architecture is to compute the plausibility of futures based on the conceptual model derived from the link between objects and actions in the cyber war scene's virtual environment.

Insufficient information integration

In the suggested architecture, low-level information integration will be used to detect, identify, and estimate the state of cyber assaults. At this level, the intrusion detection system's different sensor instruments are responsible for identifying things in the surroundings. In addition, this level defines entities (including objects and ideas), objects (including missions, hosts, and services), concepts, events (anything that occurs at a certain moment), groups (a collection of connected things), and activities (whatever is termed activity or movement). With the integration of low-level data, we aim to comprehend "us" and all that is significant to us. We can refer to our

resources (capabilities and capacities) as well as anything that is deemed to be our weaknesses. In fact, at this level, we aim to comprehend the existing environmental status in order to The degree of relevance of each action in the scenario may be judged based on the gained information. The objective of this sort of information integration, which consists of correlated, classed, enriched, and integrated data, is to develop a model that represents the situation and provides an estimate of it. Modeling a scenario requires identifying its constituent parts; in this case, the environment, objects of interest, activity, and state of objects are modeled. [29]

In addition to real elements and interactions, the environment also contains intangible entities such as vulnerabilities. In cyberspace, the environment is modeled using the cyber fight scene's virtual environment. [29]

The components and entities in the environment whose preservation is vital to the analyst are referred to as favored objects.

Activity: "activity" refers to actions made to alter the environment.

State of things: depending to the area of application, the states of objects in various domains are distinct from one another. In actuality, the condition of things is determined by the aforementioned three factors. Typically, these statuses may be targeted, revealed, partly exploited, and misused in cyberspace.

The virtual environment of the cyber battle scene

In the cyber war scene's virtual environment, it is feasible to model cyberspace for the execution of cyber assault operations. The cyber combat

scenario is a public exhibition of sensitive information about vulnerabilities, component availability, and the sensitivity of the environment, and the assault pathways suggest possible flaws that might be exploited. In this virtual environment, entities and their interactions are modeled and processed. Considered to be the most influential and complex factor in the integration of high-level information, the model presented above contains crucial information for detecting cyber incidents on the battlefield and is regarded as the most important and complex factor in the integration of high-level information.

High level information integration

The future is seen via the incorporation of high-level knowledge. This level of integration starts its work with "our knowledge" received via the integration of low-level information; but, unlike the lower level, it is crucial to learn "knowledge from them." At this stage, prediction will no longer be the primary concern; instead, visualization will be the focus. This is due to the fact that, in predicting, just the set of probable possibilities has been stated. Hence, based on prior information, events that are considerably more likely to occur are extracted from this collection and their occurrence is shown. The visualization unit enables the analyst to generate feasible futures for each scenario based on potential future states. Figures 2–6 depict the basic structure for the depiction of assaults. According to this diagram, "their knowledge" is required to picture the present scenario (which utilizes "our knowledge") in the future. The visualization of the components of

opportunity, capacity, intention, and conduct may provide this insight. This is intended by visualizing the chance to picture the potential for an attacker to launch an attack in the cyber environment. Visualization also involves visualizing the capabilities of a cyber adversary. The objective of picturing the attacker's intent is to determine what the attacker intends to do, and the purpose of visualizing the attacker's prior conduct is to identify behavioral patterns. [29]

So, in order to reach the future with a vision of the present condition, it is necessary to see each of these components and assess the future based on each of these qualities. Figure (2-6) depicts further information based on the combination of visualization components according to the descriptive architecture of Figure (2-5), which is the assignment of a credibility score by current algorithms and the integration of this data in order to anticipate the future. accountable for As seen in Figure (2-6), the algorithms construct a believability score for future events in parallel.

Using the Dempster-Shafer theory, these scores are added for each circumstance. The obtained ratings for each assault phase are recombined to provide a list of probable futures for each environment item. In high-level information integration, the vulnerability assessment unit is responsible for evaluating the effect of a set of operations on network entities. In this model, each entity is given a score between zero and one based on the damage assessment algorithm.[29]

