2015

Selection of penetration testing methodologies: A comparison and evaluation

Aleatha Shanley
Edith Cowan University

Michael N. Johnstone
Security Research Institute, Edith Cowan University

DOI: 10.4225/75/57b69c4ed938d
13th Australian Information Security Management Conference, held from the 30 November – 2 December, 2015 (pp. 65-72), Edith Cowan University Joondalup Campus, Perth, Western Australia.

SELECTION OF PENETRATION TESTING METHODOLOGIES: A COMPARISON AND EVALUATION

Aleatha Shanley[1], Michael N. Johnstone[1,2]
[1]School of Computer and Security Science, [2]Security Research Institute
Edith Cowan University, Perth, Australia
{a.shanley, m.johnstone}@ecu.edu.au

Abstract
Cyber security is fast becoming a strategic priority across both governments and private organisations. With technology abundantly available, and the unbridled growth in the size and complexity of information systems, cyber criminals have a multitude of targets. Therefore, cyber security assessments are becoming common practice as concerns about information security grow. Penetration testing is one strategy used to mitigate the risk of cyber-attack. Penetration testers attempt to compromise systems using the same tools and techniques as malicious attackers thus attempting to identify vulnerabilities before an attack occurs. This research details a gap analysis of the theoretical vs. the practical classification of six penetration testing frameworks and/or methodologies. Additionally, an analysis of two of the frameworks was undertaken to evaluate each against six quality characteristics. The characteristics were derived from a modified version of an ISO quality model.

Keywords
Penetration Testing, Methodology, System Security.

INTRODUCTION

The rate of cyber security threats detected for business and government is increasing, with approximately 7,300 incidents reported to CERT Australia in 2012 and approximately 8,500 incidents reported by August 2013 (CERT, 2013). Consequently, the cyber security industry is growing at a rapid rate with worldwide spending expected to reach US$86 billion by the year 2016 (Gartner, 2012). Australian security and intelligence agencies have stated that Australia is experiencing an increase in sophisticated cyber-attacks in both government and business originating from an array of sources: individuals, organised criminals and foreign intelligence services (CERT, 2013). The 2013 Cyber Crime and Security Survey showed an overall increase in reported cyber security incidents from 56 organisations reporting incidents in 2012 to 76 in 2013 (CERT, 2013). Fortunately, mitigation strategies are available for organisations, governments, and individuals to minimise risk. One mitigation strategy commonly used within the cyber security industry is penetration testing, commonly referred to as pentesting (Tang, 2014).

Pentesting aims to evaluate information security measures through the eyes of a potential attacker with the aim of testing the effectiveness of security controls (Midian, 2003). Pentesting is often employed by organisations as a mitigation strategy to reduce the risk of an attack on computer resources or in some cases, critical infrastructure. Pentesting attempts to ensure weaknesses and vulnerabilities in a networked environment are detected and can be addressed before they are exploited in a real-world attack (Tang, 2014). A security practitioner tasked with penetration testing will conduct a series of security tests in an attempt to gain access to a system and exploit security flaws that exist using the same tools and techniques that simulate a malicious attack, but do so in a controlled manner (Yeo, 2013). A properly scoped and deployed pen test can be an invaluable tool to assess the ability of a system to survive malicious attack (Valli, Woodward, Hannay, & Johnstone, 2014). The cornerstone of a successful pen test is its underlying methodology. A well-defined methodology plays a critical role in achieving results that can be verified and studied to protect data, applications and underlying infrastructure. Without an established methodology or framework within which to conduct a pen test, identifying vulnerabilities accurately can become difficult or provide a false sense of security (Frankland, 2009). Wilhelm (2009, p. 154) asserts that penetration tests are projects that need to be developed using effective and repeatable processes for improvements to be made, businesses goals to be met, and quality improved, therefore a methodology is a crucial factor. This suggests that penetration testing is achieving some level of maturity, akin to software engineering, although the lack of attention paid to software vulnerabilities in initial system releases may be due to the fixation of project managers on visible functionality as noted by Johnstone (2009).

