UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INFORMÁTICOS

MÁSTER UNIVERSITARIO EN INGENIERÍA INFORMÁTICA

Special Problems in Information Security: from Privacy to Emerging Technologies for Hyperconnected Systems

Autor: Francisco García Martínez

Director: Maurice Dawson

Chicago, Julio 2019

Resumen

Desde las últimas elecciones en Estados Unidos, Francia y otros países, las “fake news” se han convertido en una herramienta para manipular a los votantes. Esta creación de “fake news” crea un problema que se extiende por toda una sociedad creando división. Sin embargo, los medios de comunicación no han escudriñado lo suficiente en el uso indebido de los datos. A diario, parece que hay brechas de seguridad que causan que millones de usuarios vean su información personal recogida, expuesta y vendida en la “Dark Web” a cambio de criptomonedas. Recientemente, han aparecido noticias en las redes sociales anunciando casos de correos electrónicos leídos sin el consentimiento del usuario. Estas cuestiones suscitan preocupación por el uso indebido de los datos personales y, lo que es más importante, por la forma en que pueden utilizarse para la guerra de la información y la explotación de grupos específicos mediante el uso de Internet. Es esencial que las organizaciones revisen continuamente las políticas de datos actuales para asegurarse de que no se conviertan en víctimas de la guerra de la información.

Sin embargo, no sólo es responsabilidad de las organizaciones preservar los derechos y libertades de sus empleados y clientes, sino también de los gobiernos y las naciones. Por eso, actualmente se están elaborando normativas en materia de protección de datos. En Europa, la creación del Reglamento General de Protección de Datos (RGPD) ha constituido un enorme avance en la privacidad de datos, otorgando mayor poder a los consumidores online, que estaban condenados a la pérdida total del control de su información personal. Aunque a primera vista pueda parecer que sólo afecta a las empresas de la Unión Europea, el Reglamento establece claramente que toda empresa que tenga negocios en la UE debe cumplir con el GDPR. Otros países no pertenecientes a la UE, como los Estados Unidos, han visto los beneficios del RGPD y ya están desarrollando sus propias leyes de privacidad. Inicialmente, estas regulaciones eran exclusivamente a nivel estatal, pero últimamente se están proponiendo iniciativas nacionales. Presentaremos algunos ejemplos representativos y los compararemos con el GDPR.

Además, nada de esto tendría sentido sin el desarrollo de tecnologías y aplicaciones seguras que preserven la privacidad y la confidencialidad. Numerosos estándares y códigos de buenas prácticas recomiendan la implementación de prácticas de seguridad desde las primeras etapas del proceso de desarrollo de aplicaciones. Comúnmente conocidas como "seguridad por defecto", estas prácticas consisten en programar teniendo en cuenta la seguridad para empezar a abordar las amenazas y vulnerabilidades antes de que la integración de las medidas de seguridad resulte demasiado tediosa. Una manera efectiva de averiguar si la aplicación es segura es mediante técnicas de análisis del código fuente. Estos análisis reportan debilidades o malas prácticas en el código una vez que se evalúan frente un conjunto predefinido de reglas. De esa forma, los programadores son capaces de detectar y corregir estos problemas, evitando la explotación de dichas vulnerabilidades en el futuro y garantizando la protección de los datos personales de los usuarios. En este proyecto, se ha realizado un análisis estático del código fuente para evaluar la aplicación de escritorio de un Sistema Nacional de Denuncias para la Policía Nacional de un país latinoamericano.

Finalmente, para educar a los usuarios en estos temas, proponemos un marco que permita el desarrollo de un laboratorio virtual que incorpore tecnologías emergentes, como la Internet Industrial de los Objetos y los sistemas hiperconectados, incorporando componentes de código abierto. Este laboratorio virtual proporcionaría un entorno de aprendizaje en el que los conceptos de ciberseguridad y seguridad de la información podrían enseñarse en un entorno exploratorio.

Abstract

Since the last elections in the United States, France, and other nations, fake news has become a tool to manipulate voters. This creation of fake news creates a problem that ripples through an entire society creating division. However, the media has not scrutinized enough on data misuse. Daily it appears that there are breaches causing millions of users to have their personal information taken, exposed, and sold on the Dark Web in exchange of encrypted currencies. Recently, news has surfaced of major social media sites allowing emails to be read without user consent. These issues bring upon concern for the misuse of data and more importantly, how can this be used for information warfare and the exploitation of targeted groups through the use of the Internet. It is essential that organizations continuously review current data policies to ensure that they do not become victims of information warfare.

Nevertheless, it is not only the organizations' responsibility to preserve the rights and freedoms of their employees and customers, but also governments and nations should have a word in this matter. That is why data protection regulations are being developed. In Europe, the creation of the General Data Protection Regulation (GDPR) constituted an enormous advance in data privacy, empowering the online consumers, who were doomed to the complete loss of control of their personal information. Although it may first seem that it only affects companies within the European Union, the regulation clearly states that every company who has businesses in the EU must be compliant with the GDPR. Other non-EU countries, like the United States, have seen the benefits of the GDPR and are already developing their own privacy laws. Initially, these regulations were exclusively at the state level, but national initiatives are lately being proposed. We will present some representative examples and compare them to the GDPR.

Furthermore, none of this would make sense without the development of secure technologies and applications that preserve privacy and confidentiality. Numerous standards and codes of best practice recommend the implementation of security practices since the early stages of the application development process. Commonly known as 'security-as-default', these practices consist in coding while keeping security in mind to start addressing threats and vulnerabilities before integrating security measures becomes too laborious. An effective way to find out whether the application is secure is by performing source code analysis. These analysis report weaknesses or poor practices in the code once run against predefined set of rules. Consequently, developers are able to detect and correct these issues, preventing the application from future exploits and ensuring individuals' data is protected. A static source code analysis was performed to assess the desktop application of a National Crime Reporting System for a Latin American country's National Police.

Finally, to educate users in these issues, we are proposing a framework that allows for the development of a virtual lab that incorporates emerging technologies, such as the Industrial Internet of Things and hyperconnected systems while incorporating open source components. This virtual lab would provide a learning environment where cybersecurity and information security concepts can be taught in an exploratory environment.

Table of Contents

Table of Figures

Table of Tables

1. INTRODUCTION

In the battlefield, there is a type of warfare known as psychological operations. This aspect of warfare, used to create a favorable image, gaining adherents, and undermining opponents had already become a significant weapon of 20th-century warfare (Headquarters Department of the Army, 1979). However, “they are neither a substitute for power nor a panacea,” but employed correctly they can be instrumental, making the difference between success or failure in military operations. And not exclusively military operations, but also in numerous other fields, such as technology or marketing. Information warfare is, in general terms, a way of protecting one's information infrastructure while attacking someone else's by using computers.

Although it may appear that the misuse of data for information warfare is something new, it has been employed for centuries. Nonetheless, the term has lately become popular due to the information availability with the Internet and the media. Specially, when we see thousands of data breaches in the news almost daily. With the widespread use of newer technologies, infinite quantities of data are released to the Internet every day. Finding the value in all this data and use it to gain a favorable position over competitors is crucial for myriad organizations. Thus, millions of dollars are being invested in tools and technologies to effectively collect the most valuable data, many times in an illegitimate way, without the users’ consent. There is where privacy regulations in the United States, following the General Data Protection Regulation (GDPR), try to come in place. At this point, whereas there exist privacy laws at the state level, there are no national, centralized laws in the United States that regulate the collection and use of personal data. Federal statutes are primarily focused on specific sectors, while state statutes are the ones aiming the privacy rights of individuals. The European Union has taken the first step in relation to users´ privacy in adapting to the widely used technologies in the times being. In the US, California, following the EU, is the first of many states that have already taken the initiative and are developing their own privacy laws.

One of the updates introduced by the GDPR is, in fact, the principle of ‘Privacy-by-default’, which indicates that, in order to preserve data protection, applications must be developed following best security practices since the early stages, always keeping privacy and security in mind. The best way to prevent data breaches is by creating systems that protect the exposure of personal information. Therefore, though humans wanted to unauthorizedly access or disclose this information, it would not be possible. In this project, we try to evaluate this principle using a real-world example using code analysis techniques. A static source code analysis was performed to assess a National Crime Reporting System for a Latin American country’s National Police. The system is comprised of three main modules: the mobile app, the desktop app and the backend service. For the scope of this project, the assessment focuses in the desktop part.

The project merges four different research projects in the areas previously exposed: misuse of data for information warfare, GDPR and US privacy, information security assessment of a real-world scenario and presentation of a virtual lab that provides a learning environment for hyperconnected systems to efficiently teach cybersecurity and data protection concepts. Thus, the document separately presents each of the research projects as a whole following a sequential structure, grouping some general conclusions at the end.

2. STATE OF ART

While the General Data Protection Regulation (GDPR) unifies the privacy laws in the European Union (UE), at this moment, there are not any national privacy or data protection laws in the United States of America. Nonetheless, various states, lead by California and its California Consumer Privacy Act (CCPA or "The Act") have begun to develop their own privacy regulations. The city of Chicago represents another great example with its Chicago Personal Data Collection and Protection Ordinance. Besides, it appears that the US is lately raising data privacy awareness at a national level when, on November 1, 2018, Senator Wyden released The Consumer Data Protection Act of 2018 Discussion Draft (Wyden, 2018). This project compares these representative state and national privacy initiatives in the US to the GDPR, highlighting their similarities and disparities.

Considered one of the most violent regions in the world, high crime rates in Latin America is a reality. Moreover, countless unreported crimes would need to be added to these elevated violence rates. Several initiatives are being carried out to provide Latin American citizens with the means to report and reduce the amount of criminal activities. However, these solutions are limited by the lack of adequate and reliable data and reporting methods (Sutton et al., 2017). The National Crime Reporting System provides the means for citizens to report accidents or suspicious events to the National Police of a Latin American country. Thanks to this system, a message along with a photo can be easily and quickly sent to the authorities by the citizens for investigation. This process needs to be reliable and secure, so that the National Police can confirm legitimate reports and analyze the collected data to assess the severity of the situation and thus, establish a response plan. There are two main approaches of source code analysis: static and dynamic (Bergeron et al., 2001). Static source code analysis consists in the examination of the program code to determine properties of the dynamic execution of this program without running it. On the other hand, dynamic analysis mainly consists in monitoring the execution of the program to detect malicious behavior due to, for instance, the passing of crafted inputs or extreme cases. We are conducting a static source code analysis for the purpose of this project.

In terms of virtual or remote labs for education, existing solutions often have to deal with memory resources problems when used by large groups of students, and lack systems' compatibility, particularly with new global technology shifts, like IoT and smart devices. As a consequence, this remote lab aims to be accessible from any mobile device allowing the ability to access a remote desktop tailored environment.

3. MISUSE OF DATA FOR INFORMATION WARFARE

3.1. Introduction

Since the last elections in the United States, France, and other nations, fake news has become a tool to manipulate voters. This creation of fake news creates a problem that ripples through an entire society creating division. However, the media has not scrutinized enough on data misuse. Daily it appears that there are breaches causing millions of users to have their personal information taken, exposed, and sold on the Dark Web in exchange of encrypted currencies. Recently, news has surfaced of major social media sites allowing emails to be read without user consent. These issues bring upon concern for the misuse of data and more importantly, how can this be used for information warfare and the exploitation of targeted groups through the use of the Internet. It is essential that organizations continuously review current data policies to ensure that they do not become victims of information warfare.

3.2. Information Warfare

In the battlefield, there is a type of warfare known as psychological operations. This aspect of warfare, used to create a favorable image, gaining adherents, and undermining opponents had already become a significant weapon of 20th-century warfare (Headquarters Department of the Army, 1979). However, “they are neither a substitute for power nor a panacea,” but employed correctly they can be instrumental, making the difference between success or failure in military operations. And not exclusively military operations, but also in numerous other fields, such as technology or marketing.

Information warfare is, in general terms, a way of protecting one's information infrastructure while attacking someone else's by using computers. In the past century, it was commonly considered how future wars would take place and, more importantly, the mean by they would be won (Aldrich, 1996). Consequently, information warfare has become a significant issue in recent decades for both governments and private companies, who have often joined forces to strengthen their economies over their adversaries. For instance, in the United States, government agencies such as the Central Intelligence Agency (CIA), the Federal Bureau of Investigation (FBI) or the National Security Agency (NSA) have cooperated with private organizations to create infrastructure protection programs (Elbirt, 2003). The term, whose first recorded use was by Thomas P. Rona of the Boeing Corporation in 1976, is of such a great importance that AJ Elbirt remarks in his paper that "the International Trade Commission estimated the loss in the United States due to economic espionage at $23.8 billion in 1987 and $40 billion in 1989". Besides, a study conducted by the University of Illinois in 1988 concluded that 48% of the companies surveyed admitted to being industrial espionage victims (Schawartau, 1997).

Typical information warfare attacks are a Denial of Service attack (DoS attack), phishing, social engineering, or deletion, manipulation or modification of data. All with the common goal of gaining a favorable position over their opponents by disrupting their services or stealing classified data. As a consequence, threatened by this information warfare attacks, the United States created in 1998 the National Infrastructure Protection Centre (NIPC). Typically, an attacker performs three steps when conducting an attack: information gathering, attack planning, and attack execution (Elbirt, 2003).

