Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

A network based document management model to prevent data extrusion [1]

Kamran Morovati [a,*], Sanjay Kadam [b], Ali Ghorbani [a]

[a] Information Security Center of Excellence (ISCX), Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada

[b] Center for Development of Advance Computing (C-DAC), University of Pune, Pune, India

ARTICLE INFO

Article history:
Received 4 February 2014
Received in revised form 25 January 2016
Accepted 4 February 2016
Available online 15 February 2016

Keywords:
Data leak prevention
Insider threat
XML security
Public key infrastructure (PKI)
Data confidentiality

ABSTRACT

This paper presents a network-based document management model to protect sensitive data from unauthorized disclosure in organizations. The presented framework utilizes XML security concept, PKI cryptosystem and an XML-based metadata embedding approach mainly to address the problem of information leakage that may happen through employees of a company (also known as the insider threat).

By utilizing the above-mentioned techniques, our suggested framework wraps the confidential documents with an additional layer of security meta-data that can be investigated in critical network points of an organization to prevent illegal data distribution. This solution can also guarantee data confidentiality, integrity, and support for non-repudiation. We believe our suggested model can effectively prevent intentional or unintentional data leak incidents.

1. Introduction

In cyber security, data leak means the unauthorized release of secure information from within an organization to an untrusted environment. Data leakage can happen intentionally by the company's employees who have authorized access to an organization's network, system, or sensitive data (also known as insider threat). It may also happen inadvertently due to an employee oversight or through a malware attack. Insider threat is considered as a significant security risk to the organization's data confidentiality. According to 2015 "Data Breach Investigations" report published by Verizon (Data Breach Investigations Report, 2015) internal employees are considered as one of the main reasons of data leak incidents in organizations. Additionally, this report provides a model to determine expected financial loss based on the number of disclosed records. Table 1 shows the record count and the corresponding expected financial loss (the "Expected column" in the middle) (Data Breach Investigations Report, 2015).

Simply put, an "insider threat" means the exposure of proprietary or confidential information of the company or organization by employees from within the company. Valuable data such as the source codes, financial or medical records, trade secrets, political issues or more mundane items such as company strategies and future business initiatives can be exposed, allowing competitors to gain an unfair advantage in business operations. For instance, in May 2013, Edward Joseph Snowden, who was a former CIA employee and NSA contractor, disclosed classified details of several top-secret United States and British government mass surveillance programs to the press. Snowden's release of NSA material was called the most

[1] Preprint submitted to Elsevier journal of Computer & Security on Jan 13, 2015.

* *Corresponding author.* Tel.: +15064523313.

E-mail address: kamran.morovati@unb.ca (K. Morovati).

http://dx.doi.org/10.1016/j.cose.2016.02.003

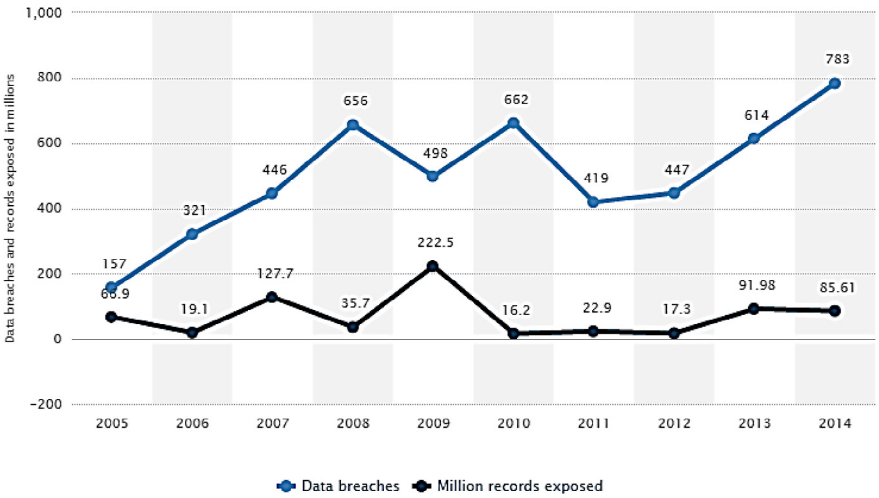

Fig. 1 – Data breaches incidents in U.S. recorded by Statista.com.