Avison and Fitzgerald (2006, p. 418) discuss in detail the loose but extensive use of the term “methodology” and argue that there is very little agreement as to what it means other than at a very general level. Furthermore there is little in the literature addressing frameworks and methodologies for the purposes of penetration testing specifically. Consequently pentesting methodologies and frameworks appear to be poorly defined. Despite this

confusion of terms there are many pentesting methodologies/frameworks available. Certain frameworks or methodologies are free to use whereas others require some form of membership, payment or contribution, for example; technical input to the framework or methodology. Several pentesting methodologies and frameworks widely available in particular include: Open Source Security Testing Methodology Manual (OSSTMM), Information Systems Security Assessment Framework (ISSAF), Open Web Application Security Project (OWASP), Metasploit Framework (MSF), and Building Security in Maturity Model (BSIMM) Penetration Testing Execution Standard (PTES).

The purpose of this research is to evaluate a selection of currently available pentesting methodologies and frameworks (see above). We perform a gap analysis to determine if a pentesting framework is actually a framework, i.e., it has a sound underlying ontology. A subset of these frameworks is evaluated against quality criteria, which will determine their suitability for real world applications.

RELATED WORK

The ISO/IEC 25010:2013 quality standard (Standards Australia, 2013) defines a product quality model composed of eight characteristics, further divided into sub-characteristics that relate to static properties of software. We consider the suitability of six different penetration frameworks and methodologies and discuss the above-mentioned quality standard as a means of selecting evaluation criteria for the frameworks.

Penetration Testing Frameworks and Methodologies

ISSAF is an Open Source, peer-reviewed, penetration testing framework created by the Open Information Systems Security Group (OISSG). ISSAF is described as a framework and encapsulates multiple methodologies in draft 0.2.1B. ISSAF attempts to cover all possible domains of a penetration test from conception to completion. The authors suggest that is easier to remove information rather than develop it from the ground up (OISSG, 2005). The penetration testing methodology embedded within the framework is divided into three primary phases, namely; planning and preparation, assessment, and, reporting and clean up. One advantage of ISSAF in particular is that the distinct relationship between the tasks and their associated tools for each task are shown.

OSSTMM is an open source security testing methodology introduced in 2000 by the Institute for Security and Open Methodologies (ISECOM). OSSTMM was developed under peer-review and benefits from open source licensing, however, access to the latest version (v4), requires paid membership. OSSTMM (v3) is defined as a methodology that encapsulates modules and channels whereby channels represent different domain areas (ISECOM, 2000). OSSTMM is primarily an auditing methodology thus is not as comprehensive as ISSAF and does not provide tools or methods for completing modules however it is a valuable auditing resource that can be used to satisfy regulatory requirements for corporate assets provided security auditors have sufficient skills to complete each phase.

OWASP is a not-for-profit organisation focused on improving software security. OWASP provides numerous tools, guides and testing methodologies for cyber security under open source licenses, in particular, the OWASP Testing Guide (OTG). OTG is divided into three primary sections, namely; the OWASP testing framework for web application development, the web application testing methodology, and reporting. The web application methodology can be used independently, or in conjunction with the testing framework; a developer can use the framework to build a web application with security in mind followed by a penetration test (web application methodology) to test the design. Therefore, OTG has a strong focus on web application security throughout the entire software development lifecycle as opposed to the ISSAF and OSSTMM, both of which are aimed at security testing an implementation. OTG is targeted specifically to a single domain area, that of web applications.

Building Security in Maturity Model (BSIMM) is a software security framework licensed under Creative Commons and authored by McGraw, Migues, and West (2009). In developing BSIMM, its authors observed the security practices implemented in sixty-seven highly successful companies. BSIMM consists of 112 activities divided into twelve practices, supporting four domains mainly; governance, intelligence, SSDL touch points, and deployment. In comparison to ISSAF and OSTMM, BSIMM does not specify what tools to use or how to use them, but describes practices used by successful companies. Pentesting is one of the practices identified within BSIMM however pentesting is only one process of many recommended activities.

Penetration Testing Execution Standard (PTES) is a penetration testing standard that was originally created in 2009 by Nickerson et al. (n.d). PTES includes pre-engagement interactions, intelligence gathering, threat

modelling, vulnerability analysis, exploitation, post exploitation, and reporting. PTES takes advantages of other resources with the approach of not reinventing the wheel, rather, incorporates other frameworks within it, for example; OWASP for web application testing is referenced and recommended for use when testing web applications. PTES attempts to create a baseline for penetration tests whereby a security practitioner and/or organisation have a reference for what to expect at a minimum concerning penetration testing requirements.

