



## Security and Privacy Concerns in Big Data

INTIKHAB AHMAD

School of Journalism and Communication,

Wuhan University, Wuhan, PR China;

TITUS HAIDUWA

International Software Institute, Wuhan University,

Wuhan, PR China;

ATTAULLAH KIANI

KAMRAN SADDIQUE

School of Journalism and Communication,

Wuhan University, Wuhan, PR China

### Abstract:

*Almost every organization anywhere in the world has some sort of data of all shapes and sizes. Due to the rising importance and benefits of the so called Big Data, many corporations and government agencies have started paying more attention to such delicate data. Big Data helps improving business and government operational efficiency, save on their expenditures, lead to grow their revenues, give them closer insights and unfold hidden patterns while empowering new business models. Those who recognize its importance and opportunities will bypass those who ignore it very soon in future in terms of competition advantage and service deliveries. Knowing that information is very expensive, it is also undeniable that it is the biggest weapon that one can use to destroy you or your business. Data is never get old and it can be used for various purposes, one of them being interpreted to give information. Information is a result of analyzed and interpreted*

*data to give meaningful information that lead to knowledge and wisdom. As we have approached digital age, where everything is done online there are now more questions regarding security and privacy in the IT the industry. People want data protection and should only be used for the intended purpose only. Any other use without our consent will always violate our right to privacy. For this reason, more than 75 countries now have data privacy laws. This requires that your organization be compliant to certain laws before you deeply get involved into big data collection. Some of the set national standards for healthcare privacy is the “Health Insurance Portability and Accountability Act” (HIPAA) in U.S and the 7 “safe harbor privacy principles” set for E.U. citizens’ personal data protection. This definitely spell out that big data require adhering to governance and privacy standards depending where you are.*

*In this paper we have discussed some few privacy and security issues related to “Big Data”. It is a must read paper especially by IT professionals, company executives, departmental and project managers who are interested in understanding the risks and challenges for working with and maintaining data security and privacy in Big Data World.*

**Key words:** big data, security concerns, privacy concerns

## INTRODUCTION

Our world has changed. What we thought was impossible in the late 1960s is now possible and all that was offline is now online. Enabled by Internet, which started in mid-20th century with 4 computers distributed between different universities and 2 years later increased to 40, it even got boosted by the creation of TCP / IP by researchers at a later stage (which became the standard for communications within computer networks today).

The first network was developed in 1969, called ARPANET (Advanced Research Projects Agency Network) by then as a network of geographically distributed computers designed to protect the flow of information between military installations. ARPANET continued to grow and went open to the world and today Internet is a network of networks and a home to millions of private, public, government, corporates and academic networks, all linked together carrying vast amounts of digital information for easy sharing and communication.

As far as the internet became an essential part of our life in a sense that we heavily rely on it and hardly do anything without it, it came with a price and this price is addiction and privacy challenges. Today, we use internet to connect to the other parts of the world through various online activities such as sending business emails, share views and share files on Facebook, Twitter, LinkedIn, Skype, Weibo, and other social platforms and blog posts sites, gamble and do online auctions and do order material and shop online such as from Amazon.com, e-Bay.com, Taobao.com, JD.com etc., read news online, watch our favorite shows online, do banking online, apply job online, Interview and Writes Exams online, find our date online, do conference online and almost everything is interlinked.

Therefore, resulted in the world creating around 2.5 quintillion bytes of data today in which 90% of it has been just generated in the last 2 years alone according to IBM Analytics Site. "This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data." (IBM, 2016)

We are continuously generating a huge volume of data every day that significantly increases from megabytes to gigabytes, gigabytes to terabytes and gradually to petabytes of data. According to statistical reports, at least we make 350,000 tweets, 300 video footage are uploaded to YouTube, 171

million emails are send and 330 GBs of sensordata are gathered from a single jet engine in every in 60s [2].

Big Data is not only limited to consumer facts from merchants or information and facts posted on social networking sites or enterprise data of manufacturing and sales though. Data can also be generated anywhere on our local machines, personal devices and home appliances. Such digital footprint is so sophisticated, huge in volume, large in velocity, coming from variety of sources and holds various values therefore making such data hard to store, share, analyze or process in real time and at appropriate speed using available tools (both Hardware and Software). This technology challenge in computer system is now what many researchers termed as “Big Data”.

## WHAT EXACTLY IS BIG DATA?

Let us face it, at the time of writing this article; the universal definition of what is “Big Data” remain unambiguous. The term seems rather vague as even the term itself doesn’t directly implies to the originally meanings of its two terms “Big” and “Data” that made up its phrase: Big Data. For someone who judges the book by its cover, may definitely refer big data to something that is so large in size and full of data/information. In fact, that person is not completely wrong. However, Big data is mostly “described as extremely large data sets that have grown beyond the ability to manage and analyze them with traditional data processing tools” [10]. Its commonly repeated definition cites its four characteristics (4Vs): volume, velocity, variety and value but many academic researchers and industry experts constantly argued that, Big Data cannot narrowly be defined in terms of its size of data it holds but rather be defined in terms of technology (tools) required to process such data at a right speed and in real time or the technical insights that can be drawn from such datasets.

Yes, it is correct that Big Data's current situation is that "datasets have grown to such enormous sizes that conventional information technologies can no longer effectively handle, either the size of the dataset or the scale and growth of the dataset" [10], but size is not the only concern. In other words, the dataset has grown so large that it is very difficult to manage and even harder to garner value out of it. Therefore, such primary difficulties in big data acquisition, storage, searching, sharing, analytics, and visualization needs advanced technologies to management them.

Actually, Big Data is a coined term of which no industry or academic expert is directly linked to its origin but has been frequently be used by some well-known researchers and writers until the notion of "big data" gained popularity in recent years and it attracted more researchers than other emerging technologies and research areas such as Internet of Things(IoT), Business Intelligence, Machine Learning, Computer Vision, Artificial Intelligence today. More credits are thus far given to scientists (David Ellsworth & Michael Cox) at NASA for having the first documented use of the term "Big Data" in their paper: "VIS '97 Proceedings of the 8th conference on Visualization", that was published by IEEE Computer Society Press Los Alamitos, USA in 1997.

