Analysis and recommendations for standardization in penetration testing and vulnerability assessment

Penetration testing market survey

William Knowles, Alistair Baron and Tim McGarr

Security Lancaster | Lancaster University

bsi.

...making excellence a habit.™

Executive summary

The threat of cyber attacks has led to an increase of simulated and controlled cyber security evaluations of IT infrastructures. Such evaluations are frequently referred to as penetration testing; however, in practice, the nomenclature encompasses a variety of other labels, including vulnerability assessments, IT Health Checks, red team exercises, and, as it is commonly referred to outside the industry, ethical hacking.

The fundamental characteristics of each are similar: an attacker (or team of attackers) will perform a security assessment of an IT infrastructure using the same techniques as a malicious attacker. A report is delivered detailing any vulnerabilities found. This enables clients to test their security posture and prioritize the implementation of security controls based on a simulated attack.

Penetration testing in its various forms is increasingly becoming a requirement for organizations of all sizes and industries. Given this rising emergence of penetration testing as an assurance technique, this study set out to answer three questions:

1. What standards currently exist for penetration testing?
2. What industry bodies exist for organizations and individuals engaging in penetration testing?
3. Is there a need for additional standards, or to modify existing standards?

Any recommendations for standardization must be grounded in the realities of industry practices and client experiences. Therefore, in addition to a comprehensive desk research phase, a series of stakeholder interviews was conducted. In total, 32 penetration testing providers were interviewed (22 organizations), 15 clients (12 organizations), and 7 industry stakeholders (7 organizations, including technical bodies, government departments and a standards body).

The IT Health Check Service (CHECK) scheme, along with technical bodies, such as the Council of Registered Ethical Security Testers (CREST) and Tigerscheme, have successfully defined the technical capabilities of individuals who perform penetration tests, and can be seen to be making great efforts to encourage evolution within the industry. In addition, both CHECK and CREST have laid the foundations for the assessment of organizational processes that support engagements (e.g. information security practices for the protection of client data). However, this study highlights issues at the start and end of engagements that the industry has failed to address. This is not through a lack of awareness of these issues; both providers and clients were found to be dissatisfied by the lack of transparency and consistency in industry offerings. Given the importance (and rapid growth) of penetration testing, resolving these needs for best practice quickly would aid both providers and buyers. Three recommendations for standardization are proposed:

1. **Standardization of terminology for the different levels of simulated security evaluations.** This would enable clients to understand the service they are purchasing, and enable the provider to compete based on the quality of that service.
2. **Guidelines for reporting structure and content.** Stakeholders felt providers should supply greater consistency and depth in their use of metrics and recommendations (e.g. analysing root causes, rather than offering spot fixes), while also empowering clients to understand the security threats facing their environments (e.g. through attack narratives).
3. **Guidelines for the use of penetration testing as audit evidence**, notably within an ISO/IEC 27001 audit. The current implementation within ISO/IEC 27001 isolates penetration testing as a single control, negating its potential for assessing other security controls within the standard.

Introduction

This document describes a study for modified or new standardization relating to penetration testing. This work was undertaken as a collaborative project between Security Lancaster, an Engineering and Physical Sciences Research Council (EPSRC) and Government Communications Headquarters (GCHQ) recognized Academic Centre of Excellence for Cyber Security Research, and BSI.

Penetration testing is the process of conducting a simulated attack on an IT infrastructure to determine any weaknesses, using the methodologies, techniques and tools that provide the best representation of what a real-world malicious attacker would do. An organization's internal staff may conduct their own penetration tests; however, the majority do not have the necessary expertise, meaning engagements are typically conducted by third parties. Given the sensitive nature of what a penetration test consists of (e.g. often an attempt to obtain sensitive information), a framework of requirements must exist to ensure that those conducting these engagements do so in a safe, secure and effective manner. Naturally, this includes the practical aspects of an engagement, such as having the necessary expertise to identify vulnerabilities or minimize operational impact from testing. However, it also includes the organizational processes that support a safe, secure and effective environment, during the engagement and long after it has finished (e.g. the secure storage of client data).

Penetration testing within the UK is predominantly led by consortia/private standards. Part of this is governmental, such as the UK's National Technical Authority for Information Assurance (CESG), which is seen as a de facto regulator, while operating its own schemes to assess the competencies of both individuals and organizations. The largest industry body is CREST, which also assesses individual and organizational competencies, while offering higher-level qualifications for those working in high assurance environments, such as the banking sector through its CBEST[1] and other Simulated Target Attack and Response (STAR) schemes. Other industry bodies specialize only in individual qualifications, such as Tigerscheme and more recently, the Cyber Scheme.

