International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-2, Issue-6, January 2014

Data Privacy in Banking and Insurance Sector: A Case Study Approach

Debaswapna Mishra

Abstract- Now a Days business and their customer's alike collect, store and transmit vast amount of information electronically and they want to believe that this information is secure. At the customer level, the concern for data privacy has resulted in a growing number of laws and regulations that address issues including what information can be collected and maintained, how the information should be stored, how and where information can be transmitted, and required actions in the event of security breach. Notwithstanding the proliferation of requirements, reports of identity theft, inadvertent release of customer and proprietary business information, and successful attempts by hackers to penetrate systems and steal information continue to command headlines in media. This is more acute in banking and financial sector where a lot of financial transactions occur on a day-to-day basis. This study aims at providing robust solutions to address various data privacy issues, particularly in Banking and Insurance sector. The current set of algorithms to handle data privacy appears to be inadequate while the need of the hour is to provide a solution that encompasses a host of algorithms and houses a set of business rules. A good privacy solution enables de-identification of confidential data that can be readily used in business software development and testing, maintenance, statistical analysis, marketing research, training and other outsourcing solutions.

Keywords: Notwithstanding the proliferation of requirements

I.INTRODUCTION

1. The Business Case -1

A leading international bank decided to outsource its testing activities as part of a cost management exercise. However, as per the provisions of the European Union Data Privacy Act, the bank was required to mask sensitive client data before sharing it with an external partner. [1]

The partner had to devise a robust solution to mask confidential client data before sharing it with testing teams.

The data masking exercise was complex and posed the following challenges:

- Ensuring data integrity and consistency across heterogeneous systems
- Complicated file feeds arriving at different times in a dynamic environment

The Business Case – 2

- Multiple TIBCO XML messages arriving simultaneously with limited queue storage requiring 24/7 masking
- Multiple programming languages and operating systems
- Multiple databases with over 1,000 tables and over 1,200 views
- About a dozen upstream and 4-5 downstream applications
- Over a dozen type of real-time TIBCO messages on queue and about 50 file feeds
- One of the world's largest insurance companies headquartered in the U.S.A. One of the largest global providers of insurance, employee benefits and annuities operating in over 50 countries. [38]
- There are close to 800 business applications that the company uses across the globe for its business functions in a variety of technologies.
- Presently a significant amount of sensitive data exists (like Name, SSN) in non-production region, which contributes to significant risk

Objective:

- Enterprise-wide initiative to reduce risk across non-production environments.
- Improve the protection of the customer data by not using it in non-production environment
- Reduce the probability of a data breach and the associated financial and customer loyalty implications.
- Protect sensitive data in non-production environments from Internal and external threats.

Mitigate risk of 300 applications across all Line of Businesses as per Senior Management commitment to the Company Board by de-sensitizing the data in the non production regions

Problem Space:

- There need is to mask the data in non-production environment which are primarily used for development, testing and support purposes. Considering that this has to been done across the application, there were multiple challenges

Technical Challenges:

- Data quality issues within the applications needed a solution that could accommodate and perform
- Limited application knowledge especially in case of product platforms or old legacy system which no longer had any development required
- Moving and masking large piece of data required due design consideration on performance and load aspect
- The data masking algorithm needed to be consistent and irreversible, so that it is robust and also able to address dependencies within and across applications
- The solution needed to address both data at rest and the inter-application data interfaces.
- Limited documentation on the flow of the data/interfaces between the application and subsystem making it difficult to

Revised Manuscript Received on 30 January 2014.
* Correspondence Author
Debaswapna Mishra*, University of Agricultural and Technology, Bhubaneswar, India.

assess impact on the integration testing

Non-technical challenges

- There needed a culture change within the AD and testing teams of the organization to adapt to data masking – This needed huge co-ordination effort
- Strong resistance to change from the application teams due to impact to their existing processes in the areas of application development, testing, data security and data management. Hence the execution had to be planned and executed with utmost caution and accuracy.
- Data masking being a whole new paradigm for the organization, communication efforts were expected to be extensive
- With the data masking service functioning as a horizontal 'across' all applications of the organization, standards and best practices in development, implementation, documentation and other areas needed extensive planning.

