Winter 12-15-2012

An Activity Theory Approach to Leak Detection and Mitigation in Personal Health Information (PHI)

Rohit Valecha
University at Buffalo

Shambhu Upadhyaya
University at Buffalo

Raghav Rao
University at Buffalo, mgmtrao@buffalo.edu

Arun Keepanasseril
McMaster University

An Activity Theory Approach to Leak Detection and Mitigation in Personal Health Information (PHI)

Rohit Valecha
Department of MSS, University at Buffalo,
Williamsville, New York, USA

Shambhu Upadhyaya
Department of CSE, University at Buffalo,
Williamsville, New York, USA

H. R. Rao[1]
Department of GSM, Sogang University,
Seoul, South Korea
Department of MSS, University at Buffalo,
Williamsville, New York, USA

Arun Keepanasseril
Hospital & Health Care, McMaster University,
Hamilton, Ontario, Canada

ABSTRACT

The migration to Electronic Health Records (EHR) has raised issues with respect to security and privacy. One such issue that has become a concern for the healthcare providers, insurance companies and pharmacies is Patient Health Information (PHI) leak. Borrowing from Document Control Domain (DCD) literature, in this paper, we develop a methodology for detection and mitigation of PHI leaks by employing Activity Theory to elucidate the complex activities in the transitive workflow.

Keywords: Patient health information, Information leak detection and mitigation, Activity Theory, Security policies

[1] Corresponding author. mgmtrao@buffalo.edu +1 716 645 3425

INTRODUCTION

As a cost-reduction measure, many healthcare providers are moving the personal health information (PHI) into electronic format, also known as Electronic Health Records (EHR). When these electronic means become the prime instrument for storage and exchange of personal health data, the risks of inadvertent disclosure of PHI increase (Sokolova et al. 2009). Migration to EHR for efficient health care service opens up issues with respect to security and privacy (Rengamani et al. 2010). These security and privacy issues lead to user workarounds. In many areas of healthcare, workarounds can be infectious creating new information risks and patient data leaks (Johnson et al. 2012).

PHI leak has become a concern for the healthcare providers, insurance companies and pharmacies – that deal with the confidential information of the patients – and the consequences have been severe. These PHI leaks could be happening due to a variety of reasons (Mishra et al. 2011). Rengamani et al. (2010) identify routine events such as data sharing, file organization, software installation, amongst others as leak factors, while (Sokolova et al. 2012) identify political events and advertisements resulting in PHI leaks. In addition, Cox (2010) identifies the various channels of the Internet resulting in PHI leak, and Lewis (2010) identifies data theft and fraud as the factors.

In the recent years, the document control domain (DCD) has provided numerous frameworks for detecting and mitigating information leaks, such as originator-based access control (Krohn et al. 2004), propagated access control (Jajodia et al. 2001), role-based access control (Sandhu et al. 2000), enforceable security policy (Schneider, 2000), Harrison, Russo and Ullman (HRU) model (Zhang et al. 2005), typed access matrix model (Samarati and De Capitani Di Vimercati 2001), etc. While the work done in DCD has been extensive, it has not been

utilized in addressing PHI leaks in a healthcare setting. Thus, borrowing from the DCD literature, and following the guidelines provided by Pramanik et al. (2004), as a solution to PHI leak problem, we develop a methodology for detection and mitigation of PHI leaks.

Hospitals in the healthcare systems are complex because of the intricacy of their information workflows. The transitive nature of the information workflow is a case in point, as the data flows from one point to the other in the workflow. The workflow consists of complex activities typically involving multiple agencies. Shanker et al. (2009) suggested Activity Theory to understand the various activities in the workflow. Activity theory is a powerful and clarifying descriptive tool with the objective to understand the activity (Nardi 1995). It gives the flexibility of breaking up complex tasks into activities that are easy to interpret and manage. It has been utilized for human-computer interaction. Thus, in this paper, we use activity theory as a lens for the development of our methodology. We adapt activity theory to align its components to capture the rights of the user in an organization on the patient data set and provide rules to access patient data set based on user's context. Consequently, the adapted activity theory provides guidance to detect and mitigate leak.

There are three major contributions of our work to existing literature: (a) we develop a methodology to detect and mitigate PHI leaks, (b) we advocate access control for transitive health workflows, and (c) we adapt activity theory to enforce security policies.

