ID 666

Big Data Analytics: Best Practices from Singapore in the Context of Sri Lanka's Digital Defence Requirements

R Amarasinghe[1] and M Ranmuthugala

Dynatech International Private Limited

[1]rosharn@dtech-int.com

Abstract— The information age has resulted in massive amounts of data being shared online and data being created over multiple platforms including smart devices. Such amounts of data create big data that can be used to understand usage patterns and internet behaviour. Many companies and countries collate such information to provide a better service to its customers or to guard its citizens. It is especially important for governments to collate, utilize, and analyse such big data to protect its borders. Analysing the chatter on the cyber space can help avert terror attacks and safeguard citizens from unscrupulous people. Sri Lankan defence has traditionally left the cyber domain vulnerable although maintaining extremely efficient protection mechanisms for traditional boundaries. It has faced many cyber threats over the past few decades. Thus, it is imperative that the country invest heavily in technology and big data. Singapore has consistently proven itself to be capable of safeguarding its borders and economy through use of technology and continues to innovate and invest in technology such as AI and big data. This paper provides a practical model for use by the Sri Lankan government based on best practices from Singapore that will help the island nation increase its security for its virtual borders. This will help it stave off security threats and economic threats.

Keywords: *Big Data, Cyber Terrorism, Data Sets*

Introduction

The growth of data that we create, and share has increased exponentially over the years. This growth is most prominent in the two decades starting from 2000 and 2010. In 2011, the data warehouse of Yahoo! was storing a total of 170 petabytes up from eBay's 8 petabytes just two years before (IMDA Gov). This is due to the amount of information we share with each other. Over the years, we have gone from storing only governmental, defence-related, or commercial information to sharing personal information online on social media, from Facebook to Instagram. According to Milenkovic (2020), "People are generating 2.5 quintillion bytes of data each day" (Lynkova, 2019) and "Nearly 90% of all data has been created in the last two years" (Lynkova, 2019).

A. What is big data?

Big data means any set of data whose size is larger than the database software tools can handle (IBM, 2020; Lynkova, 2019). Handling such data includes extracting (from), "storing, managing, and analysing" (IMDA Gov, p. 2) these datasets. If the dataset exceeds the processing capacity of conventional database systems," it is called big data (IMDA Gov, p. 2). Technology that manages such big data is a "new generation of technologies" that are specifically "designed to extract value economically from very large volumes of a wide variety of data by enabling high velocity capture, discovery and analysis" (IMDA Gov, p. 2).

Any company or large entity could have big data, where they store all the information that the entity generates. This could be a customer's personal data (name and email address), financial data (customer financial data includes their

spending patterns and payment preferences while company financial data includes expenditure, income, and sales information of the company), or operational data. To store such types of information, the company would need a warehouse or system that can handle petabytes or exabytes of information. This information comes from a variety of sources including sensors and smart devices, and can be unstructured, semi-structured or structured information (IMDA Gov).

B. Why is big data important?

In a commercial enterprise, big data becomes important because of the interconnectedness of business verticals (IMDA Gov). In the defence activities of a country, big data merges a larger data set: it will collate such information as network information (network usage), personal information, and usage data (information that is searched for and what sites are visited frequently).

The effective use of big data helps with proactive national defence policies that focus on identifying anti-nationalist activities such as the Easter attacks of 2019 where terrorist activities could have been picked up through chatter on social media and digital platforms if it had been available and used at the time. The Indian government's NIST alerted the Sri Lankan government of chatter in India and had the information been acted upon, the attacks might have been averted.

C. Global Statistics and situation report from various industries

Big data can be a boon to any economy. According to Milenkovic (2020), "By 2023, the big data industry will be worth an estimated $77 billion." According to hostingtribunal.com (2020), "big data has created 8 million jobs in the US alone and 6 million more worldwide." It can help navigate a country to safety. Terrorist attacks can be prevented with the use of big data and economies can be pushed in the right direction with the use of the correct information. Additionally, big data can help a country maintain its digital health. Countries can collate information (such as which organizations have been hacked, where the attacks have come from, and where the vulnerabilities lie in the system) to understand the risks that it might be facing as a country. This is a significant and important consideration to any nation: cyber threats can negatively affect the economic security of a country as with Bangladesh Bank attack in 2016 (Amarasinghe and Ranmuthugala, 2020) or its national security as happened in Sri Lanka with the attacks carried out by the Liberation Tigers of Tamil Eelam (LTTE) (Amarasinghe and Ranmuthugala, 2020). Big data is not only useful in one sphere. It pervades every area that matters to any nation and using information generated through daily use of technologies can help countries formulate better plans of defence or attack as the need may be.

