Charles D Freeman. Describing the Impact of Document Content Variance on Access Control Efficiency and A Proposed Solution for Improving Efficiency: Fine-grained, Redactive Access Control Models. A Master's Paper for the M.S. in IS degree. April, 2013. 49 pages. Advisor: Arcot Rajasekar

Access control is a fundamental safeguard for protecting corporate information contained in computer files (herein documents). In the de facto industry standard implementation, the most granular access control specification available is to define the entry on the entire document object. Given the proliferation of data privacy regulations mandating varying degrees of information security controls and practices, one must consider to what extent inefficiency arises due to the withholding from a collaborative workflow relevant information entwined with discrete sensitive information of which not all members of the workflow are authorized to access. Contemporary research in the field proposes a digital metamorphosis of the analogue document redaction paradigm. This research explores the aforementioned inefficiency phenomenon as well as purported utility of redaction through examination of survey data from faculty and staff at a public University.

Headings:

      Access control

      Selective dissemination of information

      Access to information

      Computer Security

DESCRIBING THE IMPACT OF DOCUMENT CONTENT VARIANCE ON ACCESS
CONTROL EFFICIENCY AND A PROPOSED SOLUTION FOR IMPROVING
EFFICIENCY: FINE-GRAINED, REDACTIVE ACCESS CONTROL MODELS

by
Charles D Freeman

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2013

Approved by

_____

Arcot Rajasekar

Table of Contents

Introduction

Executives are scrutinizing the value of their business' information assets with increasing vigilance. The ubiquity of information technology (IT) as a critical dependency appearing at some, if not all, nodes in a supply chain engenders the concern for the integrity and availability of the IT so as to avoid undue interruption to service. Furthermore, organizations handling data deemed highly sensitive assume culpability for assuring the confidentiality, integrity and authorized availability of these assets under the perils of severe sanctions. Finally, societal concern over the uncertainty of digital privacy fosters an atmosphere of anomie in the handling of sensitive data and reticence for sharing information with stakeholders.

These observations contextualize the role of information security in the corporate environment; access control is the keystone to any information security initiative. Conventional access control models for unstructured and semi-structured data describe the document as the indivisible boundary around which controls are established and enforced. This paradigm doesn't account for variance of content sensitivity within a document. As disparate organizational roles collaborate on a document the potential increases for a modicum of information deemed confidential by one of the roles to poison the document channel, rendering it inaccessible to stakeholders who would otherwise have a legitimate operational need for accessing the document. A scenario such as this

casts doubt on the efficiency of current document access control models. Deductively we posit the following hypotheses:

H_1: *Fine grained, redactive mechanisms will improve access control efficiency.*
H_2: *Document content uniformity positively correlates with access control efficiency.*

In order to explore the validity of these suppositions we proceed with a review of relevant literature.

Literature Review

Businesses are under increasing pressures to assure the security of their information assets. A survey conducted by Ponemon Institute LLC concludes "the costs to notify victims of [a data] breach increased … from approximately \$510,000 to \$560,000. A key factor is the increase in laws and regulations governing data breach notification" (March 2012, p. 3). Additionally, the residual effects of reputation loss can be catastrophic to organizations both in terms of reduced productivity and customer loyalty (Ponemon Institute LCC, January 2012). Thus, if an organization seeks to establish and maintain a propitious posture in today's marketplace, threat vectors leading to egregious privacy violations or confidentiality breaches must be carefully considered.

Assessing risks and implementing appropriate protections for information assets requires significant interdisciplinary cooperation and a long-term commitment of resources. This can be evidenced in the information security concept of least privilege. To illustrate the importance of this concept Mutch and Anderson observe, "it is generally accepted that a central goal of HIPAA as well as every other industry, governmental, and regulatory compliance statute is the implementation of least privilege" (2011, p. 149). However as Hu, Ferraiolo and Kuhn point out least privilege can be difficult and costly to

achieve as it "requires identifying the user's job functions, determining the minimum set of privileges required to perform those functions, and restricting a user to a domain with those privileges and nothing more" (2006, p. 8). At a minimum these activities demand the input of management, to develop procedure based on prioritization of vital assets under their stewardship, legal council, to interpret policies within the scope of the organization's activities and ensure procedures fulfill the organization's obligations, IT experts, to implement and maintain information systems in compliance with regulatory standards, and all stakeholders involved in the lifecycle of the asset, to operate in accordance with established procedures.

Nevertheless, in recognition of the alarming trend towards insider threats and vulnerabilities originating inside the network perimeter, measures such as least privilege are incontrovertibly necessary and worth the high initial investment. Sanyal, Shelat and Gupta (2010) cite "It is a well established fact that 70%+ of threats to an organization's network and network-based infrastructure originate from inside" to challenge the common practice of corporate information security measures which focus solely on defending the network perimeter from incoming threats and intrinsically trusting traffic originating from within the network (p. 63). Baracaldo and Joshi (2010) cite a survey estimating that 33% of computer crimes reported in 2010 involved insider attacks and comment that some of the attacks could be pre-empted if the access controls "were able to react when a user is performing actions that are not appropriate for their normal job functions" (p. 167). Furthermore, the Ponemon Institute LLC (January 2012) survey finds that, in the cases where respondents were able to identify the source of the breach, 34% were attributed to negligent insiders, 19% to outsourcing data to a third party, 16%

to malicious insiders and 6% for failure to shred documents (p. 5). These findings not only demonstrate the potential breadth of vulnerabilities lurking within and throughout an organization's human and technological assets, they stress the need to re-evaluate foundational information security practices, specifically access control.

According to the Official (ISC)² Guide to the CISSP CBK, "Access control provides the basic building blocks for enabling information security and is the foundation upon which all security efforts...are based" (Tiller & Fried, 2010, p. 2). The two seminal electronic data access control models are discretionary access control (DAC) and mandatory access control (MAC). DACs are characterized by the data owner's specification of the access control on the resource (i.e., access control is instantiated at the data owner's discretion). This model was available in some of the earliest multi-user, network computing systems and is currently supported in nearly every mainstream, commercial computing system (Tiller & Fried, 2010, p. 116).

In contrast to DACs, "access control policy decisions [in MACs] are made by a central authority, not by the individual owner of an object, and the owner cannot change access rights" (Hu et al., 2006, p. 7). The formalization of MACs applied to information systems was conceptualized at least since the 1970s in the subject clearance levels and object classification levels from the Bell-LaPadula (1973) multilevel security (MLS) model and the integrity classes expounded in the Biba (1977) model (Hu, et al. 2006, p. 7). However, for many decades MACs were deployed exclusively in military and national security information systems; only within the last decade have implementations of MACs manifested in commercially available systems. Security-Enhanced Linux (SELinux), originally conceived by the National Security Agency's (NSA) National

Information Assurance Research Laboratory (NIARL) as a flexible version of MAC to support dynamic security policies then open-sourced and subsequently incorporated into several Linux distributions, is one such example (NSA, 2009). A similar MAC subsystem, Windows Integrity Mechanism, based on the Biba model was developed for the Microsoft Windows Vista operating system (Microsoft Developer Network, 2007).

While DACs are advantageous due to their flexibility with respect to custodial control this facet also leads to the primary disadvantage that DACs "do not impose any control on how information is propagated and used once it has been accessed by users authorized to do so" and are therefore open to compromise by malicious agents acting under the authorized users' context (Elmsari & Navathe, 2007, p 807-8). Additionally access privileges assigned through DACs are transitive; an individual to whom the owner delegates access may copy data to a less secure destination whereupon the data is exposed to a wider audience, possibly unbeknownst to the owner (Hu et al., 2006, p. 6). For instance, from a sample of 3328 shared documents retrieved from various P2P networks Johnson (2009) discovered 389 unique files leaked from the healthcare industry of which approximately 5% could be used for identity theft (p. 10). The figure may seem relatively small however it becomes alarmingly significant when one considers the volume and gravity of information within the documents, one of which contained over 9000 patient identifiers (p. 11). Extrapolating the industry average cost of $198 per record breached, the compromise of this single document would result in approximately $1,782,000 damages to the organization (Ponemon, March 2012).

By contrast MACs "prevent any illegal flow of information" though at the expense of a rigid classification system which is conducive to few environments outside

of the military thus "in many practical situations, discretionary policies are preferred because they offer a better trade-off between security and applicability" (Elmsari & Navathe, 2007, p. 808). For example, in the Bell-LaPadula MLS model, every asset in the system (e.g., employee, system, document) is designated a mutually exclusive security classification (e.g., unclassified, confidential, secret, top secret). Two rules, the simple security rule and *-property enforce the authorized flow of information in the system. The former states that if a subject (e.g., employee or process) wants to access an object (e.g., document) their security classification must be equivalent or higher to the object's. The latter states that a subject cannot write to an object whose classification is lower than the subject's. Such a model would present enormous challenge for businesses, where timely exchange of information throughout the supply chain is critical. A notable approach advocated in current research is the concept of "sticky policies" that transfer with the data (Agrawal and Johnson, 2007, p. 278). Such technology could supplement either the DAC or MAC approach by enforcing any business logic introduced by stakeholders outside the immediate purview of the steward or delegated consumers of the information asset, regardless of the logical location of the document in a file system.

DACs and MACs represent the extreme poles along an access control continuum in terms of authority to prescribe least privilege. Alternatives have been developed to ameliorate these models. For instance role based access control (RBAC), an extension of the MAC model better suited for corporate environments, provides assignment of privileges to roles defined as "a collection of permissions to use resources appropriate to a person's job function." Flexibility akin to the DAC model is achieved in that "users are given authorization to adopt roles" and through delegation of object ownership, though

both come at the expense of centralized control (Hu, et al, 2006, p 16, 25). Despite the significant perceived cost associated with implementing RBAC, a survey conducted by O'Conner and Loomis (2010) demonstrated an overall positive economic impact from implementation of RBAC, quantified at $142.92 per employee, as a result of operational efficiencies in terms of access provisioning, policy maintenance and certification (p. 8-9). Moreover, RBAC has seen significant adoption since its formalization in 1992, up from 4% to 13% in 2004 and 41% in 2009 (p. 5).

Baracaldo & Joshi (2012) voice a strong criticism to RBAC in that it assumes trustworthiness is a static property since it doesn't evaluate potential precursors to an insider threat, such as inappropriate Internet activity or successive unauthorized access attempts (p. 167). They extend the RBAC model by including a criterion whereupon a user's trust value, collated from sub-systems which monitor relevant user activity, must be validated during the role activation phase (p. 169). If the user's trust factor falls below the trust threshold established by the system for the designated role set, the validation fails.