The paper is organized as follows. In the next section, we provide the literature on transitive workflows. Then, we describe activity theory. In the subsequent section, we provide the leak scenario. Next, we state the design considerations of security policies for PHI leak detection and mitigation. Then, we detail the leak detection and mitigation using activity theory.

TRANSITIVE WORKFLOWS

In a healthcare system, the data flows from multiple sources, which in turn receive data from other such sources. This type of relationship is referred to as transitive relationship. A transitive relationship is one where multiple sources exchange information with multiple others over several steps in the workflow (Lechler et al. 2011). The transitive workflow is the most general case of data exchange (Weber-Jahnke and Obry 2011).

Table 1 highlights the data typically shared between organizations. It also depicts the various data sharing authorizations that the patient data is a part of. Finally, the data set is simplified from various discrete data points.

Table 1. Patient Data Set

Type of Information	**Data Set**	**Authorization**	**Optional Control**
Identifying Information	Name SSN ID	Governmental, Insurance, Research	Commercial: (vendor supported) General: (service applications) Other: (unspecified)
Demographics Information	Gender Race Ethnicity	Governmental, Insurance	
Personal Information	Email Phone Fax Address Date of Birth	Insurance	
Medical Information	Diagnosis Treatment	Governmental, Insurance	
Other Information	Health Plan Account Numbers	Insurance	

ACTIVITY THEORY

Activity theory is more of a descriptive theory (than a prescriptive one) that provides a lens to describe or analyze the activity of a group or an organization. It involves the concepts of subject, object and community supported by tools, rules and division of labor. It suggests that an activity is directed towards an object, mediated by the instrument and socially constituted within

the environment (Bertelsen and Bodker 2003). The subject is the individual or the group/organization performing the activity supported by instruments (such as obligations). The object can be either an ideal or a material object. During this interaction, the subjects confine their understanding of the relationship between them and environment consisting of rules, responsibilities and communities (Chen et al. 2012).

Engestrom extends the concept of activity theory, and gives a specific example of its usage in a hospital setting. The activity considered is that of a doctor diagnosing a patient. In this activity, the subject is the physician, the diagnosis is directed towards the objective of patient's preliminary assessment, and is supported by the instruments such as stethoscope. The community in which the activity is placed is physician and nurse, with constraints of patient authorization before disclosure to any entity, and responsibility of assisting the patient.

In our paper, activity theory is used to detect and mitigate information leaks in a hospital setting, which is discussed in detail in Section 6.

LEAK SCENARIOS

In this section, we focus on potential leaks in a PHI. Based on these, we will formulate design considerations of security policies necessary to overcome these leaks. The healthcare domain can be visualized as a corporate network of users. Each user belongs to a role with a specific function, usually dictated by the nature of the organization. During the course of work, the role utilizes and shares a variety of information related to the patient. In order to provide this patient information to the users, organizations request this information from other health organizations in the network. The transfer of health information results in transitive information workflows. The most common case of transitive information workflow is data sharing supported

by software vendors that have a complete patient data access in about three quarters of hospitals (Johnson 1998), and is further explained next.

Table 2 shows a scenario adapted from (Lechler et al. 2011) where the research facility uses software vendor for functions involving identifiable data, personal data and medical data, and the service organization uses a software vendor for supporting personal data and medical data. This data might be saved in vendor's possession (as is the case with cloud service providing vendors). The vendor may now be able to match patient data provided by the research facilities with that provided by the service organizations and derive more information than a single entity (research affiliate or service organization) intended for the vendor. In such a case, the vendors could leak information by aggregating the data obtained from different sources.

Table 2. Leak Scenario

From **To**	**Hospital** **Research**	**Hospital** **Service**	**Research** **Vendor**	**Service** **Vendor**
Identifying Data	X	X	X	X
Demographics Data		X		
Personal Data	X	X	X	
Medical Data	X	X	X	X
Other Data		X		X

The above example demonstrates that owing to transitive workflows, the risk of information leakage increases in that the vendors might maliciously use patient treatment information, whose compromise to the outside world could lead to reputation loss for the patient. Thus, it is critical to detect leak and mitigate leak at the system level in order to achieve higher levels of information security. In addition, if there are no constraints on the privileges in the form of access control, then a malicious user or organization is capable of inflicting serious damage to the patient information.