a. Data Privacy Solution for this business case

i. Solution Objectives

- To conceptualize, plan, design and implement a company-wide data masking program to be run as a managed services model, as part of the overall information risk reduction initiative.
- Develop a factory model to enable expedited masking of the applications.
- Implementation partner to mask data in 200+ application across the enterprise.
- Partner with client in bringing in transformation to adapt to data privacy principles

ii. Details about solution offering

- Considering the scale of the program, a data masking factory model concept needs to be adopted for the implementation
- A proof of concept needs to be carried out to establish the technology, process and project execution model
- Conceptualize and implement a centralized Data Masking Services for planning, executing, controlling and maintaining the solution
- Foundational modules and frameworks needs to be built and tested before executing the factory model
- Need to develop Business As Usual (BAU) process and guidelines to ensure that the Data continues to remain masked after the application is processed through the factory

Requirements analysis

The requirement analysis is primarily divided into 2 sections:

1. Default Data Masking rule definition: This is primarily a part of discover phase and then refined further into the execution phase.
2. The requirements of each sensitive data element needs to be analyzed in detailed to understand data anomalies, data dependency needs and data consistency. For e.g. only last four digits of SSN should be masked as the first few digits have business rules associated to it. Special characters like -,etc. must retain its position as they have significance in logic.

   A detailed Design cookbook needs to be established which will be used by all the applications for implementation
2. Requirement gathering using Standard Template: Application teams are expected to provide the specific data element and their refresh methods in predefined template. The analysis involves understand the specifics of data, discover any requirement inconsistency or specific business rules

Methodology

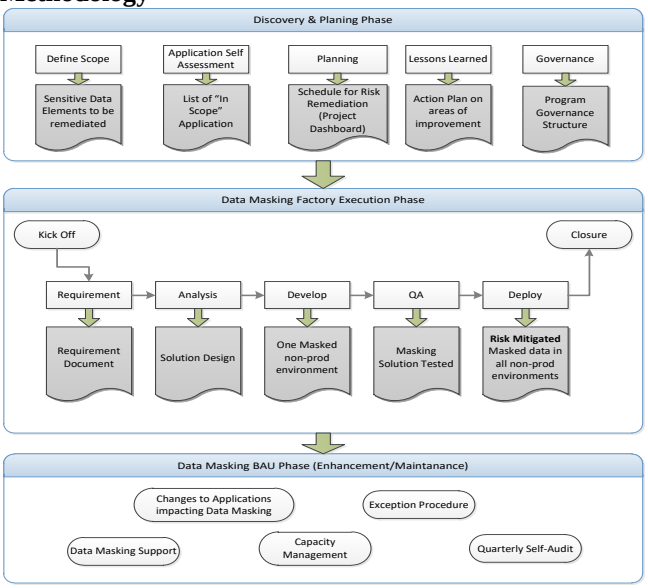

Figure 1.

The overall project model can be defined over into three sections as illustrated in above figure

Discovery Phase:

The primary purpose of the discovery phase was to assess the overall risk exposed due to sensitive data in non-production environments. The application teams are required to fill a self-Assessment questionnaire which would determine if the application needs to be remediated to mitigate the risk. Also, identify if there are already any risk mitigation in place. This phase would help finalize the high level scope for Risk Mitigation

The Discovery phase also determines the technology and the solution to be used for Risk Mitigation. This phase was used to evaluate technology, tools and methodology to be used for the next phases

Data Masking Factory Execution Phase: The execution phase was established after the Discovery phase and it has been decided to use a Factory approach for Data masking. The Factory approach defines series of steps, tools and processes to be used to mask each application.