Although formal standards from traditional standards bodies (e.g. BSI directly or ISO/IEC[2]) include some technical requirements for conducting penetration testing activities (e.g. the Common Criteria), these receive limited formal use beyond acting as informative documents. Instead, the main adoption of standards by providers has been of those which assess operational processes that support engagements (e.g. ISO/IEC 27001 to demonstrate efforts to protect client data).

The remainder of this paper summarizes a comprehensive literature review and a 54-member stakeholder survey on the requirement for additional standardization activities within the area of penetration testing. The discussion begins in Section 2.1 on qualifications for individuals, before moving onto standards from consortia and formal standards bodies (e.g. BSI directly or ISO/IEC) in Section 2.2. The penetration testing engagement process is then broken into phases in Section 3, which includes a discussion on stakeholder practices and experiences. Recommendations for future standardization activities are concluded in Section 4. Annex A outlines the composition of interviewed stakeholders, and Annex B provides the key points for recommended standards. Annex C contains a list of referenced standards.

1 http://www.bankofengland.co.uk/financialstability/fsc/Pages/cbest.aspx

2 BSI provides the "UK voice" in the development of International Standards in ISO and IEC (along with CEN and CENELEC for European Standards)

Existing penetration testing standards and qualifications

Competence requirements can be established at both the organizational and individual level. This section provides a discussion of the current state of the market for both, along with the views of stakeholders on the requirement for modified or new standards in these areas.

Qualifications for individuals

Budding and established professionals are now faced with a multitude of choices for qualifications across a range of skill levels and topic scopes. UK qualifications for penetration testing primarily arise from four providers: CESG, CREST, Tigerscheme and the Cyber Scheme.

CESG has established a qualification scheme called the IT Health Check Service (CHECK), which has been in operation for over a decade. The two levels of CHECK qualification are the CHECK Team Member and the CHECK Team Leader, the latter of which is split over two qualifications for infrastructure and web applications. Historically, CESG provided qualification assessment through an internally operated 'practical assault course'; however, it was unable to meet demand, and moved to an alternative format. The current format requires candidates to have obtained a certain type and level of industry qualification and Security Clearance (SC) to allow them to handle sensitive data, amongst other publicly undisclosed factors. The three remaining qualification bodies, CREST, Tigerscheme and the Cyber Scheme, provide the industry qualification. Their content, level and equivalence to CHECK qualifications are shown in Table 1.

Table 1 – Qualifications

Level	Qualification Bodies and Qualifications				
	CESG		CREST	Tigerscheme	Cyber Scheme
	CHECK	CCP			
Entry	N/A	SFIA Responsibility Level 3	Practitioner (CPSA)	AST	CSA
Intermediate	CHECK Team Member	Level 4	Registered (CRT)	QSTM	CSTM
Advanced	CHECK Team Leader	Level 5 Level 6	Certified (CCT)	SST	CSTL
Red Team	N/A	N/A	STAR (CCSAM and CCSAS)	N/A	N/A

CREST has since emerged as the predominant industry-led professional qualification body for UK penetration testers, and its qualifications can be seen to span four tiers. In order of required proficiency, they are Practitioner (requiring an estimated 2,500 hours of experience), Registered (6,000 hours), Certified (10,000 hours) and STAR. It is at the Certified tier that specialism occurs in the areas of infrastructure or web application security. STAR is a framework created to provide intelligence-driven red team penetration tests for the critical infrastructure sectors. Currently the main implementation of STAR is CBEST, which specifically targets the financial sector and was developed in partnership with the UK Financial Authorities (the Bank of England, Her Majesty's Treasury and the Financial Conduct Authority). STAR qualifications ensure that competency requirements are met to perform such engagements. Two forms of STAR qualifications exist: those for managers (i.e. those who lead STAR teams) and those for technical specialists. The Tigerscheme[3] and Cyber Scheme[4] qualifications follow a similar structure to the lower three CREST qualifications, each with a beginner, intermediate and advanced qualification. An equivalent to CREST's STAR qualifications is not available from other qualification bodies.