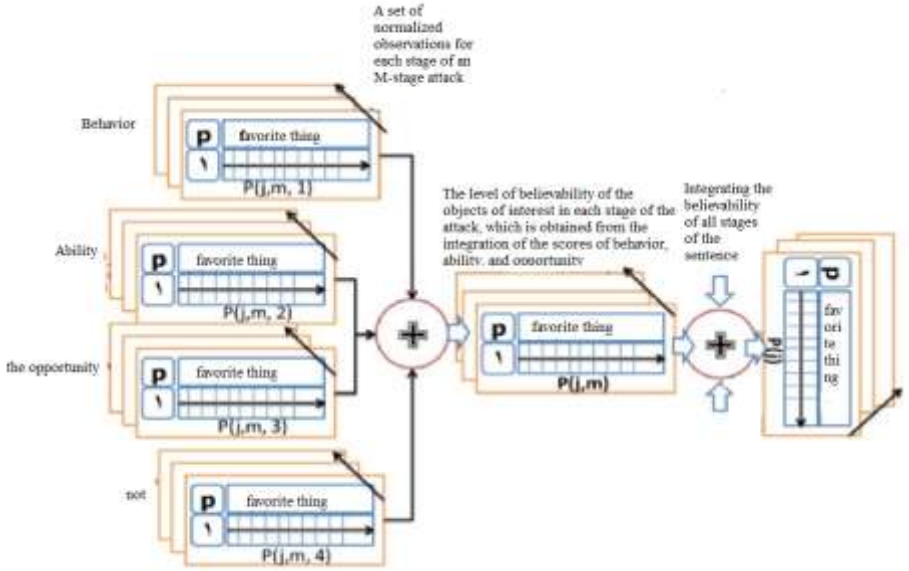

Figure 4: visualization of multi-stage cyber attacks based on the combination of visualization components

Some examples of cyber attacks

Deception offensives Deception attacks are assaults in which the purpose of the attacker is not to fool intrusion detection systems using a manner that can be detected and recognized. [30]

Network layer attacks

Typically, network-layer attacks are carried out by altering the values in the packet header in a variety of ways. Attacks on TTL fields, packet control fields, and the breaking of smaller packets are examples of network layer attacks. [30]

Transport layer attacks

As is well known, in this class of attacks, owing to their classification in the transmission layer, the protocol headers of this layer, particularly TCP, are

exploited so that unauthorized users may get access to the targeted systems. [31]

Denial of service attacks

Many intrusion detection systems include central logging servers that are used only to store notifications from the intrusion detection system. The central server is responsible for centralizing alert data. If an intruder knows the IP address of the central register server, he may use a denial-of-service attack to slow it down or even deactivate it. The attack might stay hidden after the server is shut down since the warning data is no longer collected [32, 33].

Definition of intrusion detection system (IDS)

IDS is an acronym for Intrusion Detection System; it is meant to monitor all network input and output activities and detect any suspicious behavior. These suspicious activity may suggest that someone is attempting to breach the security system by attacking the system or network. IDS is called a passive monitoring system due to the fact that its primary duty is to notify of suspicious actions but does not actively prevent them. Essentially, an IDS analyzes network traffic and data to identify attacks, exploits, and other security flaws. IDS may notify suspicious occurrences in a variety of methods, such as by flashing an alert, recording in the events section (Logs), and contacting with the administrator (such as a phone call to the system administrator). In certain instances, IDS asks reconfiguration of the system in an effort to decrease intrusions that are suspect. One of the functions of

IDS is to identify and report suspicious network activity to the system administrator. IDS actively searches for abnormal behaviors and events that may be the result of viruses, worms, or hackers. This is accomplished by looking for intrusion signatures (logs stored from login information) or assault signatures (attack signatures) that identify different worms and viruses. IDS modification encompasses a vast array of distinct items. An IDS solution might be made available as free open source software or as pricey commercial security software. In addition, some IDS incorporate software and hardware applications that are deployed and used in various network locations. [28 and 29]

Conclusion

In conclusion, the web graph and host systems are integral components of the digital landscape, serving as the foundation for information dissemination and online interactions. However, these interconnected structures are vulnerable to cyber attacks, posing significant risks to their security and integrity.

Cyber attacks, including phishing, malware injections, DDoS attacks, and data breaches, exploit the vulnerabilities present in the web graph and host systems. These attacks can result in financial loss, reputational damage, privacy violations, and disruptions in online services. Therefore, it is crucial to implement robust cybersecurity measures to protect against these threats.

Security protocols such as encryption, regular software updates, and intrusion detection systems play a vital role in safeguarding the web graph

and host systems. Additionally, user awareness and education are essential in preventing cyber attacks through safe browsing practices and cautious online behavior.

Collaboration and information sharing among stakeholders, including governments, organizations, and individuals, are key to addressing cyber threats effectively. Sharing threat intelligence, developing best practices, and establishing regulations and standards contribute to a more secure digital environment.

By prioritizing cybersecurity and implementing proactive measures, we can mitigate the risks associated with cyber attacks and ensure a resilient web graph and host systems. This will help maintain the trust of users, protect sensitive information, and foster a safer online experience for individuals and organizations worldwide

Research margins

[1] Verma, V., Muttoo, S. K., & Singh, V. B. (2020). Multiclass malware classification via first-and second-order texture statistics. Computers & Security, 97, 101895.