Another recent variation is exemplified in the Aeolus architecture of Cheng et al. (2012) that proposes an intriguing synthesis of DAC and MAC models in the respective tag and label functionality. Tags provide principals (i.e., users and applications) methods for classifying information (e.g., public, financial, medical) (p. 3). Delegation of the principals' authority on a tag functions in an acyclic graph structure. This paradigm allows for dynamic adaptation in response to evolving business logic in that delegation and revocation of access can cascade based on any established edges stemming from the vertex at which the privilege is executed (p. 4). For instance, if a data steward designates

custodianship to a stakeholder and this steward's access is later revoked, the custodian's

access is also revoked unless the custodian inherits authorization from a different steward

in the hierarchy. Labels prevent unauthorized information leakage between principals

based on adherence to a logic flowing from the MLS principle of MACs (p. 3). Each

principal has two labels: a secrecy label and integrity label. The secrecy label roughly

functions under the MLS simple security rule (i.e., a source principle's secrecy label must

be a subset of the destination's) while the integrity rule accomplishes the intent of the *-

property (i.e., a source principal's integrity level must be the superset of the

destination's). User principals can declassify data by removing a tag from a security

label or endorse data by adding a tag to an integrity label only if the rules are met and the

user has the appropriate authorization described in the tag authorization structure (p. 3 –

4). These recent examples illustrate the scope of improvements yet to be realized in this

domain.

Yet, industry is left to contend with the exponentially compounding concerns of

information classification and efficient knowledge management. Porter and Millar

(1985) portended, "So pervasive is the impact of information technology that it confronts

executives with a tough problem: too much information" (p. 154). Commercial software

offerings have devoted significant attention to isolation and obfuscation of information

outside a user's designated role. For instance, SQL standards compliant DBMSs support

permutations of access control within a single document structure (i.e., table) through the

creation of a viewed table using the CREATE VIEW statement and assignment of

privileges on the viewed table. For flat file systems hosting unstructured and semi-

structured document types (i.e., network file servers), mechanisms such as Microsoft

Corporation's (2005) Access Based Enumeration and Samba (2010) "hide" configuration

directives support the creation of directory views derived from access control lists such

that individuals are only permitted to view in a directory listing those documents for

which they are granted access. The feature is marketed based upon both its security

benefits and potential to increase productivity since "end users see only what files and

folders they need for their responsibilities rather than spending time looking through lists

of inaccessible folders and files" (Microsoft Corporation, 2010, p. 1). This technology

underscores a subtle facet of access control: the exclusion of individuals who do not

possess a need to know for the information asset reduces information sprawl for these

individuals, thereby improving the rate in locating information relevant to the user's role.

Early attempts at electronic document redaction have proven cumbersome, ad-hoc

and unreliable. Numerous incidents of failed document redaction leading to unauthorized

information leakage have been publicized. For example Liptak (2006) reports the

leakage of grand jury testimony on steroid use in professional baseball when reporters

were able to recover ostensibly redacted text from a PDF file simply by copying the text

area into a word processor. A similar exposure attributed to ineffective redaction is

reported by Wiley (2005) though, in this case, the leak had significant ramifications on an

international stage by revealing incriminating US military procedures that may have led

to the death of an Italian agent. Forrester and Irwin (2005) enumerate several additional

examples illuminating the scope of the problem (p. 4). In all of the cited cases,

information leakage is attributed to the redacting party's negligence in sanitizing

metadata and other hidden artifacts embedded in the electronic record (e.g., covering text

with a black box in a PDF file is not an effective method for sanitizing the data from the

document; c.f., Forrester and Irwin, 2005, 5 – 6). An NSA (2005) report details an effective, though highly elaborate, procedure for eliminating hidden data by converting a Microsoft Word document to PDF though the language hints at equivocation in the statement,

> "This document does not address all the issues that can arise when distributing or downgrading original document formats…Using original source format… can entail exceptional risk; the lengthy and complicated procedures for mitigating such risks are outside the scope of this note." (p. 2)

One can envision inordinate opportunities of omission in the eight-step, detailed procedure outlined by the document, spanning as many pages.

Developments in the field of document redaction promise an effective method for supporting fine-grained, content-based access control views with comparatively minimal effort on the part of the end-user. Staddon, Golle, Gagné and Rasmussen (2008) propose an attribute based encryption protocol for provisioning access control on document attributes (i.e., sensitive information identified through natural language processing or user generated tags) through authorized distribution of the attribute's associated decryption key. Staddon et al. contend this approach is efficient since it allows for public circulation of a single version of the document while ensuring each user is able to view only that information for which they're authorized (p. 27).   Bier et al. (2009) extend this approach by fusing natural language processing methodologies, both for identification of explicit sensitive information as well as those attributes, which in combination may lead to inference of the information, and user interaction as a compensating control to facilitate high precision through iterative refinement. Both prototypes demonstrate a functional framework promising unprecedented capacity for disclosing information while maintaining compliance under least privilege mandates. They improve on previous

methods, such as the NSA (2005) report, through seamless integration of redaction and

sanitization capability along with semi-automatic approaches to identification of sensitive

elements. One observation worth noting is that, especially in consideration of the

potential high cost of failure for overlooking false negatives, it may be more prudent for

the redaction mechanism to redact all information and allow the data steward to

deliberately select subsets of data within the document appropriate for sharing among

stakeholders tangential to the operational need associated with the document.

Recalling the prior attestation to the insider threat dilemma, it is instructive to

consider the attitudes of end users in estimating efficiency of a document access control

system. The Hassell (2005) trust model emphasizes affective factors such as

commitment to a social group and frame of references in the trustee's cognitive

evaluation whether to place trust in the entity under scrutiny. Furthermore, qualities such

as perceived ease of use and attitude toward using are relevant in an end-users

consideration of trusting a technology (p. 136 – 138). If these conditions are not

satisfactorily met by the end-user's estimation, the entity will not be trusted (i.e., the end-

user will attempt to evade the access control system by misusing it, working around it or

ignoring it entirely). Albrechsten's (2007) survey of employee satisfaction of

information security measures taken at an IT firm and bank elucidate this theory. Based

on interviews with the employees, Albrechsten detects a latent conflict between

information security and functionality fulminating in the employees' perception as daily

operational demands increase as well as a tendency, at least among the bank employees,

to distinguish between their individual responsibilities and the responsibilities of

information security, which in their mind are relegated to the domain of specialists (p.

281). Albrechsten concludes, "the interviewed users consider the costs of cautious

behavior to be higher than the perceived benefits…Benefits on other areas such as

usability, efficiency and functionality are achieved by a risky behavior" (p. 286).

The sentiments of our hypotheses are best exemplified in the following statement

from Hu et al. (2006):

> "The objectives of an access control system are often described in terms of
> protecting system resources against inappropriate or undesired user access. From
> a business perspective, this objective could just as well be described in terms of
> the optimal sharing of information. After all, the main objective of IT is to make
> information available to users and applications. A greater degree of sharing may
> get in the way of resource protection; in reality, a well-managed and effective
> access control system actually facilitates sharing. A sufficiently fine-grained
> access control mechanism can enable selective sharing of information where in its
> absence, sharing may be considered too risky altogether" (p. 3).

These assertions allude to the ostensibly opposing objectives of IT and the information

security domain of access controls to simultaneously share and protect data. Hu et al.

affirm these are in fact mutually supporting aims in that assurance of authorized and

sufficient access controls mitigate risk to the organization associated with information

sharing, thereby facilitating information sharing.

One notes that even MACs, the strictest extremity of the access control spectrum,

are susceptible to data breach as a result of targeted attacks; for instance, Rjaibi and Bird

(2004) cite how colluding parties in different hierarchical clearance levels can take

advantage of locks placed on a resource using a predetermined protocol to establish a

covert channel (p. 1013). Recognizing that access controls are fallible and therefore

sharing of information entails some risk, it is necessary to instill stakeholder confidence

that the access control will operate sufficiently to "enable selective sharing" according to

the least privilege principle. Otherwise as a result of privacy concerns, stakeholders may

withhold or falsify information. El Emam et al. (2009) invoke several statistics

underscoring this phenomenon among physicians and patients in the US and Canadian

health-care industry (Introduction, para 1). Providing stakeholders a means for explicitly

defining access within the document may alleviate concerns through the participatory act.

This accords with Albrechsten's (2007) assessment, "Improving individuals' knowledge,

familiarity and control of risk should influence users' perception of risk, which in turn

can affect individual behavior" (p. 286).

Building on Hu et al.'s aforementioned assertions, the research outlined in the

following sections looks at the effect of access control on information sharing from the

perspective of whether insufficiently fine-grained access controls unduly exclude

stakeholders in a network

Methodology

It is the purpose of this research to describe the impact of the document-centric

access control model on an organization's ability to conduct necessary operations while

complying with regulatory information security mandates. This is accomplished through

the endeavor to measure the efficiency of the University of North Carolina at Chapel Hill

(UNC)'s institutional document repositories in terms of appropriate application of access

control as it relates to document content and compliance with the least privilege principle.

These systems ordinarily operate under DACs verbally expressed by the principal data

steward, generally a department head or manager, and instantiated with the aid of an IT

administrator.   To form a more detailed impression of the environment under

consideration the following brief analysis of the policies and procedures governing the

appropriate disposition of the organization's information resources is presented.

Significantly, according to the UNC Chief Information Office (CIO) "All Users must be aware of the classification of the various types of University information to which they have access in order to determine the proper controls [for the information]" and "…unauthorized disclosure of Sensitive Information to individuals without a business need for access may violate laws or University policies and have significant ramifications for [the University]" (2011, p. 3). These statements solidify the burden of responsibility on stakeholders to recognize the security implications of each discrete piece of information they process and validate the legitimacy of the business need claim to sensitive information. Additionally, "Decisions about the provision of access to Sensitive Information must always be made by the Steward […] of that Sensitive Information," implying a DAC model however the depth of delegation authorized is ambiguous (p. 3 – 4; c.f., "Responsibility of Sensitive Information may be delegated by the Dean, Division Head, or their designee [and this] must be clearly identified in writing as such", p. 9). Further, "Information must be classified according to the most sensitive detail it includes" (p. 3). This final statement conjures the specter of our original fear that optimal information sharing may be interdicted by the presence of highly sensitive information in the midst of a document.

For the purpose of this research, a stakeholder is defined as anyone with a need to know any part of the content of a document as determined by the document's data steward. This distinction essentially aligns with the definition of a Consumer/User in the UNC Information Security Policy (c.f., UNC CIO, 2011, p. 4). The document's data steward is the manager or department chair in whom operational authority and discretion for defining access control is invested (i.e., the seed individual from whom transitively

defined access may germinate; subsumes the definitions of "Steward of Information or Data" and "User Managers") as well as any authorized designees (c.f., p. 4 – 5). A document is defined as a digital file object stored on an institutional file repository (i.e., University managed file server) which is ascribed a set of access control rules.