CESG has also launched a separate scheme, which forms a competence framework that is described as a certification rather than a qualification: the CESG Certified Professional (CCP). The CCP is a framework that defines seven

3 http://www.tigerscheme.org

4 http://www.thecyberscheme.org

roles, one of which is 'Penetration Tester'. Each role has differing levels of competence, which are aligned with the responsibility levels defined by the Skills Framework for the Information Age[5] (SFIA) and the skill levels defined by the Institute of Information Security Professionals (IISP). Candidates provide evidence to support their assessment, which is reviewed by one of three certification bodies: (a) the APM Group; (b) BCS, the Chartered Institute for IT; or (c) a consortium of the IISP, CREST and Royal Holloway, University of London. Four levels are defined for the Penetration Tester role: SFIA Responsibility Level 3, 4, 5 and 6. CCP, while listed in Table 1, does not currently contribute to a CHECK qualification assessment.

Although many non-UK training and qualification providers have been omitted from this research, there are three qualifications worthy of note, all of which are from US-based providers: the International Council of Electronic Commerce Consultants' (EC-Council) Certified Ethical Hacker (CEH); SANS' Global Information Assurance Certification (GIAC) Penetration Tester (GPEN); and the Offensive Security Certified Professional (OSCP). However, during the stakeholder interviews there was found to be a greater emphasis on UK qualifications for recruitment.

Findings

The consensus amongst stakeholders was a strong opposition to any form of new standard for individuals around penetration testing. Opposition was twofold. Firstly, the techniques and skills used for penetration tests evolve at a rapid pace, which would be infeasible to capture and keep current within a standards type document. Secondly, while the current system is not without fault, existing consortia providers within the UK have done an exemplary job of raising, setting and assessing the competence of individuals that conduct penetration testing, and furthermore, the UK is ahead of the rest of the world in this regard. Many stakeholders, however, did feel that there was a growing need for an independent penetration testing body, modelled in the same vein as medicine or law, in order to continue its professionalization. Taking medicine as an example, key indicators of its professionalization are the internationally recognized standards for its practices, and regulation bodies for individuals with powers such as being able to revoke the right to practise. Some providers felt that standards bodies, such as BSI, could facilitate the internationalization process by working with technical assessors such as CREST and Tigerscheme. However, not all stakeholders were positive about such an endeavour. It was noticeable that those supportive were predominantly in positions of management, whose natural proclivity is one of control. Those engaging in the practical elements of security assessments, the practitioners, had fears about the potential future exploitation of such a scheme to regulate those who wish to conduct cyber security research. It was felt that such a situation would negatively impact the industry as a whole, and lead to the loss of the UK's competitive advantage.

Standards for organizations

Organizational standards provide confidence in a provider's process readiness for penetration testing activities. Such standards fall broadly into two areas: those from consortia and those from formal standards organizations (e.g. BSI directly or ISO/IEC).

Consortia/private standards

The CHECK scheme, a governmental initiative operated by CESG, enables approved organizations to perform penetration tests for government departments and public sector bodies (including Her Majesty's Government). Organizational approval requires evidence submission in two areas: (a) capability (e.g. testing methodology and examples of previous reports) and (b) the composition of a CHECK team (at least one CHECK Team Leader and one CHECK Team Member).

The most well-established industry equivalent to CHECK is CREST, a not-for-profit organization that attempts to regulate services for penetration testing and incident response. Organizations can apply to become approved CREST 'Member Companies'.[6] The application process covers four domains of organizational capability: (a) information security; (b) operating procedures and standards; (c) methodology; and (d) personnel security, training and development. Both the ISO/IEC 27001 and ISO 9001 management system standards are referenced in CREST's guidance for applicants, but not mandated. However, CREST does require evidence of operational commitment

5 http://www.sfia-online.org

6 http://www.crest-approved.org/wp-content/uploads/App-Form-FAQs-v3.0.pdf

to the implementation of an information security management system (ISMS) and a quality management system (QMS). Furthermore, CREST requires a clear and documented complaints procedure for Member Companies, with an escalation path that makes direct reference to the CREST complaints process for independent arbitration.

CHECK and CREST focus on the competence of organizations that conduct penetration tests; however, consortia/ private standards also follow the traditional model of outlining security controls to protect against cyber attacks. Two such standards make explicit reference to the use of vulnerability assessment and penetration testing.

The first is the Payment Card Industry Data Security Standard (PCI DSS), which enforces a business requirement (rather than a legal requirement) for the information security of organizations handling payment transactions, including those by credit and debit card. Requirement 11.2 mandates quarterly vulnerability scans, while requirement 11.3 mandates penetration tests. Vulnerability assessments must be conducted by Approved Scanning Vendors (ASVs), but no competency requirements are mandated for penetration testing. However, PCI DSS have released supplementary penetration testing guidance, which includes provider competencies (e.g. qualifications), methodologies (notably including emphasis on the importance of exploitation), and reporting.