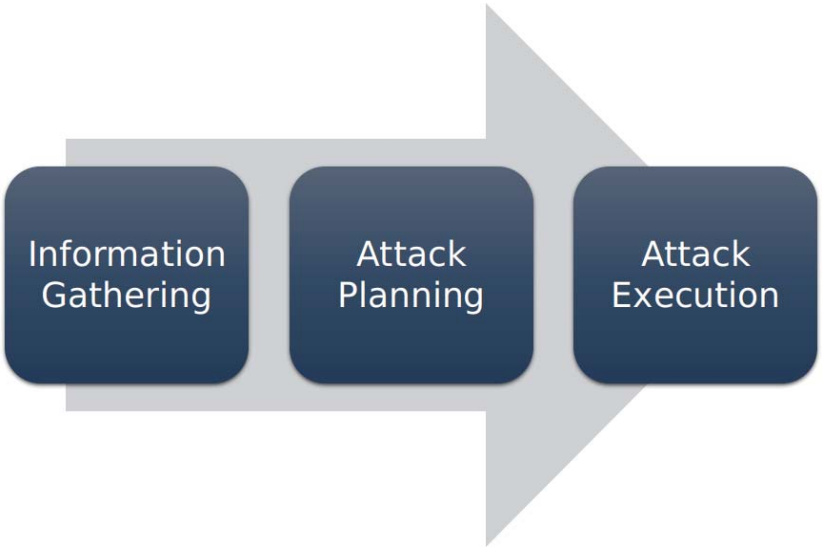

Figure 3.1. Information warfare process

The primary purpose is to retrieve as much information as possible from the adversary while protecting one's information infrastructure, thus ensuring confidentiality, integrity, and availability. From the economic point of view, there are three levels of impact in information warfare, being these personal, corporate, and global levels. The personal level affects a single individual or group of individuals electronic privacy. Attacks directed to the personal level include, but are not limited to, harassment, extortion, personal data theft or blackmailing. These kinds of attacks often consist of an individual's information gathering to, later on, perform a specific targeted campaign of blackmail or ransomware (O'Gorman & McDonald, 2012). Some other personal level attacks focus on unauthorizedly altering someone's private information. Thought this misinformation could be easily removed or corrected, due to the extremely fast spreading of data across the network, once an individual's privacy has been invaded, and the malicious activity has been done, the resulting damage is often irreparable.

When the conducted information warfare attacks are elevated to companies or organizations, they are often referred to as industrial espionage or the corporate level. The usual corporate level attacks involve competitor information theft, the release of their proprietary information, or disrupting an adversary's activity. Additionally, there have also been cases of governments making use of Information Warfare tactics to provide information to a private organization within the country from a competitor of a foreign country. Elbirt gives an example of this kind of activity in his paper: "Hitachi paid IBM a reported $300 million in a settlement agreement after being caught spying on a new generation of IBM computer equipment and that French intelligence was proven to have spied on Boeing to help Airbus" (Elbirt, 2003).

Economic espionage, or global level attacks, refer to the government's use of Information Warfare techniques to combat other countries or their allies in the desire of improving their economy or obtaining a better combative position. Nonetheless, these attacks are not limited to government activities, but they also include terrorist groups, such as Anonymous, Al-Qa'ida, or the famous Chinese cyber espionage group, Axiom. However, they require a large number of people involved and a significant monetary investment. A key aspect of being successful at the global level relays on being capable of organizing this vast number of people while maintaining a high level of privacy.

Concerning data collection, databases can represent a great source of useful data within the information warfare. Numerous access control countermeasures have been developed and are implemented, preventing unauthorized users from accessing and retrieving confidential information. Nevertheless, those techniques do not address the inference control problem, where a user could

perform legitimate general queries to the database as a whole while restricting him from extracting individual's private information (Elmasri, 2008). Clifton and Marks introduce some possible solutions in their paper (Clifton & Marks, 2010) to ensure that a company cannot infer private data from public data to, later, use it to gain a better position than its competitors in the information warfare.

3.3. All Source Intelligence

Analyzing data could provide valuable information regarding an organization's or individual's activity with the use of Open Source Intelligence (OSINT) tools. OSINT data is unclassified information or data that is publicly available. OSINT is not to be a substitute for other sources of intelligence but rather complement existing methods to collect information such as Geospatial Intelligence (GEOINT), Signal Intelligence (SIGINT), Human Intelligence (HUMINT), and Measurement Intelligence (MASINT). This data collection method relies on information that is found publicly without the need to request access to it, and it can be used to generate reports (Stalder and Hirsh, 2002). Having access to this data allows an attacker to develop an intelligence analysis on the target. This analysis can be a culmination of information about the target's movements, online behaviors, technical data, and more. With the Internet, several applications such as Maltego can make the profession of an OSINT analyst done with ease. This means they can create transforms, perform sentiment analysis of words, and review other public databases smoothly.

3.4. Misuse of Data

The widespread use of newer technologies and their correspondent tools and apps leads to infinite quantities of data released to the Internet. However, the most critical finding in the last recent years is that all this data has a value. All this information which was practically discarded was a source of intelligence that traditionally took a significant work effort to collect. Hence, enterprises have increased their investments in software, hardware, staff, education, and other associated items that constitute the digital world, by 50%, to $4 trillion (Gantz & Reinsel, 2018). Grantz and Reinsel state in their paper that "the number of information individuals create themselves - writing documents, taking pictures, downloading music, etc. - is far less than the amount of information being created about them in the digital universe". Therefore, we cannot imagine how significant this amount of data is, and even less wonder how to handle it. That is why companies are putting all their efforts to be able to generate value by extracting just the right information, or even by misusing the data for different purposes for what it was collected. Being capable of doing so would enormously help to position themselves in the "pole position" of the information warfare.

It cannot be denied that the new features included in popular apps usually make someone's life easier. However, the actual goal of the company for developing that new functionality remains unthinkable and unknown to the end user. These goals can range from the selling of data to third parties or collecting data to sell other products to the end user (Ahmed, 2004). It was probably not to make everyone's lives more comfortable but to know more about them; to gather more useful information about the people which can later be transformed into personal-oriented marketing strategies and, eventually, more revenues to the corporation. What enterprises usually achieve with these techniques is to get more private information about their users' data, or metadata, which, as a result, is growing extremely faster than the actual data itself. In recent years several patents can be found that deal with mobile data collection to (Sinisi, 2007). Facebook's new "face recognition" or "tag suggestion" feature is an excellent example of this. This functionality identifies a user's face in a picture and notifies him of the uploaded photo. Thus, the user can decide whether to be tagged in the photo or, even more, report someone who has uploaded a picture of him without consent. Although several privacy experts

claim that it is an excellent advance in protecting someone's privacy preventing fraud and identity theft, what Facebook does is maintaining what it is called a "template" (Fussell, 2018). This template is a string of numbers that is unique for each user, which could be considered similar to a fingerprint. As a consequence, Facebook becomes the owner of extremely protected biometric data of its customers, that could later be tasked for malicious purposes.

According to John T. Soma et al. personally identifiable information (PII) "is now a commodity that companies trade and sell" (Soma, Couson, & Cadkin, 2009). Furthermore, it is equaling or even surpassing the value of traditional financial assets in large corporations. Nevertheless, the question is: are companies benefitting from the use and trade of PII without protecting the privacy interests of those PII owners? This entails consequences for commercial and technological sectors.

In the marketing industry, the benefits of using PII are double (Soma, Couson, & Cadkin, 2009). Imagine that an online store sells alcohol to its consumers. Collecting data such as gender or nationality may not make any difference, but, if it also collected age values, it could significantly narrow its target to old enough consumers. Thus, the store would not only increase its revenues by approaching more likely possible buyers but also reduce costs by discarding underage consumers. Moreover, consumers can also benefit from companies keeping their PII, tailoring them future activity.

Cloud computing is becoming an excellent solution for many small and medium companies since it represents a great way of saving money by sharing resources with other organizations and avoid buying and maintaining their servers. However, regarding security, cloud providers may have to face different risks and challenges to the ones in conventional IT environments. From the end user's point of view, they are still reticent to cloud computing technologies, concerned about their data privacy and security issues — even more after knowing about the most significant cloud computing providers security breaches. Google Gmail was exposed to a severe vulnerability up to 4 hours in its VMware virtualization for Mac version in 2009, where attackers could take advantage of this vulnerability to execute malicious code on the host (Chen & Zhao, 2012). Microsoft Azure also suffered a severe outage accident on its cloud services for 22 hours earlier this year.

Concerning the health sector, due to the augment of health information available in the Internet, patients tend to look for their symptoms online, sharing especially private data to everyone, without considering its associated security risks. Researchers comment that "Both specialists and patients can benefit from linking family health profiles so that all relevant information is available for reference when the need arises," obviously, developing a safe and private environment (Gajanayake, Iannella, and Sahama, 2011). The access of illegitimate persons to one's health information can have critical consequences when later being disclosed or misused since it contains sensitive data tremendously useful for ransomware or social engineering attacks. Thus, they propose an information accountability mechanism as the solution to information misuse in the health field. Moreover, they claim that with their approach "when inappropriate misuse is detected, the agent defines methods of holding the users accountable for misuse."

3.5.PII Exploits

Krishnamurthy and Wills define personally Identifiable Information (PII) as "information which can be used to distinguish or trace an individual's identity either alone or when combined with other public information that is linkable to a specific individual" (Krishnamurthy and Willis, 2009). The term encompasses any information that can uniquely identify an individual, such as name, birthday, address, phone number, social security number, fingerprints, or a face photo.

Social networking sites are web-based services that allow their members to build a public or semi-public profile and connect with other strangers based on shared interests, hobbies, or political thoughts (Boyd and Ellison, 2007). We could say that social media is an expansion of traditional media, offering individuals highly capable and nearly unlimited ways of communicating and networking with others. There are many different kinds of social media business models, varying from sharing live-photos of places you are currently visiting activities focused on growing your professional network and seek jobs. Nevertheless, just like everything in this world, social networking sites also have their drawbacks. Users do not often realize the massive amounts of personal data that they are sharing with their network and thus, how they are being exposed to exploits of these data.

All social networks offer a wide range of possibilities concerning the privacy settings of their members. If an individual leaves these settings public by default, this can constitute a breach of privacy. Consequently, a malicious user can perform a reconnaissance attack and gather as much possible information to conduct a successful social engineering attack later. However, having a public profile is not the only vulnerability to private information on a social networking site (Gundecha, Barbier, & Liu, 2011). In their paper, P. Gundecha et al. discuss how a social media user can become way more exposed to exploits of his data by merely adding a vulnerable friend. They define a vulnerable friend "from an individual user's perspective is dependent on whether or not the user's friends' privacy settings protect the friend and the individual's network of friends (which includes the user)." Hence, a single user's privacy settings can compromise its entire network.

Frequently, social media websites partner with third-party servers to provide content and advertisements to their users. Although these websites claim in their privacy policies that they share cookies to third parties to offer a better user experience to their members, these cookies do not exclusively consist of Internet Protocol (IP) addresses (Symantec Corporation, n.d.). What is more, some third-party servers are in fact trackers or aggregators, that follow the user habits before, while and after the user's interaction with the social media application (Krishnamurthy & Willis, 2009). Krishnamurthy and Wills define this action of combining this PII with other information and sharing it to external websites as "leakage." In their paper, they present a study demonstrating how Online Social Networks (OSN) often provide information linked to a particular person to third parties via a combination of HTTP headers and cookies.

Most of the times, when a person publishes a document or picture on the Internet, he is not aware of the PII or other identifiers attached to it, even less how to remove them. There are countless situations in which personal information is retrieved from documents with inappropriate security. Therefore, this private data can further be used to commit malicious activities. An example of information leak caused by inadequate attempts to secure protected information took place in 2000 when a secret CIA document about a coup in Iran was published in The New York Times website (Aura, Kuhn, & Roe, 2006). The company unsuccessfully tried to erase the names of the persons involved by just painting white squares over their names. Consequently, the names were still in the publication's metadata and could easily be retrieved.

3.6.Where Stolen Data Can Be Found: Dark and Deep Web

The types of data captured through poor security practices and improper coding techniques provide not only side channels into the organizations but a plethora of details. For example, a photo provides lots of metadata that can give insight into camera type, specific detailed information of photo taken, latitude, and longitude coordinates. These items can be used to create an intelligence analysis of a target with the number of connected devices and those on the Web with a lack of security protections. However, the key is where do these stolen data and information end up.

The definition of the Internet as the mainstream perceives does not entirely represent what the entity is. Because of an increasing number of static HyperText Markup Language (HTML) pages, there is an enormous amount of information hidden in the layers of deep and dark Web where most search engines cannot have access (see Figure 3.2). The pathway to these remote Web locations is provided through static Uniform Resource Locator (URL) links due to their existence being depended on responses to queries submitted through the query interface of an underlying database. It is estimated that 43,000 to 96,000 deep Web sites exist along with 7,500 terabytes of data (He, Patel, Zhang, and Chang, 2007).

