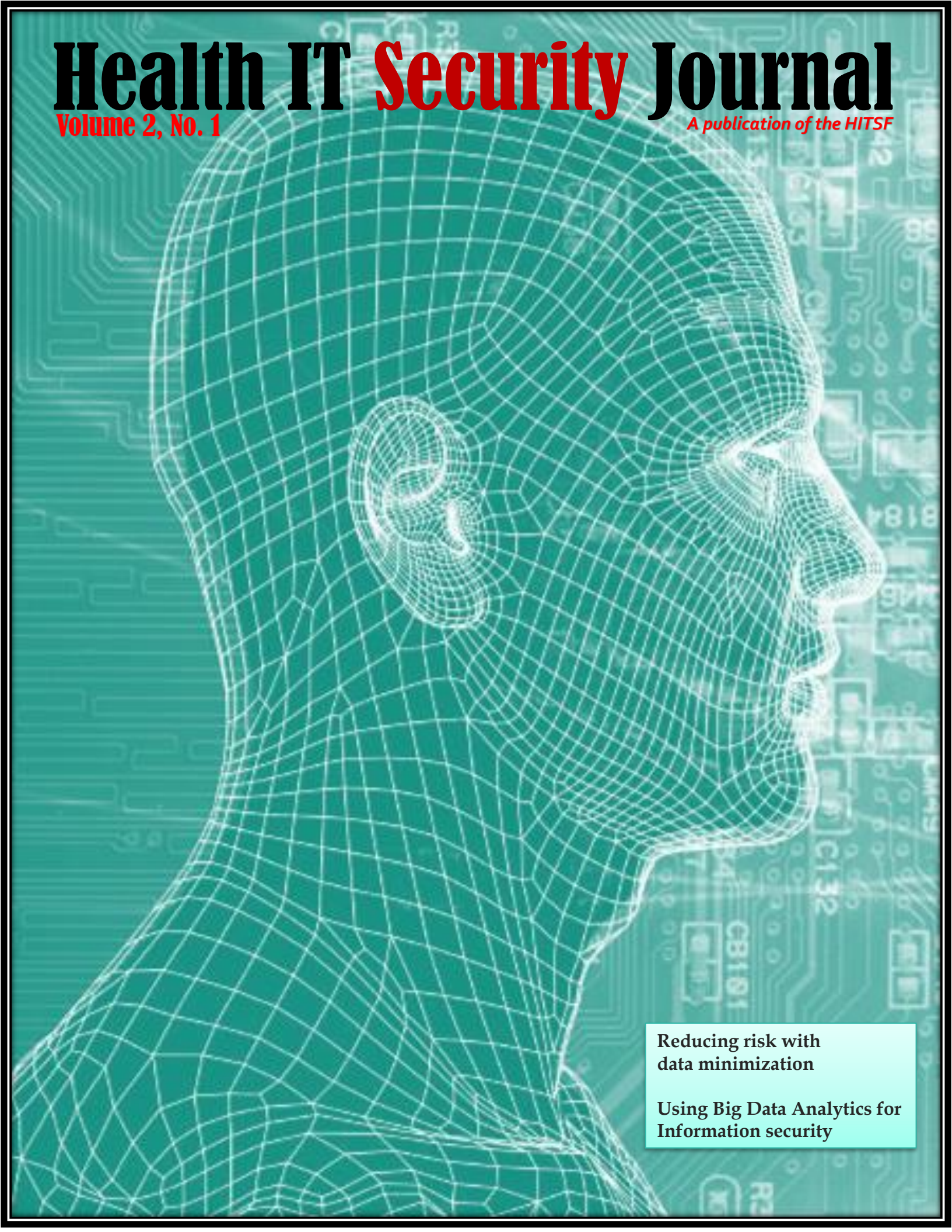
Health IT Security Journal

Volume 2, No. 1

A publication of the HITSF



Reducing risk with data minimization

Using Big Data Analytics for Information security

Health IT Security Journal

Volume 2, No. 1

HITSF

IN THIS ISSUE

This journal is a publication of the Health Information Technology Security Forum.