McKinsey has also widely-quoted "Big Data" in his 2011 Big Data Studies. According to McKinsey Global Institute, "Big Data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze." While some sources say, Big Data is "a field dedicated to the analysis, processing, and storage of large collections of data that frequently originate from disparate sources" [2]. "It has the potential for dramatically changing the way organizations use information to enhance the customer experience and transform their business models" [2]. In recent efforts, Oxford English Dictionary has added the term big data

in 2013 and it also appeared in Merriam Webster's Collegiate Dictionary in 2014.

After extensive researches and in an effort to define big data, we can finally say Big data is a combination of both old and new data management technologies and techniques used by organizations to collect, store, process, manage, and manipulate a vast amounts of data at an accurate speed and in real time in order to determine hidden patterns and enable them to gain a far insight into something of their interest and make more informed decisions according to the outcome of analyzed data.

We agreed that Big Data shall not be defined in terms of its data size only but both, the size and the technology required and its capability to manage a huge volume of disparate data of various types and format and from different sources of information at a right speed and within the required time in order to produce meaningful information and improve the subject knowledge. Disparate data are any data that are not alike or are distinctly different in kind, quality or characters. They are normally unequal and cannot be readily used or integrated to meet the business information demand. Data management involves the capturing of data and its organization, integration, analyzation and taking of actionable insights [5]. The notion of Big Data is relatively new but its evolution is not something new. It has been there for decades when people started with paper records via census and any other recordings and it has only grown and attracted attentions all over the world in recent years due to massive data sets that are created daily as a result of new technology efforts and our addiction to such.

## **TYPES OF BIG DATA**

Well, the first thing first when dealing with big data is to know exactly what type of data you are dealing with and where did it came from. This requires you to verify the sources of

information cited and to do a thoroughly investigation of the original work in order to validate your data and prove your findings. Knowing the kind of data helps when you are to locate a specific type of data in your records in near future. As learned, Big Data can be generated from different sources and such sources are classified into 3 categories [2], mainly:

### **1. Structured data.**

Structured data is commonly generated by Research & Development indicatives, enterprise applications and information systems like ERP and CRM systems, EFM databases and 3rd party market databases). Structured data is often stored in tabular format and is basically used to capture relationships between different entities and is therefore most often stored in relational databases. Examples includes: banking transactions, customer invoices, and customer records [2].

### **2. Unstructured Data.**

Unstructured Data is data that does not conform to any data model or data schema. It makes up to 80% of the data generated within any given enterprise. They are normally in textual file (e.g. tweets, blog posts etc.) or binary (e.g.images/video/audio) format, converted to self-contained and non-relational files/data. [2]. It can be further classified as:

*Unstructured human generated data.* The data that humans in interaction with computers supplies such as gaming data, text messages, web content, customer feedback, social media likes/number of views, photos, video, audio etc.

Or

*Machine generated unstructured data.* The data created by a machine without human intervention such as radio frequency ID (RFID) tags, satellite images, surveillance photos and traffic videos, Radar or sonar data, weblogs, transactional

information, call records, Wi-Fi, smart meters, medical devices, and Global Positioning System (GPS)).

### **3. Semi-structured Data**

They are called semi, as they have a certain level of structure and consistency but not relational in nature. They are hierarchical or graph-based [2]. Examples: Json and XML file.

## **MAIN DRIVERS OF BIG DATA**

Big data growth is far driven by various factors:

Firstly, the invention of world wide web (www) in 1989 by the English Scientist Tim Berners-Lee. This single decision taken by Berners-Lee to undertake such project has brought many changes. It has turned that was thought impossible to be now possible. Everything we do today is either fully or partial done online. Can you write a 200-page paper today without googling anything on the internet or reading others conference papers and/or articles? Definitely NO, else you are not a resourceful writer. More resources are now freely available online and in millions of national databases which are made easily accessible by search engines such google, yahoo and baindu.

Secondly, advanced development in Information and Communications Technology (ICT) resources [2]. Communication (Mobile phones and Network providers), this invention by AT& T Lab Researchers Richard H. Frankiel and Joel S. Engel has enabled Big Data to grow consistently. Is undeniable that cellphones are the most contributors to this generation of huge volume of data. Everyone is addicted to using a phone for whatever purpose: calling, texting, sharing, reading, files, taking pictures, banking, shopping, searching/surfing internet, finding directions, recruiting, flight and hotel booking etc. with these inter-connected devices we collect and generate trillions of delicate Data.

According to statistics, major global player messaging apps used today are the major generators of a huge volume of desperate data. For Examples includes: The Chinese WeChat, which worth 111.5 Billion USD as of 2016 it has around 762 Million monthly active users. It is a free app, owned by Tencent as a parent company and it also own another office oriented app called QQi. WeChat is used for instant messaging, video and voice calls, social gaming, host of other apps, payments, sights and other media sharing as “moments” among others [15]. It makes money from selling games, stickers and handling mobile payments. This App is available both in English and Chinese Simplified and is mainly used in China Mainland.

From WeChat to WhatsApp. Worth around 19 billion USD, WhatsApp has become Facebook's largest acquisition to date after it was sold at \$22 billion late in 2014. Even though it was set charge 1 USD after a one year of use, it is still free and most messaging app globally with 1.65 billion active users. It also allows file transfer after few updates in 2015 as to compete with Chinese 's ever improving, WeChat. It generates no money at this point of writing this article.

From WhatsApp we have a Japanese App, “Line”. With 218 million monthly active users globally is used beyond messaging such as making payments, content, advertisements and helps businesses to build their own apps from within Line. It also makes money from selling stickers, social games and advertising. Is mostly known and used in Japan, Taiwan, Thailand and Indonesia [15].