Data Masking BAU Phase: Once the application has been completely masked and the ongoing solution built, the application will be moved into Maintenance phase. The processes and procedure for Maintenance and enhancement will be outlined in the Release document for that application. A standard process document has been developed for the maintenance and enhancement of Data Masking routines

Testing:

Unit Testing: Testing was automated to check if the data was masked.

Downstream Testing: The application teams executed the integrated testing to evaluate the impact their downstream and upstream

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Delivery:

The delivery constituted completion of 2 activities

- Mask the sensitive data in all the non-production environments
- Build a solution to keep the application masked for ongoing data refresh

Key benefits

- Reduce the Risk exposure by preserving client data confidentiality
- Successfully masked more than 100+ applications in one year timeframe
- Reduction in Data Masking cost due to reusability
- Single solution addressing wide range of technologies
- Progressive benefits in terms of cost reduction and masking throughput due to the factory approach
- Establish the foundation to enable masking of International application
- Establish a program model to be used for complex enterprise programs

Key learnings

- Planning and dependency analysis plays a vital part in a large scale program such as this
- Communication and user orientation towards data masking needs substantial efforts. Such programs need to be designed as 'Over-communicated' from the beginning.
- It is very important to understand the dependency of application functionality on data, and tailor the masking solution accordingly

Solution Limitations

- Use of third party products like IBM Optim did limit the quick resolution for Data Masking, this was overcome by developing custom routine
- Application data quality needs to be given special consideration importance while developing masking algorithms. At the same 'correcting' data quality it may not be necessary for the success of the data masking program.

Generic Challenges and need for a robust solution

Enterprise production environment data contains sensitive personal information that must to be protected from data thefts and frauds. Data privacy regulations have made it mandatory that private data must be managed in a secure manner and should not be exposed to unauthorized users. At the same time, legitimate business activities like application development and testing require realistic data for successful testing. Information systems also need to hide confidential data from less privileged users for outsourcing needs. Enterprises face a daunting task of balancing out the need of protecting confidential data and at the same time making it available to entities that need them.

A good data making technology enables protection of sensitive data. So there is a need for a powerful data privacy solution that helps companies to automate masking of personally identifiable information to prevent its abuse and improve compliance.

Challenges and Opportunities

More than ever, intangible such as customers, systems and information form the foundation on which corporate value is built. In today's highly competitive business environment, organizations are placing increased emphasis on customer relationship management as a means of gaining market share and differentiating service quality. Customers want choices and ease of access, which requires them to provide personal information and preferences; business want to be able to gather, data mine and share this information efficiently.

Certain industries, such as financial services and health care, often draw the most attention in the privacy discussion because of the personal information they possess. However, all industries are affected by privacy and data protection requirements.

Protecting sensitive information is, in fact, one objective that governments and businesses share. Many industry associations, such as the payment Card Industry (PCI), the Healthcare information trust alliance (HITRUST) Telecommunications Service Companies Privacy Regulation (Germany), Information Commissioners Office (United Kingdom), and Privacy and electronic Communications Regulations (United Kingdom), have issued their own standards to supplement existing laws and regulations.

Data Privacy has Serious Consequences

Information is the oxygen of organizations. Enterprises need to store and process private information in their production environment databases. This information can coexist of employee, customer, partner and vendor records containing sensitive details like names of individuals, addresses, telephones, emails, social security numbers, credit card information, health insurance details, financial records etc.

Individuals whose data is maintained by the companies expect that their personally identifiable information (PII) will be protected using proper measures. This data is required to be shared inside and outside the organizational periphery for legitimate purposes like business application development, testing, training, statistical analysis, business process outsourcing and market research. Such a shared data is vulnerable to data privacy violations if appropriate measures are not taken.

In case a sensitive data leak materializes and results in exploitation, the ramifications for the organization can be disastrous, often resulting in lawsuits, loss of customer confidence, brand damage, erosion of share price, bad press and loss of revenue or in worst case a possibility of business closure.