Though one might anticipate that information classification should be a highly nuanced and therefore a more difficult concept to operationally define, the University Information Security Policy mandates a binary classification scheme: information is either public or sensitive (UNC CIO, 2011). The latter is defined generally as regulated information while the former is all information that is not sensitive, which must be rendered available for public inspection upon request under North Carolina General Statute, Chapter 132 (p. 2 – 4). In practice the DACs facilitate a third type of classification: they describe least privilege for the document (i.e., the minimal set of individuals who have a direct operational need to access the document) as a function of convenience (i.e., preventing users from being overwhelmed with information not relevant to their role when browsing the directory structure) and assurance (i.e., the scope of users who may alter the document in a manner unintended by the data steward is minimal).

Divining an encompassing efficiency metric for information security procedures is confronted contentiously in the literature as demonstrated by Boehmer's (2008) analysis. Two trends are identified: the Return of Security Investment (ROSI) approach and calculations evaluating loss of productiveness (p. 227). ROSI is essentially the ratio of the cost associated with a successful attack on a given asset to the cost associated with the proposed security countermeasures to protect that asset. If the cost of protecting the

asset outweighs the cost of an attack then the investment is considered inefficient. However, as alluded in the beginning paragraph of the Literature Review Section, it is increasingly difficult to measure cost associated with regulatory sanctions and loss in consumer confidence. The second approach, the evaluation of productivity loss, attempts to derive capital losses arising from a potential breach incident with respect to work stoppage. Boehmer gives the example of a file server being disrupted and the number of effected employees being considered, though the contention is made that a suitable benchmark does not exist (2008, p. 227).

While ROSI and productivity loss are useful factors to consider when evaluating the decision to pursue a regimen of security measures, they neglect to consider the influence of these technologies on quotidian operations. In this regard, the Porter Value Chain Model is highly instructive. Porter and Millar (1985) summarize the concept by stating "A company's value chain is a system of interdependent activities, which are connected by linkages [established according to the interrelated activities' influence on each other with respect to cost and effectiveness]" (p.150). These links create a complex network consisting of the company's primary, or external, and support, or internal, activities as well as edges connecting affiliates, suppliers and customers. Through optimization of these linkages, either through cost reduction or differentiation with respect to the offerings of its rivals, an organization can create competitive advantage. Authorized information sharing, especially among an organization's primary and support activities, is crucial in modern enterprises for efficient operations in the relentlessly escalating knowledge driven economy. In consideration of this work, we propose the following model for efficiency.

Efficiency is the property whereby all stakeholders who may access some part of the content of a document are allowed to do so by the access control. The DAC is efficient if $\beta - \alpha$ results in an empty set where α is the set of stakeholders defined in the DAC (i.e., those who have privileges to the entire document) and β is the set of stakeholders who have an operational need, as determined by the data steward, to access some part of the document. Efficiency of access control lists can thus be defined as

$$\frac{\alpha \cap \beta}{\alpha}$$
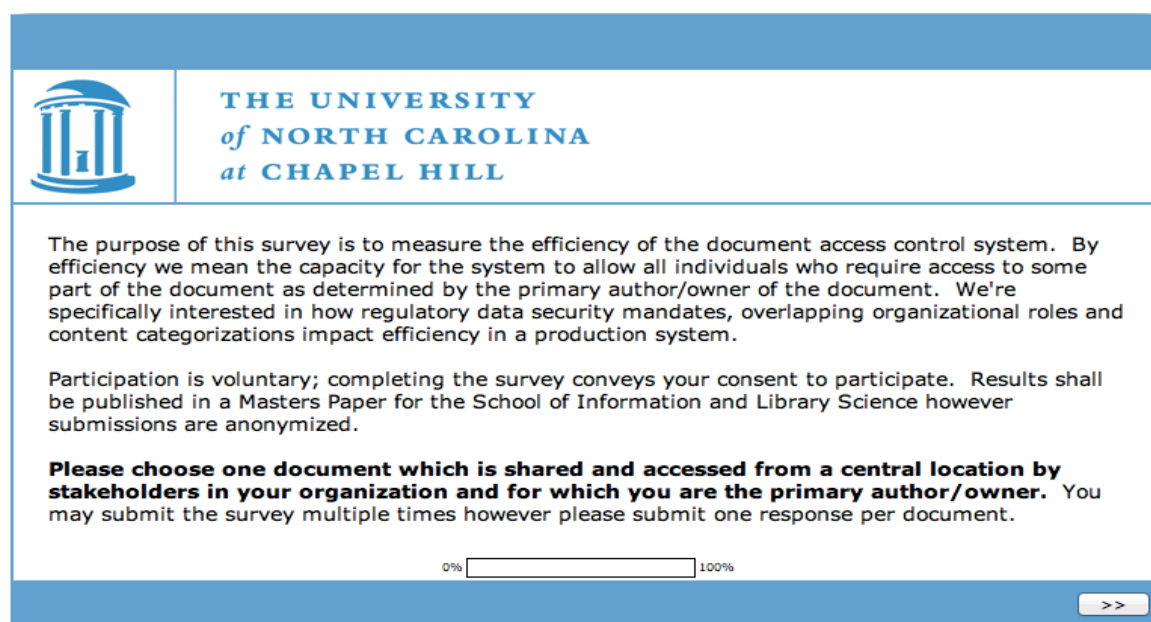
where a value of zero represents complete inefficiency and one represents perfect efficiency.

Several assumptions are made in order to carry out the research. First, the axiom $\beta > 0$ is adopted (i.e., at the very least the data steward has a need to know the information). Additionally, the assumption is made that $\beta \subseteq \alpha$ (i.e., the DAC is effective; no stakeholder outside the set of individuals for whom an operational need exists to access the document is authorized to do so). While consideration of the effectiveness of DACs is a worthy topic of future research, it is beyond the scope of the current inquiry. Finally, from $\beta > 0$ and $\beta \subseteq \alpha$ one may infer $\alpha > 0$.

To obtain the requisite data for the proposed research, an electronic survey (Qualtrics, Provo, UT) is distributed to the population of UNC faculty and staff via the University's mass mail protocol (UNC CIO, 2012). An initial invitation email, Addendum I, is sent on 2/6/2013 to the population to recruit participants and a second invitation, Addendum II, is sent on 2/13/2013. As of September 2012, the size of the population is estimated at 11,900 individuals (UNC Office of Institutional Research and Assessment, 2012).

To protect the privacy of participants and alleviate any burden of liability to disclose any behavior in violation of organizational policy potentially discovered during the course of the research, submissions are recorded anonymously; the only information collected from the survey besides the responses are the time at which the survey is retrieved from the Qualtrics system by the participant, the time at which the participant completes the survey and a unique session code to identify the survey. The survey attempts to measure four principal variables: efficiency, content variance as a function of organizational role, content variance as a function of organizational or regulatory classifications and willingness of the steward to use fine-grained access controls for the particular document. The survey consists of six questions, including an optional opportunity to leave feedback. A discussion follows of the survey content and measurement of variables.

Figure 1

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

The purpose of this survey is to measure the efficiency of the document access control system. By efficiency we mean the capacity for the system to allow all individuals who require access to some part of the document as determined by the primary author/owner of the document. We're specifically interested in how regulatory data security mandates, overlapping organizational roles and content categorizations impact efficiency in a production system.

Participation is voluntary; completing the survey conveys your consent to participate. Results shall be published in a Masters Paper for the School of Information and Library Science however submissions are anonymized.

Please choose one document which is shared and accessed from a central location by stakeholders in your organization and for which you are the primary author/owner. You may submit the survey multiple times however please submit one response per document.
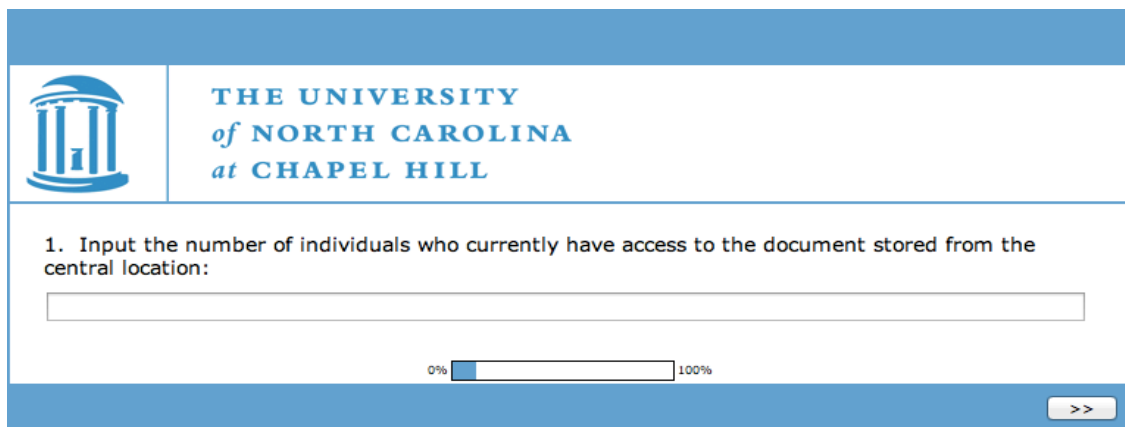
0% _____ 100%

>>

Prospective participants are greeted with the prompt in Figure 1 upon clicking the URL to the survey and authenticating with the shared password. The intent of the greeting is

threefold: to provide some background information to inform the participant of the

purpose of the study, to convey an informed consent notification and to instruct the

participant on how to select a relevant document for responding to questions in

subsequent sections. Upon clicking the arrow button in the bottom right corner of the

window, the survey proceeds to the next screen, in this instance Figure 2 (herein Q1).