We are having a lot of apps that generate data daily such as Facebook messenger, Snapchat, Viber, Imo, QQi, Twitter, LinkedIn, etc. and I have only highlighted the best 3 on a global scene that creates thousands of data daily.

Technological change such as affordable hardware infrastructures, cloud computing services, various social media platforms and mobile applications, and specialized IT infrastructures such as data warehouse, data marts, Extract

Transform and Loads, Enterprise Resource Planning, Databases and data frameworks such as MapReduce and the rise of Internet of Things(IoT) has also dramatically boosted the way we generate data today and is expected to double by 2020.

Thirdly, is globalization. Globalization is a process of international integrations through diplomatic relations. It is a practices of interchanging world views, ideas, products and other aspects of cultures. At first or let me say at the time of world war I and world war II, there was no a concept of “open arms” or brotherhood between various nations. And for those who did somehow became enemies later for political reasons and/or desires for world dominance. Today the world is like a one village. We can communicate daily to people close to us anywhere any time around the world.

Fourth, advanced researches. Improvement in data mining, Data Science and Data Analytics, has led to a common understanding of various data concepts, challenges and its impacts to our daily lives. The availability of cloud processing, analytics and storage services for example has undoubtedly opened doors for making Big Data driven analytics accessible to businesses of all sizes across many industries.

This has also significantly contributed to big data through developments of various techniques and methods on how to capture, store, analyze and process data to solve certain problems. such as the developments includes: “text mining software to help analyze non structured data types” (e.g. materials written in natural languages) [8]. “Data analytics has developed methods that allow data analysis to occur through the use of highly scalable distributed technologies and frameworks that are capable of analyzing large volumes of data from different sources” [2].Analytics helps making future predictions and forecast for common trends based on statistical and quantitative analyses [8]. Therefore, is very useful for business performances.

Firth, business digitization. Today we rely more on internet and we complete most of our daily task online. The use and more desire of “Business Process Management Systems (BPMS)” [2], has also contributed much to data collections and processing as more and more businesses today got aware of advantages of accessible data, they now apply process excellence techniques to improve their corporate execution and customer services. This has led to the rapidly growing level of demand to access customer data. The use of information in business processes do help exposes of problematic aspects of current business process to stakeholders and let them be aware of them and let them informed decisions regarding the situations [8].

## BENEFITS OF BIG DATA

Big Data has undeniable benefits to our lives. Big data solutions can have used anywhere and it can come in a form of Software such enterprise-class analytics platform and information management systems, Hardware such as HPE servers, data discovery tools and data analytics, Services and Support such as business value delivered from your data and possible world class support from international experts. In General, Data sets (groups of related data) can be analyzed computationally to reveal hidden patterns, business trends and associations relating to human behavior and interactions. Big Data technologies makes everything for us simple and quick to solve.

Some small companies may have not yet identified what big data entails for them due to their limited services and outreach but big companies like Facebook, Amazon, Google, Alibaba Group, Tencent, eBay, Twitter etc. are already benefiting and in most cases relies on Big Data analytics as part of their primary marketing schemes as well as a means of servicing their customers better [10]. Do you know that customers get excited about hearing their own data? It feels so

good to hear about your statistics such as according to Gabriele Weimann, “554 750 000 twitter users twitted 9100 messages every seconds, or 58 Million per day in mid-2013”. It is so encouraging. If you are a Chinese fan, hearing that Chinese e-commerce giant Alibaba has become the world ‘s largest Retailer super passing Walmart will make you feel very proud of your little support as a user. According to Zhang Yong, CEO of Alibaba said: “By 2024, we want to be a business platform serving 2 billion consumers and tens of millions enterprises at home and abroad”. Guess how many data will be generated by such a number of users in future.

For example, most e-commerce companies use big data analytics techniques store each customer’s searches(cookies) when visiting their sites and save previous purchases and any other piece of information available about their potential clients and then apply certain algorithms to such information and compare it to one customer’s information on their visit. Probably you have seen ads that bring back exactly what you have previously searched on a specific site or simply presented with links related to other related articles.

In traditional business approach, big data is also very much important. Based on information obtained from customer data the company’s top management can make informed decisions. Collecting and accessing data enables companies to check, control, and know its customers, employers, and visitor’s details. It can also help to prevent fraudulent and criminal activities by customers or bad guys out there, especially in insurance companies and banking industry.

A good example of Big data use or big data analytics in this case is the analysis of data for Panama Papers companies, a network of Tax haven by thousands of secretly interconnected offshore companies that were starving various nations from its important source of incomes as revealed by The International Consortium of Investigative Journalists (ICIJ) early 2016. With the help of big data technologies available, as searchable

database that strips away the secrecy of nearly 214,000 offshore entities created in 21 jurisdictions, from Nevada to Hong Kong and the British Virgin Islands was developed to explore and disclose such information to the general public and affected parties. This discovery has awakened many organizations and governments therefore we can conclude that with data available we can easily prevent, predict, identify, monitor or investigate and report unwelcomed activities in business dealings in order to fight corruption and fraudulent activities around us. By using massive amounts of historical data available, companies can distinguish between anomaly and normal activity in their computer systems in real time.

Information collected can be further used to find potential cyber threats active on the company's networks. At health care centers, big data is used to determine the cause of an illness and provide guidance on diseases treatment options. It also Reduce health care costs and improve the quality of patient treatment and save lives. Through thoroughly analysis of gathered data we can simultaneous monitor, fight and counter attack terrorism activities worldwide and increase national security, and public safety for our people. It can also help in predicting the occurrences of Natural disasters. Big Data have a large impact on the way public services are shared such as in hospitals, schools and other relevant departments or places respectively.

## **BEST PRACTICES OF USING BIG DATA TECHNOLOGY TODAY**

With a help of IoT, “Ford Fusion sends 250GB of data back to Ford, who in turn lets you know that something is wrong with your car” [13].