The Health IT Security Forum is an International Security Organization that dedicated to help healthcare organization in securing their privacy data, confidential information, medical devices and will be recognized for the passion of its members in conducting multidisciplinary research and development in the area of Healthcare IT Security and e-Health.

Letter from the editor

Dear readers,

This is the second year of the Journal and we are proud to present another issue. This issue of the HITSF journal offers insight into information security for healthcare practitioners. We hope you will enjoy it and we welcome your feedback. Please send questions and feedback to editor@healthitsecurity.org

- Hadi Syahrial

Disclaimer

The author(s) of each article appearing in this Journal is/are solely responsible for the content thereof; the publication of an article shall not constitute or be deemed to constitute any representation by the Editors that the data presented therein are correct or sufficient to support the conclusions reached or that the experiment design or methodology is adequate.

www.healthitsecurity.org

Reducing Risk with Data Minimization

by Eric Vanderburg

What if I told you that you could reduce risk and costs at the same time? Skeptical? I would be. It sounds like some cheesy marketing ploy chuck full of hidden costs or high upfront costs with low ROI. No, I am not pitching a product or trying to sell you a solution. I am however

trying to get your attention. I am talking about data minimization.

Companies collect millions of gigabytes of information, all of which has to be stored, maintained, and secured. There is a general fear of removing data lest it be needed some day but this practice is quickly becoming a problem. Some call it "data hoarding" and I am here to help you clean your closet of unnecessary bits and bytes.

Risk and costs

The news is full of examples of companies losing data. These companies incur significant cost to shore up their information security and their reputations. In a study by the Ponemon Institute, the estimated cost per record for a data breach in 2009 was $204. Based on this, losing 100,000 records would cost a company over twenty million dollars. It is no wonder that companies are concerned. Those that are not in the news are spending a great deal of money to protect the information they collect.

So why are we collecting this information in the first place? Like abstinence campaigns, the best way to avoid a data breach is to not store the data in the first place. Organizations need to ask themselves three questions:

1. Do I really need to keep this data?
2. Would a part of the data be as useful as the whole for my purposes?
3. Could less sensitive data be used in place of this data?

Do I really need to keep this data?

The first question to ask is: do I really need to keep this data? Some data is transitive in nature. It is needed in the moment but it is not needed in the long-term. Transitive data should not be stored or archived. It can simply be removed as soon as the transaction is complete. Optimally, this data should not be stored on the hard disk, but rather be kept in memory while processing the transaction and then flushed.

Other information such as buying preferences or survey data is collected to be used in aggregation and reporting. The individual responses may not be needed once the data has been aggregated so it should be purged. When analyzing business workflows, it is worth considering implementing a purge process following the aggregation and reporting process.

Effort should be made to periodically remove any records that are no longer relevant. After all, information has a shelf life, an expiration date if you will. The plain fact is that information that is no longer useful to the organization should be removed.

Another instance where you should ask if you really need to keep data is when you have a copy of the data elsewhere. In this case, you do not need to keep the data because it is a duplicate. I understand the need for redundancy but build that into a centralized database system. In this way you can protect a single area but still provide high availability. If you absolutely need distributed systems, consider segmenting the database so that distributed systems only contain the portion of the data you need.

Would a part of the data be as useful as the whole for my purposes?

The second question to ask is: would a part of the data be as useful as the whole for my purposes? Sometimes a part of the data can be as useful as the whole. Take a Social Security Number (SSN) for example. Storing the last four digits of the social may be as useful as storing the entire number and the damage associated with the disclosure of just those digits is minimal compared to the entire SSN. Similarly, a company could store just the last few digits of a credit card number rather than the entire thing.

This area of data minimization is extremely important when working with credit cards and PCI compliance as places where numbers are stored need to be in full compliance with the regulation.

Could less sensitive data be used in place of this data?

The third question you should ask is: could less sensitive data be used in place of this data? Instead of storing a value that is global in nature, like a driver's license number or SSN, consider storing a customer ID that is only used by your company. This will allow you to identify the customer without needing to store personal information.

Another option would be to store a security question such as a place of birth or mother's maiden name instead of a password. If passwords must be stored, make sure they are stored as a hash value rather than plain text. Passwords should never be stored as plain text.