Figure 2 / Q1
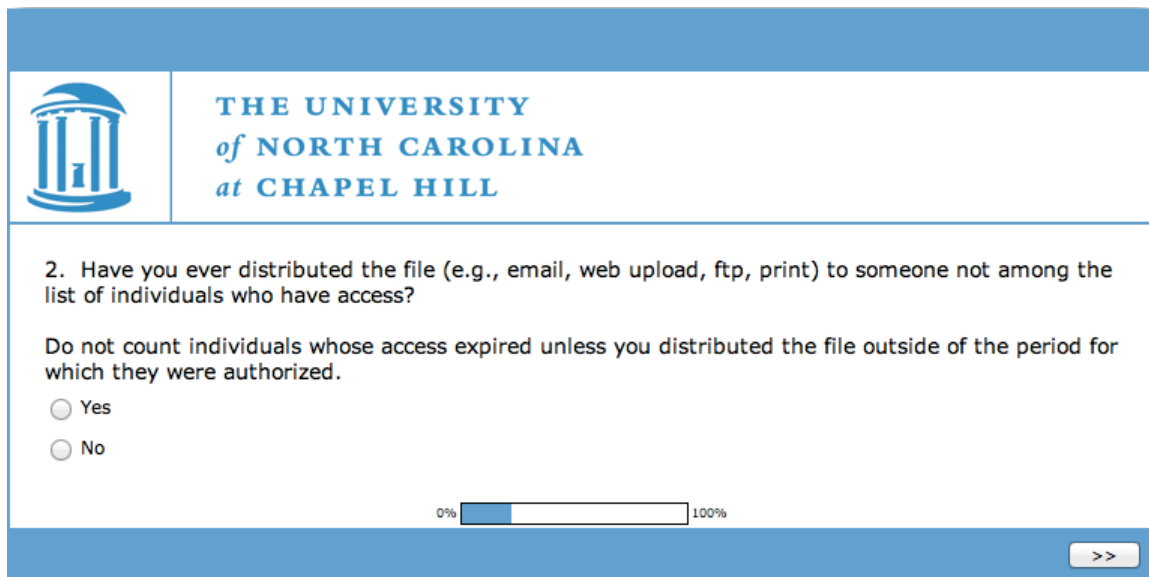
Q1 asks the participant to count the number of individuals with access to the file

and record the value in the input field as a whole number. A potential dilemma affecting

the validity of this approach is discussed in the Limitations section. Briefly, due to

inadequacies with the user interface and the segregation of the data steward and custodian

(i.e., individual responsible for implementing the access control) roles (c.f., "Access

controls must be defined by the Steward and implemented by the Custodian…" UNC

CIO, 2011, p. 12) the data steward may not be able to directly interpret the access control

and may rather rely on memory of their definition of the access control from a point in

time. To mitigate any inaccuracy deriving from this limitation it is stressed in Figure 1

that the participant selects a document for which they're the "primary author/owner" to

better ensure their familiarity and continuous involvement with the business and access

requirements.

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

2. Have you ever distributed the file (e.g., email, web upload, ftp, print) to someone not among the list of individuals who have access?

Do not count individuals whose access expired unless you distributed the file outside of the period for which they were authorized.

○ Yes

○ No
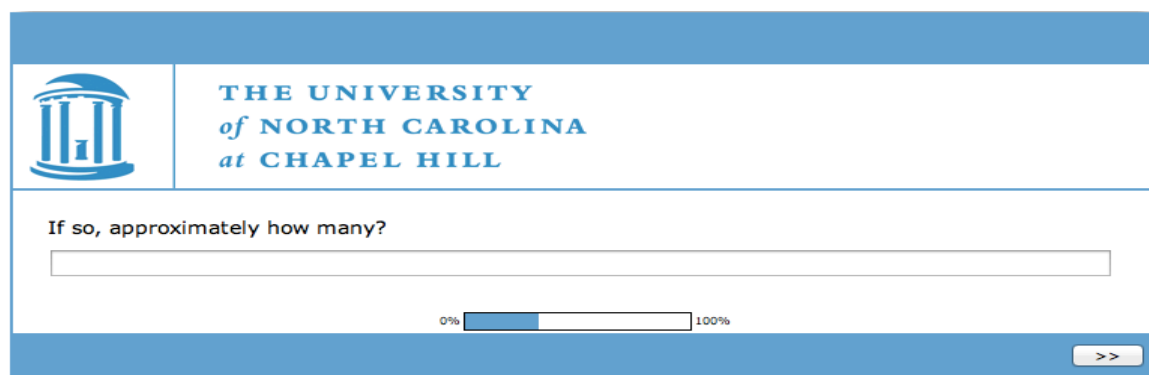
The objective of Figure 3 (herein Q2a) is to elicit evidence of access control

efficiency, or lack thereof, from an indicative behavior. The act of distributing the file

through the enumerated alternative methods amounts to circumvention of the access

control implementation and harkens back to the disadvantage expounded in the Literature

Review that the DAC model supports transitive delegation of authority to potentially less

secure destinations.

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL
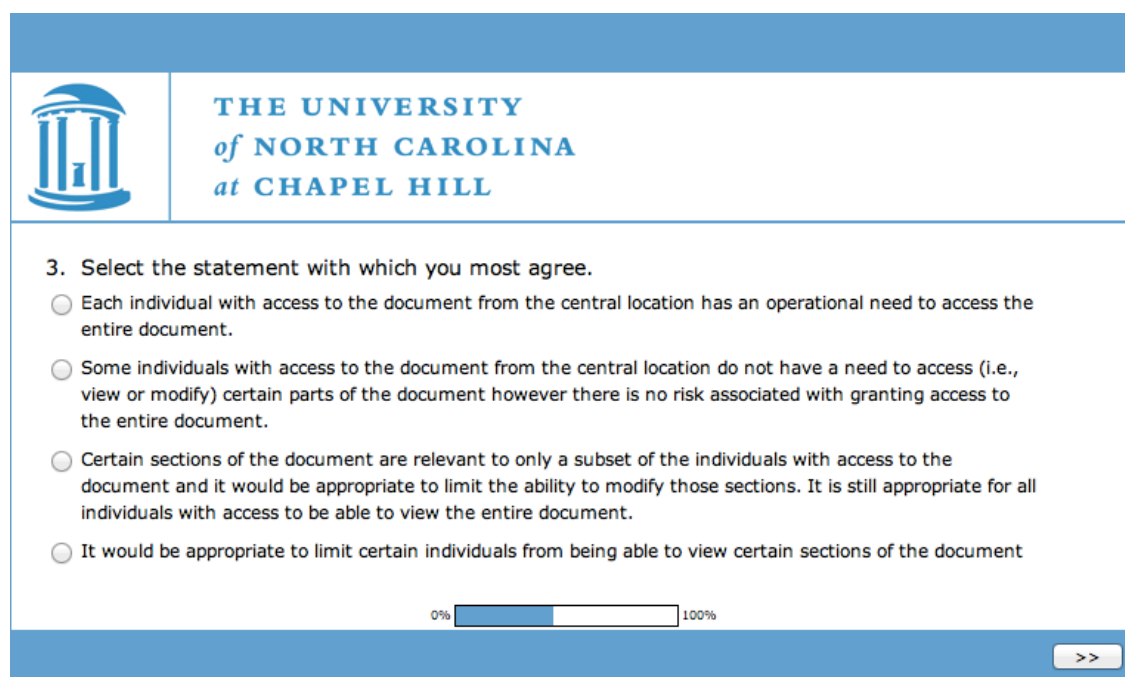
If so, approximately how many?

If the 'Yes' response is recorded the survey directs to Figure 4 (herein Q2b). Q2b allows the participant to further expand the answer by giving an approximation of how many people the file has been distributed to outside of those authorized by the access control in the form of a whole number. If the 'No' response is recorded the survey skips Q2b, coding the response as zero, and proceeds to Figure 5 (herein Q3).

Figure 5 / Q3

Q3 attempts to capture the aspect of content variation as a function of stakeholder roles. This is measured in terms of an ordinal scale. The first choice represents entire uniformity of content with respect to roles (i.e., weighted response is coded as zero) and the fourth represents high variation (i.e., weighted response is coded as three). The statements are meant to invoke the aspect of least privilege in relation to the individual's accessing the document and the information contained within the document.

Figure 6 / Q4

Figure 6 (herein Q4) conveys the document classification aspect of the research study in the form of an interval scale. Each classification is assigned a value of one and the summation is taken to represent the variance of content with respect to data classification. Thus, the value may range from one, since all University data is classified as either public or one of the five enumerated sensitive classifications, to six (i.e., the chosen document contains data covering all five sensitive categories as well as information serving public interest).

Additionally it should be observed the classifications are not mutually exclusive (e.g., a transcript of a teaching clinician's interview with a patient may contain PHI of the patient, FERPA if a student clinician is assisting in diagnosis and Non-public information if research is being conducted). Indeed it is the objective of the question to measure to what extent information assets exhibit multiple classifications within a single entity.

However it is presumed to be a rare occurrence that all six should be found in the same document.

Figure 7 / Q5a

The response from Figure 7 (herein Q5a) represents on a binary scale the willingness of the steward to utilize fine-grained access control for the document under scrutiny. The responses are coded identically to the scheme used in Q2. If the 'Yes' response is recorded the survey directs to Figure 8 (herein Q5b).

Figure 8 / Q5b
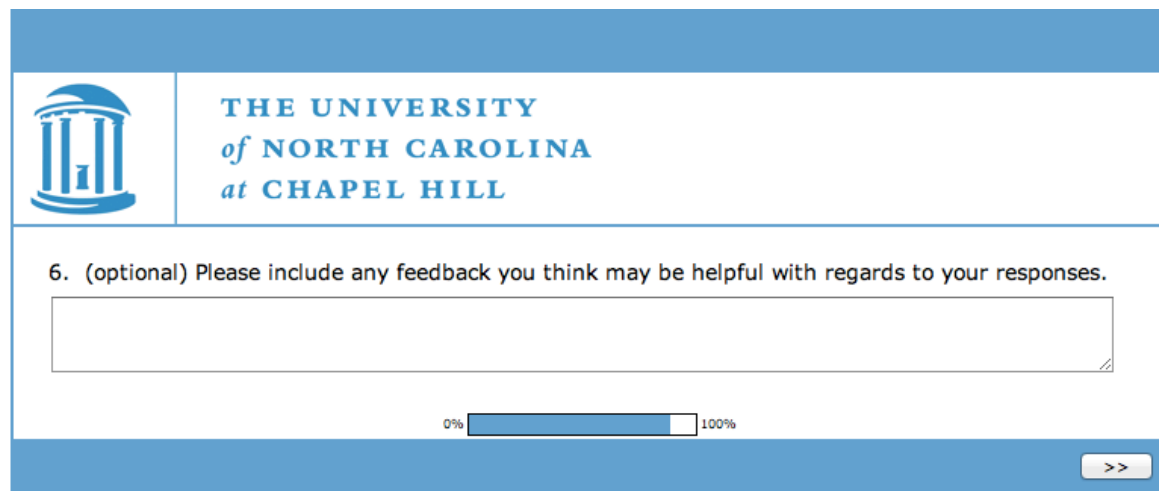
Figure 9 / Q6

Finally, the participant is presented with Figure 9 (herein Q6) and the conclusion of the survey. Q1, Q2 and Q5b represent ratio data for computing the efficiency measure (herein DAC ES).

$$DAC\ ES = \frac{Q1 - (Q2b + Q5b)}{Q1}$$

Q2 and Q5 attempt to couch the inquiry in terms of a behavioral response in order to reduce bias and encourage verisimilitude. The summation of the ratio values associated with these questions results in an approximation of β. Significantly, there is no explicit

lower bound to the efficiency score since the survey reduces the information from a comparison of access control sets to raw integers.