The use of unmanned aerial vehicles (UAVs), popularly known as drones have gained public interest in early 2014/15. They have been reportedly doing fantastic job in many part of

the world, particularly in Africa where they are used to help police officers in their daily special operation to combat and fight against crimes and speed chasing in accidents such as “hit run and run” tactics happening on public roads.

Drones are also effectively used to fight poaching in some few sub-Saharan Africa regions like in Uganda and Congo Republic. Drones have better capabilities as can they fly closer to the land and can bypass visual obstructions such as trees, clouds and buildings in order to get high-resolution pictures. They are also being used to deliver vital agricultural information to farmers through high-resolution pictures in Tanzania and other parts of Europe. Information is obtained from high resolution pictures such as the exact acreage of land, land usage and signs of pesticides or plant diseases that can potentially destroy the crops [4].

Small data out of the big data is the game changer. It allows top executives to make informed decision based on interpreted data. With the help of third party applications such as Microsoft’s spreadsheet, Oracle databases, HR ERP, data warehouse, open source file systems (e.g. Hadoop), Big Data is now made smaller therefore quickly giving us flexibility to connect to our data, consolidate it. And give a feed back quickly just while just seated with a laptop in front of you.

This aerial vehicles or drones designed primarily for spying or military purposes are also reportedly used to monitor conflict areas around the world such as in the middle east and other northern part of Africa at conflict as a pre-programmed flight plane or by remote control to gather intelligence and help to identify pertinent information. Such practices degenerate data and a proper analysis of these data acquired from drones on daily basis can be used to develop new mechanism on how to tackle certain challenges by using collected data to develop, predict and reveal patterns of recurring activities such as poaching, terrorism, diseases, natural events and many more. Despite their useful purpose, drones also face strict and

sometimes unfair regulations and in some cases banned or prohibited in countries like Namibia and South Africa due to the absence of adequate International Agreed standards and policies of flying drones [4].

Not only in Africa though but even in its largest countries of origin like China and French they limit. The U.S. Federal Aviation Administration has thus released a new set of rules that can make it much easier to fly drones by commercial operators for profit. This will boost its use by businesses, farmers, nonprofit organization and government agencies. Drones are faced with challenges such as unsatisfactory use of a drone for the first time due to the fact that it crushed right away on the first day of use due to less experience therefore concluding that they are terrible or useless. buyers also know no difference as to which drone is better or good therefore limiting its uses and potential customers.

## **SECURITY CONCERNS IN BIG DATA**

As we have learned that Big Data Technologies enables more and more data to be collected, stored, and analyzed by various companies, organizations, governments, and individual consumers, such practices have brought many advantages to our daily life yet it also requires that new IT regulations and policies related to privacy, security and intellectual property are needed to address many issues concerned to the types and use of information collected about us, no matter what advantages Big Data brings to people, organizations ,governments and business. We cannot overlook the privacy factor involved in collecting all such information.

All companies involved in collecting information about its clients must be obliged to develop Big Data strategies and policies in order to protect its consumers' information by ensuring that their organizations are not vulnerable to serious data breaches or do unsafe data disposal. And although the

commercial benefits of the use of Big Data are apparent. In the context of limiting risk, companies must have a definite purpose for the use of Big Data [9]. This way, we will be able to know when the sole purpose of collecting data about us and allow us to claim the violation of our privacy if such data is not used as intended and according to our consent or agreement. It can also give protections to companies during litigation in cases it is sued regarding how the data is being used.

Addressing the question of what security concerns are in big data. Given that we use internet so excessively, has led us to create more data every day and as a result we leave digital footprints everywhere online. So, while at this point, ask yourself a questions. How safe is my personal information while on the internet? Who can see and cannot see what I do online? Well, you will never know until you are assured.

Let me tell you a secret, your online activities may be tracked down by someone else out there or a group of people or a certain organization authorized to do so. Your privacy can be invaded by various entities without your consent. Such people could be: Your Internet Service Provider(ISP), Governments, Commercial agents (Advertisers & Online Service Providers), Employers and the bad guys so called Hackers or in other words Cybercriminals.

How they do it? It is simple. Normally every country has 1 or more ISP. The ISP are the telecommunication companies that connects you to the Internet. For every device you are connecting with to the Internet is assigned a unique name called an IP address that identifies you over their network. With such information they are able to identify your unique name and your location as well as filters your internet traffic in and out. At some point you will be denied access to a certain site or content through proxy services or censorships. In China, North Korean and other places it has become popularly known that your internet activities will be filtered and censored by the ISP on the request of the government authorities unless using a

VPN services. Particularly in China where I have been and studied, you are automatically censored and blocked from accessing certain western based internet services such as Google, Facebook, Twitter, YouTube etc. by the Cyberspace Administration of China the moment you enter their territory and using one of one their ISP such as China Telecom and China Unicom. Reasons why is known to them but I can speculate that such sites are violating the best interest of Chinese government and its people specifically government secrets such military activities and new innovations. When more information is their language might be a key advantage in terms of competitions. It can also be a tactic to promote local contents and allow similar innovations. It is also undeniable that various Chinese companies copied and developed their own Chinese versions or replicas such as Weibo, the Chinese social network platform like Facebook, Du Map a replica of Google map, Baidu search engine like google and WeChat similar to WhatsApp.

It is not only governments that can spy on internet users, ecommerce companies too. For every time you visit a certain site they require you allow them to keep cookies on your personal computers or machines. It will then be used for advertisement purposes. It brings back your browsing history; most frequently visited sites and the products you like most or that similar to that or mostly viewed or downloaded. Some sites sell ads. Hopefully you have come across to words like "Others also downloaded ...then the product list". Such activities result in increased product revenues and improve internet browsing experience. Especially for lazy people who hate punching keyboards for long like me.