Table 1 – Ranges of expected loss by number of records (report provided by Verizon).

No. of records	Average (lower)	Expected	Average (upper)
100	$18,120	$25,450	$35,730
1000	$52,260	$67,480	$87,140
10,000	$143,360	$178,960	$223,400
100,000	$366,500	$474,600	$614,600
1,000,000	$892,400	$1,258,670	$1,775,350
10,000,000	$2,125,900	$3,338,020	$5,241,300
100,000,000	$5,016,200	$8,852,540	$15,622,700

significant leak in US history by the Pentagon (Greenwald et al., 2013). As another example, Chelsea Manning,[1] a US army soldier, allegedly downloaded (in 2010) classified files from military networks and leaked them to the anti-secrecy website, WikiLeaks (Chelsea Manning, 2013).

The insider threat is a major problem that organizations must be aware of and proactively prevent since it may result in data leakage and security breaches. The threat is attributed to legitimate users who abuse their privileges, and given their familiarity and proximity to the sensitive data, can easily cause a significant security violation. In an organization's network, the focus is mainly on guarding the perimeter since external attacks have a higher incidence. Consequently, internal resources are left largely exposed with only the basic access control mechanisms. Due to the lack of tools and techniques, security analysts do not correctly perceive the threat, and hence consider the attacks as unpreventable. Data leakage perpetrated by insiders is a long-standing security problem. According to CERT Insider Threat Center (Silowash et al., 2012), many insiders who stole or modified information were actually recruited by outsiders, including organized crime, foreign organizations or governments.

According to the threat report generated by McAfee, the number of data breaches with extreme political and/or financial effects, which have been made public, is increasing every year (Labs, 2013). Fig. 1 illustrates the number of data breaches recorded by Statista (Statista, 2015). According to this graph, since 2005, the number of cyber-attacks is increasing. In 2005, for instance, 157 data breaches were reported in the U.S., with 66.9 million records exposed, but in 2014, 783 data breaches were reported, with at least 85.61 million total records exposed, which indicates an increase of nearly 500 percent over 2005.

In (Symantec, 2013), Symantec has reported that healthcare, education and government sectors accounted for nearly two-thirds of data breached in 2012. Fig. 2 shows the data breaches by different sectors in 2012. At 36%, the healthcare industry continues to be the sector responsible for the largest percentage of disclosed data breaches by industry. Based on the Symantec report, hackers and insiders are responsible for a large number of data breaches. Theft or loss of computers or drives and making the sensitive information accidentally public are some other reasons. In another report (Randazzo et al., 2005), after examining 23 incidents carried out by 26 insiders in the banking and finance sector between 1996 and 2002,

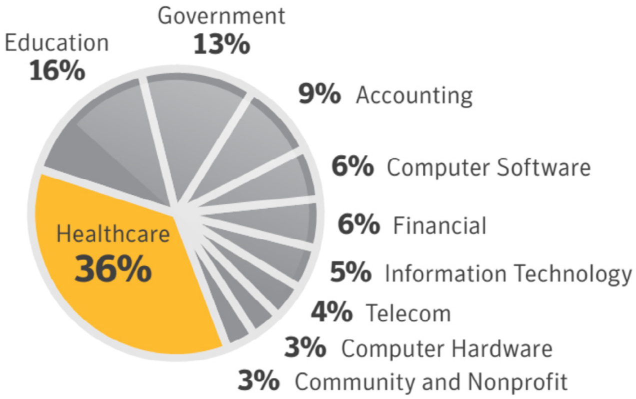

Fig. 2 – Data breaches by different sectors in 2012.

[1] Born Bradley Edward Manning, December 17, 1987.