Application Test Data: A Major source of Data Thefts

Businesses need applications to function and these applications require maintenance and testing. New applications get developed as companies grow. These activities are managed by internal development team or external contractors or outsourcing vendors.

Application development and testing activities need realistic data for validations. Usually, copies of production environment databases are created using internally developed scripts and given to development teams. However, this method is risky since real data with sensitive information could fall in wrong hands. This also applies to analysis or trainers who need such a data for their work.

This increases the chances of data getting stolen. For example, whole testing an online banking system, application tester can update customer records and as a result can view names, addresses, social security numbers and other private information of individuals.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

It is possible that this tester can steal this data without anyone being aware. This risk is further elevated due to the fact that just a single copy of master data is not sufficient for application development.

Business applications run in multiple environments and on multiple platforms. To achieve productivity, development teams need separate copies for each platform to perform the work in parallel by separate teams. This means there would be multiple copies of production database instead of just one. As infrastructure security technologies have matured, it has become quite common to implement well known and standard security measures like access permissions, setting up firewalls, enabling database native security mechanisms, conducting audits and deployed automated monitoring tools along with physical security like access cards to protect production environment. However, such measures are absent in application development environments. There is a lack of audits or monitoring of application testing data. This can result in master data copies being abused. For example an external contractor could misuse stolen retail customer data or an employee might store such data on a detachable hard disk that gets stolen.

The application development data without adequate security measures can also be stolen by a hacker if an entry inside company infrastructure is achieved.

Complying Privacy Laws

Worldwide, there have been numerous data privacy laws introduced by the governments to make sure that private data is responsible handled and proper care is taken to avoid leaking of this data. These laws have come into existence as a result of many well published data frauds by the hackers and concerned consumers pressurizing governments to take steps to keep their personal information safe. These legislations aim to improve data privacy protection and compliance to safety requirements.

The Payment Card Industry Data Security Standard (PCI DSS) act makes it mandatory for credit card payment processing companies to maintain data confidentiality while stroring processing and exchanging credit card data. Its section 6.3.4 specifies that production data should not be used for testing or development.

This allows companies to focus on core competencies, lower the cost of operations, tap remove talent and reduce geopolitical risks by diversification.

Many companies would like to take full advantage of outsourcing opportunities but they are afraid to do so as a result of data privacy concerns. For example, it is possible that an employee of an outsourced organization might be able to view sensitive details though information systems while performing duties which can result in data theft and fraud.

- Data from numerous highly integrated systems needed to be worked upon
- Lack of uniformity in frequency of data refresh from different sources
- Presence of heterogeneous platforms, including Windows, UNIX, Mainframe and Linux
- Need for data integrity across all type of applications
- It has to deal with complex Facets integrated environment with multiple upstream (20+) and multiple (20+) downstream applications

i.Data Masking is the Best Alternative

Protecting privacy is an important ingredient of overall enterprise data security plan. It improves customer trust in the brand. The challenge is to provide an appropriate level of protection and at the same time meet business objectives, thus ensuring that data is made available on a need-to-know basis. There are two approaches to solve this problem. First option is to create synthetic test data using a software tool. This can be done quickly and works well for simple test cases. However, typical application backend data has complex relationships and consists of complicated scenarios. This makes synthetic data creation a complicated task requiring significant manual efforts and a deeper understanding of application internal and business domain. Besides, this approach addresses only a part of the problem and does not obfuscate data at user interface points of business applications which is an important need for outsourcing activities. The other alternative is to implement data masking. You can either develop a generalized masking solution or purchase an off-the-shelf solution. Developing an in-house system can take months and will require maintenance efforts and specialist staff. An out-of-box product can be customized and deployed in a much shorter time and at a much lesser cost.

ii.Definition of Data Masking

Data masking is the technique of modifying sensitive data to create life-like but false values using masking software. Its effectiveness depends on whether you can create source values from the target or not. Data masking retains look and feel of data so that correct testing is possible and also less privileged users can continue to perform their work with rest of the unmasked data.