Bivariate analysis shall be conducted to determine relationships between the dependent variable, the DAC ES, and the independent variables represented by Q3, Q4 and Q5a. To reject the null hypothesis for H_2, significance is considered at the $p \leq .05$ level.

Limitations

The limitations of this research are manifold. Babbie (2010) discusses several relevant concerns related to survey research (p. 257 – 262). The dimensions of participant's willingness to respond and the perceived relevance of the topic are difficult to gauge conclusively though studies such as Albrechsten (2007) suggest security is a tertiary concern to end users next to convenience and operational efficiency. Data breaches at UNC such as the Carolina Mammography Registry incident lend credence to this assertion's applicability in the population under scrutiny (c.f., Barber, 2010, "[steward] was negligent in assigning security duties without granting additional training to [the custodian]"). Such circumstances may cultivate consternation for proffering information related to security practices though it is intended that the precautions taken to ensure anonymity mitigate this concern in prospective participants.

There are certainly reliability problems related to document classification. Different individuals may classify the same document differently or make different determinations of appropriate access as regards the least privilege doctrine based on a variety of factors: familiarity with internal policy and regulatory mandates, operational deadlines or other external pressures, misappraisal of associated risks, etc. This may

indicate a fundamental problem with the DAC model altogether; most assuredly the successful implementation of the model depends on the discretionary reliability of data stewards. It would be interesting to look at an effectiveness measure in future research considering accuracy of classification.

In the Methodology section, a validity limitation for Q1 was referenced. Specifically, it should be noted that if the individual responsible for implementing the access control follows published best practices for group nesting on Windows Server systems (e.g., Holme, Ruest, Ruest & Northrup, 2008, p. 153 – 155; Microsoft TechNet, 2005), the data steward may not be able to interpret who currently has access to the particular file since only the domain local groups for the resource (i.e., those groups dedicated to expressing a distinct access privilege for the respective object) are visible to the data steward from the access control entry interface.

For instance, Figure 10 displays the common access control entry user interface on Microsoft Windows systems; in order to accurately interpret who has modify access to the file solely based on information obtained from this view it would be necessary for the data steward to issue a query such as the one depicted in Figure 11. Dyché (2009) posits that the exact nature of data stewardship will differ across organizations given the specific problems the governance initiative is attempting to address though "as IT educates on the value of data stewardship, it's important the role is seen as a bridge between IT and the business…deploying information in a consistent and structured way…" (p. 12). How organizations deal with the practical issues of security group maintenance and align this with a unified data governance strategy may be a valuable topic for future research though it is not fully considered in this paper. In order to not

presume a particular procedure on any of the varied participant pools surveyed by this

research, the ideal DAC recalled by the participant is held as authoritative.

Figure 10
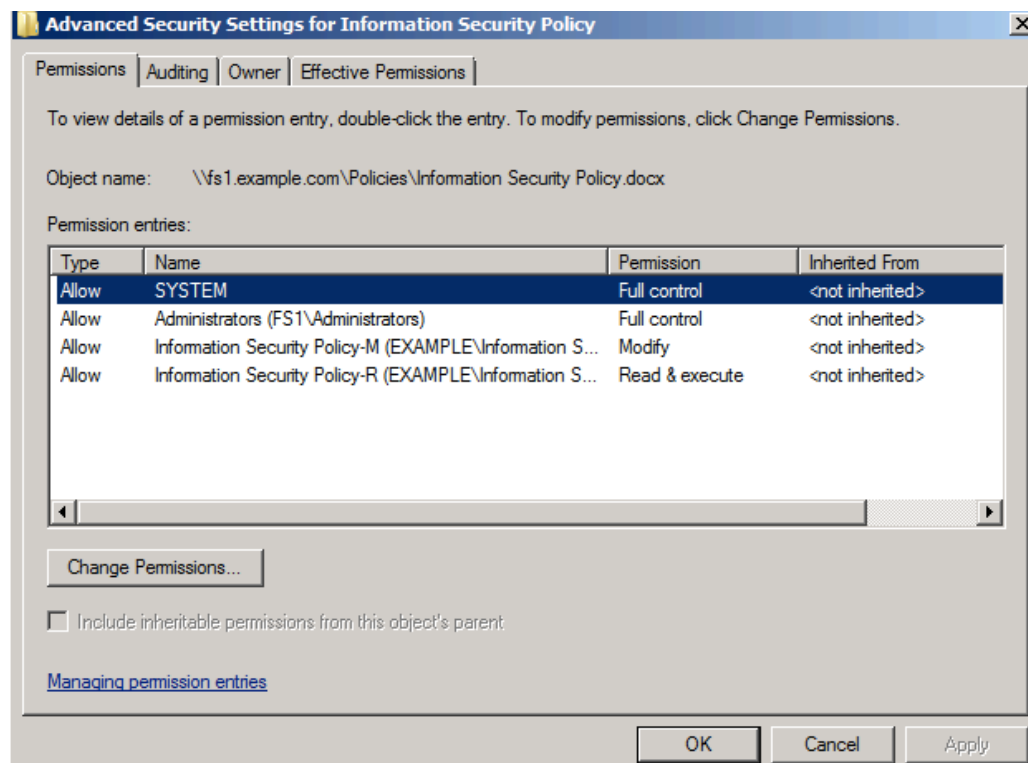
Figure 11

External validity is another limitation inherent in research related to this topic. Although information classification and access control practices are influenced by generally applicable variables such as regulatory standards and commercially available software features, the character and quality of these concepts is perhaps determined to a greater extent by the organizational culture: a small firm founded from venture capital may decide to grant everyone access to everything since loss of efficiency can result in the death of the company while a national intelligence agency may operate at a high loss of access control efficiency since leaked information can result in the death of an employee. A meta-analysis of data collected from a variety of organizations may begin to converge on generalizable results however this information would attenuate the characteristics of document classification and content variance particular to any given organization and therefore be of less utility in terms of developing an appropriate access control protocol for the organization.

Analysis

Our hypotheses predict the survey data should demonstrate that access control applied to an entire document is inefficient when multiple information classifications are subsumed and stakeholders of varying security clearances, or organization roles, are served by the document. The research stands to benefit UNC directly by offering better insight into access control practices. The claims may be extended to other NC higher education, public institutions though, as mentioned earlier, generalizability is problematic due to individual variations with respect to organizational culture. For instance, a particular NC higher education, public institution may not deal with HIPAA regulated

information to the extent UNC does if the institution lacks a hospital or medical research department.

The survey received 91 submissions total however only 52 were complete. The 39 incomplete responses were discarded from the response set. Table 1 summarizes the queries and DAC ES expounded in the Methodology section while Figure 12 depicts the distribution of the responses.

Table 1

Item	Description
Q3	To what degree does the respondent consider the chosen document's content to conform to the least privilege requirement with respect to all stakeholder roles accessing the document? Full Uniformity implies no benefit from redaction. Insignificant Disparity implies redaction would accomplish least privilege but the perceived risk value of content is low so benefit is minimal. Marginal Disparity indicates that certain individuals should be prevented from modifying parts of the document and implies an integrity benefit. Significant Disparity indicates certain individuals should be prevented from viewing parts of the document and implies a benefit to confidentiality.
Q4	How many different types of distinct regulatory or institutional classifications?
Q5a	Would the respondent use a redaction system to share parts of the document content with individuals who do not currently have access?
DAC ES	Subtract the number of individuals who have been granted access to the document through circumvention of the document access control (i.e., Q2a) plus any individuals counted in Q5a from the total number of individuals with access (i.e., Q1) and divide by Q1 to produce ratio of relative efficiency reflecting the degree to which stakeholders are denied access to relevant content due to lack of precision of the document access control.

Figure 12
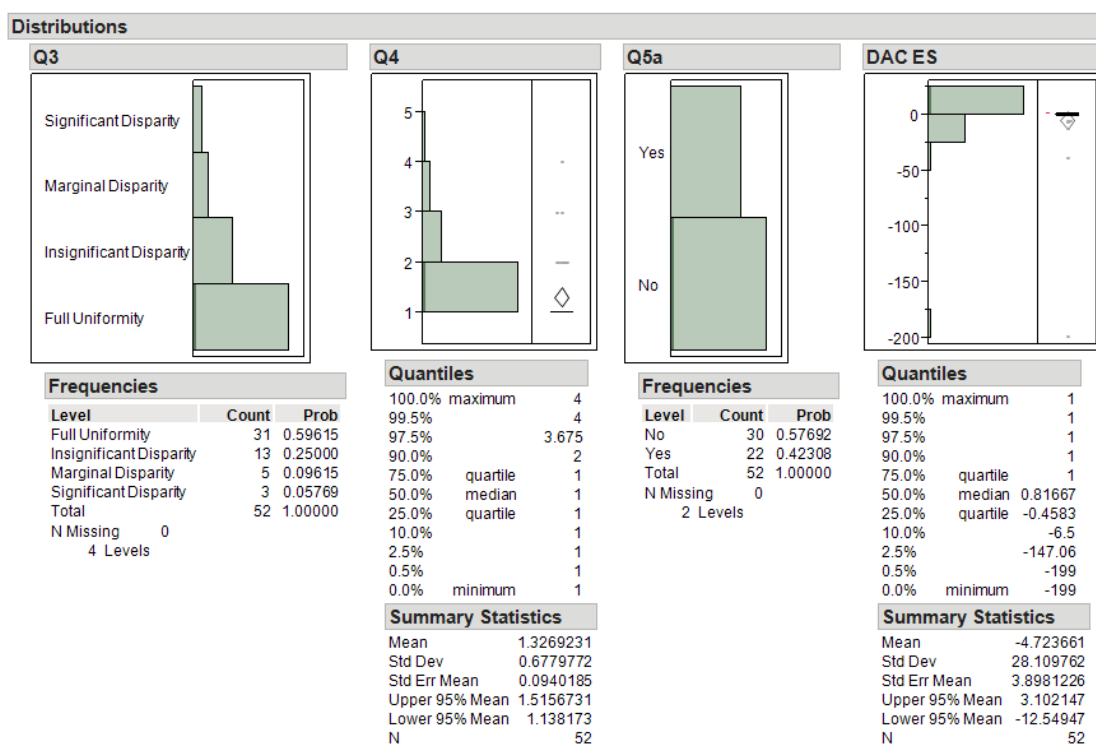
The responses for Q3 were re-labeled for clarity. The Full Uniformity label corresponds

with response one, Insignificant Disparity with response two and so on. Further,

collapsing the responses of Q3 into a binary nominal scale between uniformity and

disparity the data shows that, in the majority of surveyed cases, the steward judged the

content to be uniform in consideration of the business need to know of the various

stakeholders with whom the document is shared by a ratio of 31:21. As anticipated Q4

conveys that of the surveyed documents, the majority of cases were reported to contain

only one regulatory classification of information. None of the surveyed cases were

reported to contain more than four of the classifications. The DAC ES resulted in two

outliers, which skewed the data by conveying a negative mean (i.e., majority of cases

were inefficient). In fact, about 42% of the cases were purported to be efficient (i.e.,

DAC ES = 1).