Since the pandemic of 2020, cyber-attacks on commercial and development-sector organizations have increased (Cincinnati Business Courier, 2020). Some of the major agencies of the world have reported increases in attacks on their networks and companies. The World Health Organization (WHO) reports a "five-fold increase" of cyber-attacks (WHO, 2020) on its staff with 450 email addresses being leaked online. This could have resulted in scam mail reaching the public purporting to be from the WHO. These emails are not merely irritating in content. It can be phishing or hacking emails and can be criminally intended.

According to Vanguard (2020), Robert Rizika (Head of North American Operations - Naval Dome) says that, "in 2017 there were 50 significant OT hacks reported, increasing to 120 in 2018 and more than 310 last years." He warns that in 2020, it will rise to "more than 500 major cyber security breaches, with substantially more going unreported" (Vanguard, 2020). This is an unprecedented increase and is cause for concern.

The global pandemic resulted in staff members working from home, which led to significantly increased vulnerabilities. Enterprise Times (2020) reports that "attack numbers are up with 94% of organisations admitting to a data breach in the last year" and "80% of respondent say attacks are more sophisticated, while 18% say they are significantly more advanced." Forbes (2020) says that, "Microsoft alone is observing around 12 million attacks every day. That's an increase of roughly 20% over February of this year."

It is as important for a country to protect its virtual borders as its physical borders. Such

cyber-attacks can create havoc in a country because scams attack the citizens of a country as well as the economic institutions of a country, allowing for the siphoning of hard-earned cash to other countries and unscrupulous and criminal persons. Thus, it is important that countries invest heavily in safeguarding its virtual border. To do so, countries require a practical model or framework that takes into consideration any local issues that may affect its effective use. To propose such a framework, the authors have analysed the best practices of Singapore, given its stellar track record and innovative approaches in cyber security (Amarasinghe and Ranmuthugala, 2020).

Methodology

This paper uses desk research, addressing the best practices of Singapore in terms of usage of big data. The paper will then draw from these best practices to propose possible models and approaches for the government of Sri Lanka that will help better use the big data generated on a daily basis. The research primarily aims to provide relevant examples of best practices from the region for use by concerned governments and secondly to highlight the need for effective collection and use of big data by governments. The paper answers the following research questions: What factors drive Singapore's national cyber security policy? What are the key regional risks within Asia? What is Sri Lanka's cyber security policy and how can data analytics be used for national defence?

Limitations
The paper is limited to desk research because of the nature of the research that calls for information from a foreign government. To see the larger picture, it is important to analyse available desk research.

Results

Singapore has shown itself to be extremely capable in defending its virtual borders against cyber-attacks (Amarasinghe and Ranmuthugala, 2020). It has also integrated big data into the Singapore Armed Forces. The integration of data is expected to give it a competitive advantage in war and minimise the loss of citizen lives (En, 2016). The requirement is "intelligence that is accurate, geographically-precise and real-time" (En, 2016). The government invests heavily into all aspects of defence including "Intelligence/Command, Control, Communications, Computers and Intelligence (C4I) structures, capabilities and human resources" (En, 2016, p. 52).

However, analysis of the information that is collated is difficult due to limited human resources that will take long to produce more and better intelligence. Thus, the Singapore government uses specialised data management and data analytics from labs run by National University of Singapore and associated tech hubs at NTU. The Singapore model uses a centralised data aggregation mechanism where the armed forces work with the Ministry of Defence in collaboration with the state technological services to focus on monitoring and identifying attacks on the Singaporean government ad related entities. The attacks on the Singaporean agencies over the past two years and the responses of the government that arise from the information provided by the cyber security operations centre and cyber threat identification, management and resolution mechanism showcase Singapore's efficiency as a country that can identify potential risks and carry out effective action. This efficiency arises from analyses based on the digital platforms that have been integrated to the country's national defence networks. The Ministry of Defence has also focused on creating a specialised cyber task force and cyber warrior division to protect the digital boundaries of Singapore against cyber terrorism and organisation attacks.

The model used by the Singaporean government acts as a benchmark for regional governments highlighting the integration of public/private cooperation coupled with a systematic integration of national agencies to ensure that the country is protected from cyber hacks and to ensure that the country continues operations unhindered by cyber warfare/terrorism.

Discussion

Based on the highlighted Singapore case study, an opportunity exists for Sri Lanka to utilise its focused digital transformation efforts to set up a centralised data analytics platform. This type of platform would be placed at a centralised data aggregation/processing level where the platform

would be capable of receiving information/data sets from various highlighted authorities. A national data analytics platform of this nature can provide valuable insights across multiple areas of focus such as national defence, governance, cyber security, and global threat identification.

This is relevant and necessary for a country of Sri Lanka's nature, size, and history, because of the threats it has already faced and the potential threats that it can face in the future. Such a platform can collate and analyse information such as health risks (i.e.: predict COVID-19 hotspots and potential risk areas), terrorist activities (predict and avert any terrorist action that can compromise the safety of the country's citizens), and financial risks (spam, scams, and phishing and hacking mail).