Metasploit is a suite of penetration testing and intrusion detection tools designed to identify and exploit vulnerabilities on a target system. Metasploit was originally an open source project developed in 2003 but was acquired in 2009 by Rapid7 which is now responsible for its development and support (Holik, Horalek, Marik, Neradova, & Zitta, 2014). Metasploit, or the Metasploit Framework (MSF), is available in four different versions. MSF is suitable for the advanced security professional who has a solid understanding of penetration testing and is competent using command line pentesting tools. In comparison to ISSAF and OSSTMM, MSF is a practical solution that provides a suite of tools rather than a documented outline of process and methods to follow. MSF could be considered an application that encompasses a suite of tools that facilitate a penetration test.

In summary, there are a diverse range of methodologies and frameworks available. Each has unique characteristics and takes a distinct approach to penetration testing. The literature suggests a difference in the way terminology is applied to each concept, thus terms are used interchangeably (or incorrectly). For instance, ISSAF is defined as a framework however throughout the documentation it refers to methodology as the primary approach. MSF, on the other hand describes itself as a framework whereas it is a software application encompassing a suite of tools, therefore clarification on the classification of methodology vs. framework is essential to avoid confusion.

Quality Models

Figure 1 describes a generic quality model. Standards Australia (2013) clearly state that "It is not practically possible to specify or measure all sub-characteristics for all parts of a large computer system or software product. Similarly it is not usually practical to specify or measure quality in use for all possible user-task scenarios. The relative importance of quality characteristics will depend on the high-level goals and objectives for the project. Therefore the model should be tailored before use as part of the decomposition of requirements to identify those characteristics and sub-characteristics that are most important, and resources allocated between the different types of measure depending on the stakeholder goals and objectives for the product." Therefore, we have amended the ISO model to focus less on software quality evaluation and more on aspects of penetration testing framework evaluation (see figure 2).

The ISO9126 standard contains a taxonomy that defines software by its functionality, reliability, usability, efficiency, maintainability and portability. Security, which is the primary focus of this research, is defined as a sub-characteristic of functionality. This is a departure from the commonly held belief that security is solely within the domain of non-functional requirements. The ISO25010 standard (Standards Australia, 2013) extends upon this idea and considers security a characteristic in its own right. ISO25010 is, therefore, a replacement for ISO9126.

Figure 1: Abstract Quality Model (adapted from ISO/IEC 25010:2013).

EVALUATION OF FRAMEWORKS

First, six methodologies/frameworks were selected for potential use. The selections were classified into either a methodology or framework category respectively by way of gap analysis to establish whether or not framework or methodology characteristics were present focusing primarily on the field of penetration testing (see table 2). Table 1 summarises the gap analysis. Original classification (pre-evaluation) and post evaluation classifications are shown. The evaluation classifications illustrated in table 1 are an ordinal scale, for example; framework encapsulates methodology and methodology encapsulates tools, techniques, and resources. In addition frameworks that have included methodology in a particular framework are identified. We arrived at an individual evaluation based on theoretical analysis of each framework/methodology. Each framework or methodology underwent analysis to determine whether or not framework or methodology characteristics were present (or absent).

Table 1: Evaluation Matrix

	Standard or Guide	Framework	Frameowrk encapsulates Methodology	Methodology	Application Suite	Manual or Resource
ISSAF		* +	yes			
OSSTMM	*			* +		*
OTG	*	+	yes			
BSIMM		* +				
PTES	*					+
MSF		*			+	

Legend	
Pre-evaluation Classification	*
Post-evaluation Classification	+

From the gap analysis it was possible to develop a taxonomy of penetration testing specific frameworks or methodologies whereby suitable candidate methodologies or frameworks can be identified to facilitate their use in practice or research. We found that some frameworks were indeed frameworks in the accepted sense of the word, whereas others were simple collections of tools without a discernible underlying ontology. We felt that this was an important distinction because novice pen testers would benefit from the additional support provided by a mature framework.

Table 2: Classification of Penetration Testing Frameworks and Methodologies.