Figure 13
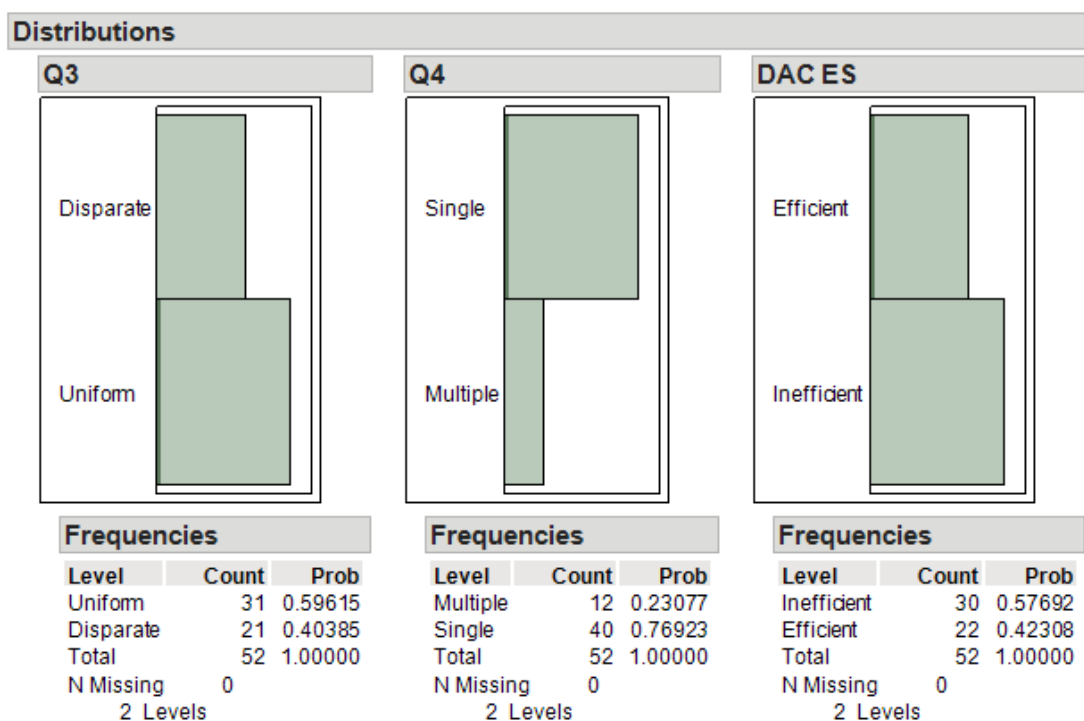
Reflecting on the distribution of responses the DAC ES, Q3 and Q4 variables are reduced to binary nominal data. In DAC ES, all values below one are grouped into the Inefficient category and all values equal to one are Efficient. In Q4, all values above one are grouped as Multiple Classification while all values equal to one are grouped as Single Classification. For Q3 all of the disparity categories are grouped together. The results of this are displayed in Figure 13.

Thus, with all variables represented in a binary nominal form the x^2 test is used to determine if a relationship exists between the DAC ES and the independent variables represented by Q3, Q4 and Q5a. The results are represented in Figure 14. A discussion and interpretation of the data follows.

Figure 14
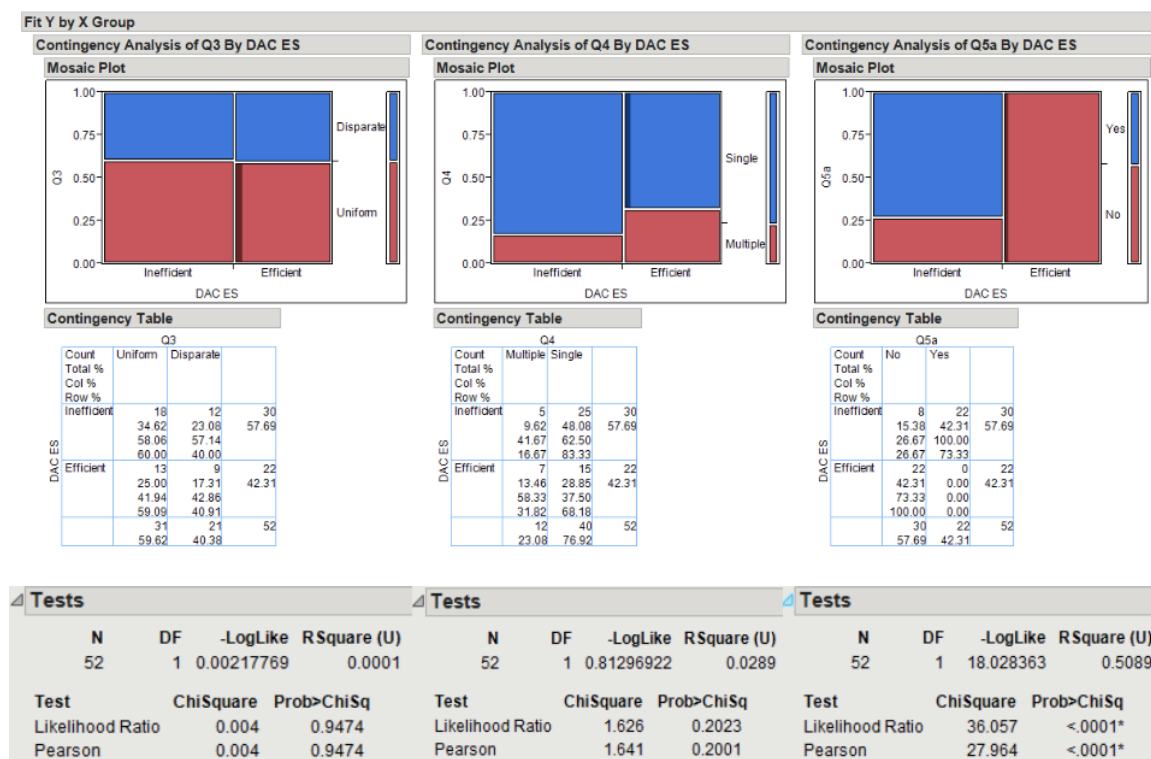
The survey data supports a positive relationship between Q5a, the steward's

willingness to use a redaction mechanism and the efficiency of the current access control

associated with the document (i.e., the steward is likely to state they would use redaction

if the current document access control prevents stakeholders from accessing relevant

information while they're unlikely to express willingness where the document access

control is judged to be efficient). One would logically anticipate such an observation.

Interestingly, though no steward purported they'd use redaction controls on a document

with an efficient DAC, eight cases claimed they'd not use redaction with an inefficient

DAC. This may be interpreted in several ways: it may indicate the latent conflict

between security and functionality noted by Albrechsten (2007), a mismatch between the

steward's perceived risk of sharing information and the benefit of sharing the information

with other stakeholders, or perhaps these are simply cases where the utility of sharing additional information is judged not to be worth the expected cost in effort of marking up the document at a more granular degree.

Figure 15 aggregates the Q5a and DAC ES contingency table by Q3 and Q4 respectively to further investigate this observation. The intent is to show among the respondents who answered they would not use redaction on an inefficient DAC to share information with stakeholders precluded from accessing the document, how many of these documents reportedly contained multiple regulatory classifications or disparate role versus content alignment. If the majority of responses did not fall into either category one might draw the conclusion that the effort versus perceived utility explains the response since these factors do not necessarily indicate a need to specify further access control. While the distribution primarily falls into the Single regulatory classification category, at a ratio of 7:1, the responses are evenly distributed along the dimension of role alignment and document content. However, one may argue that neither presents a persuasive result to explain the underlying contributing factor given the low number of responses.

Figure 15

Contingency Analysis of Q5a By DAC ES Q3=Uniform

Contingency Table

Q5a			
Count / Total % / Col % / Row %	No	Yes	
Inefficient	4 / 12.90 / 23.53 / 22.22	14 / 45.16 / 100.00 / 77.78	18 / 58.06
Efficient	13 / 41.94 / 76.47 / 100.00	0 / 0.00 / 0.00 / 0.00	13 / 41.94
	17 / 54.84	14 / 45.16	31

Contingency Analysis of Q5a By DAC ES Q4=Multiple

Contingency Table

Q5a			
Count / Total % / Col % / Row %	No	Yes	
Inefficient	1 / 8.33 / 12.50 / 20.00	4 / 33.33 / 100.00 / 80.00	5 / 41.67
Efficient	7 / 58.33 / 87.50 / 100.00	0 / 0.00 / 0.00 / 0.00	7 / 58.33
	8 / 66.67	4 / 33.33	12

Contingency Analysis of Q5a By DAC ES Q3=Disparate

Contingency Table

Q5a			
Count / Total % / Col % / Row %	No	Yes	
Inefficient	4 / 19.05 / 30.77 / 33.33	8 / 38.10 / 100.00 / 66.67	12 / 57.14
Efficient	9 / 42.86 / 69.23 / 100.00	0 / 0.00 / 0.00 / 0.00	9 / 42.86
	13 / 61.90	8 / 38.10	21

Contingency Analysis of Q5a By DAC ES Q4=Single

Contingency Table

Q5a			
Count / Total % / Col % / Row %	No	Yes	
Inefficient	7 / 17.50 / 31.82 / 28.00	18 / 45.00 / 100.00 / 72.00	25 / 62.50
Efficient	15 / 37.50 / 68.18 / 100.00	0 / 0.00 / 0.00 / 0.00	15 / 37.50
	22 / 55.00	18 / 45.00	40

The data does not support rejection of the null hypothesis for Q3 or Q4. Tertiary factors such as sampling error based on the fact that the responses only reflect individuals in the population willing to complete the survey, the ability of the data steward to accurately classify information, or the failure of the survey to not capture valid responses cast doubt on the findings. Sentiment analysis of Q6 indicates at least four of the respondents found the intent or presentation of the survey confusing. Of these, two answered they would use redaction to share the document with a wider audience but didn't indicate with how many individuals they'd include (i.e., Q5b for these respondents was zero). This inconsistency supports Albrechsten's (2007) interpretation of a gap

between security talk and action among employees where few security tasks are performed in their current roles yet the employees express motivation to perform such actions (p. 284).