The second, Cyber Essentials, is an entry-level organizational standard that provides basic assurance that an organization is meeting minimum cyber security control requirements. It was developed in 2014 through a joint industry-UK government approach. Cyber Essentials is targeted at private, not-for-profit and public organizations of all sizes, although it has particular relevance for small and medium enterprises (SMEs). It outlines two levels of certification: basic (no formal label) and 'Plus'. The Cyber Essentials standard requires the completion of a self-assessment form for basic certification, and self-assessment plus a third-party security assessment for Plus (including an external and internal vulnerability scan). In practice, however, the three accreditation bodies implementing the scheme have adapted the standard to their own business requirements. The Cyber Essentials accreditation bodies are: CREST, the Information Assurance for Small and Medium Enterprises Consortium (IASME) and QG Management Standards.

Formal standards

Technical standards exist that describe activities relating to penetration testing, or mandate its use. One widely used standard for security evaluations is ISO/IEC 15408 (Common Criteria). This standard outlines the requirements for

the 'security functionality of IT products and for assurance measures applied to these IT products during a security evaluation', while the methodology for such an evaluation is outlined in ISO/IEC 18045. A number of challenges prevents the widespread application of ISO/IEC 15408 to penetration testing (e.g. high information requirements about the testing environment, and the challenges of applying it to a dynamic, and live, system); however, attempts to mitigate this have been provided in supplemental standards. For example, PD ISO/IEC TR 19791 extends ISO/IEC 15408 to cover operational environments, while ISO/IEC TR 20004 uses the Common Weakness Enumeration (CWE) and the Common Attack Pattern Enumeration and Classification (CAPEC) frameworks to support ISO/IEC 18045 vulnerability assessments.

Vulnerability assessments and penetration tests can also contribute to standards and certifications as audit evidence towards security controls. ISO/IEC 27001 is one example where penetration testing is explicitly referred to in a security control, under the umbrella of a 'technical compliance review'.

Findings

Isolated criticisms were raised against consortia/private standards (e.g. a lack of independence from industry in their governance); however, the predominant voice amongst stakeholders was that they have done much to raise the standard of operational readiness and professionalism of penetration testing within the UK. Indeed, approval was frequently cited as a motivation for the adoption of management system standards by providers, predominantly ISO/IEC 27001 and ISO 9001.

Although the benefits of some technical standards were espoused (e.g. the thoroughness of Common Criteria), the consensus was that it would prove difficult to implement these standards for the types of services and timescales for testing that clients were demanding. A small number of providers suggested standardizing the methodology; however, this was mostly only seen as an option if standardization was forced. Other stakeholders questioned the benefit of such a 'high-level' standard, feeling there was already a significant quantity of information in the public domain on this topic. Efforts to create such a standard have been considered (and continue to be) by the subcommittee that developed the ISO/IEC 27000 series (ISO/IEC JTC 1/SC 27), although there is not yet any standard published or in development on this topic.

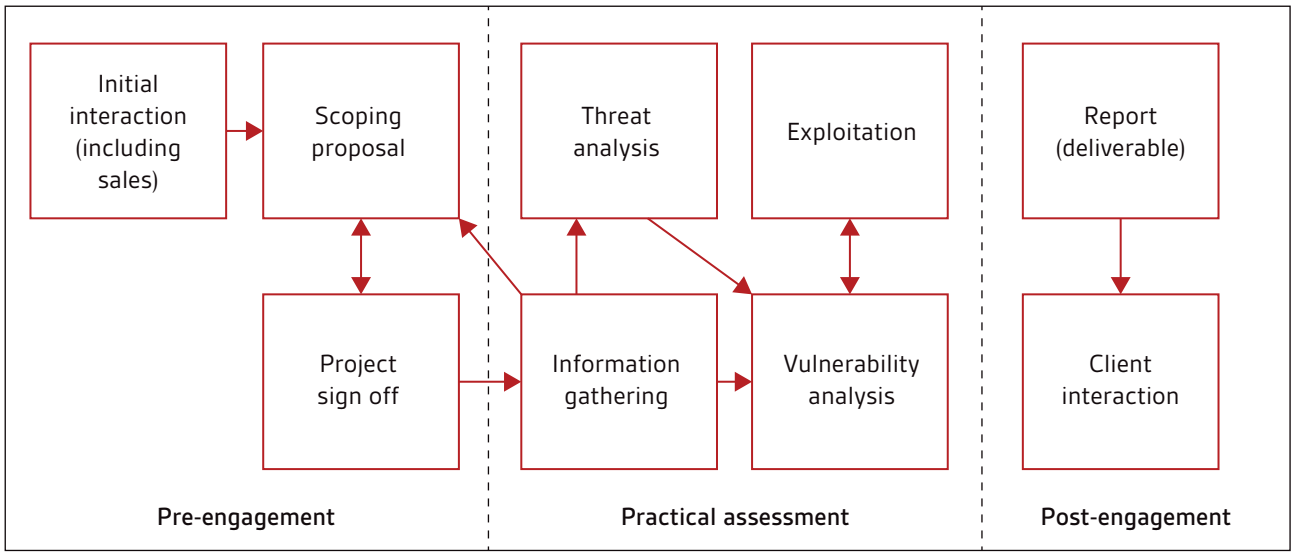
The implications of formal standards and certifications for providers will be discussed later in Section 3; at this juncture, however, two important survey findings were made regarding the use of penetration testing for compliance where security controls are established.