POLICY DESIGN CONSIDERATIONS

To design a policy specification to prevent the leak in a health organization, we need to consider both the context and information flow between requests. We take an approach where multiple policies are specified on the same data set. The policies differ in the context when they become applicable. For example, a policy might allow access to a data set on a machine owned by the organization but not on other personal computers. The current context is contained in the request for access or is maintained on the policy server. We refer to such policies as contextual policies. Setting up a highly granular access control mechanism can be arduous. So, for ease of implementation, we consider higher level of structure, for instance, the role of users instead of each user, the organization instead of each department and the data set instead of each data value.

Consider a healthcare setting with a set of roles S and a set of patient data D that they want to protect. Each role s∈S has some attributes that can be represented as a tuple <s1, s2, …, sn>. The attributes can be the name of the role, its classification in the organization, its credentials and so on. Each data set d∈D also has attributes <d1, d2, …, dm>, representing features such as name of the data set, the category of the data set (e.g., sensitive, public, etc.), the type of data set (e.g., demographics, personal, etc.) and so on. The data set will have a set of actions A that can be performed on them. Each organization z∈Z also has attributes <z1, z2, …, zk>. These attributes can be the name of the organization, its classification in the healthcare setting, etc. Each authorization a∈A is specified as a tuple <s, z>. Here s∈S specifies the role that has action (e.g., view data) allowed in an organization z∈Z. We denote the set of policies as P. Each access control policy p∈P is specified as a tuple <d, {Rule1, Rule2, …}>. Here d∈D specifies the target of the policy and rules specify the actions allowed/denied to roles. The rules are the conditions under the current context.

The critical parts of the policy are the contexts and the obligations. The contexts are conditions that are based on the known configurations. The obligations are requirements that are already specified in the policy statement. Both the contexts and obligations are specified in propositional logic on the subject attributes, object attributes and other system attributes. For all data actions performed, a request is generated which contains the role, data set, organization and other client side information. The response contains policy enforcement and the obligations enforcement. Table 3 shows the various entities in our security policy.

Table 3. Entities in Security Policy

Entity	**Description**
S	Set of roles of subjects
$<s_1, s_2, \ldots s_n>$	Attributes of roles
D	Data Set
$<d_1, d_2, \ldots d_m>$	Attributes of data set
Z	Organizations
$<z_1, z_2, \ldots z_k>$	Attributes of organizations
A	Set of authorizations
a = <S, Z>	Authorization on roles in organizations
C	Context
O	Obligation
Q	Set of request
q = <S, D, Z>	Request of roles in organizations on data set
V	Response action
v = <permit/deny>	Action of permit or deny
R	Set of rules
r = <S, A, V, C>	Rule specification
P	Set of policies
p = <D, R>	Policy of rules on data set

INFORMATION LEAK DETECTION AND MITIGATION

Activity theory recognizes that changing conditions can realign the constituents of an activity (Shanker et al. 2009). The alignment of activity theory components with information leak can be made as follows (see Table 4): inner triangle comprising of subject, object and community can be used as a framework for leak detection, and outer triangle comprising of

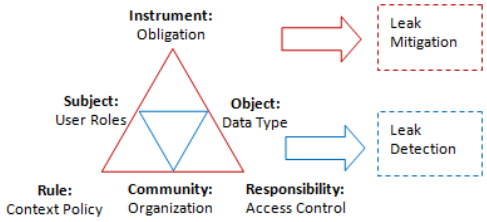
instrument, rule and responsibility can be used as a framework for leak mitigation. This is explained in detail.

Definition 1 – Information Flow: There is an information flow between organizations og1 and og2 represented as og1–>og2, if og2 requests patient data from og1.

Definition 2 – Privilege Set: A privilege set is a tuple based on activity theory's inner triangle <subject, object, community>, and represented as <s, d, z>, where s is a user role in the organization z that has access to data type d.

Definition 3 – Policy Set: A policy set is a tuple based on activity theory's outer triangle <rule, tool, responsibility>, and represented as <p, O, N>, where p is a policy containing rules to provide the access N based on obligation O. In order to map the policy set to activity theory's outer triangle, consider the contextual policy p as the specification of the set of rules, obligation O as the specification of requirements in the form of tools, and access N as specification responsibility that users have on a data set.

Table 4. Leak Detection and Mitigation using Activity Theory

Activity theory adapted from (Shanker et al. 2009)			
	Leak Detection	Subject: Object: Community:	User roles including physician, nurse, staff, administrator, etc. Data type such as personal, demographics, medical, etc. Organizations such as hospital, research, commercial, etc.
	Leak Mitigation	Instrument: Rule: Responsibility:	Obligations such as network login, registered machine use, etc. Contextual policies including rules such as password change, etc. Access (permit/deny) on privilege such as view, edit, delete, etc.