Masking is also known by other names like data obfuscation, data de-identification, data depersonalization, data scrubbing or data scrambling.

iii.Dynamic Masking: The key differentiator

Most of the masking product vendors provide static masking which scrambles data in production environment before handing over to software testers. However, the key differentiator for a masking solution is the dynamic masking capability also known as on-the-fly masking.

Dynamic masking is critical for the privacy initiative and without it, achieving objectives of data protection are almost impossible.

Data masking techniques

1. Substitution

Substitution is one of the most effective methods of applying data masking and being able to preserve the authentic look and feel of the data records. [2]

It allows the masking to be performed in such a manner that another authentic looking value can be substituted for the existing value. There are several data field types where this approach provides optimal benefit in disguising the overall data sub set as to whether or not it is a masked data set. For example, if dealing with source data which contains customer records, real life surname or first name can be randomly substituted from a supplied or customized look up file.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

If the first pass of the substitution allows for applying a male first name to all first names, then the second pass would need to allow for applying a female first name to all first names where gender equals "F". Using this approach we could easily maintain the gender mix within the data structure, apply anonymity to the data records but also maintain a realistic looking database which could not easily be identified as a database consisting of masked data.

This substitution method needs to be applied for many of the fields that are in DB structures across the world, such as telephone numbers, zip codes and postcodes, as well as credit card numbers and other card type numbers like Social Security numbers and Medicare numbers where these numbers actually need to conform to a checksum test of the Luhn algorithm.

In most cases the substitution files will need to be fairly extensive so having large substitution datasets as well the http://youtu.be/P1H6iODKUsYability to apply customized data substitution sets should be a key element of the evaluation criteria for any data masking solution.

2. *Shuffling*

The shuffling method is a very common form of data obfuscation. It is similar to the substitution method but it derives the substitution set from the same column of data that is being masked. In very simple terms, the data is randomly shuffled within the column. However if used in isolation, anyone with any knowledge of the original data can then apply a "What If" scenario to the data set and then piece back together a real identity. The shuffling method is also open be reversed if the shuffling algorithm can be deciphered. Shuffling however is a great technique to include in your overall masking approach as it has some real strengths in certain areas. If for instance, you need to maintain the end of year figures for your financial information in that test data base. You could mask the names of the suppliers and then shuffle the value of your accounts throughout your masked database. It is highly unlikely that anyone, even someone with intimate knowledge of the original data could derive a true data record back to its original values.

3. *Number and date variance*

The numeric variance method is very useful for applying to financial and date driven information fields. Effectively, a method utilizing this manner of masking can still leave a meaningful range in a financial data set such as payroll. If the variance applied is around +/- 10% then it is still a very meaningful data set in terms of the ranges of salaries that are paid to the recipients.

The same also applies to the date information. If the overall data set needs to retain demographic and actuarial data integrity then applying a random numeric variance of +/- 120 days to date fields would preserve the date distribution but still prevent traceability back to a known entity based on their known actual date or birth or a known date value of whatever record is being masked.

4. *Encryption*

Encryption is often the most complex approach to solving the data masking problem. The encryption algorithm often requires that a "key" be applied to view the data based on user rights. This often sounds like the best solution but in practice the key may then been given out to personnel without the proper rights to view the data and this then defeats the purpose of the masking exercise. Old databases may then be copied with the original credentials of the supplied key and the same uncontrolled problem lives on.