[2] Nikbakht, Elham and Moatar, Seyyed Mohammad Hossein and Wafai Jahan, Majid, 2015, Review of Malware Identification Techniques, Eighth National Conference of Electrical and Electronic Engineering of Iran, Gonabad [3] Ding, Y., Dai, W., Yan, S., & Zhang, Y. (2014). Control flow-based opcode behavior analysis for malware detection. Computers & Security, 44, 65-74.

[4] Hassen, M., & Chan, P. K. (2017, March). Scalable function call graph-based malware classification. In Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy (pp. 239-248).

[5] Gupta, S., Sharma, H., & Kaur, S. (2016, December). Malware characterization using windows API call sequences. In International Conference on Security, Privacy, and Applied Cryptography Engineering (pp. 271-280). Springer, Cham.

[6] Yuxin, D., Xuebing, Y., Di, Z., Li, D., & Zhanchao, A. (2011). Feature representation and selection in malicious code detection methods based on static system calls. Computers & Security, 30(6-7), 514-524.

[7] Cabau, G., Buhu, M., & Oprisa, C. P. (2016, September). Malware classification based on dynamic behavior. In 2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 315-318). IEEE.

[8] Han, W., Xue, J., Wang, Y., Huang, L., Kong, Z., & Mao, L. (2019). MalDAE: Detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. computers & security, 83, 208-233.

[9] Gulli, A., Signorini, A., 2015, "The indexable web is more than 11.5 billion pages", In Proceedings of the 14th World Wide Web Conference (WWW), Special interest tracks and posters, pages 902-903.

[10] Orth-Alfie, C., & Mahan, N. (2019). Off the Web and into the Fishing Hole: Simulating the Iterative Search Process through Active Learning. Brick Click &.

[11] Bar-Yossef, Z., Broder, A. Z., Kumar, R., Tomkins, A., 2014, "Sic transit Gloria telae: Towards an understanding of the web's decay", In Proceedings of the 13th World Wide Web Conference (WWW), pages 328-337. ACM Press.

[12] Berners-Lee, T., Hendler, J., Lassila, O., 2001, “The semantic web. Scientific American".

[13] Csalogány, K., 2009, "Methods for Web Spam Filtering", Technical Report, Eötvös Loránd University.

[14] Davison, B. D., 2016, "Recognizing nepotistic links on the web", In AAAI-2000 Workshop on Artificial Intelligence for Web Search, pages 23-28, Austin, TX.

[15] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J., 2017, "Graph structure in the web", In Proceedings of the 9th World Wide Web Conference (WWW), pages 309-320. North-Holland Publishing Co. .

[16] Silverstein, C., Marais, H., Henzinger, M., Moricz, M., 1999, "Analysis of a very large web search engine query log" SIGIR Forum, 33(1):6-12, 1999 .

[17] Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., Raghavan, S., August 2011, "Searching the web. ACM Transactions on Internet Technology (TOIT)", 1(1):2–43,.

[18] Alipour, Hassan, 2013, Fundamentals of Information Technology, second edition, Tehran: Khorsandi Publications

[19] Farsi, Asghar, 2013, international cyber basics and ways to deal with it in the international arena, master's thesis, Payam Noor University, Tehran.

[20] Zandi, Mohammad Reza, 2013, preliminary research on cyber, Tehran, Jungle Publications

[21] Bastani, Broumand, 2013, computer and internet crimes, new fronts of delinquency, first edition, Tehran: Behnam Publications [22] Spinello, Richard A. (2014), Cyberethics: Morality and Law in Cyberspace, fifth edition, USA, Jones & Bartlett Learning

[23] Buscaglia, C. A., & Weismann, M. F. (2012). How " Cybersafe " are the BRICs? Journal of Legal,Ethical and Regulatory Issues

[24] Dipert, R. R. (2010). The Ethics of Cyberwarfare. Journal of Military Ethic

[25] Caplan, N. (2013). Cyber War: the challenge to national security. Global Security Studies

[26] K. Dadashtabar, A. J. Rashidi and H. Shirazi,(, (2014 A new pattern for improvment of situation awareness beasd on information fusion," 6'th National conference in electronic warfar.(in persian)

[27] K. Dadashtabar, A. J. Rashidi, and H. Shirazi, (2015)A new model for projection of multi stage cyber attack," 2'th National symposium in cyber defence,. (in persian) [28] K.

Dadashtabar, A. J. Rashidi, and H. Shirazi, (2015)a new architecture for impact projection of cyber attacks based on high level information fusion in cyber command and control," journal of electronical & cyber defence, vol. 2, no. 4, (in Persia

[29] J. Holsopple and S. Yang, "FuSIA(2008) Future situation and impact awareness," in Proceedings of 11th International Conference on Information Fusion, pp. 1–8.