Twenty responses included a comment for Q6. Figure 16 depicts a sentiment analysis of the comments along the following dimensions: redaction is not relevant to the respondent's role, the respondent expresses a positive opinion of using redaction in a workflow, the respondent expresses a negative opinion towards redaction. A value of one indicates the sentiment is expressed in the comment while zero indicates the sentiment is not expressed by the comment. The sum attribute represents overlapping sentiments in the individual comments.

Distributions

Not Relevant			Positive			Negative			sum		

Frequencies

Not Relevant			Positive			Negative			sum		
Level	Count	Prob	Level	Count	Prob	Level	Count	Prob	Level	Count	Prob
0	16	0.80000	0	11	0.55000	0	16	0.80000	0	6	0.30000
1	4	0.20000	1	9	0.45000	1	4	0.20000	1	11	0.55000
Total	20	1.00000	Total	20	1.00000	Total	20	1.00000	2	3	0.15000
N Missing	32		N Missing	32		N Missing	32		Total	20	1.00000
2 Levels			2 Levels			2 Levels			N Missing	32	
									3 Levels		

Among the six comments expressing none of the enumerated sentiments (i.e., sum = 0), three convey confusion over the survey, one lauds the sharing capability of the Dropbox service while mentioning they take care not to use it for "protected information," one cites a procedure for sending a document for manual redaction to the legal council's office in the event of public information request and one indicates that

while the data in the document under consideration is not "confidential" they "tend not to share it outside of the committee." The latter respondent categorized the document as public in Q4 and indicated that if redaction were available the document could be shared with approximately ten additional individuals. One assumes then this document represents the class of using access control as a way of suppressing information that is judged to be non-relevant to roles outside of a group yet the indication that redaction may be considered suggests that some of the information in the document may be classified as "company confidential" (i.e., the ten individuals should not be able to view the redacted information). This implies a further need to extend the classification scheme to include a category for "company confidential" (i.e., information that wouldn't cause harm if exposed though is generally not relevant to positions outside of a given set of roles).

Of the three comments expressing two of the enumerated sentiments, one falls into the not relevant to the respondent's role and positive categories: "Not necessary for my work but very cool feature :)." The other two express both positive and negative opinions. The first states, "I think I'd simply rather not share a document than redact portions and share it, but if I could go back to the previous question I'd probably change my answer to yes." This comment was also counted among the comments expressing confusion over the survey since they indicated a desire to change an answer upon completing the survey; the other three comments conveying confusion with the survey were more explicit (e.g., "I didn't feel this survey was very clear as to the intent of the questions."). The second of these comments reads,

> "I wouldn't use a tool that required me to manually redact each item. If there were a tool to do this semi-automatically (afterwhich I could do quality control) and also could check to make sure information would no longer lead to identification of individuals, I would consider it - depending on how much time would be

required to process the data and how much time were available to me to learn how to use tools and apply them."

The latter comment encapsulates many important criteria for a redaction access control to meet in order to better ensure stakeholder satisfaction and underscores features expounded in Bier et al. (2009). For example, the respondent indicates the system should provide some functionality for assisting with the detection of candidate elements for redaction though ultimately the utility of such a feature is dependent on time involved in using the system and the learning curve associated with the new technology.

Another interesting trend is the discussion of workarounds to the problem. Five comments allude to the practice of manually partitioning documents based on sensitive information and utility of sharing subsets of classified data with certain stakeholders. Two of these cite the opinion that redaction has a negative connotation, that it is distracting and may cause resentment among stakeholders, and they'd prefer either not to share a document or manual partition it into multiple documents so that no data is ostensibly obfuscated. In contrast one respondent cites this practice and laments "Right now we have multiple versions of files and it gets confusing/potential for errors and inadvertent disclosure." Yet another describes a process of affixing confidential notes to an electronic document after it has been printed to keep the two separate. The fifth comment in this category is the individual who mentioned sending the document to legal council for redaction. This comment didn't necessarily have a negative or positive opinion of the process.

These distinctions reveal some interesting differences in organizational cultural which deserve consideration prior to deployment of a redaction access control system. To address the negative sentiments, alternative approaches such as the Active

Enforcement technology of Agrawal and Johnson's (2007) Hippocratic Database System may be used for internal stakeholders to re-write queries posed to an institutional repository search engine based on role and business rules so that redaction is accomplished in a more subtle way. Further, the analysis suggests many avenues for future research. Clearly prototyping of a system among stakeholders in sensitive roles that routinely interface with other stakeholder roles which do not possess a need to know for the types of information accessed by the primary group. Doctors who work with researchers and students are one example. Several different implementation approaches are suggested from the survey data including an approach for redacting documents by section to produce multiple versions, dynamic views of a document to improve upon the manual process of partitioning and managing multiple versions of a document. More data on measuring the aptitude of various stakeholder groups to accurately classify documents based on all relevant regulations, statutes and organizational policies would also be invaluable for tuning such a system.

It is the position of this research to support the accepted practice of granting least privilege to information assets however the contention is made that this practice is untenable in the current environment. Data stewards are placed in the intractable position of reconciling layers of policy legalese with quotidian operational needs. If it is the case that the access control mechanism does not facilitate reconciliation in every case, one may expect inconsistent results placing the organization at risk of lost efficiency through interruption to workflow confluence.

Conclusion

Information classification is a critical function for organizations, specifically those processing sensitive information. Access control is the primary method for instantiating classification for information assets. Through a case study of a DAC implementation, this research explores whether the widely implemented practice of defining access control at the document layer is efficient. An efficiency model is proposed which accounts for individuals who have a need to access discrete segments of a document in proportion to individuals who have a need to access the entire document. Data stewards are surveyed in order to determine to what extent they consider the document as one logical unit or an amalgamation of classifications with multifaceted segments of varying utility to diverse sets of stakeholders. The initial hypothesis anticipated that the document-centric approach, though perhaps sufficient for certain innocuous information or documents of strict content uniformity, is not efficient when content varies in classification. Our research does not cogently support such a claim in an absolute sense, however it does suggest that there is awareness among stewards as to inefficiencies where potentially valuable information is withheld from stakeholders due to current access controls and in such cases the stewards are willing to use a redaction access control. Integrating redaction into document processing stands to efficiently facilitate granular data classification while improving the availability of relevant information to key stakeholders. The challenge remains to introduce such a system and cultivate acceptance among data stewards.

Bibliography

Albrechtsen, E. (2007). "A qualitative study of users' view on information security." *Computers & Security*, 26(4), 276-289. http://dx.doi.org/10.1016/j.cose.2006.11.004

Agrawal, R., & Johnson, C. (2007). "Securing electronic health records without impeding the flow of information." *International Journal of Medical Informatics*, 76(5), 471-479.

Babbie, E. R. (2010). "Survey Research" in *The Practice of Social Research* (12th ed.). Belmont, CA: Wadsworth Cengage.

Baracaldo, N. & Joshi, J. (2012). "A trust-and-risk aware RBAC framework: tackling insider threat." In Proceedings of the 17th ACM symposium on Access Control Models and Technologies (SACMAT '12). ACM, New York, NY, USA, 167-176. DOI=10.1145/2295136.2295168 http://doi.acm.org.libproxy.lib.unc.edu/10.1145/2295136.2295168

Barber, R. C. (8 October 2010). "Researcher Yankaskas appeals pay cut, demotion." *The Daily Tar Heel*. Retrieved from: http://www.dailytarheel.com/article/2010/10/researcher_appeals_pay_cut_demotion

Bier, E., Chow, R., Golle, P., King, T.H., Staddon, J.; (2009) "The Rules of Redaction: Identify, Protect, Review (and Repeat)." Security & Privacy, IEEE , vol.7, no.6, pp.46-53, Nov.-Dec. 2009 doi: 10.1109/MSP.2009.183 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5370699&isnumber=5370689

Boehmer, W. (August 2008). Appraisal of the effectiveness and efficiency of an Information Security Management System based on ISO 27001. In Emerging Security Information, Systems and Technologies, 2008. SECURWARE'08. Second International Conference on (pp. 224-231). IEEE.

Cheng, W., et al. (June 2012) "Abstractions for Usable Information Flow Control in Aeolus" In Proceedings of the 2012 USENIX Annual Technical Conference, (Boston, MA, USA). Retrieved from http://pmg.csail.mit.edu/papers/aeolus-usenix.pdf

Dych, J. (1 May 2009). "DATA STEWARDSHIP STRATEGY: 6 Keys to Success." *Information & Management*, 19(4), 10.

El Emam, K., et al. (2009) "The Inadvertent Disclosure of Personal Health Information Through Peer-to-Peer Sharing Programs." *Journal of the American Medical Informatics Association*, 17, pp 148-158.

Elmasri, R. & Navathe, S. B. (2007). "Database Security" in *Fundamentals of Database Systems*. (p. 795-820). Pearson Addison Wesley.

Forrester, J., & Irwin, B. (2005). "An Investigation into Unintentional Information Leakage through Electronic Publication." *Information Security South Africa*. Accessed from: http://icsa.cs.up.ac.za.libproxy.lib.unc.edu/issa/2005/Proceedings/Poster/012_Article.pdf

Hassell, L. (2005). "Affect and Trust." *Trust Management*, (p. 131-145). Accessed from: http://www.springerlink.com/index/yk5drc2yxv2b7pkm.pdf

Holme, D., Ruest, N., Ruest, D., & Northrup, T. (2008). "Developing a Group Management Strategy" in *MCTS Exam 70-640 Configuring Windows Server 2008 Active Directory*. (p. 153 – 155). Redmond, WA: Microsoft Press.

Hu, V.C., Ferraiolo, D.F. & Kuhn D.R.; (September 2006) "Assessment of Access Control Systems." Interagency Report 7316. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology. Gaithersburg, MD. Retrieved from http://csrc.nist.gov/publications/nistir/7316/NISTIR-7316.pdf

JMP, Version 10. SAS Institute Inc., Cary, NC, 1989-2013.

Johnson, M.E., (2009) "Data Hemorrhages in the Health-Care Sector" in Lecture Notes in Computer Science: Financial Cryptography and Data Security (p. 71-89). Springer Berlin / Heidelberg.