3. *Solution Approach*

Organizations seeking to build and maintain an effective and compliant privacy and data protection program should undertake the following activities:

- Conduct a comprehensive risk assessment that, among other considerations, identifies the nature of information collected, where it is stored, how and where it is transmitted, and the laws, regulations and standards that govern handling of the information. It is important to note that the risk assessment process can be especially challenging for large global organizations that have a multitude of systems and must consider the impact of myriad, and sometimes conflicting, laws and regulations.
- Establish privacy and data protection policies that are monitored and enforced continuously within the organization.
- Ensure there is clear organizational accountability for privacy and data protection, as well as strong coordination among key players – compliance, information technology, security, business lines and internal audit, among other process owners.

i.4. *Key Features*

Following figure illustrates various components of iMask

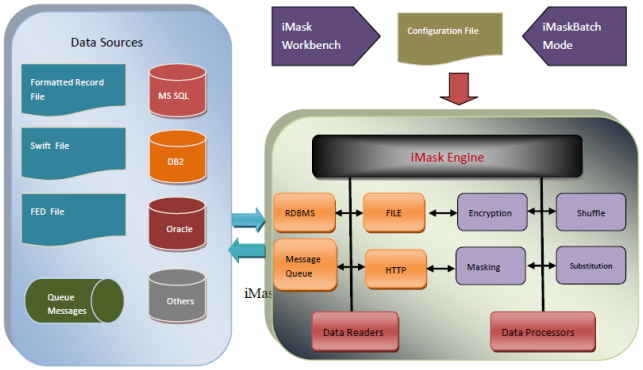

Figure 2. iMask: High Level Architectural view

Organizations are increasingly becoming aware of data privacy needs and challenges. They are in need of a right masking solution.

Data masking must be an integral part of enterprise data security initiative to avoid costly downsides. A good solution is critical for the successful implementation of ever changing masking requirements.

iMask, with its key differentiator as dynamic masking is a proven masking product that can be quickly deployed to successfully automate complex and diverse data privacy and protection needs of an enterprise.Incremental Masking: Configure the product to mask only the newly added records to save time and resources.

Dynamic Masking: Mask the production data on-the-fly as it moves from secure production environment to business applications residing locally or remotely. Mask the data based on location and role of the business application user.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Platform Independence: Run the product on multiple platforms enabling you to perform masking for entire organizational needs.
Selective Masking: Mask only the subset of data using regular expressions for speed and efficiency.
Multiple Algorithm Support: Multiple transformations to cater wide ranging needs. These include shuffling substitution, masking, random value, encryption and number variance.
Workbench: Powerful graphical workbench to model masking configurations containing data sources and rules.
Key Features of this solution

Figure 3. Key Features

Ensured Data Protection: iMask provides de-identified data that can be safely used in production, test and maintenance situations to ensure protection of sensitive data. Maintaining data privacy avoids costly consequences of data breaches and frauds.
Consistency and Standardization: iMask follows a consistent and standard method of masking across all databases in the enterprise. This results in uniform appearance of de-identified data.
Legal Compliance: using scrubbed data helps you comply with privacy regulations and avoid penalties and other serious consequences.
Cost Savings: iMask can be quickly deployed to automate masking of confidential data. Its user friendly features assist in achieving high productivity that allows to free resources for other business purposes thus saving costs.
We have compared the expected results from iMask solution with other products existing in data masking space. There is a considerable improvement in both response time and accuracy as illustrated in below figures.

