Security and Privacy Mini Case 1

Group A

Emmanuel Casalino
Andrew Doan
Enhong Ma
Andriy Trofimov

June 9, 2014

Question 1: The evolution of Big Data analytics raises questions and concerns over the management of Big Data. There are numerous accounts of security related risks when government organisations and companies collect, store, analyze and distribute large volumes of information.

1. **Concentration of Information Assets:** By its very name 'Big Data' suggests that more data is being collected, managed, and analyzed to support business / organizational objectives. It can be used to establish sensitive intellectual capital, build competitive advantages, etc. The protection of sensitive data applies to normal operations, but the sheer volume / variety of data increases security concerns in a Big Data environment.
2. **Open Source:** The reliance on open source applications / frameworks is consistent with a Big Data environment. Open Source has key benefits such as tapping in the skills and experience of developers across the globe, rapid application development / enhancements, etc. But, it also has some downsides such as the need to lean on internal resources to support application / data security - there is no software vendor there to develop patches, warn clients of emerging / real threats, etc. Normal operations can also rely on Open Source applications / software, but HADOOP - is founded on open source making the concern of particular relevance in a Big Data environment.
3. **Amount of Data:** The perception that hackers lack the capability to "steal" petabytes of big-data stores and the analytical tools in order to extract meaningful information perpetuates the belief that Big Data in itself has its own, inherent security measures. As a result, the need for additional security measures may be overlooked. This security risk is more specific to Big Data than normal application and operations.
4. **Uncategorized Risk :** The cost associated with a security breach, loss of corporate trust and brand image is not easily calculable. It is difficult to thoroughly evaluate security coverage resulting from the inability to assess security weaknesses - hence it is difficult to determine how much to invest in security protection.. This often leads to uncategorized risks and security vulnerabilities that are difficult to prevent and manage. While this security risk primarily affects Big Data, it still remains a security risk for normal application and operations.
5. **Lack of Technical Support:** Big data that uses the Hadoop platform tend to suffer from the lack of robust end-to-end technical support e.g., on and between data nodes. Authentication / authorization policy enablement may be limited, posing a significant security risk due to associated exposures.. This security risk is more specific towards Big Data - as normal operations tend to have more robust enterprise-wide controls.
6. **Security Breach Identification / Logging:** Organizations that are compromised through security breaches tend to have difficulties identifying when the threat had occurred. Whether the threat is discovered within days (27%), weeks (24%), months (39%) or years (9%), the amount of harvestable data can be substantial and immeasurably valuable. IT decision makers and professionals sometimes fail to respond to threats or lack the knowledge of how to properly respond to a threat. The longer a team requires responding to a threat, the more damage that is done the corporation. This security risk affects normal applications and operations, however, the relative size and impact of breaches is likely to be larger in a Big Data environment.
7. **Poor Training and Supervision:** Data breaches can occur from within the organization, whether inadvertently or maliciously. System and database administrators may abuse their permissions. Their willingness to execute malicious error or fraud will immensely affect internal users and the systems' functionality. In contrast, lack of training and/or ignorance may increase the system's susceptibility to external attacks through inadvertent errors, cause internal disruption. This security risk affects normal applications and operations as well.
8. **Nascent Big Data COTS Security Solutions:** There are currently no off-the-shelf or standardized solutions that can address all the challenges that Big Data poses. As such, traditional methods of security and rules of engagement are not applicable and thus customized tools are necessary. Currently, this security risk is more specific to Big Data than normal application and operations.
9. **Increasing Level of Criminal Activity:** The sheer volume of threats has grown exponentially over the past decades. With the increase in the availability of hacking tools to the public and the financial incentives associated with

acquiring Big Data, threat levels to security have reached an all-time high. This security risk affects normal applications and operations as well.

10. **Traceability:** The shift towards IPv6 from IPv4 will expand the availability of unique IP addresses from 4 billion to an almost unlimited amount. As a result, it is easier for cybercriminals and hackers to hide their location and thus any means of traceability. Furthermore, due to the infancy of Big Data, which utilizes historical data (or lack thereof) to predict future attacks, it is easier for hackers to exploit this security without being identified. The ability to predict cyber-attacks is specific to Big Data. However, the challenges of tracing cyber attackers due to the expansion of IP addresses affect normal applications and operations.

Question 2: The evolution of information technology or "Big Data" has resulted in the exponential growth of new digital data. Data can be collected, stored and analyzed at a lower cost, paving new avenues for innovation. However, this evolution has raised various new, unintended, and unplanned privacy concerns.