Wait a moment. Have you googled your name? If you didn't for sure you are very behind and it might be the reason why you hardly come out of the interviews and secure a promising job. As a result of good use of data analytics it can bring everything associated with your footprints. Images,

videos and other things you have shared online publicly. As a result, Interview panels tends to google you first before they meet you in order to understand you better by visiting your social networks or other familiar platforms. If you are such a radical person who speaks out so loudly and share everything on your mind online, here they have a catch like fish in the lake. Bosses can also spy you on you while you are on the sick leave to see if you are real sick or you simply wanted to take a day off and go enjoy a honeymoon somewhere else. Picture backgrounds and time shared always share more you never thought it does.

Hackers, are the most feared guys when it comes to Information Technology. This internet espionage can do a lot of damages in attempt to gain access to your personal information and black mail you or simply ask for ram some money which you sometimes kind afford to give to such croakers. To them is a profession and they have a bunch of tools to use against you such as Trojans viruses, malwares, and spywares just to collect your personal info without your knowledge [16]. They can also disrupt services through social engineering, brute of force, and denial of services. Their motives vary.

Knowing that privacy and safety while online matters. I think it also worth mentioning that is possible to escape from the lions' mouth and avoid being a victim by conducting your own online activities assessment occasionally and avoid creating questionable foot prints because people change their minds overtime. When subscribing on sensitive sites such as online shopping and electronic banking sites, always set and use a unique strong and long password; which are a combination of letters, numbers, characters and very hard to guess. Such passwords, cellphone numbers and subscription emails should vary from each sites you visit as it can easily be used to guess an entry to the other such PayPal, Facebook, emails etc.

Be watchful not be redirected to bogus sites, sites that have no SSL connections (sites that have weird URL address and do NOT start with https://) and never open attachments to emails unless scanned and cleared by your antivirus. Mostly following such links leads to phishing website that tricks you with offers and ask you to pay certain deposits or surrender your personal information and some points exposes you to viruses that encrypt files on your computer, and then make a high demand of ransom money to decrypt it or threaten you with public expositors. Antiviruses are software installed for further protection and are used to scan your computer, files and guards you online to avoid malicious damage to your machines by Trojan viruses, spyware and affected sites.

Virtual Private Network(VPN) in a form of software as a services or VPN Routers to helps you surf the internet more safely as it builds a secure tunnel between your devices and the VPN servers and allows you to transmit your data in a well encrypted manner which makes you a vigilant and anonymous on the internet. This makes it very difficult for the ISP and possible hackers to track down your traffic as your IP address keeps changing or simply pretending to be somewhere else in another country while you are just seated in front of their noses.

Another thing we need to avoid as internet users is never send sensitive files or log into your banking sites using open or unknown public networks which are normally situated at airports, cafes etc. as free Wi-Fi hotspots. As you not required a password to get connected, it means is free for everyone and the snoopers might be sharing the same network. Therefore, you have no ideas who is spying or filtering the network traffic. Hackers can use such opportunity to intercept the information transmitted from you via this network and steal it such as emails and other personal credentials such as online bank account information. You may also want to delete unwanted cookies to avoid privacy invasion at it maybe

potentials used in cases your computer is stolen, hacked or lost. A well protected private network is a must use by both business traveler, bloggers and government agencies.

## BIG DATA AND PRIVACY

Given that a massive amount of data is collected about people every day and we do generate trillions on data daily. I think it worth asking: where is this data stored? How is such data is protected? Who own such data? Is there Regulators who can put a limit to what info can be collected about us?

Well, be reminded that Internet has no geographic boundaries and most of the data flows freely [11]. Except in China, and maybe other countries such as North Korea. I think that any company that will consider big data, shall not only be concerned with keeping out systems hackers to avoid possible data breaches but must also think about data backup to avoid data corruption and avoid a single point of failure. data distribution is also a hot topic as we are interested to know as to whom, when, where and how data is made available/accessible).

Many companies still favor end to end data encryptions, yet remember that we are already faced with storage problem in big data. Data is so huge to be handled, stored analyzed, and shared at appropriate speed and in real time using the traditional infrastructures. Therefore, encouraging higher levels of encryption will just cause processing burdens and hamper system performances. Therefore, security is one of the biggest trend in big data technology.

Having learned that Big Data is having some very serious security and privacy concerns that need to be addressed it is advisable for companies to do a thoroughly research on why and how to migrate their data as well as if they will real make use of such big data technology or services fully, such as “cloud computing”. Despite security concerns, the use of Big Data

technology in companies is also disruptive in such a way that it changes how business intelligence (BI) is used. For example, spying on other's activities on the company's network may be good for the company to save its resources and make informed decisions/steps but might not be good to individuals. in-house Big Data use and processing is also associated with high cost as it requires you to have high technology tools and expertise skills. Similarly, if is outsourced or done off-site it will again raise security concerns especially in handling of high sensitive data.

Another threat to Big Data is accessibility and liability. This involve both unauthorized access to sensitive data by hackers or any other unauthorized personnel(intruders). Think of health records in a hospital or data compromising by hackers in unsecure Banking systems or national databases. Organizations are held liable for any damages may be caused by unauthorized access to confidential information. Legal requirements, privacy issues and copyright to intellectual property is also connected to data as far is concerned.

Another thing that should worry you is that; we share too much information online despite a day hardly goes by without seeing or hearing headlines about any significant Cyber-attacks or some kind of coordinated attacks. This is a proof that we shouldn't take privacy in big data lightly. If personal accounts of bosses for top internet giants such as the hacking of email account for the CIA director John Brennan by a teen, that of FBI Deputy Director, Mark Giuliano and his wife, Google CEO's Quora as well as the Facebook's Co-founder Mark Zuckerberg, who going to say you can't be hacked?