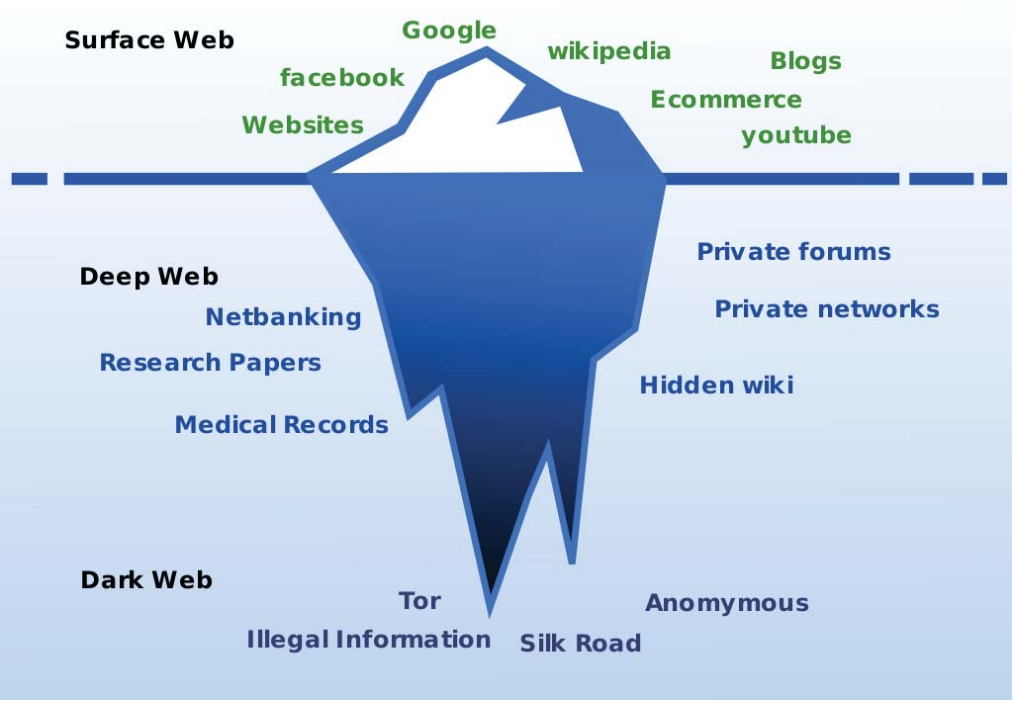

Figure 3.2.Complete Web retrieved from https://commons.wikimedia.org/wiki/File:Deepweb_graphical_representation.svg. Licensed under CC Attribution-Share Alike 4.0 International.

The issue with trying to locate deep websites is that they do not exist. That being the site is not indexed in a traditional sense like a standard search engine works. Take the search engine Google.com for example items are added to Google's database by either the website itself informing Google of their URL or the web crawlers looking for, finding, and indexing all "known" websites it finds. A deep website is not indexed in either capacity you need to know the URL of what you are searching for directly.

Now there are extensive databases that try to compile large amounts of search engine data, and these are called metasearch engines like DuckDuckGo (see Figure 3.3) and DogPile for example. These meta search engines allow you to search various standard search engines all at one time. In some cases, as many as 40-50 search engines can be searched with the entry of search terms and the press of a button. However, even these metasearch engines do not take into account the vast information that is found on the deep web.

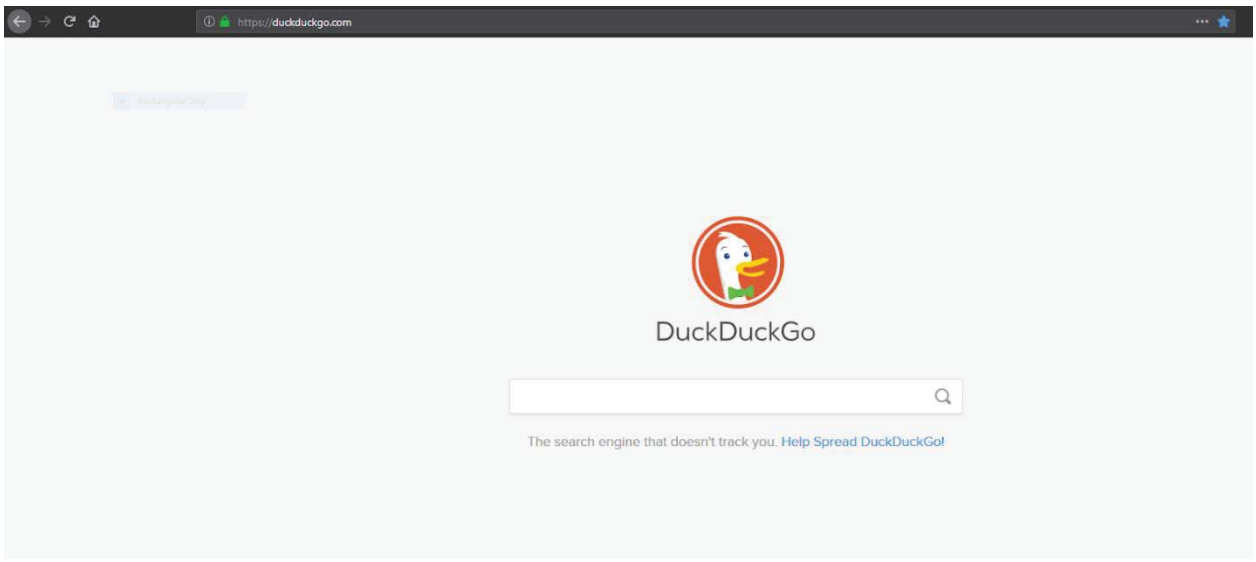

Figure 3.3. Example of the DuckDuckGo search page

There are several specialty search engines, like TORCH and the Onion URL Repository, which index as many deep websites that can be found. The key to these types of search engines is that they do not act like traditional searches. You need to have access to TOR networks which work as a semi-autonomous network that provides private browser and viewing of sites. Once you are on this network, you still be able to access repositories of different search engines usually broken down by subject matter and start digging into the deep web.

Another item of note is the deep web and the dark web is not the same thing. While you may make use of TOR to access the dark web search engines that index the deep web. Both these environments are independent of each other. Deep websites can be found using traditional browsing methods as long as you know the URL for it where dark websites leverage a software package like TOR to access the pages.

The Onion Router (TOR) gained popularity when the news was released around the globe about Edward Snowden exposing what the American government was doing with citizens' data. The tool of choice used was TOR. The Tor Browser can be used on Gnu Not Unix (GNU) Linux, Windows, and Mac without the need for installation of any software (Tor Project, n.d.). Tor was developed further by the Defense Advanced Research Projects Agency (DARPA) after the first principle of onion routing developed from a United States Naval Research Laboratory scientist. In Figure 3.4 shown are two Tor Browsers on Ubuntu Linux. The other browser shows The Uncensored Hidden Wiki and some onion links that have been verified. The first browser window displays the welcome message for anonymous exploration.

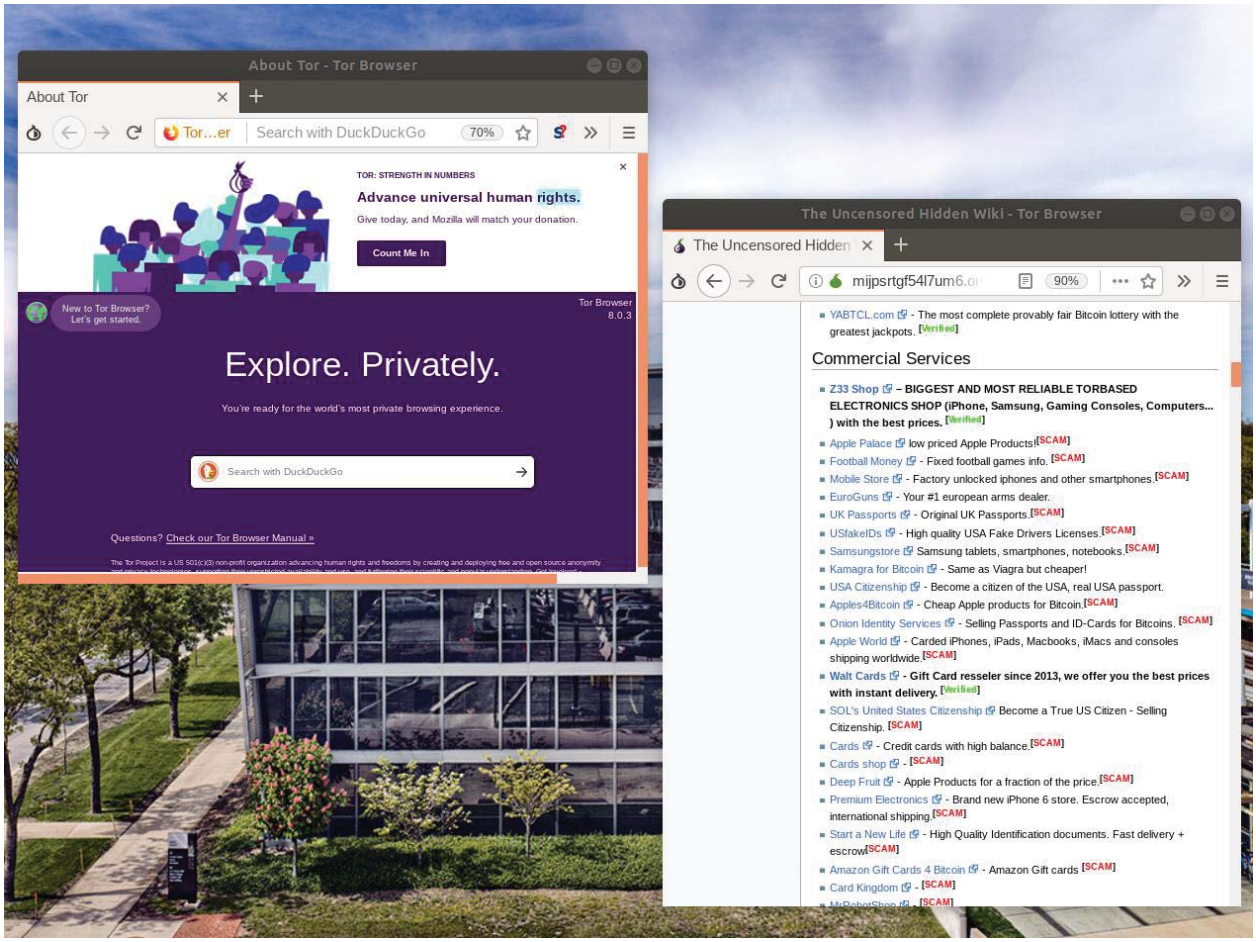

Figure 3.4. TOR browsers

TOR is native in the Tails Operating Systems (OS) Tails is a Debian based Linux distribution which primary goal is the preserve privacy and anonymity to beat surveillance.

In recent years, organizations such as the NSA have been attacking this browser. One attack revealed was the exploitation of the Tor Browser Bundle. When using the Tor Browser security that leaves a system vulnerable such as Flash become enabled in this attack (Schneier, 2013). This attack targeted the Firefox browser by identifying the Tor Users and executing attacks against the browser (Schneier, 2013). Other tools detected Hypertext Transfer Protocol (HTTP) through Capability Network Exploitation (CNE), which is the starting point for finding Tor users. Researchers at the University of Waterloo and Stony Brook University discuss active attacks for website fingerprinting to identify destination web pages by passively observing their communication traffic (Wang, Nithyanand, Johnson, & Goldberg, 2014).

However, these attacks have not deterred the use of Tor Browser. For users conducting illicit activities, this browser allows for undetected movement. One needs not to look too far to see the activities that occur on the Dark Web from the sale of illicit narcotics to human trafficking. Services from experienced hackers to assassins can be located using Tor and exploring Hidden Wiki.

Some browsers allow the user to protect their privacy. One such browser is Searx that does not share the users' IP, search history, and aggregates the results of more than seventy search engines (Tauber, n.d.). Searx browsers allow for advertisement filtering, personalization, and use of HTTP POST by default. Figure 3.5 shows the results of a search of Illinois Institute of Technology that populates that allows for files to be downloaded; pages scraped and allowed customization in terms of time.