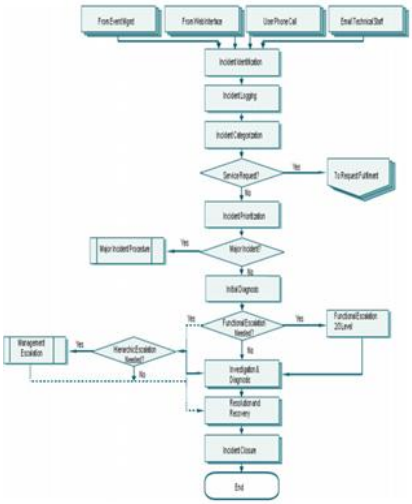
1. **Preserving Anonymity:** There are limitations associated with maintaining the anonymity of Big Data. As more information is collected, the easier it is to connect that data to an individual(s). There has been past successes in which computer scientists have leveraged non-personally identifiable information to reconstitute the person's identity. While pseudonymous are employed to help mask identity, Big Data is can make anonymity more difficult to preserve. This risk also affects normal and operational applications.
2. **Protecting Personal Preferences:** One the key benefits associated with Big Data is the ability to see and reveal previously unrecognized patterns - in customer behaviour. In certain instances individuals are agnostic as to whether such behaviour analysis is being undertaken, in other instances individuals may wish to keep such preferences private. Big Data can be a challenge to the latter. This privacy risk is more pronounced in Big Data, but can also be a risk in normal operations.
3. **Customer Segmentation:** Profiling can result in discrimination based upon age, sex, ethnicity, etc, creating undesired segmentation - typically from the individuals' perspective.. Individuals could lose out on important opportunities and/or offers due to segmentation. Since this is done "invisibly," most people are unaware of these practices. This is more pronounced in Big Data, due to enhanced information to support segmentation.
4. **Legal Requirements:** Many organization operate in multiple jurisdictions - creating a patchwork of complex privacy legislative frameworks under which it must function. At times, legislations may be contradictory - and can lead to breaches, fines, reputational harm, etc. if an organization is not able to effectively navigate in such an environment. This risk also applies normal operations / applications..
5. **Newness of the Process:** The development of new processes and how Big Data is being captured, stored and executed raises concerns on whether or not privacy measures have been consistently implemented across all entities that are controlling / managing data. Often times, gaps in privacy are not discovered and covered by the procedures and software. This is also applies to normal operations / applications.
6. **Concentration of the Data:** A principle of privacy is to limit collection of data - avoiding unintended or malicious breaches. Big Data could be seen as the antithesis of this - and increasing risks vis-a-vis privacy breaches. Due to the amount of data collection and management in a Big Data environment this risk is more pronounced in such an environment, but it does also apply to normal operations / applications. .
7. **Lack of Online Transparency:** In order for Big Data to function within the realms of ethicality, the data owners (i.e. data collectors) need to provide transparency of how the data is being used or sold. Embedded software or browsing "cookies," which track user's activity beyond their website, collect information of the user without the knowingly consent of the end user. As these cookies store personal information through form information and ad tracking, the collector can profile the end user. In turn, targeted ads ensue or the selling of personal data. This is more specific towards Big Data.
8. **Open Source Software:** The reliance on open source applications / frameworks is consistent with a Big Data environment - and despite significant benefits it has some downsides such as some privacy exposures. Open Source applications are open to all to review, examine - readily allowing for privacy vulnerabilities to be seen and exploited. Normal operations can also rely on Open Source applications / software, but HADOOP - is founded on open source making the concern of particular relevance in a Big Data environment.
9. **Government Intervention and Legislation:** The unknowingly collection of Big Data by Government organizations (i.e. NSA), whether warranted or unwarranted (illegal), have led to huge ramifications concerning our rights to privacy. The perception of protecting the country's security and the ideology of "the greater good: raises concerns over the ethicality of such initiatives. Corporations as well may be obligated to relinquish their

customers information to government bodies or be forced to gather information of their customers despite a lack of corporate need. This is specific to Big Data.

10. **Social Media Data Mining:** Hundreds of companies are using social media APIs dynamically analyze user's identity information, profile, and photo information and more without owners' consent. In turn, this data may be used and sold to advertisers for direct marketing. This is specific for Big Data.

Question 3: The case for organization operation to have a mature policies and procedures to effectively identify, manage, resolve and mitigate privacy and security incidents / breaches are certainly enhanced in a 'Big Data' environment, when compared to normal operations - considering the potential increases to Risk exposures due to:

- causes / sources of incidents
- variety of real / potential incidents
- impact of incidents
- cost / time to recover



The image represents a typical incident management process as described within ITIL which may also apply to address security and/or privacy breaches. In an event that an organization has suffered a major data breach, 10 important activities to be undertaken include:

