Information Security Management System Portfolio

Identification of security issue

Within this portfolio the security issue focused on will be the data breach Alibaba suffered in 2019. They are a multinational technology company, who are the owners of various subsidiaries. The business aims to provide marketing reach along with technology infrastructure in order to aid retailers and merchants. The organisations Chinese shopping operation, Taobao, had information stolen from the site via a web-crawler. The web-scraper disguised as a web-crawler managed to infiltrate the website and collect sensitive user information. This crawler had been developed by a marketing consultant who was affiliated with the organisation, along with the developer's employee. The scraping was carried out for multiple months before Alibaba discovered the bot.

Analysis of the issue

The web-crawler attack took place for approximately 8-9 months (November 2019 – July 2020) before Alibaba recognized the scheme taking place. The crawler had been able to successfully access data, which was further down than the human eyes perspective, and in turn acquired a billion plus points of data. The information that was successfully collected during the data breach contained customer data, including user ID's, customer comments and mobile numbers. After the culprits were convicted in May 2021 it was revealed that the data had not been shared as they collected it for their own purposes.

A web-crawler, also known as a spider, is a search engine bot which collects information about a webpage (e.g. the copy and meta tags) and then indexes pages based on their content. This is achieved by passing between links on web pages. Web crawlers are usually operated by search engines and aim to sift through all webpages on the internet, to grasp what each individual page is about. Search engines are able to supply links of relevance as feedback to the users search queries with the application of a search algorithm connected to the collection of data.

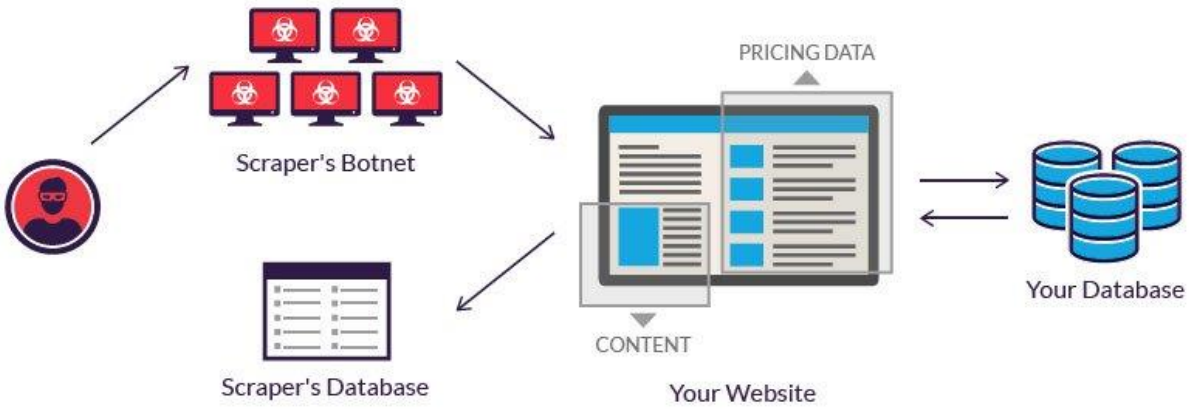
Web-scraping, or data scraping, is when the content residing on a website is downloaded using a bot without authorization from the owner. In many cases, the content has been illegally downloaded for malicious purposes. This method is a distinctly more targeted approach in comparison to web-crawling. Scraping software's are programmed to zone in on particular websites and pages instead of constantly following links or index the content. In addition, these bots ignore the robots.txt file and do not limit their requests. This action causes the web servers to become strained and overtaxed. In turn, the website may slow down or even break. Scrapers are designed to transverse through databases in order to extract information. These bots have been customized to be able to extract and transform content, store scraped data and extract data from APIs. They are also able to recognize unique HTML structures. Differentiating between whether a bot is malicious or genuine is not a simple task as bots with malicious intent are disguised. However, there are ways to distinguish the bots. When malicious bots are disguised, they create a false HTTP user agent. If the bot is legitimate, the organisation they are scraping for will be identified.