Everyone is at risk here and is only a matter of time and interest by those who can. This has put fears in people and companies involved in doing business or giving online services to its customers such as through trading/shopping, financial transactions, data storage and synchronization too. This threats comes in many forms: ransom ware attacks, malware, System

hacking (stealing data), services disruption (attacks on electric grid) and distributed denial of services (website shutdown). Such cyber-attacks are routinely done via our insecure internet connections to the internet and via our internet-connected embedded system devices such as Smart Phones, Smart TVs, Set-top boxes, Security Cameras, printers, Refrigerators, Microwaves and many more. as such devices are the one used to send out and spread malicious spam emails which in turn used for hacking or ripping off people's sensitive information. Hackers have sophisticated programs that helps them to best extract such data and turn it into meaningful and useful information which will then be used as the biggest weapon in launching cyber-attacks [11].

Are you even aware that these Intelligent devices do collect dozens valuable data about us unknowingly? We are so exposed and so vulnerable. All such information is purposely collected of which some are without your consent. Most of which is sold or rented to third parties. Some can be used for good purposes such as retail analytics, customer relationship management, demand monitoring and inventory management. Just Think about it.

Your car collects information about you (location, speed and driving patterns). Google Maps on your phones knows exactly where you are. Your phone remembers all Wi-Fi network it has been connected to previously, this can give access certain features on your phone. Facebook keeps history of places visited, the likes you have and is even able to analyses it and give you statistics according to time, gender in a graphical format. Certain sites you visit use cookies. Credit card keeps your shopping history. CCTV cameras has your moves. Internet browsers you are using to search internet remembers all passwords and usernames and do auto fill for you if you choose that in settings and at some points keep your browsing history. This is very dangerous in case your computer is infected by certain malware of DNS infections. Everything is

exposed. Your PC for instance keeps track of recent accessed file therefore making it easy to do a follow up on what you were busy doing your computer without any struggle to search your files. Whoever said we are “living dataveillance society, where our actions and communications are systematically monitored” [11] is correct because our behavior is easily tracked.

We easily sell ourselves. For instance, if you have used your account details to subscribe somewhere else to bogus sites such information can be reused to guess by this cyber crooker to guess and gain access to any possible accounts such as Facebook, pay pal and other accounts associated to your email address. Because there is a higher chance that you use will use the same email to register anywhere else. Trust me, no one is willing to memorize to email addresses. Mostly we have just one or two that we use often. One can also trick you by sending you promising emails that looks genuine and from legitimate person that in most cases redirects you to bogus sites that will prompt you to surrender your personal information or exposes you to viruses attached to such emails.

## WHAT IS PRIVACY?

Let us talk about Privacy in specific. Historically, the word “privacy” has been difficult to pin down its single definition until in the 1960s. Privacy actually has something to do with keeping something private. Something which is private to a person is something which is seen as inherently special or sensitive to them.

In U.S privacy is seen as the right to be alone and in Europe is regarded as an honor and dignity [11]. According to Wikipedia, “Privacy is the ability of an individual or group to seclude themselves, or information about themselves, and thereby express themselves selectively.” The Internet has brought new concerns about our privacy, "where every online photo, status update, Twitter post and blog entry by and

about us can be stored forever"[6]. Data privacy requires data security.

In Big Data World it has become very much difficult to put limits to how much data one government or corporation should be allowed to keep about people in order to ensure privacy. What is good to hear is that most of those that monitor us from various points of view are mostly involved in as self-regulatory policies and enforcement bodies. In digital age, privacy can be categorized into three basic types, namely [7]:

Physical privacy (freedom of intrusion into your physical person, possessions, or space).

Informational privacy (your privacy to personal information being collected, stored, and shared in any other format.)

Organizational privacy (government agencies, organizations, and/or businesses privacy to keep its activities or secrets from being revealed to others).

The right to privacy is regarded as a fundamental human right and recognized in the UN Declaration of Human Rights, the International Covenant on Civil and Political Rights and in many other international and regional treaties. "The right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, property, thoughts, feelings, secrets and identity. The right to privacy gives us the ability to choose which parts in this domain can be accessed by others, and to control the extent, manner and timing of the use of those parts we choose to disclose" [3].

## **HOW BIG DATA RISK OUR PRIVACY?**

Well, whenever you give out your information for the convenience of using a products or services such as signing up for an account online or register for certain services with a merchant store, you have definitely opened doors for the use of

that data by companies and government agencies and/or third parties. In fact, there is an endless way of how our information could be collected. For example, technological innovations has made it easy to track our online activities such as call logs, geolocations, text messages and many more. This information is always saved on various servers/ databases which can be accessed by third party application. Think about social media such as Facebook, twitter, LinkedIn, WhatsApp, WeChat etc. that we use for almost every second and you can readily share any information with your friends just on a simple click of a button. Big Data analytics market has in recent years benefited certain companies by selling raw data to concerned companies in exchange of billions of dollars. This is a big trade off. “companies not only must be sensitive to how they deal with consumer information but also must consider the market effects of providing their Big Data analytics to third parties” [2].

In some countries where security agency and Intelligence services do spy and collect intelligence over millions of citizens on daily bases through facial recognitions and other intelligence capabilities also arises questions of over privacy. Early July 2016 on “PandaGuidesOfficial” report, China’s internet regulator: Cyberspace Administration of China(CAC) has strengthen its muscles by increasing its censorship by issuing “six obligations” for mobile app providers in China to comply with by asking all app providers to do real name identifications for its users.