Firstly, stakeholders from all three categories felt the link between penetration testing and ISO/IEC 27001 was currently 'disparate' and poorly documented. Two interrelated approaches for establishing a link emerged from early interviews, and subsequent stakeholder views were widely positive for both. The first approach was to establish a clear link between the activities within a penetration test and ISO/IEC 27002 security controls, and the second was to establish greater auditor guidance for using penetration tests as audit evidence, within the larger ISMS audit. Arguably, the former must happen to enable the latter. Criticism was expressed by one stakeholder, whose views are notable due to their proximity to the standardization process. This stakeholder felt the approach was at odds with the ISO/IEC 27001 model, which was not about security in itself, but knowing insecurity and planning for continuous improvement. This stakeholder added: 'Why favour a particular method over another? Why is penetration testing better than auditing security records?' Risk must be identified to be managed, however, and for other stakeholders, the enthusiasm was focused on penetration testing's ability to assess controls in demonstrable terms.

Secondly, whilst the motivations behind Cyber Essentials were widely applauded, it was criticized for its lack of target market, while provoking explicit and widespread confusion about its implementation. Such confusion arose primarily from the heterogeneous approaches of the accreditation bodies. Frequent remarks concerned the integration of companion standards within Cyber Essentials, where vulnerability assessments were required (or where they were not), and the separation of accreditation and certification status. Some providers further questioned whether consistency could be achieved due to the ambiguity in the testing guidelines, and the subjectivity required to implement them.

The penetration testing process and the need for standardization

Figure 1 – Penetration testing phases



Stakeholders were also questioned on their practices and experiences around penetration testing. To contextualize the findings, the terminology for an engagement's subprocesses has been defined through a reference model in Figure 1. This reference model was derived from stakeholder responses. The model splits an engagement into three broad phases: pre-engagement, practical and post-engagement.

The *pre-engagement* phase is concerned with establishing the parameters of allowed activity for the practical assessment. There will be some form of *initial interaction* which may be initiated by the provider or the client. A chosen methodology (e.g. questionnaires or interviews) will be used by the provider to generate a *scoping proposal*. This proposal may go through multiple rounds of negotiation. The client will then *sign off* on the proposal before the practical assessment begins.

The *practical assessment* phase involves the exposure of a client system or component to a simulated attack. The phase begins with *information gathering*. This may uncover systems that necessitate further discussions around the engagement scope (e.g. if a client uses systems owned or operated by third parties and there are questions of testing authorization). A provider may conduct a threat analysis or move straight to its subsequent stage, a *vulnerability analysis*. *Exploitation* of identified vulnerabilities may occur in order to attempt penetration of the system, and to gain access to additional resources (e.g. sensitive data or higher system privileges). The subprocesses of the practical assessment stage may go through multiple repetitions (e.g. a compromised system may be connected to another internal network which, if under scope, can also be attacked).

The *post-engagement phase* is concerned with the delivery of findings to the client, usually in the form of a written *report*. The majority of providers will supplement this with additional forms of *client interaction* (e.g. final meetings) in order to educate them about the findings and the remedial actions that need to be undertaken.

Comments resulting from the stakeholder interviews concerning these three stages of the penetration test engagement will now be discussed in turn.