The privilege set is a representation of rights a user in an organization has on the patient data. It is created only once when a new patient is enrolled, and modified as per the patient's

requirements. The policy set is a representation of rules that provide access based on user's obligation. It is constructed at the onset of each workflow when the user requests the patient data.

As an initial setup, we start with a pool of patients and their data set. Once the system is deployed, its first task is to build the privilege set of all users in the organization. The privilege sets are generated based on the policies specified on the documents. The leak detection triangle of activity theory is used to generate the privilege sets for all the users in the organizations. The policy sets get created and deleted based on the current context with every request for patient data. Whenever a request is received, the leak mitigation triangle of activity theory is used to generate the policy sets.

Each user request is framed in the form of a privilege set by specifying the role, organization and the data set. This request set is compared with the privilege set to decide whether the rights should be granted. The primary objective in this case is to prevent illegal information flow from one organization to the other. Based on the definition of information flow, in order to prevent illegal information flows we have to provide restrictions in the form of obligations. The restriction will be enforced at the client side by executing policies containing rules to permit/deny access to the information.

In order to compute the new set of obligations, all requests that are not a part of the privilege set are added into E, where E contains the illegal information flows (in form of tuples). For each illegal information flow in E, a "deny" obligation (or restriction) is added, if the obligation is not already present. Such an obligation prevents the current user from tampering with the patient data, and setting liberal rights on the data. Also when a new patient is enrolled the privilege sets of all the users are recomputed. When a patient is deleted, the static access rights are checked and if allowed then data is deleted and privilege sets of all users are updated.

CONCLUSION

Due to migration to EHR, issues with respect to security and privacy have become amplified. One such issue is PHI leak that has become a concern for the healthcare providers, insurance companies and pharmacies. Further, the issue of PHI leak is aggravated due to the transitivity of health workflows. Borrowing from the DCD literature, and following the guidelines provided by Pramanik et al. (2004), as a solution to PHI leak problem, we develop a methodology for detection and mitigation of PHI leaks by employing activity theory to understand the various activities in the workflow. In doing so, this work offers significant advances to the document control domain as well as the healthcare literature in that it (a) illustrates methodology for leak detection and mitigation in transitive workflows (b) adapts activity theory to enforce security policies for PHI leaks.

There are few limitations of our work. First, different organizations have different role hierarchy. Our methodology does not take this into consideration. Second, the security policies are applied at the user machine, and are not transitive like the patient information in workflow. There are a few areas that we have identified to further this research. First, the security policies should be made dynamic and should be transferred along with the information. Second, a framework for prototype system should be developed using an open source system like OSCAR GI (Chan and Gallagher 2007) in the Canadian Healthcare System that can interface with existing healthcare systems. Third, a simulation based on the proposed methodology for illustrating its effectiveness.

ACKNOWLEDGEMENTS

The authors would like to thanks the reviewers for their critical comments that have greatly improved the paper. This research is supported in part by NSF Grant No. 0916612. The research of the third author was funded in part by Sogang Business School's World Class University Project (R31-20002), funded by Korea Research Foundation and by Sogang university Research Fund. Usual disclaimer applies.

REFERENCES

Bertelsen, O., and Bodker, S. 2003. "Activity Theory," In *J. M. Carroll (Ed.), HCI Models, Theories and Frameworks: Towards a Multidisciplinary Science*. San Francisco, Mogan Kaufmann, pp. 291-324.

Chen, R., Sharman, R., Rao, H. R., and Upadhyaya, S. 2012. "Data Model Development for Fire Related Extreme Events: An Activity Theory Approach". *MIS Quart*. Forthcoming

Chan, D., and Gallagher, J. 2007. "Open Source Clinical Application Resource – Canada (OSCAR)". Retrieved from http://www.oscarmcmaster.org/

Cox, B. 2010. "Senior Citizens and Identity Theft: Protecting the Elderly from Phishing Scams," in http://consumereducation.suite101.com/article.cfm/senior_citizens_and_identity_theft.

Engestrom, Y., Miettinen, R., and Punamaki, R. 1999. "Perspectives on activity theory," Cambridge Press.

