Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Risk Assessment and Analysis for Big Data

Dr. Birendra Goswami
ICFAI, Ranchi,
India

Pradip Kumar Chandra
Block Project Manager, WB,
India

Abstract and Dissertation: ***When responsible organizations identify new ways to process data, for example, when launching a new program, product, system or service, they utilize Privacy Impact Assessments (PIA) to conduct a systematic analysis to identify and address privacy issues. Current PIA practice includes detailed frameworks to help privacy professionals understand and quantify privacy risks. Yet accounting for risks is only part of a balanced value equation. Decision-makers must also assess, prioritize, and to the extent possible, quantify a project's benefits in order to understand whether assuming the risk is ethical, fair, legitimate and cost-effective.***

The phenomenon of "Big Data" exacerbates the tension between potential benefits and privacy risks by upping the ante on both sides of the equation. On the one hand, big data unleashes tremendous benefits not only to individuals but also to communities and society at large, including breakthroughs in health research, sustainable development, energy conservation and personalized marketing. On the other hand, big data introduces new privacy and civil liberties concerns including high-tech profiling, automated decision-making, discrimination, and algorithmic inaccuracies or opacities that strain traditional legal protections.

This document offers decision-makers a framework for a reasoned analysis to balance big data benefits against privacy risks. This process of identifying both benefits and risks is grounded in existing law. The Federal Trade Commission weighs benefits to consumers when evaluating the unfairness of business practices under Section 5 of the Federal Trade Commission Act. Similarly, the European Article 29 Data Protection Working Party applied a balancing test in its opinion interpreting the legitimate interest clause of the European Data Protection Directive.4 The White House Office of Science and Technology Policy, which has recently studied the social and technical ramifications of big data, recognized the need to strike an appropriate balance between new opportunities and individual values.

Key Words: quantify privacy risks; algorithmic inaccuracies; Big unfairness, legitimate interest

I. INTRODUCTION

Gaining visibility into and control over the multitude of internal and external risks is one of the top priorities of corporations today. With a recent jump in regulatory mandates and increasingly active shareholders, many organizations have become sensitized to identifying areas of risk in their business, be it financial, operational, IT, brand, or reputation related risk. They are looking to systemically identify, measure, prioritize, and respond to all types of risk in the business, and then manage any exposure based on business strategies and priorities.

Metric Stream Risk Management Solution

Metric Stream provides an integrated and flexible framework for documenting and assessing risks, defining controls, managing assessments and audits, identifying issues, and implementing recommendations and remediation plans. The risk management system also includes powerful risk analysis and monitoring tools such as configurable risk calculators and risk heat maps.

Risk Assessment and Analysis

Metric Stream Risk Management Solution provides a centralized risk framework to document and manage all risks faced by an organization. It supports risk assessment and computations based on configurable methodologies and algorithms, giving a clear view into each organization's risk profile, and enabling managers to prioritize their response strategies for optimal risk/reward outcomes.

Controls Design and Assessments

Once the key risks have been identified and prioritized, Metric Stream Risk Management Solution leverages the COSO framework to help define a set of controls that mitigate those risks. The solution also allows associated policies and procedure documents to be attached for reference. Assessment plans to evaluate and ensure the effectiveness of the controls can be designed and assigned to owners based on roles and responsibilities.

Using the solution, risk officers and process owners across the organization can efficiently manage risk assessment programs to ensure the effectiveness of controls. The system supports assessments based on predefined criteria and checklists, and has a mechanism for scoring, tabulating, and reporting results.

A repository of all assessments, with an easy search capability, ensures that users can check to see if a specific control was tested. They can also view the assessment results, and determine whether or not remedial action is required.

Metric Stream Risk Management Solution provides seamless integration with Metric Stream Internal Audit Management Solution for streamlining the audit management process in the organization. The solution provides the flexibility to manage a variety of audit-related activities, data, and processes to support risk management. It also supports the complete spectrum of audits, including internal audits, operational audits, IT audits, supplier audits, and quality audits.

The solution provides a single framework to manage the complete audit lifecycle, beginning with audit planning and scheduling, and extending to the development of standard audit plans and checklists, field data collection, development of audit reports and recommendations, review of audit recommendations, and implementation of those recommendations.