1. **Incident Identification:** The process begins when a breach has occurred, with recognition to the importance of early detection to help mitigate downstream impacts including the use of automated tools.
2. **Incident Logging:** Details (time, date, and source) and description of the breach are entered within the organization's incident management system.
3. **Incident Categorization:** Identify which service(s) within the organization is impacted by the breach
4. **Incident Prioritization:** Takes into account the urgency of the incident (how quickly a resolution is needed) and the level of impact (financial, reputational, etc.) and assigns a priority to the breach .Different response times and levels of organization attention vary based on the priority assigned.
5. **Incident Diagnosis / Analysis:** Identify and assess root cause(s) of the breach; scope of impact; affected system(s) / records.
6. **Incident Resolution:** After the breach has been properly triaged you begin the process of resolving the incident. Research is undertaken as how best to bring the matter to a close, looking at past events, and other knowledge repositories. Proposed actions are documented, and undertaken. Often, the matter is not immediately solved and several iterations of this step are necessary. Or, if it cannot be resolved it may need to be escalated.
7. **Incident Escalation: I**n the event that first lines of support cannot resolve the incident it is escalated within the organization for resolution. Higher priority issues have predefined timelines for when escalations must occur.
8. **Incident Recovery:** Taking steps to restore normal operations, as before the breach. Ensuring SLA are restored, response times are back to normal, etc.
9. **Incident Closure:** Once the breach has been resolved, key stakeholders are notified that the cause of the breach has been closed, and a description of results of any mitigation steps taken.
10. **Lesson Learned:** It is important that a critical review of : what caused the breach, how it was managed, how communications were received, efficacy of mitigation steps, improvements to response times, etc. be examined in support of continuous improvements and to prevent any similar breaches from occurring again.

Question 4: In a traditional / normal operations Privacy Impact Assessment(PIA) and Threat Risk Assessment(TRA) are effective tools at understanding the environment in which the operation exists, the types and possible sources of risk, impacts and risks - and means of mitigating risk are typically focused on the systems and information available within the organization's own 'walls'. Big Data can help fundamentally change how CIOs look at enhancing the organization's security and privacy posture. Consider the following, the 'art of the possible' if you will:

1. **Tapping into the Blogosphere:** There is an entire world of individuals monitoring, debating, and discussing new and emerging threats- using blogs as their medium of choice. Organizations could surveillance such blogs in an attempt to 'getting ahead of the problem' - taking preventative and decisive action much earlier than otherwise possible.
2. **Workflow Management:** It is generally recognized that the majority of privacy and security threats come from within the organization itself. And, the majority of inside breaches are accidental in nature e.g., mistakenly attaching personal information in an email. Big Data could be used to establish patterns of behaviour, identify potential vulnerabilities, and traditional workflow and alert individuals that they are about to: send a document that i*s classified as sensiti*ve to a person outside of the organization; send an email that contains information that contain*s PI tagged data*; or copying a person who has a role that is not permitted to view confidential data and / or a person that you have not typically communicated under this context.
3. **Geotagging / Access and Incident Logs:** With respects to internal threat sources, Big Data could also be used to conduct forensic analysis on security / privacy breaches. Employees are generally required to carry identification badges on them at all times, and allow them entry. Badges can include RFID chips that track geo-positions of employees / contractors at all times. In case of breaches, RFID data could be overlaid with incident information, access and login logs, etc. to narrow down who was the potential cause of the breach.
4. **Sentiment Tracking / Risk Management:** Continuing with the theme of mitigating internal risks, sentiment tracking can used to monitor the 'mood' of employees / staff across the organization - by monitoring emails content, email patterns, web traffic, etc. Individuals could be categorized by risk exposure, based on sentiment analysis algorithms with appropriate risk management tactics put into place to control / limit high risk individuals' access to sensitive systems / information,
5. **Risk-Based Shipping and Delivery:** In globalized world, transcontinental shipping is a reality for most multinationals and shipping in many parts of the world has become dangerous for a variety of reasons (e.g., gov't instability, pirating, and terrorism). Big data could be used to complement cost-based analysis with a risk-based approach to optimizing the combination of: shipping firms, modes of transportation, shipping times, etc. with the intent of avoiding unplanned shipping delays, insurance claims, and disruptions to supply-chain.
6. **Leveraging Social Media:** Turning to social media to help identify, counter and protect against new and emerging online or real-world threats. New viruses emerge on a regular and increasing basis - social media could be used as a 'canary in the coal mine' instrument to take preventative measures, and the same sources can be helpful in developing mitigating steps to emerging threats
7. **Turning to Crowdsourcing:** We continue to hear about true tales of how crowdsourcing has been used to solve crimes (i.e. recent kidnapping of a newborn in Trois Rivieres, PQ). The same tools could be used to help protect organizations against theft, loss, harm to brand, etc.
8. **Central Log Analysis:** Use predictive security scans of patterns to proactively block vulnerabilities - real or potential. Predictive scanning can be enhanced through event correlations from different data access points
9. **Big Data enables Analysis of Historical Trends:** By collecting data on a large scale and analyzing historical trends, you would be able identify when an attack started, and what were the steps that the attacker took to get a hold of your systems. Even if you did not detect the original attack in your systems, you can go back. So long-term historical analysis is one advantage.

10. **Efficiency of Queries:** Timely fashion to understand the data instead of carrying out complex queries. Provides the ability of advance packets filtering in firewalls.