Scraping tools can also be utilized to gain detailed information. There are a number of components that make up a website. Databases containing user information are an element of these websites. The information can be extracted and then transferred to the attacker. Another form of scraping is content scraping. With the stolen information the hacker can replicate the website it stole from, making it difficult to discern the real and fake. The hacker can use the fake site to send phishing emails to collect sensitive information. The link in the email can lead to the fake website where the victim will enter their data, for instance bank details. After gathering these various details, the attacker can gain access to multiple accounts. Content like prices are also scrapeable. Some companies deploy bots to scrape their competitions website with the goal of gaining the upper hand on their competitor. When the companies are in the same sector of business, this is a common technique used to analyse the opposition.

Threat vectors

A threat vector is the path a cybercriminal uses to bypass the systems defences and cause harm. In order to protect systems from these attacks the potential routes available must be analysed to limit and/or prevent any form of exploitation that may occur.

In this incident, the attack vector was carefully constructed as the marketing consultant themselves programmed the scraper. Scrapers are implemented with a union of tools to penetrate their chosen target. These include sequential requests, proxy Ips and multiple UAs (user-agents). In order to evade security systems, put in place such as intrusion detection and prevention systems (IDS/IPS) and web application firewalls (WAFs) attacks utilizing web-scrapers are planned in stages. For this instance, it is suspected the website was crawled to find the location of the desired information first. After confirming the location and inspecting the backend code of the site (HTML source code) to identify unique nest tags, the developer integrated these into the software. By incorporating these into the code, the bots have knowledge of where to look to find the desired details for extraction. The bots then save the data in the selected location of the developer. These files could be JSON or CSV. Furthermore, with a lack of defenses employed to counter this attack, the web scraper bot managed to successfully infiltrate the website. There were no procedures in place for the malicious bots to be detected. The developer of the scraper also took advantage of their connection to the organisation to potentially learn which defenses have been set up and then studied methods to bypass the system. By pulling metadata from documents which have been posted online it is possible to a number of pieces of information. The metadata includes usernames, employee names and email formats.



Assets affected

The first stage in the risk assessment is asset identification. To fall into the category of an asset, it has to be of value to the organisation. Assets may be intangible or tangible in relation the organisation. Identification of the assets which may have been affected due to the breach, is a fundamental step in the procedure. A major asset affected was the user information, which included sensitive data such as customer comments, user IDs and phone numbers. The leak of this private data goes against the users consent and policy agreements. This leads to the reputation of the company being affected, which is another intangible asset affected from this breach. The decline in the company's reputation leads to revenue loss, as people from users to sponsors draw away from the organisation due to lack of trust and other aspects. With revenue loss counting as an affected asset, the outcome could result in several consequences. The reduced revenue may force the company to reduce their operational output in various ways. In turn, the advantage Alibaba may have held over competitors will disappear. The repercussions of the breach will also reach other branches of the business. A key asset also affected, is the RedMart App Alibaba has ownership of. Being an online supermarket based in Singapore, providing home delivery seven days a week, an

established app owned by the company is bound to draw in concerns about the security of the information customers are inputting. The data includes numerous pieces of information, notably bank details and home addresses. RedMart's website alternative will also be affected as collateral damage. The website could possibly have an even larger collection of data in comparison to the app as its arguably more accessible. Additionally, the database holding the data is affected since it has been compromised. This confidential information being breached can potentially jeopardize the safety of individuals. As a consequence of the attack, over a long period of time the analytics, another asset, are now distorted. The marketing team and other aspects of the company depend on accurate analytics, namely the user demographic, page views and bounce rates to decide which direction to go with certain ideas. The bounce rates especially will be warped as the bots have added to the website's traffic, loading the server infrastructure and slowing down the page. When the website begins to slow down this can affect user experience, an asset to the organisation as it affects the reviews, reputation and brand image of the business.

Short term impact

The web-scraper attack on Alibaba's website lasted over 8 months. Over this course of time, the stolen consumer data reached over a billion pieces. As Alibaba is a multinational technology company which encompasses retail, e-commerce and technology, this breach affects all aspects of the company's security and relations. The stolen data included information such as usernames and phone numbers. In China, it is a requirement to register with real name identification before you can obtain a phone number. As a high number of Chinese consumers would sign up for services on the internet using their phone number, knowledge of the number could make it possible to track user's personal information. In this case the law views mobile numbers as a form of personal information. Due to this breach, clients in business with Alibaba started to question the security of the organisation. Consumers of this business, from using the internet to a website under the organisations name, began to question the safety of the data they input. Additionally, this unexpected data leak which took place under the radar for such a long time effected the company's relationships with other clients. It was stated that Alibaba and its customers did not incur any financial losses from the leak. Furthermore, the People's court of Suiyang District declared that no laws were violated. However, in accordance to China's 2017 proposed cyber laws, Alibaba was at risk of receiving sanctions due lack of security. The case concluded in the organisation receiving no financial punishment over the matter. This incident took place shortly before the new Data Security Law passed on 1st September 2021, stating authorities are to be granted unrestricted power from the government to shut down tech companies which are found mistreating core state data. On 1st November 2021, China also introduced a new Personal Information Protection Law (PIPL), which did not impact the company.