The solution is equipped with advanced capabilities such as built-in remediation workflows, time tracking, email based notifications and alerts, risk assessment methodologies, and offline functionalities for conducting audits at remote field sites. These capabilities enable organizations to implement best practices for efficient audit execution, and ensure integration of the audit process with the risk and compliance management system.

II. OBJECTIVE OF RESEARCH

For issues arising from the assessment and auditing processes or from external events such as loss events or "near misses," Metric Stream Risk Management Solution provides seamless integration with Metric Stream Issue Management and Remediation Management modules. Once issues are identified, documented, and prioritized, a systematic mechanism of investigation and remediation is triggered by the underlying workflow and collaboration engine.

The solution supports the triggering of automatic alerts and notifications to appropriate personnel for investigation and remedial task assignments. The issues remain open till the action plan is carried out and results have been verified for effectiveness. Managers can track the status of issues as they automatically move from one stage to the next based on the organization's risk management procedures.

Monitoring Risk

Executive dashboards provide enterprise-wide visibility into the risk management process and highlight issues that need to be addressed. The solution also has the ability to track risk profiles, control ownership, assessment plans, remediation status, etc. on graphical charts that can be accessed globally and display real-time information. The ability to drilldown provides an easy way to access the data at finer levels of detail.

In addition to pre-configured standard risk reports, the system provides tremendous flexibility by enabling stakeholders to configure ad-hoc or scheduled reports. These reports can be used to view metrics by a variety of parameters such as by process, by business units, or by status. The system also provides quarterly and monthly trending analyses with the ability to drill down into each report and dashboard to see the underlying details. This enables risk managers and process owners to stay in constant touch with the ground reality and progress on risk management programs. Automated alerts for events such as exceptions and failures eliminate any surprises and make the process predictable.

Big Data wasn't designed with security in mind; in fact, security has never really been the focus of distributed sciences. With these mountains of data, informing businesses on critical buyer decisions, habits, and countless other minutiae, comes a pressing need to keep this valuable information secure and protected. This is sensitive information, after all, and with so much of it comes a greater risk of breaches.

Data volumes are doubling annually, and roughly 80 percent of that captured data is unstructured, and must be formatted using a technology like Hadoop in order to be mineable for information. Considering this growth, it is clear that security concerns won't be going away anytime soon. Quite the opposite, actually.

As Hadoop becomes more widely adopted in the enterprise, its security limitations are becoming more apparent. Brian Christian, co-founder and CTO of secure Big Data management vendor **Zettaset**, explains the biggest Big Data security challenges facing the enterprise today and his thoughts on creating a unified security model for Big Data.

Big Data: Risks and Rewards

Digitization itself is not new, but the maturation and availability of the Internet; the rapid growth of mobile computing; and, more recently, the addition of sensor data (data derived from devices that sense their environment) to the mix have all pushed the boundaries of how we think about data and its uses. The term *big data* represents the need for a new way of thinking but also implies new tools and new ways of managing data. Like many things, data can be used to do positive things for the world, but it can also be used to manipulate, embarrass, or repress. Data can be highly accurate and efficiently structured or unstructured, fragmented, and highly suspect. Data can also be managed well or carelessly. Big data, in its outsized properties, amplifies those effects. It is in those extremes that the risks and rewards of big data are decided.

Three Key Big Data Trends

As the world becomes more familiar with big data, three key trends that have a significant impact on those risks and rewards are emerging. First and foremost, *big data leverages previously untapped data sources.* Those sources are of

several types. The first includes wearable devices that stream data about an individual and his or her surrounding environment on a moment by- moment basis—such sensors include the applications on a smart phone that sense movement. The sensor in a runner's shoe is a very consumer-facing example, but business-facing sensors, which track all kinds of things, are proliferating very quickly. A pacemaker is a sensor that has been around a while (the newer models give feedback to healthcare workers).