To sum it all up, data minimization can reduce the amount of data you need to protect and store, reducing IT costs and information security costs and risk. Three questions can aid in determining what data to prune. Ask yourself (1) Do I really need to keep this data? (2) Would a part of the data be as useful as the whole for my purposes? And (3) Could less sensitive data be used in place of this data?

About the Author:
Eric Vanderburg

Director, Information Systems and Security, JurInnov Ltd.

Eric Vanderburg understands the intricacies inherent in today's technology and specializes in harnessing its potential and securing its weaknesses. He directs the efforts of multiple business units including Cyber Security, eDiscovery, Computer Forensics, Software Development, IT and Litigation Support at JurInnov, an eDiscovery and eSecurity consulting firm. Vanderburg holds over thirty vendor certifications and is completing a doctorate in information assurance. He has dedicated much of his career to designing and implementing systems, policies and procedures to enhance security, increase productivity, improve communications and provide information assurance. He has been invited to speak at conferences and events on technology and information security and he is active in promoting security and technology awareness through various publications. Look for his latest book, "Storage+ Quick Review Guide", due for publication with McGraw Hill in December of 2013.

Using Big Data Analytics for Information Security

by Hadi Syahrial

The 2013 Verizon Data Breach Investigations Report revealed that in 2012, 66 percent of breaches that led to data compromise within "days" or less remained undiscovered for months or more, and that in 69 percent of the cases, a third party discovered the breach. It is statistics like this that drive security teams to remain committed to evolving analytic models that provide insight on how to better protect critical processes and sensitive information.

Big data analytics has become an extremely important and challenging problem in disciplines like computer science, biology, medicine, finance, and information security. This problem involves several aspects. First, large volumes of data must be imported and stored relying on cleansing and filtering techniques. Next, sophisticated algorithms are used to analyze the data and extract "useful" information. Finally, various user interfaces can be used to visualize and understand the data. Big data analytics is finding insight that helps across the public and private sector.

The following steps are an example how to implement big data analytics for information security:

Step 1: Preparation

There are three crucial aspects in preparation step:

- Determine the business value that can be derived from big data analytics.
- Perform sample analyses.
- Build a business case.

Step 2: Determine big data requirements

An organization needs to create a big data store in order to support a big data analytics capability. There are two crucial aspects to create a big data store:

- Identify the data it will contain.
- Create an infrastructure to store and process the data.

Step 3: Determine analytics requirements

There are two crucial aspects to determine analytics requirements:

- Assess existing big data analytics capabilities.
- Determine requirements to deliver big data analytics for information security.

Step 4: Deliver full capability

There are four crucial aspects in this step:

- Assess your capabilities in light of your needs.
- Determine your place on the maturity model.
- Determine where you want to be on the maturity model.
- Write a plan and get business case approval

Big Data Analytic Process for Specific Information Security Problem

There are five steps for performing an individual information big data analysis:

1. Create Hypothesis
   The hypothesis is a question that will help resolve the business problem. The hypothesis should consider the:
   - Business problem.
   - Objective of the analysis.
   - Type of result that will be produced.

2. Select Data
   Proving the hypothesis necessitates selecting the data sources that will be used to perform the analysis. Examples of source of data for security analysis:
   - Honeypots and Honeynets
   - Malware Collectors
   - Honeyclients and Honeymonkeys
   - Spam

- Phishing Databases
- IRC Chats
- Forum
- Logs
- DNS
- DHCP

3. Analyze Data
The type of analysis will depend on the data selected and the result desired.
A wide range of techniques can be used to perform the analysis including:
    - Cluster analysis – classifying events into smaller groups of events
    - Network analysis – identifying relationships between events, often displayed using visualization
    - Time series analysis – analyzing events over time to identify patterns of behavior.

4. Examine Result
The examination should determine whether the results answer the hypothesis.

5. Act on Insight
The results will deliver insights with which the organization can make decisions and take appropriate action.

About the Author:

Hadi Syahrial

Hadi Syahrial is a Lecturer and Researcher at Budi Luhur University, Jakarta, Indonesia.