Scope

The data breach involved the collection of users' information across the globe. With this in mind, I have decided to create a plan covering methods to mitigate web scraping from occurring. Alibaba is also a multinational organisation, which has headquarters branched across 13 different countries. Taking this factor into account, the controls will also cover methods to improve security around the branches across the world. The scope will cover physical, procedural and technical controls the company could incorporate into the security in order to decrease the probability of breaches. The vulnerabilities of Alibaba will be taken into account to form a suitable plan for the scope of the business. The technical controls will be more focused on the prevention of a scraping attack alternatively to protection against various other embodiments of hacking. Procedural controls will look into what steps can be taken to ensure the company's personnel are informed on security procedures and policies essential when working. As Alibaba holds sensitive data of users on a global scale, it is vital any possible leaks are kept to a minimum. The physical controls will also display a plan to refine the security in and around the headquarters.

Asset Register

The purpose of an asset register is to present a breakdown of all the assets affected within an organisation. Identification of all the important assets which could be affected is crucial to the construction of an effective plan.

Asset ID	Asset	Asset class/process	Likelihood of damage/loss	Impact of damage/loss	Asset Value	Owner	Comments
1	Sensitive user/customer information	Data	Very high	Medium	Very high	IT department	The IT department is responsible for the security of the information within the company. As the data is sensitive information, there should be layers of protection to prevent it from being compromised.
2	Revenue loss	Intangible	Medium	Very high	Very high	Alibaba	The loss of revenue affects the entire organisation as the money is needed to carry out organisation activities. The drop in revenue affects Alibaba directly.
3	RedMart App	Services	Low	Medium	High	Alibaba	The app is another form of income for the business. Since Alibaba is the owner of this app, they are responsible for data held on the platform. The company's profit will also be affected if the apps engagement drops.
4	Database holding the data	Data	Very High	Very high	Very High	IT department	Protection of the database, which holds the data, is the IT departments responsibility to keep secure and put measures in place. In the event of an attack on the database,

							there should be firewalls to monitor any traffic on the network. Data should also be encrypted, along with the employees regularly backing up the data.
5	User experience	People	High	High	Medium	User	The experience of the user is a core asset to the business. If the users are not receiving a smooth a trouble-free experience, the reviews left will be negative, lowering the company's reputation
6	Reputation of the company affected	Intangible	Medium	Very high	Very High	Alibaba	The reputation of the company effects all the brands connected in conjunction. Current and potential retailers and merchants could also withdraw interest from the organisation.
7	RedMart Website	Services	Low	Medium	High	Alibaba	The website is a service which would receive just as much if not more backlash from the issue due to its connection to the organisation.
8	Analytics	Data	Very high	High	High	Analytics team	The analytics team responsible for collecting the information and activity related to the website, will have been

							passing inaccurate numbers to team like the marketing team. Inaccurate analytics can affect a business's actions and decisions.

Identifying Vulnerabilities and threats

Threat Agent	Can Exploit This Vulnerability	Resulting in this Threat
Attacker	Procedures not in place to prevent things such as web scraping and other exploitation techniques	Sensitive data being leaked and stolen
Employees	Lack of training to be able to identify irregularities in data and how to handle data, making sure it is secure.	Incorrect data being provided, in turn affecting the business and operations
Malware	Aspects of potential vulnerability not covered to mitigate potential breaches	Exploitation of these areas to infect the system with viruses