The next type comprises connected sensors that instantly digitize and report what is happening in any moment and in any location. Examples of this type include the global positioning system (GPS) device that reports location back to a central computer or a user, and devices in the soil of a farm that sense when and how much to irrigate. There are also sensors in trains, for example, that watch for signals that maintenance is necessary before a human could ever see them, such as brake heat, brake wear, movement in the rails, and so on. This new breed of sensors is coming into service and is connected to the Internet, making big data even bigger than human-generated information.

The third type of sensor provides constant reporting by machines that perform the work critical to our security, health, and lifestyle. Machines can be something as large as an aircraft or locomotive or they can be components of one of those things. Some of the most interesting of these sensors are the ones that measure the way an aircraft engine is performing mid-flight. Machines used to be purely mechanical but are increasingly computer controlled. Those computer controls mean not only that data are constantly being fed into machines but that they are also coming out of machines at a quickly increasing rate. We have reached a point of information discovery
that reveals correlation before causation, leaving researchers scratching their heads to find the underlying causes for correlations that data analysis clearly demonstrates. TIBCO's chief executive officer.

III. METHODOLOGY

Ranadive, is fond of saying that we have reached a point where we may know the "what" without knowing the "why." The previously untapped information sources create a data ecosystem that can be modelled in a way that blends historical with in-the-moment information and is remarkably useful for anticipating the future. These models accurately predict such diverse outcomes as the spread of disease, the failure rate of aircraft components, and consumer behaviours. Big data's effectiveness is tightly coupled to an organization's ability to bring the right data together in the right moments that allow for the right response and outcome. Whatever we may know today, the continued discovery of previously untapped data sources will continue to change and improve our models, allowing us to better anticipate future events and to continue to increase our ability to affect desired outcomes.

The desire to affect outcomes brings about the second trend of big data: *the need for automation technologies.* Richard Hackathorn wrote about the value time curve of information back in 2004 in "Real-Time to Real-Value," just as the world was becoming broadly and acutely aware of the explosion of data.1 Hackathorn's curve describes the decreasing value of data over time as it passes through stages of. The challenge of the decreasing value of data over time has become even more meaningful in the age of big data. Today, the volume, velocity, and variety of data continue to push the curve down and to the right as organizations struggle to capture, analyze, and decide in a gradually more difficult environment. Added to this complexity is the increasing access to real-time data that leaves organizations in some industries attempting to reduce their response time to microseconds, understanding that this is a crucial part of being successful in their business.

The value-time curve challenge makes big data management a function of creating automation wherever possible. Machines have always been humanity's friend in making work more efficient, and big data follows the same path. Big data's growth in each of its dimensions eliminates the ability for humans to intervene and reprogram processes in real time, opening the door for better and better tools that can manage data far more quickly and efficiently than a human can. Data exist in a moment, ready for decision and action, but there is a higher-level purpose for information. Data comprise the digital representation of events, or things that happen in patterns that occur over time, in conjunction with other events or in isolation, and even with things that may be expected but do not occur (such as when a patient fails to pick up a prescription after being discharged from a hospital, starting a likely string of events that will lead to readmission).

The idea of keeping track of what does *not* occur is a level of complexity higher than the old ways of waiting for data to arrive or change. Automation is especially well suited to the complexity of predicting, and then anticipating, events. In many organizations, automation is also a significant part of the actions that events precipitate.

The big data conversation often centres on the use of machines as the best resource for the storage and analytic processing of vast amounts of data, but this is only a piece of the story. Automation is increasingly a logical response to the need to find, filter, and correlate each piece of data as it flows over the enterprise so that decisions can be made—some through automation and some using a hybrid approach combining human and machine. Once decisions are reached, automation becomes the path for taking action in the shortest time frame possible before the value of data decays further.

The third trend being driven by big data is the *necessity for adaptable, less fragile systems.* For big data to leverage previously untapped sources of information, organizations need to quickly adapt to the opportunities and risks represented by these new sources. Automated systems that manage big data ecosystems cannot be developed around rigid schemas that require redevelopment for each new stream of information. Instead, systems need to absorb new information in an adaptable way that also adds value to existing data that have already been collected. Adaptable systems treat new sources of data coming constantly as the means to improve analytical models, create better decisions, and drive more appropriate actions.

IV. DATA ANALYSIS

The ultimate value of a data project's benefits as well as the magnitude of its privacy asks are linked to the risk mitigation strategies that have been implemented.