First, the CAC stresses that “the mobile app providers to verify users’ identities by requiring mobile phone number or other identification details/information”. Secondly,” the app developers are also required to protect user’s information and cannot use the information without the users’ consent”. The question here is what about those start up that has little knowledge or no required technology to ensure data protection yet their services are used by millions of people daily? The third requirement is on censorship. Saying “the app providers should

improve their censorship and punish anyone releasing illicit information through warnings, suspension of services or shut down accounts.” Fourth,” app providers must inform users of their rights, but yet providers are forbidden from collecting users’ location information and reading their contacts stealthy. “and the fifth, are banned from pirating rival’s products and services. These two sound good but how about the sixth one that asks the app providers to record user logs and preserve the information for at least 60 days? This is giving enough time for hackers to collect all the information associated to us. Therefore, putting app users at a higher risk in cases of data theft. (Arif, 2016)

All I have described above will creates millions of data over time that can be used to improve certain services and such data collection is also in the interest of government to protect its people and the country’s top security interest against the spies but equally putting our privacy at stake. In the absence of secure connections such via VPN services, everything about your internet use is controlled: your searches, and communications. Besides, having such data collected about us, we are the biggest contributor to generating such huge of data as we seem like we are so addicted to using our phones, tablets and PC in doing anything. Instead of sleeping, reading or doing something else, we are connected to the internet 24 hours a day / over 7 days a week, having our eyes glued on laptops, phones, tabs, Smart TV, check emails, face booking, reading news, reading twitter feeds, read our favorite blogs, booking restaurants, making jobs application, sharing moments and reading our friends post, reading and buy books online, bidding and transfer money, using Maps for directions, and many more.

Today we are also in a world of rising drones that are being flown everywhere due to the limits or absence of proper laws against the use of drones in specific areas. They are being used to collect intelligence information through taking surveillance videos and aerial images which are more detailed

and in some circumstances violates personal privacy and other government policies. And due to this challenges is very difficult challenge drone owners in courts. Typical drones can travel at 160km or more per hour and they can fly to a height of 150 meters. Therefore, posing a big threat to aviation proclaimed areas as well as protected and private properties. Even though drones are applicable in any situation where gathering information is either too time consuming, costly or dangerous, drones are health risky as they can make quick and sharp cuts.

Reportedly, Militant group such as ISIS used DJI's Phantoms for surveillance purposes. In 2015, a DJI's Phantom reportedly crashed – landed on the White House lawn and months later on later landed on the roof of the roof of the Japanese Prime Minister's Office. In South Africa, drone rules were enacted in 2015 to impose a 10-year jail sentence for any illegal drone use. And subsequently, a total ban on drones for civilian use was imposed by Kenya in January 2015. In Namibia drone use is also limited to private use and will confiscated if used in public without the approval from the aviation offices. Many drones have been reported shot down immediately after seen hovering over private properties. However, drone use has been welcomed in Nigeria, while seeking international community support. Drones have capabilities to send and receive information therefore its intelligence can also be seen a threat to personal and national security and privacy. Some drones can also lift up certain weights therefore it can be secretly used to dispose dangerous devices or weapons such bomb etc. Currently, DJI (a Shenzhen, China based drone manufacturer) hold 70% of the market comparing to its rivals Chinese firms such as Yuneec, Ehang and the electronic giant Xiaomi.

The digital age had our world changed. What was impossible is now possible and that was offline life is now online. We are now giving away more information about ourselves and that of others than before. Due to digitalization

we are trading our personal information online like e-commerce sites such as amazons, eBay, Alibaba, JD. Com, bogus websites, via emails, job applications, and via instant communication platforms by keeping in contacts with families, friends or business colleagues when we are away from homes or work place. In some cases, our personal data are sold to third party companies for marketing purposes without our consent. This has now brought fear to us to such extend that online privacy is more dangerous than a home invasion. Yes, that is true. Google your name and see how much information will be displayed about you. Shall any of your secrets files get into wrong hands the most you shall fear is being posted somewhere else online that might be viewed by millions of people and those you never wanted to if such secrets exist. For example, your followers and friends may retweet or share such information several times until it has reached one connected your friends and as well as friends' friends in anywhere in the world. An example is "Kanye Wes's famous video", which is intentionally put up on the internet regardless its revealing content of other artist without their consent.

## **CHALLENGES IN BIG DATA**

"Despite of its emergence and advantages in various domains, big data still suffers from major disadvantages. Timeless, scalability, and privacy are the main problems that hinder the advance of big data" [12]. Lack of Human Collaboration in computer programs, heterogeneity and incompleteness of data, capturing impact from analytics and being or pointing the leader is also some of the difficult tasks in Big Data analytics.

In terms of data protection, many Modern data centers are still faced with dangerous threats such as DDoS Attacks, Web Application Attacks, DNS Infrastructure: Attack Target and Collateral Damage, SSL-Induced Security Blind Spots and

Brute Force and Weak Authentication. Therefore, leaving privacy with a BIG question mark.

## ESCALATING TRENDS IN BIG DATA

As big data changes and new ways of working with such data changes and new innovations just pop up. As big data grows the following trends are expected to be escalating in Big Data:

1st, the NoSQL is expected takeover. NoSQL technologies is commonly associated with unstructured data and since most of the unstructured data is the most one in the data we have this still lead to a high need of NoSQL Technologies. The NoSQL companies includes: MongoDB, DataStax, Redis Labs and MarkLogic. “NoSQL was founded to provide scale, flexibility, and the ability to leverage large sets of data faster. Companies like MarkLogic, Casandra, Couchbase, and MongoDB are bringing new innovation to the SQL database market...” [13].

2nd, Hadoop projects are expected to even emerge bigger. Hadoop’s big data architecture has been incorporated by many data storage vendors including the traditional database providers such as Teradata, HP, Oracle, IBM, Microsoft and SAP. This big data architecture is here to stay or at least it will be doing well in the next few years and will be concurrently be used by other various companies as an addition to their classical data architect.

3rd, is the Apache Spark. Spark provides a more increased data processing speed compared to Hadoop. As an open source project, has a potential to be one of the largest big data platform of choice for a number of enterprises in near future.

4th, Internet of Things (IoT). The development of the Internet in which everyday objects will have network connectivity, allowing them to send and receive data, The Internet of Things is said to be doing well so far and it has seen

more boost from various companies such as Ford, GE and Rolls Royce. Even though the technology is still in its early stage more companies are looking forward to invest in such technology. According to Tableau, “IoT, Cloud, and Big Data are expected converge” [13].