Risk Register

Risk ID	Risks associated with the assets	Vulnerability	Impact	Risk Level (%)	Action/Control
1	1,4	The absence of controls in place to protect the sensitive data along with the database	Hackers taking advantage of the vulnerabilities in the systems and websites defences to infiltrate	80%	When the risk takes place, Alibaba should take steps towards securing the information. Hiring a penetration tester to check the security of the websites and other systems should also take place. Various layers of

					security need to be added to mitigate the chances of a data breach happening. Policies can be imposed too to help educate employees on the importance of security
2	2,3,4,6,7	The services are vulnerable to a loss of engagement as they become collateral damage due to being possessions of the business. This would cause the revenue to decline further down in comparison to the initial decrease	The decline of the revenue and reputation can impact stock price and increase liquidity risk. The organisation will also lose customers whilst the sales drop	95%	After the risk happens, Alibaba needs to reassure the media, users and clients they have the situation under control. Technological safeguards should be put into place, so the organisation does not reach this state of affairs. To reduce the chances of this risk taking place, legal teams should be trained and educated in the actions to take if these problems may arise. Regularly changing the websites HTML can decrease the probability of these risks
3	8	The analytics have become distorted due to the 9 months of web scraping.	Alibaba would make wrong or inaccurate decisions when it comes to directions to take for the business	65%	The team in charge of analytics can look at the data before the breach took place and compare it to the current data. A comparison should make it easier to identify the issues and distortions. Implementing technical controls into the website to detect when irregular activity is taking place would lower the chances of this happening. Honeypot pages, Rate limiting, and CAPTCHAs are an effective control methods for catching malicious bots

Proposed Controls

If specific security protocols were put in place Alibaba could mitigate or prevent incidents similar to this occurring in the future. The creation of an information security management system (ISMS) would allow the Alibaba to reinforce the current system in place through identification of any deficiencies. Where web-scraping and crawling are involved it is increasingly difficult to format a plan to perfectly repel threats that may present themselves. By implementing controls into the website's framework, reducing and limiting the risk towards the organisation via breaches or various forms of faults in security is possible. The three different controls used to mitigate are procedural, technical, and physical.

Procedural Controls

As human error is one of the most common faults leading to a security breach, it is vital for user behaviour to be guided within the security context. If the values of the organisations security are held to a low standard, the policies put in place to help prevent these incidents would become meaningless.

Password use policies

Password policies are rules put into place to help increase the security of the computer and network. These rules are created in order for strong, complex passwords to be created. Maintaining a solid baseline of security is a fundamental component for all forms of security. The policy may contain the password strength required, the need for varying passwords for different systems and no reuse of old passwords. A limited number of entries should before the account is locked, or a cooldown is activated should also be included. Passwords could also be changed after a certain period of time has passed. Likewise, similar policies can be introduced to the biometric system.

Information classification

In any organisation, protecting classified information from unwanted and unnecessary access is desired. Therefore, keeping certain information classified unless the personnel has the required level of clearance, is an effective method to keep information secure. Staff will only have access to the assets and data they require to carry out their tasks. By assigning a designated asset owner, they can assign and control the levels of access others have to assets. Labelling data with an information classification policy will aid in distinguishing the rules about the use of the data and publication as well as other regulations.

Education and training programs

Educating and training staff on the security needs is as crucial as all other security controls. Procedures need to be set up for all personnel to be educated about security measures and how to keep data they interact with safe and secure. In the event of an incident it is important staff know how it should be handled and where to report. Training can also involve how to use specific systems or the procedures necessary when dealing with particular assets. Alibaba handles a range of technology and information, so it is significant for employees to maintain a certain level of awareness when carrying out tasks. This even stretches to being conscious of potential threats via email. Something which may seem insignificant such as phishing or viruses through emails or links can be detrimental towards an organisation.

Storage/Deletion Policies

The aim of policies like this are to control how data is handled and treated by personnel in the organisation. These policies are required by law to follow the data protection principles. The General Data Protection Regulation states that any information collected is to be used lawfully. Alibaba can incorporate these policies into the organisations framework to control the storage of data. The

stored data should be encrypted when it is not in use and during and transfers. This reduces the vulnerability of the data to any feasible breaches. After a certain period of time or when the data is no longer being used, it should be disposed of in a secure process to avoid any leaks or copies. To understand and identify the point of vulnerability which may appear, it is important to understand the flow of data. Therefore, data mapping should be executed to match multiple data sets. The aim of the mapping is to prepare the data for any analytical processes it may have to go through. Alibaba can improve both the flow and security of data through the implementation of these policies.