	← **Classification** →			→ **Suitability**	
Candidate	Framework	Methodology	Other	Penetration Testing Specific	Security Assessment
ISSAF	✓			✓	
OSSTMM		✓	✓		✓
OTG	✓			✓	
BSIMM	✓				✓
PTES			✓	✓	
MSF			✓	✓	

Selecting candidates for the evaluation of quality was determined by two criteria. First, whether or not a particular candidate classified as either framework or methodology. Candidates that fall under a methodology or framework were deemed in-scope, thus eliminating all other candidates. The second criterion was to examine the boundaries of a particular candidate for its scope, in other words, whether or not a particular candidate is focused entirely on penetration testing as opposed to an overall security assessment. The research being undertaken has a primary goal of evaluating penetration testing methodologies and frameworks explicitly rather than assessing the security posture of an organisation in its entirety, therefore candidates that are categorised as penetration testing explicitly are preferred over security assessment specific candidates. As a result the two remaining candidates are ISSAF and OTG.

Next, quality characteristics were nominated (see figure 2), for the purpose of evaluating the refined subset of frameworks. Two factors were taken into consideration for selection of quality characteristics. First, the field of study or context from which a characteristic definition was drawn (in particular that of information systems was preferred); and second, whether or not a particular quality characteristic was directly applicable to the field of penetration testing.

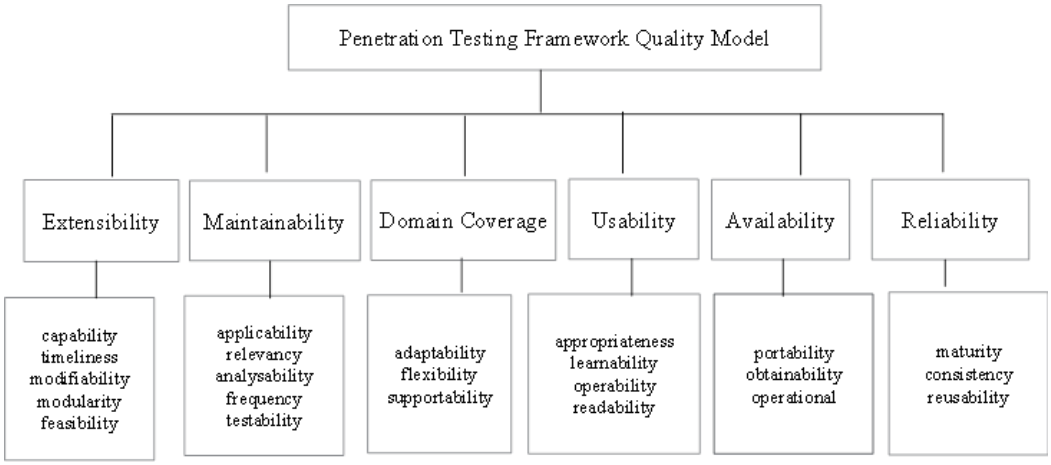

Figure 2: Penetration Testing Quality Model (adapted from ISO/IEC 25010:2013).

From the revised taxonomy shown in table 3, the selected quality metrics are applicable to both frameworks. Both frameworks display evidence of the quality characteristics applicable to penetration testing, therefore the six quality characteristics selected are considered suitable, thus will be used to facilitate this research in evaluating efficiency for the two chosen candidates, ISSAF and OTG. Note that reliability, whilst a valid characteristic, was not tested in this evaluation due to lack of delivery of an expected real-world case study.

Table 3: Quality Matrix of Selected Penetration Testing Frameworks.

Candidate ↓	Quality Metrics →					
	Extensibility	Maintainability	Domain Coverage	Useability	Availability	Reliability
ISSAF	✔	✖	✔	✔	✔	n/a
OWASP Testing Framework	✔	✔	✖	✔	✔	n/a

LEGEND	
Exhibits property	✔
Does not exhibit property	✖
Not Assessed	n/a

DISCUSSION

Gap analysis

The gap analysis showed in some cases that the classification of a particular framework and/or methodology can often be misleading, for instance; MSF is described as a framework; however, a subsequent evaluation of characteristics using factors outlined in table 1, showed that MSF more closely aligns with a suite of penetration testing tools, therefore is appropriately classified as an application suite that can facilitate a penetration test rather than a framework. In contrast, OTG was pre-classified as a standard or guide, however strong framework characteristics are identified throughout the documentation that suggest framework characteristics in contrast to its original classification, thus, the post-evaluation classification more appropriately aligns with framework. Turning to PTES, the characteristics do not illustrate enough properties to be considered either a methodology or framework, due to incomplete documentation or loose structure when compared to the more mature frameworks evaluated. It should be noted however that PTES has the potential to be further developed into a framework should future amendments be undertaken; as a consequence PTES classifies as a resource post-evaluation. From the six frameworks and/or methodologies reviewed, three (ISSAF, OSSTMM, and BSIMM), agree with the pre-evaluation classification, in other words did not change classification post-evaluation. As can be expected, the three aforementioned frameworks are considered mature, therefore it is not surprising that the classification of these three in particular, do not change post evaluation. Although not all the frameworks and methodologies evaluated show disparity with relation to classification it is important to note that some do, of which consideration needs to be given. The consequence of inappropriate classification lends itself to the possibility that penetration testing practitioners risk implementing or become reliant on a framework or methodology that might not meet an organisations goals in relation to completing a penetration test in its entirety, moreover adapting an approach that could potentially fail.

