University of Victoria | Graduate Studies

Notice of the Final Oral Examination for the Degree of Doctor of Philosophy

of

HANAN ALHINDI

MSc (King Saud University, 2013)
BSc (University of British Columbia, 2009)

"A Framework for Data Loss Prevention using Document Semantic Signature"

Department of Electrical and Computer Engineering

Tuesday, October 29, 2019
11:00 A.M.
Clearihue Building
B017

Supervisory Committee:
Dr. Issa Traore, Department of Electrical and Computer Engineering, University of Victoria (Supervisor)
Dr. Kin Fun Li, Department of Electrical and Computer Engineering, UVic (Member)
Dr. Venkatesh Srinivasan, Department of Computer Science, UVic (Outside Member)

External Examiner:
Dr. Cherie Ding, Department of Computer Science, Ryerson University

Chair of Oral Examination:
Dr. Darlene Clover, Department of Educational Psychology & Leadership Studies, UVic

Dr. David Capson, Dean, Faculty of Graduate Studies

Abstract

The theft and exfiltration of sensitive data (e.g., state secrets, trade secrets, company records, etc.) represent one of the most damaging threats that can be carried out by malicious insiders against institutions and organizations because this could seriously diminish the confidentiality, integrity, and availability of the organization’s data. Data protection and insider threat detection and prevention are significant steps for any organization to enhance its internal security. In the last decade, data loss prevention (DLP) has emerged as one of the key mechanisms currently used by organizations to detect and block unauthorized data transfer from the organization perimeter. However, existing DLP approaches face several practical challenges, such as their relatively low accuracy that in turn affects their prevention capability. Also, current DLP approaches are ineffective in handling unstructured data or searching and comparing content semantically when confronted with evasion tactics where sensitive content is rewritten without changing its semantic. In the current dissertation, we present a new DLP model that tracks sensitive data using a summarized version of the content semantic called document semantic signature (DSS). The DSS can be updated dynamically as the protected content change and it is resilient against evasion tactics, such as content rewriting. We use domain specific ontologies to capture content semantics and track conceptual similarity and relevancy using adequate metrics to identify data leak from sensitive documents. The evaluation of the DSS model on two public datasets of different domain of interests achieved very encouraging results in terms of detection effectiveness.