Data Center Security

The purpose of data center security is to enforce policies and various practices with the goal of averting any cases of unauthorised access to the facility or information. Physical threats can be as impactful as cyber threats. The cabinets where the information is being retained should be secured using locks and other means of security. Biometric systems can also be incorporated into the system at the data center. Employees who operate the center should also receive background checks. Over 30% of data breaches are caused by internal personnel. Alibaba was also breached via an affiliate so this is a step they should consider including. Security guards and CCTV can be added to add to the layers of security, along with limited entry and exit points for the building.

Technical Controls

Technical controls are the utilization of hardware and/or software solutions to aid mitigation of the risk to the organisation. In order to identify which methods to put in place, it is required for there to be an understanding of the attacks form. The majority of these controls will serve to mitigate the vulnerability of information.

Rate limiting and CAPTCHAs

Setting rate limits for IP addresses and CAPTCHAs in place should aid in deterring attackers. IP blocking can also be utilized within these techniques. When an IP address is detected making an increased number of requests, access from that address can be blocked if the system detects inhumane activity. Employing these methods should increase the complexity for any forms of scraping to be carried out. However, it is still plausible for these defences to be bypassed with the application of a list of IP addresses and CAPTCHA solving services.

Firewalls and Anti-viruses

Security rules can be set to block specific traffic if it meets the criteria. This method stops any undesired traffic passing through the network. There are three types of firewalls; software firewall, next-generation firewall and hardware firewall. Firewalls provide several benefits when implemented. They send alerts when malicious activity is identified, whilst also tracking IPs along with signatures. If the signature resembles a malicious attack the firewall will know. Alibaba can apply this to their systems to aid the detection of unprecedented attacks in the network.

Regular audits and Stripping metadata

An additional method of control for Alibaba to use, would be auditing the website on a regular basis. This action can help ensure critical information is not accidentally exposed through websites viewable to the public by identifying vulnerabilities in security. This information can be in the form of data stored in the back-end databases which are linked through the website.
By implementing a process to strip the metadata from documents before they are published, key information can be protected from potential vulnerability. This data includes print queues, usernames, software versions and file paths.

Generic message returns

This factor may appear insignificant in comparison to others; however, it is just as crucial. When carrying out tasks such as resetting a password, the messages shown in response may reveal personal information including the full name. An example would be the phone number or full name associated with the email being presented on the screen. To combat this issue, generic messages should be displayed to the user, letting the user know that if the account does exist, an email or text message will be sent.

Regularly change website's HTML

Changing the websites HTML on a consistent basis, could cause the scraper to give up their search. When scrapers are searching for data to extract, they take advantage of the patterns within the HTML markup to reach their goal. Altering the HTML does not mean the whole website is required to be redesigned. Alibaba can request the web developers to alter the classes and IDs of elements within the HTML. As the markup is frequently changing, the location of specific data becomes difficult to locate. Additionally, inserting fake data within the IDs and classes which already contain the old markup, can ruin the scraper.

Creating Honeypot pages

Another control would be honeypot pages. There are different types of honeypot traps with varying purposes (Email traps, decoy database, malware honeypot and spider honeypot). The spider honeypot is used to trap web-crawlers through the utilization of links and web pages. These are pages a human user would rarely ever visit. Only a bot passing through all links in a page would visit a honeypot page. Due to this factor, when a client has been detected visiting the page, it is almost assured they are not a human user. With that knowledge, all requests originating from that client can be blocked. Monitoring the traffic entering the honeypot, it becomes possible to grasp the level of threat the bot poses, along with where the cybercriminals are accessing the system from. Furthermore, the data they may be interested in could also be revealed.

Intrusion detection system

Intrusion detection systems are used as to monitor traffic coming in and out of the network and detect any malicious activity. This system is a passive monitoring system used to only alert and report activity. IDS's can be implemented on the network, becoming a network-based intrusion detection system (NIDS). For this variation, hardware sensors are used within the network. Detection systems are also installed onto computers, where they analyse data passing through the network. A host-based intrusion detection system (HIDS) is installed on each system individually. Host intrusion detection systems have the ability to notice when the system is altered. Using signatures to detect threats, the system provides a number of benefits. Alibaba is a global organisation with a collective of headquarters, meaning the variations in systems would be beneficial to the them company. With a high quantity of data, the detection systems can help monitor the volume

Physical Controls

Physical controls are put in place to prevent unauthorised access to the building and any sensitive information an organisation may hold.