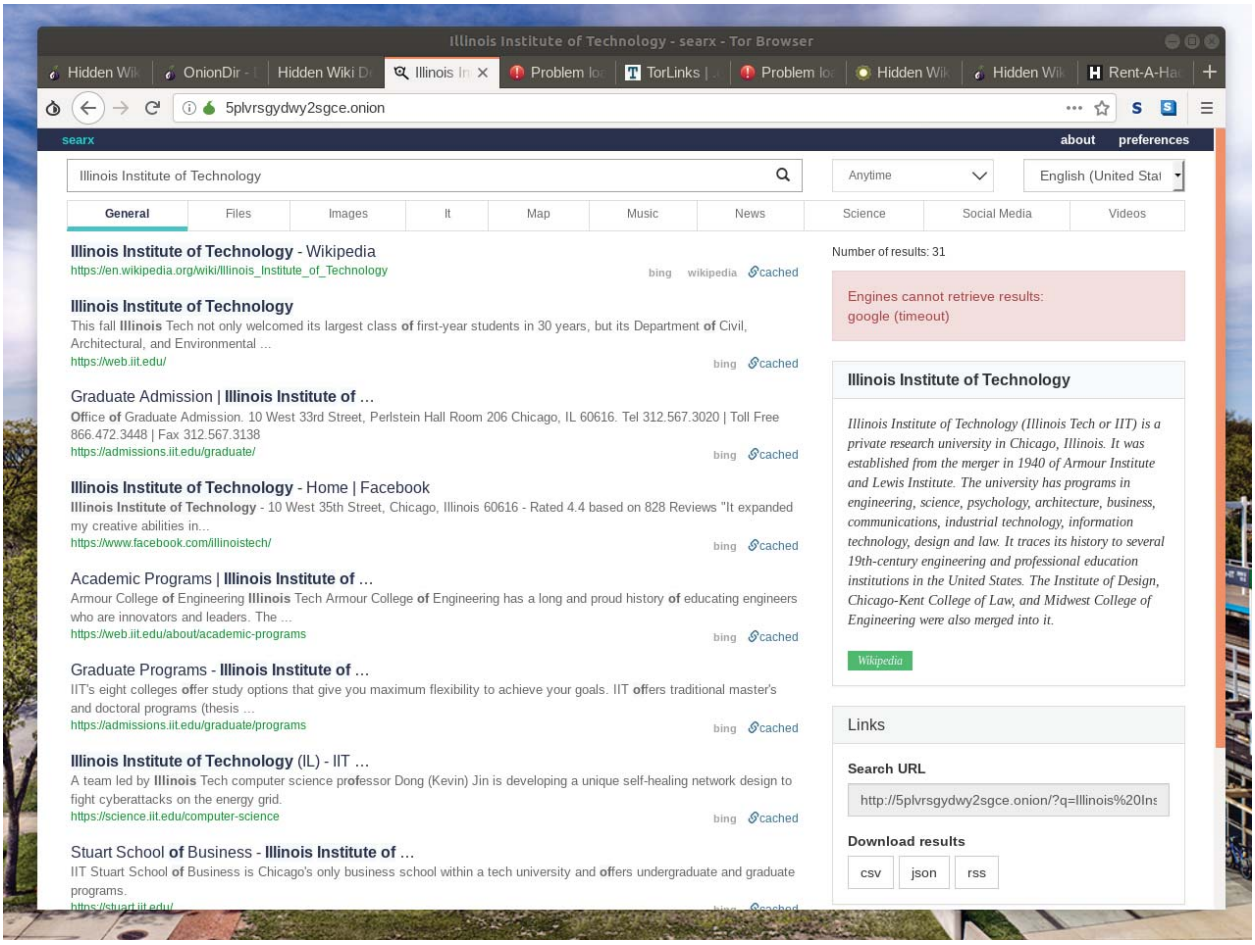

Figure 3.5. Searx browser

There have been several occasions where the Tor network has been abused for personal gain. In 2013 a Harvard University student used this mean of anonymity to send emails to the school for a hidden bomb threat to avoid a final exam (Lin, 2017). Silk Road is an online black market being accessed by nearly one million users through the exclusive access of the Onion Router. It includes illegal services like drug trafficking, child pornography, and arms trafficking; the value of its transactions has been calculated to be worth $12 billion. Its operations were shut down in October 2013 by the Federal Bureau of Investigation (Lin, 2017). “Anonymous,” the notorious worldwide hacker organization, launched a DDoS attack against Sony Corp in April 2011. They used the anonymous network and managed to steal the personal data of nearly 1 billion people. This attack had a disruptive financial impact of $171 million (Lin, 2017).

3.7.Using Web for Targeted Warfare

Researchers have discovered that Internet sites such as YouTube Kids and YouTube have detected unsafe content through nefarious promoters that target kids through psychological means (Kaushal, Saha, Bajaj, & Kumaraguru). This means that the threat landscape is altering to include all active users regardless of age or other constraints previously considered off limits. In the past, mainly adults have been the targets of individuals or national states. However, due to technological advances and increased connectivity, any connected user can be a target.

Reviewing the Open Web Application Security Project (OWASP) top 10 over the last ten years, it is apparent that the same critical web application vulnerabilities are still found (Wichers, 2013). One such vulnerability is the Common Weakness Enumeration (CWE) 89: Structured Query Language

(SQL) Injection, which is rather easy to exploit using an application called sqlmap. A simple search of php?id=[number] while bringing up several websites through a query that can be a potential target.

3.8. Conclusion

The misuse of data and deficiency of knowledge to apply security controls is a critical issue across enterprise networks. The Internet has allowed for older techniques used for warfare to be modernized at levels that make a novice intelligence analyst near a Subject Matter Expert (SME). This is a drastic change to the landscape of the current battlefield in which is still evolving with the ever expansion of networked systems such as the Internet of Things (IoT) and 5G. The apparent scarcity of applied cybersecurity protections is allowing for threat agents to take advantage of organizations and individuals that lack the necessary knowledge for ensuring protection. This, combined with laws that do not require companies to have stronger security, enable attackers to perform exploits continuously.

4. ANALYSIS OF THE US PRIVACY MODEL – IMPLICATIONS OF THE GDPR IN THE US

4.1. Introduction

Last May 25th, 2018 the General Data Protection Regulation (GDPR), the most shocking privacy law in Europe, and perhaps in the whole globe, became of full application. Although it is commonly wrongly said that last May was the date when the GDPR was enacted by the EU, truth is that it has been already effective for two years, since April 14th, 2016. The regulation had given European companies two entire years to become compliant. Thus, it is from May 25th, 2018 when every enterprise dealing with privacy data of European citizens must be fully compliant with the General Data Protection Regulation. Otherwise, they would have to face thrilling fines.

Nonetheless, one of the key things that need to be highlighted from this complex regulation is that not only European companies are affected by this new data protection law, but also foreign companies that do business in the EU. Numerous non-EU companies are now wondering whether they must follow the severe inflictions of the GDPR or if the European regulators are truly going to take serious actions.

What scares the most American companies are the penalties established by this new privacy law, which can go up to the higher of 4% of the company´s worldwide revenue or 20 million euros. However, European regulators recognize what they call good faith. This means that enterprises are not getting fined right away if they do not completely comply the regulation. Instead, the consensus is that EU regulators go slowly at first giving warnings before imposing the striking penalties (Cornock, 2018).

4.2. GDPR Most Significant Updates

Apart from the already mentioned increased penalties, the General Data Protection Regulation has included many other updates that directly affect US companies with businesses in Europe. This is, in fact, the most important update: every non-EU organization must be compliant with the regulation when they conduct activities related to the collection and treatment of private data to EU citizens ("Regulation (EU) 2016/679 of the European Parliament and of the Council", 2016).

With the goal of preserving the security and liability of the enterprise, as well as of offering guidance to technology professionals, controllers have to designate a qualified individual called Data Protection Officer (DPO) in the following scenarios ("Regulation (EU) 2016/679 of the European Parliament and of the Council", 2016):

- processing is carried out by a public authority, except a court acting in the exercise of its judicial function.
- the main activities consist of processing operations which, by reason of their nature, scope and/or purposes, require routine and systematic observation of subjects of large-scale data.
- the main activities consist of the large-scale processing of special categories of personal data and of data relating to convictions and criminal offences.

Another great way of preserving the security and minimizing risks is by performing a Privacy Impact Assessment (PIA). This is basically a risk assessment to better know the potential risks to which an

organization is exposed based upon the type of activities that it does with the personal data. Specifically, the GDPR defines that a PIA must be performed, at least, in any of the following cases ("Regulation (EU) 2016/679 of the European Parliament and of the Council", 2016):

- The company's activities involve profile elaboration.
- The company treats large scale sensitive data.
- The organization systematically observes great scale data of public areas.

The General Data Protection Regulation introduces a new principle, called "accountability", that implies a cultural and organizational change in enterprises. Thus, this principle states that companies should have a proactive and preventive attitude, instead of a reactive one, considering security as a part of their business model and demonstrating it. Besides, the regulation defines two key concepts related to this accountability principle: Privacy-by-design and Privacy-by-Default. As a consequence, personal data protection should be taken care of since the definition of the architecture or the implementation of the database, keeping privacy and data protection in mind at every step. What is more, this data protection law establishes that a register of all treatment activities must be maintained.

Whereas the previous data protection laws in Europe allowed a tacit consent to accept a user's personal data treatment, the GDPR imposes that this content has to be explicit and informed. This means that a company dealing with someone's personal data must be able to prove that it provided the user with all the required information related to his rights and purpose of uses of his data, and, only after doing so, the individual explicitly gave his consent. Additionally, this information can be presented in layers, informing about basic obligations at first, but providing a link to a deeper detailed explanation of all obligations of the law.

Notifying of data breaches within an appropriate period of time and, in any case, no later than 72 hours, to the Data Protection Authority of Control every time a data breach occurs and, to the interested individuals when the breach constitutes a high risk to their rights and freedoms, is another update introduced by the GDPR. Therefore, should any data breach take place affecting the privacy of European users, the controller or processor of the company must communicate it to them.

Previously, data protection laws in Europe provided a classification of personal data types, as well as their corresponding security measures. Nonetheless, the General Data Protection Regulation does not longer define a list of security measures, stating that corporations should be the ones implementing the appropriate technical and organizational security measures according to their previously identified risks.

This European data protection law includes an important principle: minimization of personal data. It basically refers to "exclusively gather the strictly needed personal data, only when you need them and only for the declared purposes". In this aspect, numerous US businesses are going to be affected, since this principle wants to eradicate the ongoing misuse of data problem in today's technological world (Elbirt, 2003), for instance, for information warfare. This principle leads to the user's right of treatment limitation, by which an individual can contest the use of his personal data.

Perhaps the most relevant update introduced by the General Data Protection Regulation for US companies is the international data transfer policy. This represents a key concept in today's globalization of services. One of the cases in which the regulation establishes that a corporation can transfer personal data to third companies or organizations if they are either part of the European Economic Area or they belong to a list of accepted countries provided by the European Commission,

which are considered to provide an adequate level of personal data protection (Bieker, 2017). One of these legitimate countries is United States. Concretely in this case, though there are other situations, the law affirms that international data transfer to United States is permitted if the entity is adhered to the "Privacy Shield" (Weiss and Archick, 2016). Consequently, to be GDPR compliant, enterprises should first identify and map all international data flows. If each of these flows are between a country in the EU or otherwise deemed adequate, the company can proceed with the personal data transfer. On the contrary, the company should consider whether any specific derogation apply or whether any appropriate safeguards have been put in place (Rosenberg, 2018).

4.3.GDPR vs. Privacy in the US

Since 9/11, the United States of America chose security over privacy. Therefore, at present, the US is still steps behind in data privacy. Whereas there exist many privacy laws at the state level, there are no national, centralized laws in the United States that regulate the collection and use of personal data. Federal statutes are primarily focused on specific sectors, while state statutes are the ones aiming the privacy rights of individuals. The European Union has taken the first step in relation to users´ privacy in adapting to the widely used technologies in the times being. California, following the EU, is the first of many states that have already taken the initiative and are developing their own privacy laws.

4.3.1. The California Consumer Privacy Act of 2018

Although quickly created and introduced, the California Consumer Privacy Act of 2018 (the "Act") ("The California Consumer Privacy Act", 2018) was signed into law by California Governor Jerry Brown on June 28, 2018. Like it happened with the GDPR, companies will have almost two years to become compliant with it, beginning effective in January 1, 2020. The Act, applicable to large businesses either located in the State of California or that deal with California residents' personal information, proposes as its major provision five basic rights to consumers:

"(1) The right of Californians to know what personal information is being collected about them.

(2) The right of Californians to know whether their personal information is sold or disclosed and to whom.

(3) The right of Californians to say no to the sale of personal information.

(4) The right of Californians to access their personal information.

(5) The right of Californians to equal service and price, even if they exercise their privacy rights." As it can be appreciated, the four first rights are very similar to the GDPR's premises. However, although they share general ideas, both laws are not as identical as it may seem (Mathews and Bowman, 2018). The following table shows a comparison between the two laws ("The California Consumer Privacy Act", 2018):

Table 4.1. The Act vs. GDPR

The Act	GDPR
Applies to companies outside California state borders.	Applies to companies outside EU borders.
Mainly concerned with consumers' privacy disclosures.	Regulates companies' disclosures, international transfers, data breaches notifications, security implementations...
Consumers' right-to-know. General privacy pol- icy and with more specifics available upon re- quest.	Consumers' right-to-know. Layered information. Additional requirements on how to present the data to the user.
Does not require companies to obtain user's con- sent. Only consumers' opportunity to opt-out the sale of their personal information.	Enforces companies to explicitly obtain users' consent to collect and treat their data with opt-in mechanisms.
Data subject: California resident defined under California tax law.	Data subject: any identified or identifiable person.
Cross-border data transfers not restricted. Only requirement of an agreement containing certain provisions.	Any transfer of personal data can only take place if controller and processor comply with certain conditions.
Civil penalties up to $7,500. For certain data breaches, between $100 and $750 per data subject per incident.	Penalties up to the higher of 4% of the company´s worldwide revenue or 20 million euros.
Companies cannot sell children's data if they are under 16, unless the consumer has opted-in to the sale.	Legal age children's data process 16 years old. Other EU countries may establish a lower age not less than 13.

4.3.2. Chicago Personal Data Collection and Protection Ordinance

The City of Chicago has also moved forward towards individuals' data privacy with the Amendment of Municipal Code Title 4 by adding new Chapter 4-402 entitled "Personal Data Collection and Protection Ordinance" (the "Ordinance") in April 18, 2018 ("Chicago Personal Data Collection and Protection Ordinance", 2018). The main purpose of this amendment is "to provide for the regulation of operators that collect sensitive customer personal information through the Internet about individual consumers in the City of Chicago". Differing from the California Consumer Privacy Act of 2018, the Ordinance enforces the use of prior opt-in consent from the consumers to use and share their personal data. Whereas the GDPR talked about an informed consent (sharing with the individual the use of its personal data in treatment), the Ordinance does not require companies to present the user the personal information they maintain, unless explicitly requested. Nevertheless, it shares some features with the GDPR, like the notification of data breaches without unreasonable delay, or that both apply to businesses outside the borders of the territory.