Measurement of Frameworks with Quality Characteristics

Both OTG and ISSAF were measured by the quality characteristics denoted in table 3. In some cases this measurement was not direct, as the characteristic did not have a direct mapping to a concept/artefact used in pentesting. An example would be “maintainability”.

Conventional measures of software maintainability are not suitable for pentesting frameworks as the artefacts in question are documents or risk matrices rather than software artefacts (such as code). However, some commonalities exist as penetration testing in its essence is testing software and/or hardware with software of some type in most cases.

Maintainability is the ease in which a framework can be understood, adapted, enhanced or modified. To measure this characteristic, consideration is given particularly to the number of revisions (frequency of change) a framework has undergone since inception. The frequency of change (actually a sub-characteristic of maintainability) can be quantified by the number of revisions a framework has undergone in its lifetime, the underlying assumption being that revisions reflect a measure of activity. The type of activity, whether bug fixes, documentation readability, addition of new features or other activity is not considered in isolation in this instance. Both OTG and ISSAF do not offer enough comparative information; therefore that level of detail is restricted. It is also worth mentioning that other sub-characteristics, namely; applicability, analysability, testability, and relevancy, can be incorporated as additional measures. Discussion of these other sub-characteristics is outside the scope of this particular research, therefore we focus here on one sub-characteristic that can be quantified.

Table 4: Total number of revisions (r) / framework lifetime in years (t)

	Revisions	Years	Date of Inception
ISSAF	5	2	December, 2004
OTG 3.0 + 4.0	468 (245 + 223)	7	May, 2008

Table 4 shows that OTG averaged 66 revisions per year in comparison to ISSAF which averaged 2.5 revisions per year. While it is obvious OTG had far more revisions than its counterpart it does not necessarily prove that ISSAF lacks quality. Moreover, other factors come to the fore, for example; Does OTG have more contributors than ISSAF?, Were revisions for OTG related to bad design, or does it simply suggest the product is better

overall? These questions are considerations for future research, however we conclude that OTG is more maintainable than ISSAF, due to its higher revision activity.

Similarly, Usability is measured by its sub-characteristic, Readability. Readability is significant as a sub-characteristic of usability primarily because the penetration testing frameworks evaluated in this research are documents, therefore if a document is not readable, usability is affected. Readability is concerned with the level of difficulty to read or comprehend written text (Ludger & Gottron, 2012). There are a number of ways to measure vocabulary difficulty and sentence length to predict the difficulty level of text resulting in various readability formulae in use today (DuBay, 2004). One such formula known as the Gunning Fog Index (GFI), was published by Gunning (1952, p. 36), developed specifically for adults. Fog index attempts to estimate the number of years of education that is required by the reader in order to understand the text at first reading, for example a GFI score of 10 would indicate that ten years of formal schooling is required to understand the text. The formula works with two variables; first, the average sentence length (ASL), and second, the number of words with more than two syllables for each one hundred words (PHW). Finally the result is multiplied by 0.4 (DuBay, 2004), thus produces the formula: Grade Level (GL) = 0.4 (ASL + PHW).

Table 5: Fog Index Scores

Framework	ISSAF	OTG
GFI Score	7.7	11

Results were obtained using an automated tool that calculates various readability scores, among them GFI score. The readability score automation tool is available for public use released as an open source project ("Readility-Score," 2015).

From table 5 it is clear that ISSAF is more readable than OTG. Taking into consideration the intended users of both frameworks it can be safely assumed that a security practitioner would have a minimum of 11 years formal schooling therefore readability scores for both frameworks are sufficient.