Security Guards

Employing security guards to guard the out and inwards of the organisations building can provide multiple benefits. They are an effective first line of defence when it comes to vulnerabilities. Acting as a potent method to deter criminal behaviour, security guards are able to detect abnormalities within the surroundings in real time. As seen earlier in the portfolio, employees of a business can be responsible for threats towards the company. Having security guards stationed inside the building, would mitigate the chances of any forms of theft from occurring.

c

CCTV

CCTV would further the security of the building and can be applied to blind spots that may be found around the company. Due to this implementation, if someone slips through security measures in place, the unauthorised access to areas within the building are recognizable. Furthermore, potential breaches from within Alibaba would become an increasingly controllable factor. An example would be an employee attempting to access an area of the site without a permission, will be identified via CCTV. Security guards can also be deployed to monitor the CCTV footage, further widening the scope of security. Additionally, the task of keeping the staff and clients safe becomes easier to achieve.

Biometrics and key card access

Biometric security devices require validation and the consent of individuals who choose to use the system. As these are highly important components with multifactor authentication, implementing this appliance into the security structure can become an essential element to security. Biometric systems have can be used to keep an accurate record of who has accessed specific information or areas of the organisation. There are multiple ways to incorporate this into security. These methods encompass fingerprints, facial recognition (inclusive of iris recognition) and vascular pattern recognition. As the data is recorded, if any situations do arise, staff will be put in the position to take accountability for their actions. Integrating the system with biometric access as a way to control and track who can enter and exit the building can deter unwanted circumstances. Moreover, company losses due to illegal entry or fraud can be avoided, saving money in the process. Adding biometrics as a form of security can be exceedingly expensive, especially depending on the scale of implementation. As well as biometrics, key cards can be used for the same purpose. However, there may be situations where it may be stolen or lost.

Evaluation

Throughout this portfolio I have analysed the factors which led to the breach in Alibaba's security. The complication of the attack was not out Alibaba's capabilities to prevent. With the size of the organisation reaching a global scale, the costs of implementing the security necessary to ensure the data was secure is not a major concern. The company overlooked the threat of a web-scraping attack as it does not necessarily have a direct effect on the organisation. However, the repercussions of the breach in security, far outweigh leaving the exposure. If the company had assessed the potential attack vectors against the website, they could have taken steps to reduce chances of any data being stolen. After this breach all the systems should be reinforced with security measures to protect databases and other important assets to the business. In addition, physical control measures should be carried out to protect Alibaba from any threats that may be from outside or inwards. As the attack was executed by an acquaintance of the company, it is vital Alibaba are aware of the potential threats employees may pose.

The controls which have been looked at within this portfolio serve the purpose of strengthening the defences of the organisation. The proposed controls should be able to cover more areas in comparison to prior security protocols. Implementation of the technical and procedural controls before the scraping attack could have prevented the breach from reach the scale it did. Even though there were no physical breaches, forms of security which could be set up were recommended. This was for the purpose of covering more sectors on possible holes in the protection of the organisation and personnel. The creation of the plan, after the analysis of risks, provides Alibaba with the correct methods and direction to take in terms of cyber security.

Bibliography

Ghadawala, M. (2019). *7 Key Benefits of Security with the Addition of Biometrics*. [online] www.globalsign.com. Available at: https://www.globalsign.com/en/blog/7-benefits-security-with-biometrics [Accessed 4 Mar. 2022].

Hashed Out by The SSL Store™. (2021). *What Is Data Center Security? 6 Ways to Ensure Your Interests Are Protected*. [online] Available at: https://www.thesslstore.com/blog/what-is-data-center-security-6-ways-to-ensure-your-interests-are-protected/ [Accessed 8 Mar. 2022].

Hillier, W. (n.d.). *What Is Web Scraping? [A Complete Step-by-Step Guide]*. [online] careerfoundry.com. Available at: https://careerfoundry.com/en/blog/data-analytics/web-scraping-guide/#how-to-scrape-the-web-step-by-step [Accessed 1 Mar. 2022].