4.3.3. The Consumer Data Protection Act of 2018 Discussion Draft

Besides, it appears that the US is lately raising data privacy awareness at a national level. On November 1, 2018, Senator Wyden released The Consumer Data Protection Act of 2018 Discussion Draft (Wyden, 2018). On it, it is stated that "Information about consumers' activities, including their location information and the websites they visit is tracked, sold and monetized without their knowledge by many entities" and "Consumers have no effective way to control companies' use and

sharing of their data". Therefore, it proposes that the US as a nation should establish minimum privacy and cybersecurity standards to protect consumers' privacy. Moreover, like the GDPR, it would impose monetary fines up to 4% of annual revenue and from 10 to 20 years criminal penalties for senior executives. It would also "copy" the GDPR by giving more power to consumers: "Give consumers a way to review what personal information a company has about them, learn with whom it has been shared or sold, and to challenge inaccuracies in it" (Wyden, 2018).

4.3.4. Privacy State of Art Review at the National Level

Nevertheless, as previously commented, there is not any existent data protection law at the national level. The following table shows a brief actual state of art regarding privacy, highlighting four key points that the United States of America lack in relation to the GDPR.

Table 4.2. GDPR vs. US Privacy

	GDPR	**US**
Security Measures	Implement "appropriate technical and organizational measures" to protect the data.	Encrypt data in storage and in transit ("NIST SP 800-122", 2010).
Data breach notifications	Report all breaches that constitute a risk of harm to individuals' "rights and freedoms". Notify Data Protection Authority and individuals, without unreasonable delay.	Mostly inexistent. Notify only affected individuals. Timely notification restriction non-contemplated.
Consumers' power	Explicit opt-in consent to collect and share their data, after being informed of their rights, uses, etc.	Completely ignorance of what personal data is being collected or how it is being treated and shared.
International data transfers	Any transfer of personal data can only take place if controller and processor comply with certain conditions.	No restrictions of cross-border data transfers.

4.4. US Companies Adapting to the GDPR

So, what do US companies need to do to become GDPR compliant? Firstly, it is necessary to remark that complying with the regulation takes more than just a few days. Hence, countless businesses, especially small and medium companies, did not realize that fact in advance and have encountered several problems adapting to the law on time. In fact, there are still many corporations out there which are not GDPR compliant yet. It needs to be understood that adapting to the GDPR is a process that requires resources, time and money.

Perhaps the most important update introduced by the GDPR concerning the US is the invalidation of the US Safe Harbor cross-border personal data transfer framework and its replacement by the Privacy Shield (Voss, 2017). Any US company which wants to send or receive data to or from the EU must adhere, upon certification of commitments to the Privacy Shield. This provides guarantees from US agencies and means of enforcement in case of violations. Privacy Shield requires the companies to be shelf certified (Dode, 2018).

Corporations should emphasize their commitment to compliance with the GDPR and the integrity of their customers data. Before the law was enacted, the biggest thrill of a company in the US having a data breach was being publicly disclosed. Now, however, apart from loss of brand image, other concerns arise, such as the high penalties. What the regulation wants to achieve with this is that the enterprises become more proactive instead of reactive. Ways of becoming proactive include considering privacy at every business step (privacy by design) and try to appropriately implement, depending on the personal data or technologies used, the right security measures to avoid data breaches. Whereas data breaches may still appear, these should be notified to the correspondent Data Protection Agency and, in certain cases, to the affected individuals without unreasonable delay. Furthermore, companies must keep track of all the activities concerning personal data management, detailing what specific information they are collecting and how are they using it, as well as the technical and organizational security measures in place to preserve data privacy. Like previously explained, certain corporations may also have to perform Privacy Impact Assessments (PIA) and/or name a Data Protection Officer (DPO).

From the users' perspective, the General Data Protection Regulation empowers consumers. It defends the idea that users should be aware of what personal information enterprises are collecting from them, how are they treating it and whether it is transferred to other enterprise to, aware of this, explicitly give their consent to those specific purposes. Therefore, to be GDPR compliant, US companies should provide their users with their rights, how they can make use of them and all previously mentioned information. This can be done in two layers; a first layer briefly giving minimum information, and a link to a second more detailed layer containing the full privacy policy. As a result, companies must include a mechanism to verify that the user is giving explicit consent to the collection and treatment of his personal data.

4.5.Conclusions

GDPR is not a problem. It is a regulation that adapts to the times being and the widespread use of newer technologies, that helps corporations by protecting their consumers privacy. The most transparent and secure a business is about their consumers' personal information, the most confident consumers would be to share their data. Lawyer Robert Bond commented to CNBC "there may be some short-term pain in GDPR but if it creates trust and better customer experiences, it should lead to more long-term loyalty and over time better shareholder value". Thus, do not think of GDPR as a revolution, but an evolution.

There is not any privacy law similar to the GDPR present in the United States at the national level. However, whereas various states are still reluctant to set privacy above security, some of them have begun to develop privacy regulations realizing the importance of consumers data protection in today's technological world. The California Consumer Privacy Act of 2018 or Chicago Personal Data Collection and Protection Ordinance, although not as severe as the GDPR, represent great examples of these states. It seems that the US is finally moving towards data privacy, but a lot of progress and consolidation are still needed in this field.

5. ASSESSMENT OF NATIONAL CRIME REPORTING SYSTEM: DETAILED ANALYSIS OF THE DESKTOP APPLICATION

5.1. Introduction

For decades, Latin America has been traditionally considered one of the most violent regions in the globe. Inequality, degree of repression or governments´ effectiveness are commonly explanations for the high crime rates. Although there is not a consensus around the reasons of this occurrence, studies show that mortality due to violence in Latin America is much greater than in any other region (Soares and Naritomi, 2010). According to the World Health Organization (WHO), the number of violence casualties in Latin America is around 200 percent higher than in North America and 450 percent higher than in Western Europe. Besides, countless unreported crimes would need to be added to these elevated violence rates. Several initiatives are being carried out to provide Latin American citizens with the means to report and reduce the amount of criminal activities. However, these solutions are limited by the lack of adequate and reliable data and reporting methods (Sutton et al., 2017).

The National Crime Reporting System provides the means for citizens to report accidents or suspicious events to the National Police of a Latin American country. Thanks to this system, a message along with a photo can be easily and quickly sent to the authorities by the citizens for investigation. This process needs to be reliable and secure, so that the National Police can confirm legitimate reports and analyze the collected data to assess the severity of the situation and thus, establish a response plan.

In order to be a reliable source of crime reporting, every system component needs to be fully secure. From the user mobile application that pushes a message notifying of an incident to the police's systems receiving it, sent over through secure communication channel. In our case, we are assessing the source code of the desktop application, which is responsible of receiving and displaying the reports at the National Police offices' computers.

There are two main approaches of source code analysis: static and dynamic (Bergeron et al., 2001). Static source code analysis consists in the examination of the program code to determine properties of the dynamic execution of this program without running it. This technique has been very popular for decades to conduct analysis aiming to optimize the code (Louridas, 2006). On the other hand, dynamic analysis mainly consists in monitoring the execution of the program to detect malicious behavior due to, for instance, the passing of crafted inputs or extreme cases. This practice is widely used in the present for vulnerability scanning in web applications (Petukhov and Kozlov, 2008).

The purpose of this project is to perform a static code analysis to detect programming errors, vulnerabilities and poor coding practices. Using various open-source tools, these weaknesses in the source code will be classified and prioritized. Besides, recommendations will be presented based upon the results. This project is not intended to exploit the detected vulnerabilities.

5.2. Architecture

The National Crime Reporting System could be divided into three main modules: the mobile app, the desktop app and the backend service. Figure 1 shows the system's basic functionality. The former consists of a mobile application exclusive to Android devices that allow end users to report any suspicious activity or crime to the police computer systems. Police officers analyze and manage the

reports received from the citizens through the desktop application. This is, concretely, the module where this paper is focused, where police officers can view, analyze and prioritize incidents, and act accordingly. Finally, the web backend service allows the system to function, communicating the information within the mobile and the desktop applications.

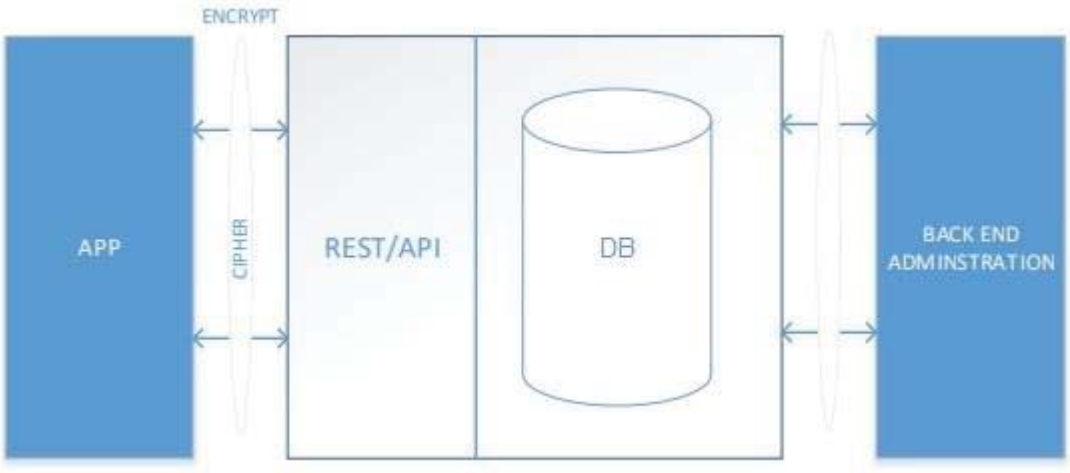

Figure 5.1. System´s basic functionality

The system follows an MVC design pattern (Leff and Rayfield, 2002). The 'Model' contains all the data definitions. Complaints' and users´, among other objects, structure and information are specified in the model, as well as, the complaints' status or the trace of the case. Database's definition and information is also contained in the model. All different screens and forms of the desktop application that the end users will face are programmed in the 'View' component. The 'Controller' contains all the responsible methods to capture the interface's inputs and pass them off to the 'Model'. When the desktop app is launched, it loads the login form view.

C# is the main programming language present in the PN-Complaint desktop app. Besides, it contains .NET modules and XML code to connect to the database.

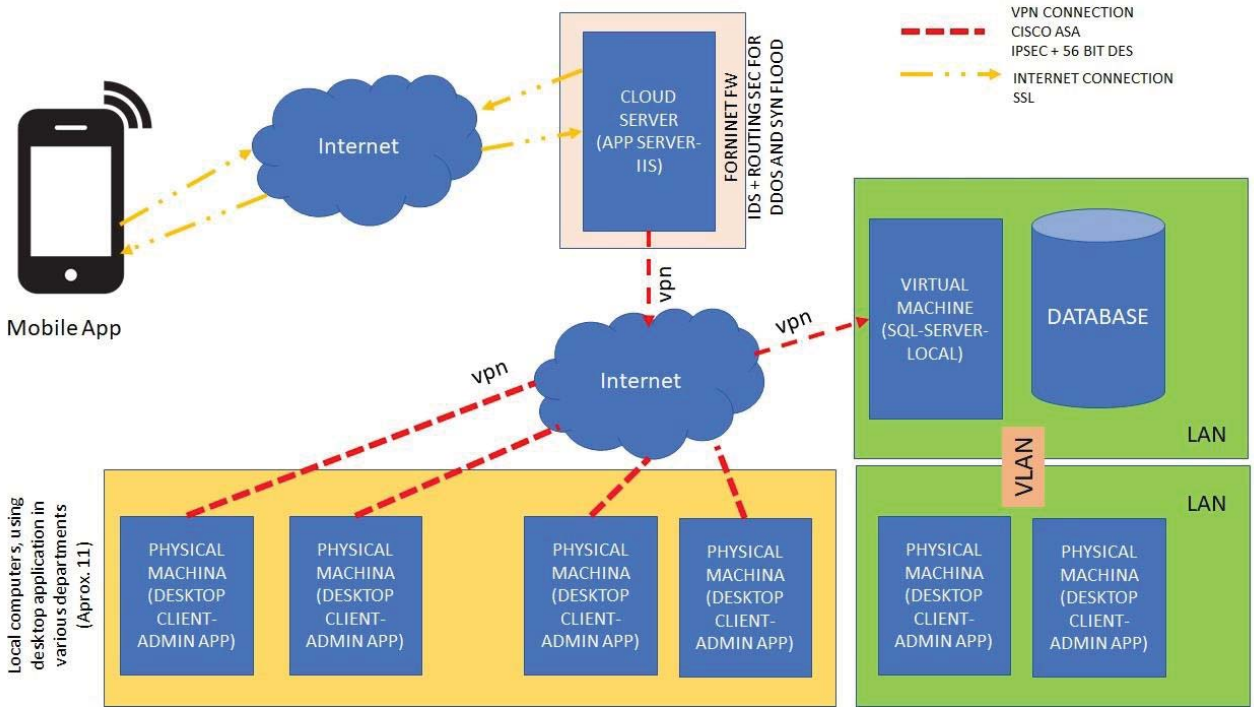

Figure 5.2. Detailed system architecture

5.3. Methods and Tools Used

As mentioned in the introduction, a static source code analysis has been performed for the purpose of this work. Thus, all weaknesses found in the PN-Complaint System were identified by automatically examining the code using open source tools before the program is run. These open source code analysis tools were selected from a list presented by the OWASP project ("OWASP Source Code Analysis Tools", 2019), according to what programming languages they were tailored.

