



Data Integrity - The Next Big Challenge in Information Security

Alun Thomas - Tru Data Integrity

Introduction

To date, data integrity has not been one of the hot topics discussed in the discipline of information security. Now, respected industry commentators^{1 2} describe data integrity as the next big challenge. This paper explores that challenge. In particular, it suggests that file-centric data integrity is a requirement if an organization is to maximize the value it extracts from its information.

About Digital Assets

Digital assets are the life blood of any organization. Key data about customers, transactions, products and other stakeholders are all valuable digital assets. In many industries, the product itself is a digital asset (entertainment, financial services, professional services etc.).

Digital assets are not just documents. Drawings, financial analyses, video recordings, voice recordings, software code are all generally digital assets.

The volume of digital information is exploding. A recent IDC White Paper³ estimates that the digital universe is some 500 billion Gigabytes in size and that it is growing at approximately 50% per annum with no discernible slowing in growth due to economic recession. The proportion of that universe that is “compliance intense” is also forecast to grow from around 20% in 2008 to 30% in 2012.

The Value of Data Integrity

Collecting, storing and retrieving data has a cost. An organization would not go to that expense unless that data had a value. If the integrity of that data is questionable, then its value rapidly diminishes. This impact manifests itself in several ways:

- Inability to rely on that data when making decisions.
- An increased probability of a poor decision because of poor data quality.
- Errors caused by “wrong paperwork”.
- Disputes, ultimately litigation, due to “different paperwork”.
- Continued use of paper-based or other physical processes because of an inability to trust electronic transactions and records.
- Increased probability of adverse legal judgments due to reduced evidential weight of information.
- Difficulties with regulators or in achieving compliance standards.

Just to take the example of mortgage fraud, in its *Filing Trends in Mortgage Loan Fraud* for fiscal year 2008, the Treasury Department’s Financial Crimes Enforcement Network (“FinCEN”) reported 62,084 cases of suspected mortgage fraud.⁴ By comparison, the total number of mortgages originated during that period was in the order of 4.5 million.⁵ Reviewing the data instead in dollar terms, for FY 2007, the FBI’s *Mortgage Fraud Report* reported that the total cost of losses in the 7 percent of reported fraud cases where a loss value was given was \$813 million.⁶ Given the relatively small size of this sample, it is difficult to extrapolate this figure to all mortgage fraud, but doing so would yield mortgage fraud losses of \$11.6 billion. Using the last two quarters of 2006 and the first two quarters of 2007 to approximate the government’s fiscal year, the total value of mortgage originations during that period was \$2.635 trillion.⁷

¹ David Lacey Honorary Fellow Jericho Forum talking at <http://www.infosecurityadviser.com/podcasts>

² <http://www.guardian.co.uk/technology/2009/jun/24/read-me-first-identity-fraud>

³ <http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm>

⁴ United States Department of Treasury, Financial Crimes Enforcement Network, *Filing Trends in Mortgage Loan Fraud*, available at http://www.fincen.gov/news_room/nr/pdf/20090225a.pdf.

⁵ MortgageDataWeb100 - Top 10 Lenders (January 2009) available at [http://mortgagedataweb.com/reports/\\$blog/mdw10Provident.htm](http://mortgagedataweb.com/reports/$blog/mdw10Provident.htm) (noting that top 10 lenders, representing 39% of total market, originated 1.8 million loans).

⁶ United States Department of Justice, Federal Bureau of Investigation, *2007 Mortgage Fraud Report*, available at http://www.fbi.gov/publications/fraud/mortgage_fraud07.htm.

⁷ Mortgage Bankers Association, *Mortgage Origination Estimates* (February 2009) available at <http://www.mortgagebankers.org/files/Research/HistoricalMortgageOriginationEstimates021109.xls>.





Thus, the value of fraud reported was in the range of 0.5% of the value of all originations - small as a percentage but huge in absolute terms. One common form of mortgage fraud involves alteration of documents after they have been presented to and executed by borrowers.

Data Integrity Can Only Get More Important

So, data integrity is an important challenge. Its importance and visibility can only continue to increase.

- Admissibility of digital evidence is a much discussed topic in the legal profession. One Judge went so far as to rule inadmissible digital evidence that could not be independently authenticated.⁸
- High profile examples of tampering with or falsifying documents like mortgage fraud or the backdating of stock options are regularly business news items.
- The inability to rely on digital information to be what it purports to be is now firmly embedded in popular culture. A recent best-selling novel opens with a video faked to imply falsely that a prisoner has been tortured and murdered in order to start a war and sell more arms.⁹
- Organizations are going to want to continue to drive paper and physical processes out in favor of electronic processing in order to be cheaper, faster, greener.
- The national security community in the UK and the US considers that organizations that form part of the national critical infrastructure should consider explicitly the risk of a data integrity attack¹⁰. The concern is that criminal elements or unfriendly nation states access data repositories not just to read key data but to alter it, undermining an ability to transact business or produce accurate reports, thereby engendering widespread disruption or worse.

This all adds up to increased reliance on electronic data and increased awareness of the scope for manipulation, translating to a need to demonstrate integrity of sensitive data.

The Traditional IT Security Model

Most IT security activity has assumed that data exists in a trusted and controlled space in the main. Occasionally, that data needs to be transferred to another trusted environment across a hostile environment.