CONCLUSION

This paper examined several penetration testing frameworks and methodologies, with particular reference to ISSAF and OWASP's OTG. It was found that many frameworks were either mis-named (i.e., were not actually frameworks) or lacked domain coverage or a sound ontological foundation and thus were restricted in their application.

The frameworks were selected for evaluation based on their focus (penetration testing specific or security general) and their ability to act as a framework (rather than a collection of techniques without a unifying theme). We found that many "frameworks" were not able to be generalised across problem domains (as would be expected for a generic pentesting framework). The quality characteristics mapped well to the selected frameworks (ISSAF and OTG), which suggests that they are appropriate candidates to evaluate penetration testing frameworks.

The next step in this research programme is to evaluate the selected frameworks with a real-world case study.

ACKNOWLEDGEMENTS

This work has been partially funded by the European Commission via grant agreement no. 611659 for the AU2EU FP7 project.

REFERENCES

Avison, D., & Fitzgerald, G. (2006). *Information Systems Development: Methodologies, Techniques and Tools*. London: McGraw-Hill.

CERT. (2013). *Cyber Crime and Security Report 2013*. C. Australia.

DuBay, W. H. (2004). The Principles of Readability. Retrieved from http://www.impact-information.com/impactinfo/readability02.pdf

Frankland, J. (2009). The importance of standardising methodology in penetration testing. *Database and Network Journal, 39*(3), 13. Retrieved from http://ecu.summon.serialssolutions.com

Gartner. (2012). *Gartner Says Worldwide Security Infrastructure Market Will Grow 8.4 Percent* [Press release]

Gunning, R. (1952). *The Technique of Clear Writing*: McGraw-Hill.

Holik, F., Horalek, J., Marik, O., Neradova, S., & Zitta, S. (2014, 2014). *Effective penetration testing with Metasploit framework and methodologies.* Paper presented at the 15th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary. doi: 10.1109/CINTI.2014.7028682

ISECOM, I. o. S. a. O. M. (2000). *Open Source Security Testing Methodology*: ISECOM.Retrieved from http://www.isecom.org

Johnstone, M.N. (2009). Security Requirements Engineering: The Reluctant Oxymoron. Proceedings of the 7th Australian Information Security Management Conference, Perth, Western Australia, 1st to 3rd December 2009.

Ludger, M., & Gottron, T. (2012). Readability and the Web. doi: 10.3390/fi4010238

McGraw, G., Migues, S., & West, J. (2009). Building Security in Maturity Model. Retrieved from https://http://www.bsimm.com

Midian, P. (2003). Perspectives on Penetration Testing — Finding the Right Supplier. *Network Security, 2003*(2), 9-11. doi: 10.1016/S1353-4858(03)00210-1

Nickerson, C., Kennedy, D., Riley, C., Smith, E., Amit, I., Rabie, A., . . . Strand, J. (n.d). Penetration Testing Execution Standard Retrieved from http://www.pentest-standard.org/index.php/PTES_Technical_Guidelines

OISSG, O. I. S. S. G. (2005). *Information Systems Security Assessment Framework* OISSG.Retrieved from http://sourceforge.net/projects/isstf/

OWASP. (2014). *OWASP Testing Guide* Retrieved from https://http://www.owasp.org/index.php/OWASP_Testing_Guide_v4_Table_of_Contents

Readility-Score. (2015). Retrieved from https://readability-score.com/

Standards Australia. (2013). AS/NZS ISO/IEC Standard 25010:2013. Systems and software engineering--Systems and software Quality Requirements and Evaluation (SQuaRE)--System and software quality models.

Tang, A. (2014). A guide to penetration testing. *Network Security, 2014*(8), 8. doi: 10.1016/S1353-4858(14)70079-0

Valli, C., Woodward, A., Hannay, P., & Johnstone, M. (2014). Why Penetration Testing Is A Limited Use Choice For Sound Cyber Security Practice. *Proceedings of the Conference on Digital Forensics, Security and Law U6* 35. Retrieved from http://ecu.summon.serialssolutions.com/

Wilhelm, T. (2009). Professional Penetration Testing : Volume 1: Creating and Learning in a Hacking Lab (Vol. 1, pp. 26). Burlington: Syngress. Retrieved from http://ecu.summon.serialssolutions.com/

Yeo, J. (2013). Using penetration testing to enhance your company's security. *Computer Fraud & Security, 2013*(4), 17-20. doi: http://dx.doi.org/10.1016/S1361-3723(13)70039-3