5.3.1. Puma Scan

Puma Scan1 is an open source static source code analyzer that runs as an IDE plugin for Visual Studio and via MSBuild in CI pipelines that provides real-time continuous analysis for C# applications. Created by Puma Security, LLC, Puma Scan version 2.1.0.0 was installed as an extension for Microsoft Visual Studio 2019. While the integrated security rules search for vulnerabilities, it immediately displays the weaknesses present in the environment as spell checker and compiler warnings in the Visual Studio Error List window. Using Puma Scan, two main weaknesses are discovered:

- Unvalidated file paths are passed to a file write API, which can allow unauthorized file system operations (e.g. read, write, delete) to be performed on unintended server files.
- Unvalidated file paths are passed to a FileStream API, which can allow unauthorized file system operations (e.g. read, write, delete) to be performed on unintended server files.

Both vulnerabilities were found in two different points of the desktop application. These types of weaknesses in the code can be associated to the CWE-20: Improper Input Validation vulnerability ("CWE-20: Improper Input Validation", 2019). An attacker could be allowed to submit a malicious

path as input and cause the application to crash, expose sensitive data, modify data or possibly alter control flow in unexpected ways.

5.3.2. .NET Security Guard

.NET Security Guard2 is another extension for Microsoft Visual Studio that performs security audits in .NET applications in a background. It has two modes: for developers and for auditors. Based on a set of predefined signatures, it detects various security vulnerability patterns. Continuous Integration (CI) through MSBuild is also provided by .NET Security Guard. Version 3.2.0 of the Visual Studio 2019 extension was able to find the following vulnerability:

- Weak Hashing Function. The program uses the MD5 hashing algorithm. MD5 or SHA1 have collision weaknesses and are no longer considered secure hashing algorithms. Instead, SHA256 or SHA512 should be used.

5.3.3. Sonar Qube (SonarLint for Visual Studio)

SonarLint3 version 4.10.0.9867 IDE extension for Visual Studio 2019 was installed. It is a code analyzer that scans the source code for more than 20 languages for bugs, vulnerabilities and code smells, so they can be fixed before committing code. Instead of reporting vulnerabilities per se, the tool identifies poor code practices. For instance, the IDE extension recommends “Remove this hardcoded path-delimiter”, “Make this field ‘private’ and encapsulate it in a ‘public’ property” or “ Remove the unnecessary Boolean literals”.

5.3.4. VisualCodeGrepper

VisualCodeGrepper (VCG)4 is an automated code security review tool for C++, C#, VB, PHP, Java and PL/SQL which is intended to drastically speed up the code review process by identifying bad/insecure code. Current version at the time of this work is 2.1.0. The fact that VCG provides vulnerabilities’ categorization and allows .cvs exportation, makes this tool handier than the previously described IDE extensions for Visual Studio 2019.

After configuring the tool for C# and performing a full scan, VCG identified approximately 35 weaknesses in the PN-Complaint desktop application. All are categorized as ‘standard’ severity, except one, which is categorized as ‘medium’ severity:

- Potentially unsafe code – Insecure storage of sensitive information. Concretely, the code uses standard strings and byte arrays to store sensitive transient data such as passwords and cryptographic private keys, instead of the more secure SecureString class. This vulnerability could be matched to the CWE-922: Insecure Storage of Sensitive Information.

5.4.Results

Table 5.1. Desktop application source code analysis

Application Name	Complaint System - Desktop Application
Review Date	23/06/2019
Objective	Security Code Review
Lines of Code	13,684
Code Review Mode	Static

Once the source code had been evaluated using the previously mentioned tools, results with the identified vulnerabilities and poor coding practices were aggregated in a .csv file. Since the severity levels description reported by each tool differed from one to another, these categories were adjusted following a common pattern. This file was later analyzed on RapidMiner Studio to find out trends and prioritize vulnerabilities based on severity levels. Figure 5.3. shows the number of times a weakness was encountered in the source code. Carrying out integer operations without enabling overflow defenses is the most common security issue found in the code. This vulnerability could enable a malicious individual to conduct an integer overflow attack, leading to fatal software errors. For instance, a truncation error on a cast of a floating-point value to a 16-bit was the responsible of the destruction of Ariane 5 flight 501 in 1996 (Dietz et al., 2015).

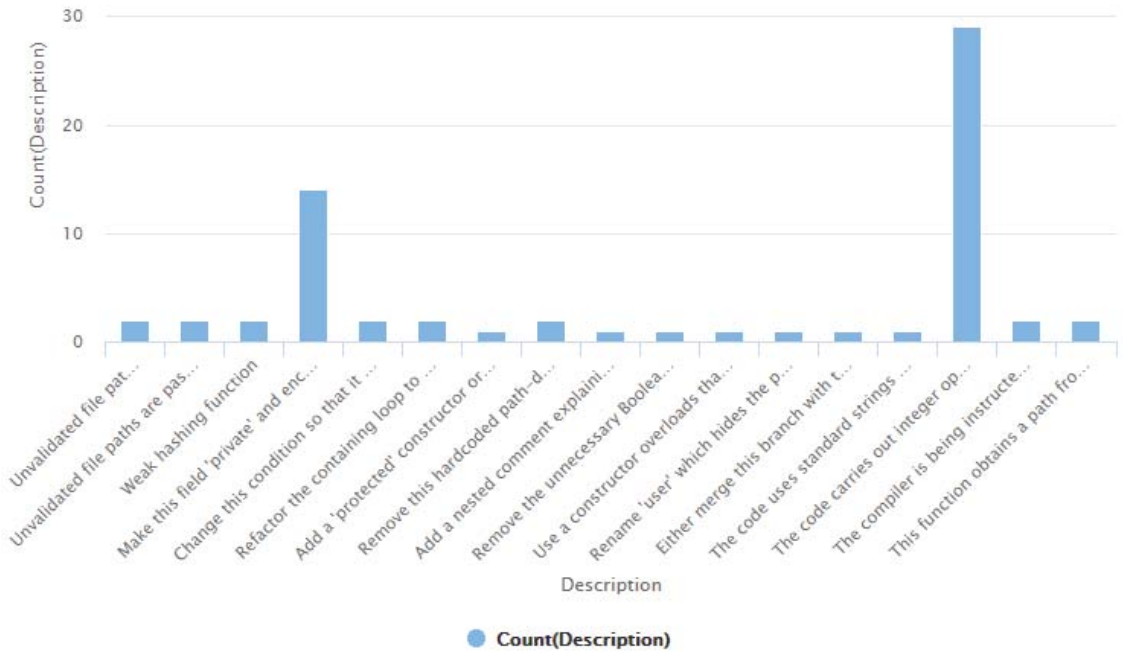

Figure 5.3. Count of vulnerabilities

Grouped by severity level, the following graph indicates how many vulnerabilities of each severity were found in the source code. Severity levels were defined as standard, medium and high. Out of the

66 identified vulnerabilities, 59 were classified as standard, 3 as medium and only 4 of them possessed a high criticality.

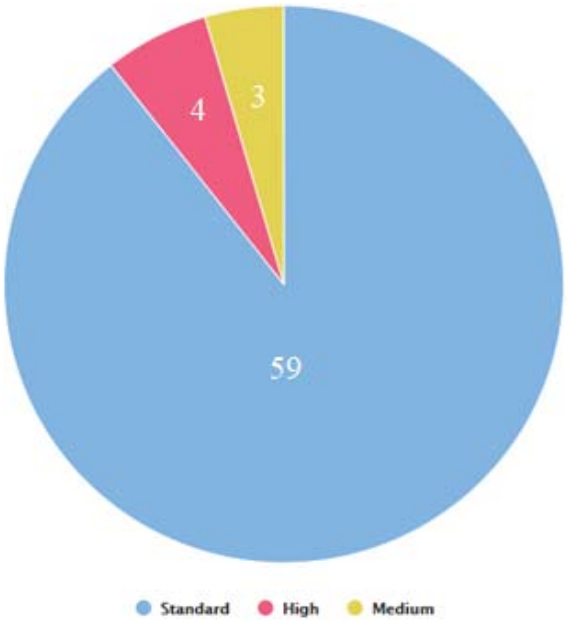

Figure 5.4. Vulnerabilities grouped by severity level

Table 5.2 highlights the vulnerabilities categorized as medium or high. Although the tools reported 7 weaknesses, it can be appreciated that, in fact, only 3 different types of high or medium severity vulnerabilities exist: unvalidated file paths passed to an API, use of weak hashing functions and use of unsecure arrays to store sensitive data.

Table 5.2. Medium and high vulnerabilities

High	Unvalidated file paths are passed to a file write API, which can allow unauthorized file system operations (e.g. read, write, delete) to be performed on unintended server files.
High	Unvalidated file paths are passed to a FileStream API, which can allow unauthorized file system operations (e.g. read, write, delete) to be performed on unintended server files.
High	Unvalidated file paths are passed to a file write API, which can allow unauthorized file system operations (e.g. read, write, delete) to be performed on unintended server files.
High	Unvalidated file paths are passed to a FileStream API, which can allow unauthorized file system operations (e.g. read, write, delete) to be performed on unintended server files.
Medium	Weak hashing function
Medium	Weak hashing function
Medium	The code uses standard strings and byte arrays to store sensitive transient data such as passwords and cryptographic private keys instead of the more secure SecureString class.

Unvalidated file paths passed to an API is the highest criticality vulnerability present in the source code. It was found to be present 4 times. These weaknesses can allow unauthorized file system operations, such as read, write or delete, to be performed on unintended server files. In other words, an attacker could pass a crafted input to the application to see sensitive information stored in the server, delete important information or configuration files, or write malware into the server. This type

of attack is commonly referred to as injection attack. Injection attacks often involve exploiting a weakness in the code to inject malicious code into an executing application and then cause the injected code to be executed (Hu et al., 2006). The simplest way to protect the application from injection attacks is by validating user input using some sort of filtering, preventing the application from accepting special characters or known malicious statements. Never trust user input is a widely recognized security practice. Nonetheless, more complex solutions are developed, like the one presented by Hu et al. based on instruction-set randomization.

The National Crime Reporting System desktop application uses the MD5 hashing function. MD5, so as SHA-1 have known collision weaknesses and are no longer considered strong hashing algorithms. Thus, it is recommended to use more secure algorithms instead, such as SHA-256 or SHA-512. Note that MD5 produces a 128-bit hash value, whereas SHA-512 produces a hash output of 256 bits. This means that finding a collision for a SHA-512 hashed value is twice as difficult than for MD5. Moreover, no collisions have yet been produced for SHA-256. This vulnerability could be classified as CWE-327: Use of a Broken or Risky Cryptographic Algorithm ("CWE-327: Use of a Broken or Risky Cryptographic Algorithm", 2019), and "may result in the exposure of sensitive information".

Finally, the last medium severity weakness reported by our static source code analysis refers to the use of unsecure programming methods. Concretely, the code uses standard string and byte arrays to store sensitive transient data, such as passwords and cryptographic keys. This is inappropriate because those methods store the data in plain text, leaving the data open for attack (Asad and Ali, 2017).Furthermore, "String class is also immutable, which leaves copies in memory on every change which could be compromised as it is impossible for a garbage collector to clear all the copies of data". Although it may look that the information is not readable by humans, it is encoded in a certain way that that some techniques can determine the encoding system in use and then, decode the information (Andreeva et al., 2016). Hence, the application should use more secure methods to protect sensitive information. Found in the System.Security namespace, C# provides a SecureString class that automatically encrypts the string and stores it in a special memory location (Asad and Ali, 2017). Asad and Ali provide an example of how to use this class. This vulnerability could be matched with CWE-312: Cleartext Storage of Sensitive Information ("CWE-312: Cleartext Storage of Sensitive Information", 2019).

Despite the great majority of the weaknesses identified in the analysis are purely poor coding practices, there are important vulnerabilities present in the code. Some of them could lead to sensitive information exposure, one of the greatest risks of the application being assessed. Following the National Institute of Standards and Technology (NIST) guidance, the system should employ a type of cryptography to support the protection of the sensitive information being transmitted ("NIST Special Publication 800-53 Rev. 4", 2015). Moreover, it is highly recommended that the application validates the users´ inputs when file paths are passed to APIs to ensure that they are not malicious and thus, prevent injection attacks.

5.5.Conclusions and Future Work

We performed a static source code analysis to assess a National Crime Reporting System desktop application. for a Latin American country's National Police. After determining that the programming languages in use were mainly C# and .NET, open-source code analysis tools for these specific languages were used to identify weaknesses or vulnerabilities in the code. Out of the 66 identified weaknesses, 59 were classified as standard, 3 as medium and only 4 of them possessed a high criticality. Most of the reported standard weaknesses referred to poor development practices.

However, the existing vulnerabilities were of high and medium severity, and constitute a great risk to the system. Among other consequences, these vulnerabilities could lead the application to expose sensitive information or allow remote code execution (Zheng and Zhang, 2013), one of the most noticeable security threats for web applications.

Each tool discovered different vulnerabilities and none of the high or medium ones were reported by more than one tool. Therefore, we could not rely on any tool by its own to assess the security of the system. Besides, one of the major limitations of the project was the budget. In the future, having more budget available to buy licenses for commercial code analysis tools, a deeper and more reliable static source code analysis of the application could be performed.