So, just like a walled medieval town, the IT security team builds a wall round the perimeter of the Corporate IT network. They deploy technology and process to identify the good guys from the bad guys. This wall prevents bad guys from penetrating the environment to read and corrupt the organization's data. The security team builds a secure data repository within this environment.

When data needs to be transferred to another trusted environment, it is coded or encrypted so that anyone illicitly intercepting the data does not have a key and cannot decrypt and read the information.

Need to Protect the Data Not the Environment

It has become increasingly obvious that this traditional security model is valuable but has its limitations.

Data needs to be shared to be of maximum value. The people with whom this data needs to be shared are often outside the organization's trusted environment. They might be employees working at home or on the road. They might be business partners involved in some form of collaborative working. They might be suppliers or customers working across a global supply chain.

No perimeter control is perfect. Hackers are increasingly resourceful (or lucky). Often, there is some form of insider element to a fraud or data leak.

The IT security community has responded to these limitations with a movement known as de-perimeterization¹¹. In summary, this model pays less attention to the walls around the organization's IT environment and more attention to protecting the data itself, whether it is at rest or in motion from one place to another.

⁸ <http://www.ediscoverylaw.com/2007/05/articles/case-summaries/chief-us-magistrate-judge-grimm-provides-detailed-analysis-of-evidentiary-issues-associated-with-electronic-evidence/>

⁹ http://www.amazon.co.uk/Whole-Truth-David-Baldacci/dp/0330456520/ref=pd_sim_b_2/276-9568031-1365103

¹⁰ <http://www.computerweekly.com/Articles/2009/02/16/234824/how-to-defend-against-data-integrity-attacks.htm>

¹¹ Jericho Forum



Defining Data Integrity

CESG is the UK national technical authority on information assurance. Their model¹² identifies five key principles, essential for safe electronic transactions; i.e. for the safe and reliable transfer of digital information between two parties.

- **Confidentiality**- keeping information private.
- **Integrity** - ensuring information has not been tampered with.
- **Authentication** - confirming the identity of the individual who undertook the transaction.
- **Non-repudiation** - the individual who undertook the transaction cannot subsequently deny it.
- **Availability** - ensuring information is available when required

Not all of these principles receive the same level of attention. A great deal of attention is given to confidentiality and to identity management solutions which achieve the properties of authentication and non-repudiation. This model provides a working definition of data integrity. Relatively speaking, integrity is the forgotten sister and the model doesn't really talk to the value of data integrity and the implied cost of it being that forgotten sister.

Data with readily demonstrable integrity is data which has the property of quality and of trustworthiness.

Data Integrity Has Been a Low Priority

As we have already seen, the typical organization is working hard to protect their IT infrastructure with perimeter defenses. They might also be protecting the confidentiality of that data at rest on laptops and storage devices and in motion over the internet. What are people doing about protecting the integrity of their data?

The truth is that many organizations - at least in the UK - pay little attention to data integrity unless and until they are party to a high profile piece of litigation or regulatory investigation. At the most security conscious end of the market e.g. organizations handling Government classified information, encryption technologies are deployed that rely on Public Key Infrastructure (PKI). This form of encryption has the additional benefit of containing an inherent integrity check when a message is decrypted. However, this is almost a by-product and subsequent integrity checks are generally not possible. Also, large scale deployments of PKI have a reputation of being problematic.

Data integrity is a requirement of many compliance and regulatory frameworks but it is often implicit in a requirement to be able to authenticate records or to use best endeavors to ensure that reports are accurate.

An indication of a latent demand for integrity is the frequency with which organizations convert word processed files to Portable Document Format (PDF) in the belief that a PDF file is protected against change. It is true that a PDF file is somewhat harder to change accidentally but a determined party with rudimentary software skills can easily change the content of a PDF format file.

Need for a Better Way

Digital assets have great value. That value is greatly reduced, if not eliminated, if the integrity is called into question. Increasingly, judges and regulators require demonstration of that integrity rather than relying on assertions. There is a need for capabilities and technologies which meet several easily identified criteria.

- The technology needs to be file-centric not system-centric, so that it goes wherever the information goes rather than simply being of value in a controlled environment.
- The proposed solution needs to be forensically pure; i.e. meeting the best practice requirements of digital forensic investigators and digital evidence specialists such that it survives intense interrogation in a court of law.
- The requirement is that this integrity can be validated for as long as the asset has value - perhaps 100 years and longer - surviving all kinds of discontinuities in computer technology and practice.
- And of course, any technology must be easy to use and cost effective.

¹² http://www.cesg.gov.uk/about_us/whoarewe.shtml



It doesn't appear that the true value of reliable data is going to appear on the balance sheets of organizations any time soon. So the question is "where is the business case for treating data integrity as a priority?"

- Immediate and direct cost savings by eliminating the production or physical movement of paper or physical storage media (for example in the cases of compliance records or criminal justice).
- Facilitation of major process re-engineering to deliver substantial benefits and which should not be pursued without demonstrable data integrity (electronic medical records, cloud computing).
- Closing the loophole through which revenue leaks (digital piracy).
- Addressing a fundamental requirement for trustworthy and reliable data in mission critical systems (cyber security).

Organizations placing a high value on their data, or relying upon the integrity of that data, must conduct a holistic data integrity review to ensure the successful, cost effective deployment of information integrity technologies.