5th, The “death” of data warehouse. As Big Data gets cloudy, a demand for cloud data warehouse has increased dramatically and many organizations have seen this platform as a game changer. Analytics Data Warehouse innovations by leading cloud and data companies such as Microsoft (with Azure SQL Data Warehouse), Google BigQuery and Amazon Redshift, Batch and Stream Data Processing such as Google Cloud Dataflow , the Managed Hadoop & Spark such as Google Cloud Dataproc, Powerful Data Exploration such as Google Cloud Datalab and Scalable Event Ingestion and Messaging Middleware such Google Cloud Pub/Sub has boosted the demand for cloud computing services and brought it to life in businesses today. Many corporations have realized the importance of effectively capture, store, manage and use the massive amounts of incoming data. It is evidently showed some achievement that led to new revenue opportunities, improved customer service, or more effective marketing.

Intel Technology has also contributed highly by creating cloud-powered innovations that will accelerate time to get meaningful insights from your data and transform company networks. With faster IT services and flexible solution readily available we can seamlessly integrate our work in clouds services that will results into major functional automation and greater efficiency.

## CONCLUSION

Big Data is continuously improving and the benefits offered by data-driven decision making are widely recognized. Big data has an ability to transform the nature of a business as it

enables innovation within an enterprise given that the enterprise acts upon its insights. Through thoroughly data analysis, we can examine collected or generated data to find the actually facts, relationships, hidden patterns, and unfold common trends. This way we can support better the decision making by governments agencies and corporate executives. Because traditionally ways of data collection, analysis, processing and storages is never being sufficient, is preferable to use Big Data storage and analysis resources and corporate performance monitoring tools together with available Big Data solutions and practices to process and analyze such data to their advantage and broaden their analytic capabilities and deepen their insights delivered by such Business Intelligence. Business Intelligence enables organizations to gain insights into the performance of an enterprise by analyzing the data generated through its business processes and information systems (IS). As a result, such analysis can be used managements to steer the business in an effort to correct detected issues and enhance business performance. Due to the rapid growth of accessible data, companies need to find ways on how to acquire, evaluate and safeguard such “Big Data” in order to ensure personal privacy and security. And for the data to be truly secure it must satisfy the three elements: confidentiality, integrity, and availability (CIA).

## REFERENCES LIST

- [1]. Judith Hurwitz, Alan Nugent, Fern Halper, & Marcia Kaufman (2013). Big Data for Dummies. John Wiley & Sons, Inc., Hoboken, New Jersey.
- [2]. Thomas Erl, Wajid Khattak, & Paul Buhler (2016). Big Data Fundamentals Concepts, Drivers & Techniques.ISBN-10: 0-13-429107-7. Arcitura Education Inc., USA.

- [3]. Yael Onn, et al., (2005). Privacy in the Digital Environment. Haifa Center of Law & Technology, pp. 1-12.
- [4]. Pavithra, R. (2016, July 07). Drones - Keeping a watchful eye on Africa. The Namibian. Retrieved July 11, 2016, from: <http://www.namibian.com.na/Drones---Keeping-a-watchful-eye-on-Africa/42683/read>.
- [5]. Judith, H., Alan, N., Fern, H., & Marcia, K. (2013). Big Data for Dummies. John Wiley & Sons, Inc., Hoboken, New Jersey, USA.
- [6]. Rosen, Jeffrey (2010).The Web Means the End of Forgetting. New York Times.  
Accessed from:  
[http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html?\\_r=3&ref=technology](http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html?_r=3&ref=technology) , since , July 19, 2016.
- [7]. Terence, C, & Mary, L. (0).Privacy and Big Data. Accessed from:  
<https://books.google.com/books?id=Y AeNPJdNo2UC&printsec=frontcover>, since July 14,2016.
- [8]. Vladlena, B. (2013). Making the Most of Big Data. Manager's Guide to Business Intelligence Success, 1st Ed. ISBN: 978-87-403-0520-3. Accessed from:  
<http://bookboon.com/en/making-the-most-of-big-data-ebook#download>, since July 16, 2016.
- [9]. James, R. & Michael, R. (2015). A Business and Legal Guide. Taylor& Francis Group, 6000 Broken Sound Parkway NW, Suite 300.
- [10]. Frank, J. Ohlhorst, (2012). Big Data Analytics: Turning Big Data into Big Money.ISBN: 978111814759.John Wiley & Sons. Accessed from: <https://www.safaribooksonline.com>, since 18 July, 2016.
- [11]. Mary E. Ludloff &Terence Craig (September 2011). Privacy and Big Data. ISBN: 9781449305000. Accessible from: <https://www.safaribooksonline.com/library/view/privacy-and-big/9781449314842/> , since July 20 2016.

- [12]. Dalei Wu, Li Yang. (January 2016). International Journal of Information Security and Privacy (IJISP) Volume 10, Issue 1. ISBN: 9781466689688. IGI Global. Accessed from: <https://www.safaribooksonline.com/library/view/international-journal-of/9781466689688/>, since 20 July 2016.
- [13]. Tableau. Top 7 Trends in Big Data for 2015. Accessed from:<http://www.tableau.com/sites/default/files/media/7-trends-for-big-data-final-mw.pdf>, since July 21 2016.
- [14]. Tableau. Top 8 Trends in Big Data for 2016. Accessed from:  
[http://www.tableau.com/sites/default/files/media/top8bigdatatrends2016\\_final\\_2.pdf](http://www.tableau.com/sites/default/files/media/top8bigdatatrends2016_final_2.pdf), since July 21 2016.
- [15]. Matt,B. & Nicky, Z. (16 July 2016). How Much is WeChat Worth? Accessed from: <https://mp.weixin.qq.com/>, since 22 July 2016.
- [16]. Qihoo 360 Technology Co.Ltd.(August 4,2016). Who's tracking your online activities? Accessed from: <https://blog.360totalsecurity.com/en/avoid-online-activities-tracking-keep-privacy/>, since August 2016.