6. FRAMEWORK FOR THE DEVELOPMENT OF VIRTUAL LABS FOR INDUSTRIAL INTERNET OF THINGS AND HYPERCONNECTED SYSTEMS

6.1. Introduction

With the advancement of technologies and the Internet, education trends have also evolved. Nowadays, online learning has become a popular solution for countless students in every education field. Even back in 2009, over 5.6 million students in the United States were taking at least one online course during the fall 2009 term; an increase of nearly one million students over the number reported the previous year (Allen and Seaman, 2010). Nevertheless, these practices become more challenging in engineering, where traditional hands-on laboratories need to be adopted to online education with guarantees.

A virtual laboratory is the representation of a place equipped with the necessary means to carry out research, experiments and works of a scientific or technical nature, produced by a computerized system, which gives the sensation of its real existence. In the academic environment, arises from the need to create student support systems for their laboratory practices. According to (Maurel and Soria, 2014), these settings serves the purpose of optimizing the time to developer such practices and minimize the demand for infrastructure resources. Among the advantages provided by a virtual laboratory are:

- Useful explanations of the theoretical concepts.
- Conducting experiments step by step, avoiding the problem of overlapping with the schedules of other educational experiences.
- It is flexible and with easy-to-use tools and minimizing risks.
- It is a low cost alternative and ability to be scalable for growth.
- It allows a greater number of students to experiment with a laboratory asynchronously, regardless of whether they coincide in space.

There are two online lab environment approaches, virtual and remote labs (Cheng et al., 2010). Virtual labs are based on software that simulates a real environment in a safe setting, allowing students to set up different parameters in their experiments and learn from failure without causing real damage. LabVIEW, Flash, Java Applet or Matlab/Simulink are examples of virtual labs.

On the other hand, remote lab is, by definition, “an experiment which is conducted and controlled remotely through the Internet”. In these labs, real components or programs are normally running at a different location from where they are being controlled. However, remote labs often have to deal with memory and resources problems when used by large groups of students. For instance, the Illinois Institute of Technology offers a Remotely-Accessible Dynamic Infrastructure for Students to Hack (RADISH) to the School of Applied Technology students (Broda et al., 2014).

Simulated environments are of key importance to satisfy the ongoing demand of new online learning trends. Not only remote desktop solutions are enough. The emerging information and communication technologies that embed a huge number of tiny computers, smartphones, contactless smart cards, Radio Frequency Identification (RFID), etc. (Sakamura and Koshikuza, 2005), defined as ubiquitous computing or u-computing, play a significant role.

Derived from u-computing, u-learning, or ubiquitous learning, represents a new learning paradigm. Moving from traditional learning to electronic learning (e-learning), and from electronic learning to mobile learning (m-learning), we are now shifting to u-learning (Yahya et al., 2010). Ubiquitous learning could be defined as the use of u-computing technologies for education purposes, allowing anyone to learn at any place at any time. Yahya et al. propose five characteristics of u-learning: permanency, accessibility, immediacy, interactivity and context-awareness.

Since 1980, computer labs have been used to increase students' ability in science and engineering fields. A great example is the Computer as Learning Partner (CLP) integrated instructional unit developed by a group of technologists, classroom teachers and education researchers ("America's Lab Report: Investigations in High School Science", 2006).

These laboratory experience solutions, in addition to theoretical lectures, have proved to enhance the mastery of science subject matter, increase understanding of the complexity and ambiguity of the empirical work, develop practical skills and cultivate interest in science among students over history. In this paper, concepts and requirements for a virtual Internet of Things (IoT) and information systems hyperconnectivity lab will be presented. The framework would take advantage of the new u-learning technologies, industrial IoT and embedded systems, allowing students to fully experience traditional hands-on laboratories using any device at anytime from anywhere.

6.2. Purpose

The world is preparing for new global technology shifts that will dramatically alter the landscape of business operations with autonomous driving, Artificial Intelligence (AI), Industry 4.0, smart living, and more. In its simplest form, IoT is a complex interconnection of hardware, such as sensors and actuators, and software, running as a middleware layer (Taveras, 2018). Therefore, the creation of a safe environment for student learning to understand the presented challenges is important for being able to safeguard critical systems. This remote lab is to be accessible from any mobile device that allows for the ability to access a remote desktop environment.

6.3. Framework

The Mission Framework serves this purpose as it is holistic with a focus on policy, education, and technology that can be tailored to any environment (Dawson, 2017).

This framework ties in three themes. The first of these themes is education as it deals with establishing cybersecurity education programs. The next theme is the role of policy in cybersecurity, and the final theme covers the role of technology in cybersecurity. The Mission Framework is shown in Fig. 6.1, which was used to develop the learning environment. Each policy required a different implementation of a particular technology. Using this process allows for educational components that can be designed that reflect current policies. Thus, when an instructor needs to customize the virtual lab to represent a specific country, then they can use this framework as a guide to ensure all components are considered to deploy a virtual lab.

Virtual labs allow for learning and experimentation that would normally require the use of expensive equipment. Virtual labs allow students to operate in a safe, secure, and sandbox. Students have the ability to practice system testing several times to learn from successes and failures without inflicting actual damage. The virtual lab uses a significant amount of Open Source Software (OSS) to alleviate costs and allow students to grasp concepts such as review engineering the source code to harden or break.

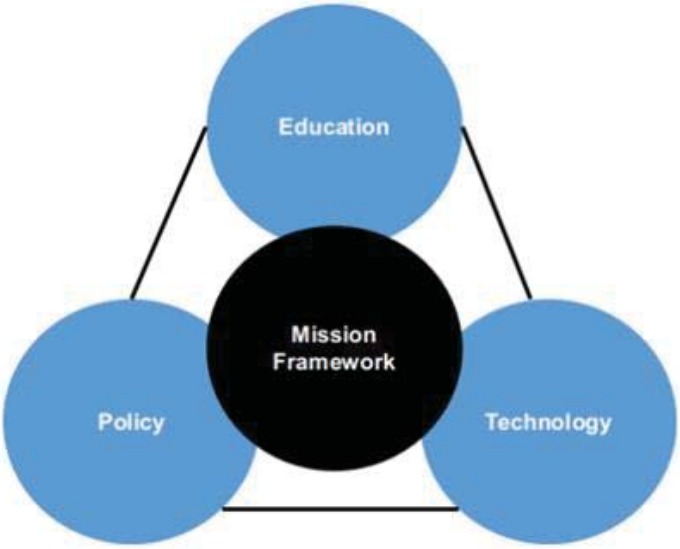

Figure 6.1. Mission framework (Taveras, 2018)

6.4. Web-based Experiment Framework

The system diagram of the School of Applied Technology Cybersecurity Lab (SAT-CL) is shown in Figure 6.2. Students can access the virtual lab from several computing devices using a remote desktop client. Once students have been authenticated, they are provided a Kali Linux Virtual Machine (VM) from where they can gain access to many systems beyond a router/Domain Name System (DNS) server. The system is scalable, in which it allows for additional components such as servers or images of IoT devices to be added or removed.

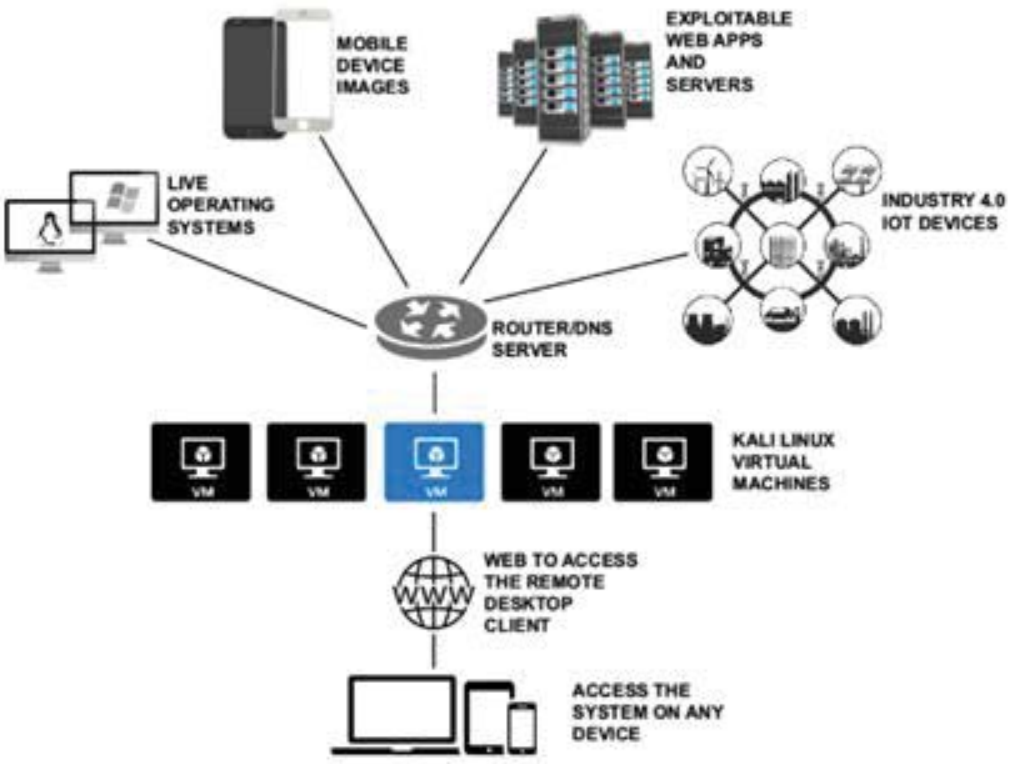

Figure 6.2. School of Applied Technology cybersecurity lab

Virtual labs evolve past the need of non-existent hardware dependencies. The use of server virtualization with other key tool, provide a set of experimentation services. The following components are present in the described concept:

- Service virtualization layer. Simulates all of the dependencies needed to run virtual instances of IoT end-points.
- Service and API layer. Provides a channel to drive the virtual IoT node and simulates the provision of services and APIs.
- Runtime monitoring. Used to trace the behavior of the simulated set ups and captures important analytics about the experiments and virtual lab.
- Core software layer. Include the operating systems and tools to provide realistic calls from the virtual nodes over the network (whether they are protocols like REST/HTTP, or IoT popular protocols like CoAP, XMPP, or MQTT).

In Fig. 6.3, the current implementation is shown that runs on any desktop or laptop. The issue with this former implementation is that it depends solely on the end user's computing environment, such as the Central Processing Unit (CPU), memory, and hard disk space. However, the applications such as the Kali VM and vulnerable web server in Fig. 6.3 are all found in Fig. 6.2, which is an expansion of the virtual environment with an ability to manage the complexity of the network. This provides flexibility in what concepts can be taught according to specific laws and guidelines for a particular mission.

Figure 6.3. Oracle Virtual Box networked environment

6.5.Issues and Future Direction

There are indeed projects that have saved money by adopting desktop virtualization, and most do not see any ROI for at least a few years. To justify the use of this type of technology for the proposed objectives, the cost should not be a central argument. This is because initially, licensing costs and initial infrastructure far exceed the initial benefits. VDI based projects generally require investment in thin clients, and improvements to storage and network infrastructure, which can make it an expensive project in the front end. To simulate basic operations such as the use of office automation

applications, 200 kilobits of network bandwidth is probably sufficient. However, for operations that need video streaming and 3D graphics rendering, network requirements can scale to hundreds of megabits. It is essential to keep in mind that latency is the enemy of desktop virtualization; therefore, special attention must be placed on the configuration of the network core to optimize the links. There are a few challenges in this implementation, such as finding a fully open source virtual desktop infrastructure. Other challenges are the Graphical User Interface (GUI) remote desktop client. Further issues are on hosting mobile device images and ensuring the ability to have them operate as if hosted on their native platform. The goal is to have a fully an open source environment so that developing countries are not burdened with the associated cost of software licenses.

In addition, according to (Taveras, 2018), an IoT based scenario must comprise a plethora of different protocols, platforms, and vertical solutions that are in the process of refining its various proposals to reproduce a coherent composition, capable of delivering open solutions that are attractive to businesses and consumers. At the present time, each middleware solution focuses on different IoT challenges, such as data management, interoperability, security, and many more. A middleware solution that addresses all the challenges posed by IoT is yet to be designed. Experimenting and testing the compliance of these protocols and testing backward/ future compatibility creates a major for industry, researchers and learners. Today there are hundreds of cloud platform providers, each using their own implementation or a particular version of the open source component ("Why IoT needs simulation of load testing", 2018).

7. CONCLUSIONS

Information warfare techniques have been employed for centuries. Nonetheless, the term has recently gained popularity and modern information warfare techniques have widespread due to the capabilities brought by the Internet. After realizing that all the data generated about individuals in the web had a great value, organizations, corporations and governments have invested significant amounts of money in intelligently using this information to obtain a better position over their competitors. Collecting PII is that easy with the current available tools that even children can play a crucial role in this information warfare. We have explained different means that make personal data fairly accessible for anyone.

While paramount data breaches have surfaced the Internet, governments and organizations are developing privacy laws to stop this data bleeding. The GDPR in Europe represents a worldwide known privacy regulation. Concretely in the United States, national level initiatives, like the California Consumer Privacy Act or the Chicago Data Collection and Protection Ordinance, as well as national ones, were presented in this project.