Liptak, A., (23 June 2006) "Prosecutors Can't Keep a Secret in Steroid Case." *The New York Times.* Retrieved from: http://www.nytimes.com/2006/06/23/us/23leak.html?_r=2&oref=slogin&

Microsoft Corporation (2005). Access-based Enumeration. Retrieved from http://download.microsoft.com/download/4/9/8/498EEEEF-97B0-450E-8E56-26105D4B092E/Accessbasedenum.doc

Microsoft Developer Network (2007). Windows Vista Integrity Mechanism Technical Reference. Retrieved from http://msdn.microsoft.com/en-us/library/bb625957.aspx

Microsoft TechNet (2005). "Best Practices for Shared Folders." Retrieved from: http://technet.microsoft.com/en-us/library/cc780313(v=ws.10).aspx

Mutch, J. & Anderson B. (2011) "Security Does Not Equal Compliance" in Preventing Good People from doing Bad Things. (p. 141 - 161). Apress. http://dx.doi.org/10.1007/978-1-4302-3922-2_9

National Security Agency (2009). Secutiry-Enhanced Linux. Retrieved from http://www.nsa.gov/research/selinux/

O'Connor, A.C. & Loomis, R.J.; (December 2010). "2010 Economic Analysis of Role-Based Access Control." Retrieved from http://csrc.nist.gov/groups/SNS/rbac/documents/20101219_RBAC2_Final_Report.pdf

Ponemon Institute LLC. (January 2012) "Aftermath of a Data Breach Study." Retrieved from http://www.experian.com/assets/data-breach/brochures/ponemon-aftermath-study.pdf

Ponemon Institute LLC. (March 2012) "2011 Cost of Data Breach Study: United States." Retrieved from http://www.symantec.com/content/en/us/about/media/pdfs/b-ponemon-2011-cost-of-data-breach-us.en-us.pdf?om_ext_cid=biz_socmed_twitter_facebook_marketwire_linkedin_2012Mar_worldwide__CODB_US

Porter, M. E., & Millar, V. E. (July – August 1985). "How information gives you competitive advantage." *Havard Business Review* (p. 149-160). Retrieved from http://www.ida.liu.se.libproxy.lib.unc.edu/~TDEI65/documents/8500002422.pdf

Rjaibi, W. & Bird, P. (2004). A multi-purpose implementation of mandatory access control in relational database management systems. VLDB '04 Proceedings of the

Thirtieth international conference on Very large data bases, Volume 30. Retrieved from http://dl.acm.org.libproxy.lib.unc.edu/citation.cfm?id=1316776

Samba (17 May 2010) "smb.conf - The configruation file for the Samba suite." Retrived from http://www.samba.org/samba/docs/man/manpages-3/smb.conf.5.html

Sanyal, S., Shelat, A. & Gupta, A. (2010). "New Frontiers of Network Security: The Threat Within." In Proceedings of the 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems (VCON '10). IEEE Computer Society, Washington, DC, USA, 63-66. DOI=10.1109/VCON.2010.19 http://dx.doi.org.libproxy.lib.unc.edu/10.1109/VCON.2010.19

Staddon, J., Golle, P., Gagné, M., & Rasmussen, P., (2008). "A content-driven access control system." In Proceedings of the 7th symposium on Identity and trust on the Internet (IDtrust '08). ACM, New York, NY, USA, 26-35. DOI=10.1145/1373290.1373296 http://doi.acm.org/10.1145/1373290.1373296

Tiller, J.S. & Fried S. (2010). "Access Control" in Tipton H. F. (Ed), *Official (ISC)² Guide To The CISSP CBK*. (p. 1-154). Boca Raton, FL: Auerbach Publications.

UNC CIO (30 June 2011). "Information Security Policy." Retrieved from http://its.unc.edu/files/2012/03/ccm1_033440.pdf

UNC CIO (11 October 2012). "UNC Chapel Hill Policy on Mass Mail." Retrieved from http://help.unc.edu/help/unc-chapel-hill-policy-on-mass-email/

UNC Office of Institutional Research and Assessment (December 2012). "Employees by Category, Fall 2012." Retrieved from: http://oira.unc.edu/employees-by-category-fall.html

Wiley, D. (1 May 2005). "Italy media reveals Iraq details." *BBC News.* Retrieved from: http://news.bbc.co.uk/2/hi/europe/4504589.stm

The gathered data for this paper was generated using Qualtrics software, Version 38761 of the Qualtrics Research Suite. Copyright © 2013 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. http://www.qualtrics.com

Addendum I

Received: from mxip1i.isis.unc.edu (152.2.0.74) by ITS-MSXHT0.ad.unc.edu
 (172.27.172.65) with Microsoft SMTP Server (TLS) id 14.2.328.9; Wed, 6 Feb
 2013 21:35:01 -0500
X-RemoteIP: 152.2.1.138
X-Group: OVERRIDELIST
X-Policy: $BYPASS_SBRS
X-MID: 930677954
X-SBRS: 1.6
X-IronPort-Anti-Spam-Filtered: true
X-IronPort-Anti-Spam-Result:
ArYGAH0SE1GYAgGKdGdsb2JhbABFDq4LkjcWDgEMFQg7gh8BAQEFAQEVbyM
LDwwKDwxDEhuHdgyuR4UqiQoEjQ2BIYMnA4hmjTuTElyBUQ
X-IronPort-AV: E=Sophos;i="4.84,619,1355115600";
  d="scan'208";a="930677955"
Received: from notify.isis.unc.edu ([152.2.1.138]) by mxip1i.isis.unc.edu
 with ESMTP; 06 Feb 2013 21:35:01 -0500
Received: from notify.isis.unc.edu (localhost [127.0.0.1])   by
 notify.isis.unc.edu (8.13.6/8.14.3) with ESMTP id r172Yx6i028726;   Wed, 6 Feb
 2013 21:34:59 -0500 (EST)
Received: (from root@localhost)   by notify.isis.unc.edu (8.13.6/8.13.6/Submit
 id r172YxBq028725; Wed, 6 Feb 2013 21:34:59 -0500 (EST)
Date: Wed, 6 Feb 2013 21:34:59 -0500
Message-ID: <201302070234.r172YxBq028725@notify.isis.unc.edu>
Subject: Message 10642 Sent: INFORMATIONAL: Seeking Participants for Information
Security Study
To: <massmail-employees@listserv.unc.edu>, <deric_freeman@unc.edu>
From: <massmail@unc.edu>
MIME-Version: 1.0
Content-Type: text/plain
Return-Path: root@notify.isis.unc.edu
X-MS-Exchange-Organization-AuthSource: ITS-MSXHT0.ad.unc.edu
X-MS-Exchange-Organization-AuthAs: Anonymous
X-MS-Exchange-Organization-AVStamp-Mailbox: MSFTFF;1;0;0 0 0

Hi,

I'm contacting you today as a graduate student of the UNC School of
Information and Library Science. Under the supervision of faculty
advisor Arcot Rajasekar, I am conducting research on the impact of

current document access control practices to influence sufficient information sharing within heavily regulated environments such as UNC's teaching and research activities. This brief, voluntary survey asks you to answer questions related to a shared document for which you're attributed the role of primary data steward or custodian (i.e., individual responsible for interpreting who has a business need to access the information).

All responses are anonymized prior to submission. The objective of our research is to form an understanding of access control decisions in a production system and use this knowledge to test our hypothesis on the use of alternative technologies which may offer improved availability of information and end-user control as compared to the current methods. This research has received approval from the UNC Non-Biomedical IRB under the title "Describing
the Impact of Document Content Variance on Access Control Efficiency and A Proposed Solution for Improving Efficiency: Fine-grained, Redactive Access Control Models", IRB #13-1008 approved on 1/31/2013 non-biomedical

Thank you for your valuable consideration in this effort.

Survey Link: https://unc.qualtrics.com/SE/?SID=SV_1F9S23d6gd8KOZ7
Password: 8capSoap#

PI: Deric Freeman
Email: deric_freeman@unc.edu
Phone: 9199669171

FA: Arcot Rajasekar
Email: rajasekar@unc.edu
Phone: 9199663611

This email is sponsored by: School of Information and Library Science

Addendum II

Received: from mxip3i.isis.unc.edu (152.2.2.195) by ITS-MSXHT0.ad.unc.edu
 (172.27.172.65) with Microsoft SMTP Server (TLS) id 14.2.328.9; Wed, 13 Feb
 2013 21:14:10 -0500
X-RemoteIP: 152.2.1.138X-Group: OVERRIDELIST
X-Policy: $BYPASS_SBRS
X-MID: 876598432
X-SBRS: 1.6
X-IronPort-Anti-Spam-Filtered: true
X-IronPort-Anti-Spam-Result:
AnYHADpHHFGYAgGKdGdsb2JhbABEDq4fkkIWDgEMFQg7gh8BAQEFAQEVWx
QjCyUPDEMSG4d3tmyJDwSNNoRKA4hmjT6TGl2BVA
X-IronPort-AV: E=Sophos;i="4.84,660,1355115600";
  d="scan'208";a="876598433"
Received: from notify.isis.unc.edu ([152.2.1.138]) by mxip3i.isis.unc.edu
 with ESMTP; 13 Feb 2013 21:14:10 -0500
Received: from notify.isis.unc.edu (localhost [127.0.0.1])   by
 notify.isis.unc.edu (8.13.6/8.14.3) with ESMTP id r1E2DtSw008259;   Wed, 13 Feb
 2013 21:13:55 -0500 (EST)
Received: (from root@localhost)   by notify.isis.unc.edu (8.13.6/8.13.6/Submit)
 id r1E2DtRS008258; Wed, 13 Feb 2013 21:13:55 -0500 (EST)
Date: Wed, 13 Feb 2013 21:13:55 -0500
Message-ID: <201302140213.r1E2DtRS008258@notify.isis.unc.edu>
Subject: Message 10667 Sent: INFORMATIONAL: Last Chance to Participate in
Information Security Study
To: <massmail-employees@listserv.unc.edu>, <deric_freeman@unc.edu>
From: <massmail@unc.edu>
MIME-Version: 1.0
Content-Type: text/plain
Return-Path: root@notify.isis.unc.edu
X-MS-Exchange-Organization-AuthSource: ITS-MSXHT0.ad.unc.edu
X-MS-Exchange-Organization-AuthAs: Anonymous
X-MS-Exchange-Organization-AVStamp-Mailbox: MSFTFF;1;0;0 0 0

This is a reminder of the impending closure of our survey on document
access control practices for the UNC Non-Biomedical IRB approved study
titled "Describing the Impact of Document Content Variance on Access
Control Efficiency and A Proposed Solution for Improving Efficiency:
Fine-grained, Redactive Access Control Models", #13-1008 approved on
1/31/2013.

If you haven't yet responded to the survey please take a moment to consider completing the short survey (average time to complete ~6 minutes, 56 seconds). If you have submitted the survey, we sincerely appreciate your participation in this effort. All responses must be in by next Wednesday, 2/20/2013.

PI: Deric Freeman
Email: deric_freeman@unc.edu
Phone: 9199669171

FA: Arcot Rajasekar
Email: rajasekar@unc.edu
Phone: 9199663611

This email is sponsored by: School of Information and Library Science

===
=====
"INFORMATIONAL:" email will only be sent to those who have indicated that they do want to receive mass email. To set your informational mass email preference, sign into MyUNC at http://my.unc.edu, and select "Update Personal Information".