In the following model, the created/appointed coordinating government agency would feed numerous data sets and provide management platforms to present the information collected in a visually comprehendible structure utilising analytics tool. A primary focus of the platform would be a user-friendly search engine enabling quick queries and customised searches across various data sets. Machine readable data injections is crucial to ensure cross device/service integration.

Another key area would be the Standard Operating Procedures (SOPs) that are essential in creating a document-driven structural guideline for the usage, maintenance, and engagement with the centralised data analytics platform. This also ensures that the data is updated regularly through operating protocol and that the collectors and inputs from the agencies are regular.

A central steering committee for project management, reporting, and decision-making is required as there will be several inputs from external consultants such as SL CERT, security companies (both private and public) and regional cooperating bodies as many countries use cyber security and data sharing as a platform for government-to-government based discussions. This ensures access to information that is non-confidential in nature and it assists government agencies in identifying potential regional risks due to terrorist, virus or other cyber-based risks.

D. Proposed Model/Process

The following highlights key areas of focus in creating a national data analytics platform to ensure national security and economic growth.

- Platform establishment

The platform will focus on two aspects of premise-based physical infrastructure and a cloud-based backup/system mirror. The reason for this is that all governments advocate for national security but data integrity is compromised when hosting on third party global platforms such as Microsoft Azure or Amazon AWS. Although the management of the platform is simplified by the use of cloud hosting, the data resides outside the physical boundaries of Sri Lanka, thus creating a risk in case of unexpected sanctions/war. The platform will be based on industry-leading hardware with failover protocols to ensure near 100% operational efficiency.

- Products/Services Selection

The platform carries two major focuses; namely, the selection of products and brands to set up the security operations centre and data analytics network as well as the services of third-party security integrators, national advisers, and security/compliance experts. These two aspects ensure a holistic approach to platform security and availability. The usage of external security experts ensures that the Sri Lankan government receives industry relevant insight and security consultation both during the initial set up and during the execution/training phases.

- SOP Creation

Standard Operating Procedures ensure that a user-friendly guideline is created to instruct and train platform users on how the system can be optimised, how to generate reports, and also the escalation process if emergency threats/risks are detected. This also ensures that an audit trail exists in case of an unlikely system/platform compromise. This provides the government with valuable insight into fine tuning operations and ensuring the integrity of the system.

Identification of Cluster Agencies

This is used to identify national agencies that are considered on a three-tier basis: Essential,

Medium Risk and Low Risk. These three categories are used to focus on the various government agencies and entities based on their importance to national development. Thereafter, the three categories are documented as clusters for phase-based engagement and data collection.

- Data collection mechanism

The data collection mechanism is based on collectors that are set in each agency/institution. A mini centralised server is established in each agency and the various networking and security devices are pointed towards this server. Data is pulled/gathered on a regular basis from these devices into a central repository. The central repository is thereafter connected to the national platform and information is processed. This ensures that if a collection point/agency server becomes inactive, easy identification/rectification of that inactivity is possible. The availability and continuity of data and information systems is a crucial factor in making the national platform a success.

- Analytics Platform

A combined platform for data analytics is recommended utilising data processing, visual reporting, and data analytics tools. A cross product/service method ensures that false positives are avoided, and false negatives are detected. The approach to platform integration would be based on the sector-specific application of the various tools with the sharing of results/processing amongst different tools. This ensures that several reports are created using the same data sets and by applying intelligent algorithms to detect anomalies.

- Reporting/Escalation

A dedicated national security team should be in place based on tier 1, tier 2, and critical tier 3 escalations. The categories are to be defined in the SOPs. This ensures that national-level threats detected based on regional data sets or other methods that have been fed in are identified. This has two benefits to the Sri Lankan national security apparatus; firstly, a competent set of skilled security experts/data scientists are created who can then train the next generation of platform specialists. Secondly, a national-level cyber security response protocol is created taking into account the tri forces, various defence agencies, and coordinating entities. This creates an actionable platform to support national security and cybersecurity development.

- Future Expansions

Once the national platform is created, future expansions can focus on regional information sharing, international platform creation, and regional security partnerships. This ensures that governments within the region use cyber security and data analytics as tools for diplomatic and trade agreements where physical resources, scripts, and information are shared to ensure that the region is secure against unexpected cyber-crimes, financial fraud or digital malpractices. The platform can serve as a tool to strengthen regional ties and create an inclusive cyber security defence mechanism.

Conclusion

Based on the research carried out on the Singapore model and the security provided across the cyber domain against terrorism, fraud, and hacks, the proposed Sri Lankan data analytics model can create a stronger platform, which can integrate digital channels and ensure that the government (in collaboration with defense agencies) ensures the continuous protection of the country's cyber domain.