One of the considerations brought up by these regulations is developing secure applications by coding following best security practices. This must be done since the early stages of the application development. In this project, we have assessed the security of the desktop application of a National Crime Reporting System for a Latin American country's National Police. The reported vulnerabilities, if exploited, could lead to PII being employed in information warfare or exposed, constituting another data breach. Therefore, it is of key importance to secure applications.

Finally, we have proposed a virtual lab framework whose purpose is to educate users in data protection issues. It is holistic with a focus on policy, education and technology that can be tailored for any environment. Thus, when an instructor needs to customize a virtual lab to teach specific information security requirements for a determined country or scenario, he or she can use this framework as a guide to ensure that all needed components in scope are considered to deploy the virtual lab.

8. FUTURE WORK

Significant work needs to be done to improve privacy and prevent data breaches from appearing on everyday news. In the United States, a national privacy law is required to unify all the different state initiatives and proved a common guidance framework for corporations and individuals. Besides, technical frameworks and standards should also be developed to help developers address security concerns since the early stages of coding. This would greatly reduce the number of PII exploits produced due to a vulnerability in the application.

In terms of virtual labs, there are still challenges that need to be addressed, like finding a fully open source virtual desktop infrastructure or lighter GUIs that mitigate latency problems. In addition, at the present time, each middleware solution focuses on different IoT challenges, such as data management, interoperability, security, and many more. A middleware solution that addresses all the challenges posed by IoT is yet to be designed.

9. REFERENCES

1. Ahmed, S. R. (2004, April). *Applications of data mining in retail business. In International Conference on Information Technology*: Coding and Computing, 2004. Proceedings. ITCC 2004. (Vol. 2, pp. 455-459). IEEE.
2. Aldrich, R. W. (1996). *The international legal implications of information warfare* (No. INSS-OP-9). AIR FORCE ACADEMY COLORADO SPRINGS CO.
3. Allen, I.E., J. Seaman. (2010). *Class Differences: Online Education in the United States*. Retrieved from https://files.eric.ed.gov/fulltext/ED529952.pdf.
4. America's Lab Report: Investigations in High School Science (2006). *Chapter 3: Laboratory Experiences and student Learning* . Retrieved from https://www.nap.edu/read/11311/chapter/5.
5. Andreeva, O. et al. (2016). *Industrial Control Systems Vulnerabilities Statistics.* Retrieved from https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2016/07/07190426/KL_REPORT_ICS_Statistic_vulnerabilities.pdf
6. Asad, A., Ali, H. (2017). *Working with Cryptography.* Retrieved from https://link.springer.com/chapter/10.1007/978-1-4842-2860-9_13
7. Aura, T., Kuhn, T. A., & Roe, M. (2006, October). *Scanning electronic documents for personally identifiable information.* In Proceedings of the 5th ACM workshop on Privacy in electronic society (pp. 41-50). ACM.
8. Bergeron, J. et al. (2001). *Static Detection of Malicious Code in Executable Programs*. Retrieved from https://nnt.es/Static%20Detection%20of%20Malicious%20Code%20in%20Executable%20Programs.pdf
9. Bieker, F. (2017). Enforcing Data Protection Law – *The Role of the Supervisory Authorities in Theory and Practice*. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-55783-0_10.
10. Boyd, D. M., & Ellison, N. B. (2007). *Social network sites: Definition, history, and scholarship*. Journal of computer-mediated Communication, 13(1), 210-230.
11. Broda, D., Khodja, B, Lidisky, B. (2014). *Remotely-Accessible Dynamic Infrastructure for Students to Hack (RADISH).* https://appliedtech.iit.edu/sites/sat/files/pdfs/ITM/Remotely-AccessibleDynamicInfrastructureforStudentstoHackRADISH.pdf.
12. California Consumer Privacy Act of 2018 (2018).
13. Chen, D., & Zhao, H. (2012, March). *Data security and privacy protection issues in cloud computing*. In 2012 International Conference on Computer Science and Electronics Engineering (Vol. 1, pp. 647-651). IEEE.
14. Cheng, X., Song, G., Zhang, Y. (2010*). Virtual and Remote Laboratory Development: A review*. https://www.researchgate.net/profile/Xuemin_Chen4/publication/228988059_Virtual_and_Remote_Laboratory_Development_A_Review/links/55b773fe08aed621de046178/Virtual-and-Remote-Laboratory-Development-A-Review.pdf.
15. Chicago Personal Data Collection and Protection Ordinance (2018).
16. Clifton, C., & Marks, D. (1996, May). *Security and privacy implications of data mining*. In ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (pp. 15-19).
17. Common Weakness Enumeration. *CWE-20: Improper Input Validation.* Retrieved June 23, 2019 from https://cwe.mitre.org/data/definitions/20.html
18. Common Weakness Enumeration. *CWE-312: Cleartext Storage of Sensitive Information.* Retrieved July 13, 2019 from https://cwe.mitre.org/data/definitions/312.html

19. Common Weakness Enumeration. *CWE-327: Use of a Broken or Risky Cryptographic Algorithm.* Retrieved July 7, 2019 from https://cwe.mitre.org/data/definitions/327.html
20. Cornock, M. (2018). *General Data Protection Regulation (GDPR) and Implications for Research*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378512218300367.
21. Dawson, M. (2017). *Hyper-connectivity: Intricacies of national and international cyber securities*.
22. Dietz, W. et al. (2015). *Understanding Integer Overflow in C/C++*. Retrieved from https://dl.acm.org/citation.cfm?id=2743019
23. Dode, A. (2018). *The challenges of implementing General Data Protection Law (GDPR)*. Retrieved from http://www.academia.edu/37461999/The_challenges_of_implementing_General_Data_Protection_Law_GDPR.
24. Elbirt, A. J. (2003). *Information Warfare: Are You At Risk?* Retrieved from https://pdfs.semanticscholar.org/3efc/0039ba4cd44b96c63e91c00d3d4320101fb0.pdf
25. Elbirt, A. J. (2003). *Information Warfare: Are you at risk?*. IEEE Technology and Society Magazine, 22(4), 13-19.
26. Elmasri, R. (2008). *Fundamentals of database systems*. Pearson Education India.
27. Fussell, S. (2018, March 01). *Facebook's New Face Recognition Features: What We Do (and Don't) Know [Updated]*. Retrieved March 18, 2019, from https://gizmodo.com/facebooks-new-face-recognition-features-what-we-do-an-1823359911
28. Gajanayake, R., Iannella, R., & Sahama, T. (2011). *Sharing with care: An information accountability perspective*. IEEE Internet Computing, 15(4), 31-38.
29. Gantz, J., & Reinsel, D. (2011). *Extracting value from chaos.* IDC iview, 1142(2011), 1-12.
30. Gundecha, P., Barbier, G., & Liu, H. (2011, August). *Exploiting vulnerability to secure user privacy on a social networking site*. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 511-519). ACM.
31. Headquarters Department of the Army (1979, August 31). *Psychological Operations*.
32. Hu, W. et al. (2006). *Secure and Practical Defense against Code-Injection Attacks Using Software Dynamic Translation.* Retrieved from https://dl.acm.org/citation.cfm?id=1134764
33. IoTfy (2018). *Why IoT needs simulation of load testing*. Retrieved from https://iotify.io/why-iot-needs-simulation-instead-of-load-testing/.
34. Kaushal, R., Saha, S., Bajaj, P., & Kumaraguru, P. (2016, December). *KidsTube: Detection, characterization and analysis of child unsafe content & promoters on YouTube.* In 2016 14th Annual Conference on Privacy, Security and Trust (PST) (pp. 157-164). IEEE
35. Krishnamurthy, B., & Wills, C. E. (2009, August). *On the leakage of personally identifiable information via online social networks.* In Proceedings of the 2nd ACM workshop on Online social networks (pp. 7-12). ACM.
36. Leff, A. and Rayfield, J. T. (2002). *Web Application Development Using the Model/View/Controller Design Pattern*. Retrieved from https://ieeexplore.ieee.org/abstract/document/950428
37. Lin, Z., Tong, L., Zhijie, M., & Zhen, L. (2017). *Research on Cyber Crime Threats and Countermeasures about Tor Anonymous Network Based on Meek Confusion Plug-in*. 2017 International Conference on Robots & Intelligent System (ICRIS). doi:10.1109/icris.2017.69
38. Louridas, P. (2006). *Static Code Analysis*. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1657940

39. Mathews, K. J., Bowman, C. M. (2018). *The California Consumer Privacy Act of 2018*. Retrieved from https://privacylaw.proskauer.com/2018/07/articles/data-privacy-laws/the-california-consumer-privacy-act-of-2018/

40. Maurel, M., Soria, N. (2014). *The virtual laboratory: a tool to face the shelling. Ibero-American Congress on Science, Technology, Innovation and Education.* Retrieved from https://www.oei.es/historico/congreso2014/memoriactei/677.pdf.

41. NIST Special Publication 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) (2010). Retrieved from https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf

42. NIST Special Publication 800-53: Security Controls and Assessment Procedures for Federal Information Systems and Organizations, Rev. 4 (2015). Retrieved July 2, 2019 from https://nvd.nist.gov/800-53/Rev4/control/SC-13

43. O'Gorman, G., & McDonald, G. (2012, Nov 8). *Ransomware: A Growing Menace*. Retrieved from http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/ransomware-a-growing-menace.pdf

44. OWASP Source Code Analysis Tools. Retrieved July 13, 2019 from https://www.owasp.org/index.php/Source_Code_Analysis_Tools

45. Petukhov, A., Kozlov, D. (2008). *Detecting Security Vulnerabilities in Web Applications Using Dynamic Analysis with Penetration Testing*. Retrieved from https://pdfs.semanticscholar.org/9d33/19d49a52395e37bc6ba29c1e3282c0f0a06a.pdf

46. Regulation (EU) 2016/679 of the European Parliament and of the Council (2016).

47. Rosenberg, M. (2018) *Cross-Border Transfers of Personal Data in Light of GDPR*. Retrieved from https://dataprivacy.foxrothschild.com/2018/03/articles/european-union/gdpr/cross-border-transfers-of-personal-data-in-light-of-gdpr/

48. Sakamura, K., Koshikuza, N. (2005). *Ubiquitous Computing Technologies for Ubiquitous Learning.* Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1579224.

49. Schawartau, W. (1997, October 16). *What Exactly is Information Warfare? - Part 2*.

50. Senator Wyden (2018). *The Consumer Data Protection Act of 2018 Discussion Draft*. Retrieved from https://www.wyden.senate.gov/imo/media/doc/Wyden%20Privacy%20Bill%20one%20pager%20Nov%201.pdf

51. Sinisi, J. P. (2007). U.S. Patent No. 7,313,759. Washington, DC: U.S. Patent and Trademark Office.

52. Soares, R. R., Naritomi, M. (2010). *Understanding High Crime Rates in Latin America: The Role of Social and Policy Factors.* Retrieved from https://www.nber.org/chapters/c11831.pdf

53. Soma, J. T., Courson, J. Z., & Cadkin, J. (2009*). Corporate Privacy Trend: The "Value" of Personally Identifiable Information ("PII") Equals the "Value" of Financial Assets*. Richmond Journal of Law & Technology, 15(4), 11.

54. Stalder, F., & Hirsh, J. (2002). *Open source intelligence*. First Monday, 7(6), 1–8

55. Sutton, H. et al. (2017). *Restoring Paradise in the Caribbean: Combatting Violence with Numbers.* Retrieved from https://publications.iadb.org/en/restoring-paradise-caribbean-combatting-violence-numbers

56. Symantec Corporation. (n.d.). *What Are Cookies?* Retrieved July 7, 2019, from https://us.norton.com/internetsecurity-how-to-what-are-cookies.html

57. Tauber, A. (n.d.). *Welcome to Searx*. Retrieved December 3, 2018, from https://asciimoo.github.io/searx/
58. Taveras, P. (2018). *A systematic exploration on challenges and limitations in middleware programming for IoT technology.* International Journal of Hyperconnectivity and IoT, vol. 2, no. 2, Dec-2018.
59. Thunder Experience Cloud. *GDPR vs California Consumer Privacy Act (CaCPA) Detailed Comparison*. Retrieved October 28, 2018 from https://www.makethunder.com/gdpr-vs-california-consumer-privacy-act-cacpa-detailed-comparison/
60. Tor Project. (n.d.). *What is Tor Browser?* Retrieved December 3, 2018, from https://www.torproject.org/projects/torbrowser.html.en
61. Voss, W.G. (2017). *European Union Data Privacy Law Reform: General Data Protection Regulation, Privacy Shield, and the Right to Delisting.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2894571
62. Wang, P., Dawson, M., & Williams, K. L. (2018). *Improving Cyber Defense Education through National Standard Alignment: Case Studies*. International Journal of Hyperconnectivity and the Internet of Things (IJHIoT), 2(1), 12-28.
63. Weiss, M.A., Archick, K. (2016). *U.S.-EU Data Privacy: From Safe Harbor to Privacy Shield*. Retrieved from https://epic.org/crs/R44257.pdf
64. Wichers, D. (2013). *Owasp top-10 2013*. OWASP Foundation, February.
65. Yahya, S., Ahmad, E. A., Adb Jalil, K. (2010). *The definition and characteristics of ubiquitous learning: A discussion*. Retrieved from https://www.learntechlib.org/p/188069/.
66. Zheng, Y., Zhang, X. (2013). *Path Sensitive Static Analysis of Web Applications for Remote Code Execution Vulnerability Detection*. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6606611