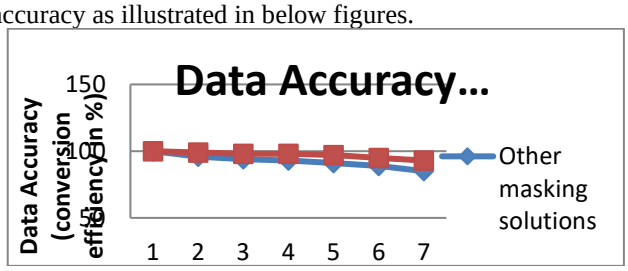

Figure 4. Data Accuracy improvements

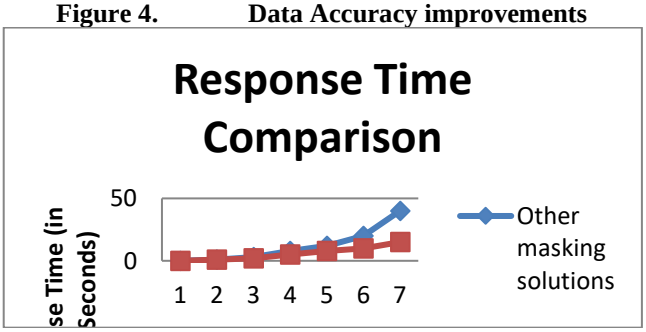

Figure 5. Response Time Improvements

II.CONCLUSION

Organizations are increasingly becoming aware of data privacy needs and challenges. They are in need of a right masking solution. This study offers the current limitations in the banking and insurance sector and comes up with a robust solution to address those weaknesses.
Data masking must be an integral part of enterprise data security initiative to avoid costly downsides. A good solution is critical for the successful implementation of ever changing masking requirements. iMask, with its key differentiator as dynamic masking would be a proven masking product that can be quickly deployed to successfully automate complex and diverse data privacy and protection needs of an enterprise.
It will address the following deficiencies in other products:

- Address Test Data Management Features, Data Anonymization and Data subseting
- Support wide variety of Data Sources and domain specific entries
- Flexible and extensible to support specific needs
- Rich features to access Data Privatization Implementations

✓ Data Discovery
✓ Multiple pre-defined algorithms that support various kinds of data obfuscation
✓ Data Subsetting
✓ Flexible deployment and execution options

- Centralized Configuration and Management of the data Anonymization rules
- Support for packaged applications

It is highly customizable and configurable. It has a rich set of transformation techniques, including encryption, hashing, substitution, shuffling, lookup transformation, number masking, data masking, character masking, string translation and masking special fields like SSN, credit card and CUSIP. Loaded with deterministic, selective, dynamic and static masking tools, it can be deployed on any platform and supports all major databases as well as multiple operating systems.

REFERENCES

1. ‘‘Estimates Put T.J. Maxx Security Fiasco At $4.5 Billion,’’ Information Week, May 2, 2007, citing research conducted by IP Locks, a compliance and database security company, and the Ponemon Institute, an independent research company. www.informationweek.com/news/security/showArticle.jhtml?articleID=199203 277.
2. www.finance.yahoo.com, February 24, 2009.
3. ‘‘Fifth Annual U.S. Cost of Data Breach Study,’’ Ponemon Institute, 2010, p. 4.
4. The President’s Identity Theft Task Force, ‘‘Combating Identity Theft: A Strategic Plan’’ (April 2007), 2–3, www.idtheft.gov/reports/StrategicPlan.pdf.
5. Ibid., 3.
6. United States General Accounting Office, ‘‘Identity Theft: Prevalence and Cost Appear to be Growing, Report to Congressional Requestors,’’ March 2002, 35.
7. Byron Acohido and Jon Swartz, ‘‘Meth Addicts’ Other Habit: Online Theft,’’ USA Today, December 15, 2005, 10A.
8. Cassell Bryan-Low, ‘‘Growing Number of Hackers Attack Web Sites for Cash,’’ Wall Street Journal, November 30, 2004, A1.
9. Microsoft Corporation, ‘‘Zombies and Botnets: Help Keep Your Computer under Your Control,’’ January 3, 2007, www.microsoft.com/protect/computer/viruses/ zombies.mspx.
10. Bryan-Low, ‘‘Growing Number of Hackers.’’
11. President’s Identity Theft Task Force, ‘‘Combating Identity Theft,’’ 15.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

12. Yochi J. Dreazen, ‘‘Identity Theft as an Inside Job Is Increasing,’’ Wall Street Journal, July 31, 2002, B1.
13. Federal Trade Commission, ‘‘2006 Identity Theft Survey Report’’ (November2007),31–32, www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf.
14. Martin T. Biegelman and Joel T. Bartow, Executive Roadmap to Fraud Prevention and Internal Control: Creating a Culture of Compliance, (Hoboken, NJ: John Wiley & Sons, 2006), 219.
IJRTE_PAPER_F0930012614 sachinandan09@gmail.com

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)