The recommended model takes into consideration the changing Sri Lankan landscape along with regional influences to create a sustainable digital model that can be expanded to other areas such as healthcare, governance, and digitization/automation of government services. The data analytics component will provide the Sri Lankan government and its associated agencies with actionable intelligence in the form of data-driven trend analysis, identification of potential risks, and mitigation strategies based on the intelligent scripts written.

The overall landscape of Sri Lankan defense has traditionally left the cyber domain vulnerable whilst maintaining extremely efficient protection mechanisms for traditional boundaries. The cyber domain has become a key area of attack globally with hackers continuously testing the security measures utilised by governments in the hope of penetrating national defenses and securing nation critical information of holding different government/essential agencies to ransom by encrypting confidential data. The

model recommended takes into consideration the various agencies and the integration required to ensure that the model receives relevant information efficiently. The model also covers a comprehensive approach for the identification, categorization and protection of key agencies and government establishments. The data analytics platform coupled with a strong cyber security operations platform will ensure the security, reliability, and availability of information. It is also recommended by the authors that the data analytics platform be hosted within the boundaries of Sri Lanka to ensure that any future sanctions or threats from war or other unforeseen issues will not affect the security of the nation. Information hosted and indexed onsite on physical servers with locally available backups will ensure that the systems continuously run smoothly and provide the government with a clear perspective of Sri Lankan security operations, information that can be used to create changes within the Sri Lankan governance and defense industries and to most importantly ensure that the data analytics platform brings positive change to all agencies that have been granted access. The platform will create the change required within the nation to achieve long-term sustainability through a focused effort on strengthening national security and the utilisation of data analytics for nation building purposes.

References

Amarasinghe, R. & Ranmuthugala, M. E. P. (2020). "Cyber Security in the Modern World: an Analysis of Cyber Security and Legal Framework in Three Asian Countries." Available at: http://ir.kdu.ac.lk/handle/345/2529?show=full

Cincinnati Business Courier. (2020). "Cyberattacks on the rise during the Covid-19 pandemic" Available at: https://www.bizjournals.com/cincinnati/news/2020/06/01/cyberattacks-on-the-rise-during-covid-19.htmlEnterprise Times. (2020). "Cyberattacks increase on the extended enterprise." Available at: https://www.enterprisetimes.co.uk/2020/07/14/cyberattacks-increase-on-the-extended-enterprise/

En, T. B. (2016). Swimming In Sensors, Drowning In Data - Big Data Analytics For Military Intelligence. Pointer - Journal Of The Singapore Armed Forces 42 (1).

Forbes (2020) "Microsoft: COVID-19 Cyber Attacks Peaked In March And Fell Off Quickly" Available at: https://www.forbes.com/sites/leemathews/2020/06/17/microsoft-covid-19-cyber-attacks-peaked-in-march-and-fell-off-quickly/#42297fcec9aa

Hostingtribunal.com. (2020). 77+ Big Data Stats for the Big Future Ahead | Updated 2020. Available at: https://hostingtribunal.com/blog/big-data-stats/

IBM. (2020). "Big data analytics." Available at: https://www.ibm.com/analytics/hadoop/big-data-analytics

IMDA Gov. (2020). Available at: https://www.imda.gov.sg/-/media/Imda/Files/Industry-Development/Infrastructure/Technology/BigData.pdf?la=en

Lynkova, D. (2019). "39+ Big Data Statistics for 2020." Leftronic. Available at: https://leftronic.com/big-data-statistics/

Milenkovic, J. (2020). "30 Eye-Opening Big Data Statistics for 2020: Patterns Are Everywhere." Kommando Tech. Available at: https://kommandotech.com/statistics/big-data-statistics/

Vanguard. (2020). "Maritime cyber-attacks increase by 900% in three years." Available at: https://www.vanguardngr.com/2020/07/maritime-cyber-attacks-increase-by-900-in-three-years/

WHO. (2020). "WHO reports fivefold increase in cyber-attacks, urges vigilance." Available at: https://www.who.int/news-room/detail/23-04-2020-who-reports-fivefold-increase-in-cyber-attacks-urges-vigilance

Author Biographies

Rosharn Amarasinghe is an Entrepreneur and Cyber Security Consultant. He holds an BSc Electrical and Electronic Engineerng from Northumbria University, MSc Strategic Marketing from AEU University and an MBA from Cardiff Metropolitan University and is currently reading for his DBA. He represents eSec Forte Technologies Singapore in the capacity of Director for Sri Lankan operations. He has extensively written on topics related to robotics, cyber security, and national cyber defence initiatives

Madara Ranmuthugala is a social researcher working on the areas of HIV/AIDS, gender-based violence, women's education rights, and use of technology in social issues. A freelance researcher, writer, and editor, she is currently working on her PhD.