In many cases, mitigation techniques may impact data utility by reducing the potential benefit. This means that the **Data Benefit Analysis** is a dynamic process, through which mitigation techniques are carefully calibrated to optimize the risk-benefit equation in order to reach the apex point. The OMB calls this exercise "sensitivity analysis," noting that "major assumptions should be varied and net present value and other outcomes recomputed to determine how sensitive outcomes are to changes in the assumptions." Of course, in many cases, a baseline level of protection against risk will mandatory under regulation in order to support the legitimacy of the data processing.

Once an organization has a better understanding of a project's benefits, it can map the **discounted benefit value** against privacy risks identified through a PIA. By doing so, it can now visualize a complete picture to inform decision-making weighing both benefits and risks. By mapping benefits against risks, an organization evaluates the merits of a big data project. To do so, an organization must elucidate where a project falls on the a risk-benefit continuum.

Mapped in this way, a contemplated project is placed on a continuum ranging from projects that the FTC and the Article 29 Working Party may view as unfair to project that the regulators regard as being within the legitimate interest of the organization or the public at large.

While some of the assessments proposed in this framework can be standardized and quantified, others require value judgments and input from experts other than privacy professionals or data regulators. For example, assessing the scientific likelihood of capturing a benefit in a specialized area cannot be made solely based on privacy expertise.

Furthermore, this framework cannot achieve mathematical accuracy given the inherent degree of subjectivity in assessing the relative merits of various benefits. However, this has not stopped policymakers in other arenas from proposing structured processes to measure project benefits against risks. For example, the OMB states, "Although net present value is not always computable … efforts to measure it can produce useful insights even when the monetary values of some benefits or costs cannot be determined."

This highlights the importance of determining *who* will be tasked with undertaking the Data Benefit Analysis. Moving forward, organizations will need to create or expand accountable data ethics review processes to engender trust and address privacy concerns. Many companies have already laid the groundwork to address these decision-making challenges by appointing Chief Privacy Officers or building internal ethical review programs. Further efforts are needed to understand the most effective structures for different organizations and different types of data. Models may range from a formal Institutional Review Board-type process to empowering Chief Privacy Officers through cross-functioning privacy committees, or involve building structures such as external advisory boards or opportunities for policy maker or regulator input.

V. CONCLUSIONS AND IMPLICATION

A common element in most security best practices is the need for the support of senior management, but few documents clarify how that support is to be given. This may represent the biggest challenge for the organization's ongoing security initiatives, as it addresses or prioritizes its risks.

An information security framework is important because it provides a road map for the implementation, evaluation and improvement of information security practices. As an organization implements its framework, it will be able to articulate goals and drive ownership of them, evaluate the security of information over time, and determine the need for additional measures.

Specifically, an enterprise security risk assessment is intended to be suitable for the following, which could be specific to any organization:

- A process framework for the implementation and management of controls to ensure that the specific security objectives of an organization are met
- A definition of new information security management processes
- Use by management to determine the status of information security management activities
- Use by internal and external auditors to determine the degree of compliance with the policies, directives and standards adopted by the organization
- For implementation of business-enabling information security
- To provide relevant information about information security to customers

Overall, an organization must have a solid base for its information security framework. The risks and vulnerabilities to the organization will change over time; however, if the organization continues to follow its framework, it will be in a good position to address any new risks and/or vulnerabilities that arise.

REFERENCES

[1] http://www.futureofprivacy.org/wp-content/uploads/FPF_DataBenefitAnalysis_FINAL.pdf
[2] http://www3.weforum.org/docs/GITR/2014/GITR_Chapter1.5_2014.pdf
[3] http://www.ey.com/Publication/vwLUAssets/EY-Global-Forensic-Data-Analytics-Survey-2014/$FILE/EY-Global-Forensic-Data-Analytics-Survey-2014.pdf
[4] http://www.cpni.gov.uk/advice/Personnel-security1/risk-assessment/
[5] https://www.google.co.in/?gfe_rd=cr&ei=rMvqVaDxDLPG8Afdx7mIDA&gws_rd=ssl#q=security+risk+assessment