Engestrom, Y. 1999. "Outline of three generations of activity theory," Cambridge University Press.

Jajodia, S., Samarati, P., and Sapino, M., Subrahmanian, V. 2001. "Flexible support for multiple access control policies", *ACM Transactions on Database Systems*, (26:2), pp. 214–260.

Johnson, E., Agarwal, R., Cowperthwaite, E., El Emam, K., and Connelly, P. 2012. "Usability and Healthcare Data Breaches", Retrieved from http://www.ists.dartmouth.edu/events/sith2/panel1.html

Johnson, R. 1998. "Trends in health care and health care systems," *In Proceedings of the 1998 Annual HIMSS Conference*, pp. 323-330.

Kuutti, K. 1995. "Activity Theory as a potential framework for human computer interaction research," In *B. A. Nardi (Ed.), Context and consciousness: activity theory and human-computer interaction*, pp. 17 – 44.

Krohn, A., Beigl, M., Decker, C., Robinson, P., and Zimmer, T. 2004. "ConCom – A language and Protocol for Communication of Context", *Technical Report* ISSN 1432-7864 2004/19

Lechler T., Wetzel S., and Jankowski R. 2011. "Identifying and Evaluating the Threat of Transitive Information Leakage in Healthcare Systems," *In the Proceedings of the 44th HICSS*, pp. 1-10.

Lewis, N. 2010. "EMR Data Theft Booming," *Information Week.*

Markus, M., Majchrzak, A., and Gasser, L. 2002. "A design theory for systems that support emergent knowledge processes," *MIS Quarterly,* (26:3), pp. 179–212.

Mishra, S., Leone, G., Caputo, D., and Calabrisi, R. 2011. "Security Awareness for Health Care Information Systems: A Hipaa Compliance Perspective," *Issues in Information Systems*, (12:1), pp. 224-236.

Nardi, B. 1995. "Activity Theory and Human-Computer Interaction," In *B. A. Nardi (Ed.), Context and consciousness: activity theory and human-computer interaction*: MIT Cambridge, MA, USA.

Pramanik, S., Sankaranarayanan, V., and Upadhyaya, S. 2004. "Security policies to mitigate insider threat in the document control domain," *in Proc of the 20th ACSAC'04*, Tucson, Arizona, USA, pp. 304–313.

Rengamani, H., Upadhyaya, S., Rao, H., and Kumaraguru, P. 2011. "Protecting senior citizens from cyber security attacks in the e-health scenario: an international perspective," *Proceedings of the 6th AWCSII*, NY.

Samarati, P., De Capitani Di Vimercati, S. 2001. "Access control: Policies, models, and mechanisms", In *Foundations of Security Analysis and Design*, Lecture Notes in CS, 2171, Springer-Verlag, NY.

Sandhu, R., Ferraiolo, D., and Kuhn, R. 2000. "The NIST model for role-based access control: Towards a unified standard", *In Proc. of 5th ACM on Role-based Access Control*, Germany, pp. 47–63.

Schneider, F. 2000. "Enforceable security policies", *ACM Transactions on Info and System Security,* (3:1), 30-50.

Shanker, D., Agrawal, M., and Rao, H.R. 2009. "Emergency response of Mumbai terror attacks: An activity theory analysis," *In Proceedings of ICSCF 09*, Kochi, India

Sokolova, M., El Emam, K., Rose, S., Chowdhury, S., Neri, E., Jonker, E., and Peyton, L. 2009. "Personal health information leak prevention in heterogeneous texts," *Biomedical Information Extraction International Workshop with the 7th International Conference on Recent Advances in Natural Language Processing,* pp. 58-69.

Sokolova, M., El Emam, K., Arbuckle, L., Neri, E., Rose, S., and Jonker, E. 2009. "P2P Watch: Personal Health Information Detection in Peer-to-Peer File-Sharing Networks," *Journal of Med Internet Research,* (14: 4), pp. e95.

Walls, J., Widmeyer, G., and Sawy, O. 1992. "Building an information system design theory for vigilant EIS," *Information Systems Research,* (3:1), pp. 36–59.

Weber-Jahnke, J., and Obry, C. 2011. "Protecting privacy during peer-to-peer exchange of medical documents," in *Information Systems Frontier*, (14), pp. 87-104

Zhang, X., Li, Y., and Nalla, D. 2005. "An attribute-based access matrix model", *In SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, New York, NY, USA, pp. 359–363.