[30]Fakulta informatiky ,Masarykova uviverzita(2010)IDS system evasion techniques

Reference

[1] Verma, V., Muttoo, S. K., & Singh, V. B. (2020). Multiclass malware classification via first-and second-order texture statistics. Computers & Security

[2] Nikbakht, Elham and Moatar, Seyyed Mohammad Hossein and Wafai Jahan, Majid, 2015, Review of Malware Identification Techniques, Eighth National Conference of Electrical and Electronic Engineering of Iran, Gonabad

[3] Ding, Y., Dai, W., Yan, S., & Zhang, Y. (2014). Control flow-based opcode behavior analysis for malware detection. Computers & Security

[4] Hassen, M., & Chan, P. K. (2017, March). Scalable function call graph-based malware classification. In Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy

[5] Gupta, S., Sharma, H., & Kaur, S. (2016, December). Malware characterization using windows API call sequences. In International Conference on Security, Privacy, and Applied Cryptography Engineering.

[6] Yuxin, D., Xuebing, Y., Di, Z., Li, D., & Zhanchao, A. (2011). Feature representation and selection in malicious code detection methods based on static system calls. Computers & Security

[7] Cabau, G., Buhu, M., & Oprisa, C. P. (2016, September). Malware classification based on dynamic behavior. In 2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)

[8] Han, W., Xue, J., Wang, Y., Huang, L., Kong, Z., & Mao, L. (2019). MalDAE: Detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. computers & security

[9] Gulli, A., Signorini, A., 2015, "The indexable web is more than 11.5 billion pages", In Proceedings of the 14th World Wide Web Conference (WWW), Special interest tracks and posters, pages

[10] Orth-Alfie, C., & Mahan, N. (2019). Off the Web and into the Fishing Hole: Simulating the Iterative Search Process through Active Learning. Brick Click.

[11] Bar-Yossef, Z., Broder, A. Z., Kumar, R., Tomkins, A., 2014, "Sic transit Gloria telae: Towards an understanding of the web's decay", In Proceedings of the 13th World Wide Web Conference (WWW)

[12] Berners-Lee, T., Hendler, J., Lassila, O., 2001, “The semantic web. Scientific American".

[13] Csalogány, K., 2009, "Methods for Web Spam Filtering", Technical Report, Eötvös Loránd University.

[14] Davison, B. D., 2016, "Recognizing nepotistic links on the web", In AAAI-2000 Workshop on Artificial Intelligence for Web Search,

[15] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J., 2017, "Graph structure in the web", In Proceedings of the 9th World Wide Web Conference (WWW), pages 309-320. North-Holland Publishing Co.

[16] Silverstein, C., Marais, H., Henzinger, M., Moricz, M., 1999, "Analysis of a very large web search engine query log" SIGIR Forum,

[17] Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., Raghavan, S., August 2011, "Searching the web. ACM Transactions on Internet Technology (TOIT)”

[18] Alipour, Hassan, 2013, Fundamentals of Information Technology, second edition, Tehran: Khorsandi Publications

[19] Farsi, Asghar, 2013, international cyber basics and ways to deal with it in the international arena, master's thesis, Payam Noor University, Tehran.

[20] Zandi, Mohammad Reza, 2013, preliminary research on cyber, Tehran, Jungle Publications

[21] Bastani, Broumand, 2013, computer and internet crimes, new fronts of delinquency, first edition, Tehran: Behnam Publications

[22] Spinello, Richard A. (2014), Cyberethics: Morality and Law in Cyberspace, fifth edition, USA, Jones & Bartlett Learning

[23] Buscaglia, C. A., & Weismann, M. F. (2012). How " Cybersafe " are the BRICs? Journal of Legal,Ethical and Regulatory Issues

[24] Dipert, R. R. (2010). The Ethics of Cyberwarfare. Journal of Military Ethic

[25] Caplan, N. (2013). Cyber War: the challenge to national security. Global Security Studies

[26] K. Dadashtabar, A. J. Rashidi and H. Shirazi,(, (2014 A new pattern for improvment of situation awareness beasd on information fusion," 6'th National conference in electronic warfar.(in persian)

[27] K. Dadashtabar, A. J. Rashidi, and H. Shirazi, (2015)A new model for projection of multi stage cyber attack," 2'th National symposium in cyber defence,. (in persian) [28] K. Dadashtabar, A. J. Rashidi, and H. Shirazi, (2015)a new architecture for impact projection of cyber attacks based on high level information fusion in cyber command and control," journal of electronical & cyber defence, vol. 2, no. 4, (in Persia

[29] J. Holsopple and S. Yang, "FuSIA(2008) Future situation and impact awareness," in Proceedings of 11th International Conference on Information Fusion

[30]Fakulta informatiky ,Masarykova uviverzita(2010)IDS system evasion technique