Hope, A. (2021). *Web Scraping on Alibaba's Taobao Resulted in Data Leak of 1.1 Billion Records*. [online] CPO Magazine. Available at: https://www.cpomagazine.com/cyber-security/web-scraping-on-alibabas-taobao-resulted-in-data-leak-of-1-1-billion-records/ [Accessed 5 Mar. 2022].

Jie, Y. and Lin, L. (2021). Alibaba Falls Victim to Chinese Web Crawler in Large Data Leak. *Wall Street Journal*. [online] 16 Jun. Available at: https://www.wsj.com/articles/alibaba-falls-victim-to-chinese-web-crawler-in-large-data-leak-11623774850 [Accessed 5 Mar. 2022].

Kaspersky (2020). *What is a honeypot?* [online] www.kaspersky.co.uk. Available at: https://www.kaspersky.co.uk/resource-center/threats/what-is-a-honeypot [Accessed 6 Mar. 2022].

Learning Center. (n.d.). *What Is Scraping | About Price & Web Scraping Tools | Imperva*. [online] Available at: https://www.imperva.com/learn/application-security/web-scraping-attack/#:~:text=Web%20scraping%20is%20the%20process [Accessed 7 Mar. 2022].

Liu, C. (n.d.). *Defending Against Web Scraping Attacks*. [online] Dark Reading. Available at: https://www.darkreading.com/endpoint/defending-against-web-scraping-attacks/a/d-id/1340846 [Accessed 2 Mar. 2022].

Marzouk, Z. (n.d.). *Alibaba data breach exposes 1.1 billion pieces of data*. [online] IT PRO. Available at: https://www.itpro.co.uk/security/data-breaches/359897/alibaba-data-breach-exposes-11-billion-pieces-of-data [Accessed 1 Mar. 2022].

PYMNTS (2021). *Alibaba's Data Leak Exposed User Information*. [online] www.pymnts.com. Available at: https://www.pymnts.com/data/2021/data-leak-at-alibabas-shopping-site-taobao-exposed-billion-bits-of-user-info/ [Accessed 6 Mar. 2022].

Sahin, K. (2021). *Web Scraping vs Web Crawling: Ultimate Guide*. [online] www.scrapingbee.com. Available at: https://www.scrapingbee.com/blog/scraping-vs-crawling/ [Accessed 7 Mar. 2022].

Seqrite (2018). *Benefits of having Intrusion Prevention/Detection System in your enterprise*. [online] Seqrite Blog. Available at: https://www.seqrite.com/blog/benefits-of-having-intrusion-preventiondetection-system-in-your-enterprise/ [Accessed 9 Mar. 2022].

Singhal, G. (2020). *Advanced Python Web Scraping Tactics | Pluralsight*. [online] www.pluralsight.com. Available at: https://www.pluralsight.com/guides/advanced-web-scraping-tactics-python-playbook [Accessed 4 Mar. 2022].

SolarwindsMSP (2020). *Why a Strong Password Policy Matters: Tips for Service Providers*. [online] Passportal MSP. Available at: https://www.passportalmsp.com/blog/how-to-make-a-strong-password-policy-tips-for-service-providers [Accessed 5 Mar. 2022].

Thatha, R. (2018). *Scraping Attacks: Compromising Web Security, Impacting Business Continuity*. [online] Security Boulevard. Available at: https://securityboulevard.com/2018/10/how-scraping-attacks-can-compromise-web-security-and-impact-business-continuity/ [Accessed 3 Mar. 2022].

www.bloomberg.com. (n.d.). *Bloomberg - Are you a robot?* [online] Available at: https://www.bloomberg.com/news/articles/2021-06-16/alibaba-victim-of-huge-data-leak-as-china-tightens-security [Accessed 1 Mar. 2022].

Yuen, S. (2021). *Alibaba falls victim to data leak, insists no customer data was sold*. [online] www.marketing-interactive.com. Available at: https://www.marketing-interactive.com/alibaba-falls-victim-to-data-leak-insists-no-customer-data-was-sold [Accessed 7 Mar. 2022].

Imperva.com. (2022). [online] Available at: https://www.imperva.com/learn/wp-content/uploads/sites/13/2019/01/web-scraping-attack.jpg [Accessed 10 Mar. 2022].