Pre-engagement

> *'The quality of the marketing collateral of penetration testing companies leaves a lot to be desired. I think it's a marketplace that's shrouded in mystery and myth. It's very difficult as a person wishing to purchase penetration*

> *testing and IT Health Check services... to assess the marketplace and find out whether or not your potential vendors will satisfy what you require, other than them being able to say that they're CREST or CHECK registered... it almost feels like you need to be an expert yourself to buy an expert to come in and help you... Being able to come up with a framework with which you can engage these suppliers, and understand the nature of the different tests that they will do, and how they will treat that information in terms of reporting it back, and there being some consistency across the marketplace... I think that would be a very welcome development.'*
>
> A client of penetration tests

There was a predominant sense of confusion and frustration amongst stakeholders about the ambiguity in what constitutes a penetration testing service. Such ambiguity was evident from the varied service definitions of providers, in particular around the level of exploitation that occurs during engagements. A number of providers stated that vulnerabilities within engagements were not exploited by default, with additional value provided through theorized exploitation and/or false negative and positive verification. This caused a commonly cited issue during the tender process, which was found to be increasingly common for the procurement of penetration tests. Clients were often found to be unable to differentiate between providers, even amongst some of those that had approved CHECK or CREST status, while providers argued that clients often failed to understand their requirements, provided limited opportunities for consultation, and made procurement decisions based predominantly on economic factors. Providers argued that this could lead to clients failing to procure a level of testing rigour appropriate to the requirements of their environment. Some providers felt this was in part because clients are not concerned with the quality of the test: 'clients are just looking for a tick in the box and resent any issues found'. The issue from client and provider perspectives is related, and can arguably be reduced to issues with terminology for defining services.

The definition of consistent terminology was widely supported amongst providers (18) and clients (5).[7] Another two providers expressed support at a conceptual level, but argued practical definitions were difficult to determine; the subjective nature of exploitation was cited as an issue by one provider. Questions around terminology arose from an early interview with a provider who suggested that the market would benefit from BSI working with partners to define testing types. This provider argued that 'it might not be right; it might get slaughtered, but it's a stake in the ground', where clients can say they want a peer defined simulated security evaluation (e.g. vulnerability assessment or penetration test) and have a clearer understanding of the service that they desire and what will be delivered.

As the support for terminology definitions suggests, the industry is acutely aware of the issues caused by the lack of precise terminology. One initiative with potential industry impact is the community standard, the Penetration Testing Execution Standard (PTES).[8] The upcoming version of the PTES is stated to have 'levels' of testing (potentially four, according to one of three providers who feature in the PTES 'group'[9] that were interviewed as part of this study). Supportive providers felt levels would empower clients and facilitate the process of procuring a certain type or level of test. If a provider was then to fail to deliver the requirements of that level, they would be in breach of contract. One non-PTES provider was supportive of a level approach; however, they urged caution with definitions, as part of the process of adding value is having the power to deviate. Such issues could easily be addressed through clarifying testing requirements in pre-engagement negotiations. If a standard is too specific, however, it could cause issues if clients do not need an aspect of that test, and do not understand why they do not need it. Providers would then need to deliver unnecessary services to meet that level.

An alternative solution proposed by one industry stakeholder used a measure of the client's risk appetite to map onto industry services; however, it received strong opposition due to the difficulty in computing risk appetite (even amongst those versed in its specialism) and the lack of potential for internationalization where many providers wished to focus their efforts.

The scoping procedures of providers were found to have formed a de facto standard, with strong commonalities in the basic stages (see Figure 1), methodologies to derive client requirements (predominantly questionnaires), the structure of scoping proposals, and types of testing proposed (almost wholly white box). Client views on scoping were polemic. Providers were largely seen as providing adequate assistance; however, in some cases they were criticized for their excessive use of questionnaires and lack of face-to-face meetings, even by larger clients. One CESG Listed Advisors Scheme (CLAS) consultant argued that you 'often can't interact with penetration testing providers'. For some providers, especially the microenterprises, face-to-face meetings were pursued at all opportunities to differentiate themselves against the rise of 'faceless' providers.

7 These figures were reached without all stakeholders being questioned on the topic due to time restraints.

8 http://www.pentest-standard.org/index.php/Main_Page

9 http://www.pentest-standard.org/index.php/FAQ

Larger enterprises were considered to be the only clients who understood their requirements, and this often manifested in engagements with strict guidelines, goals and deliverables. Small providers were deemed to require greater levels of assistance, but have grown increasingly knowledgeable over the past three to five years through their periodic audits. A common area of contention amongst stakeholders was found in industries that mandate penetration tests; notably in the CHECK scheme for Public Services Network (PSN) compliance and PCI DSS. Multiple CHECK requiring clients expressed the opinion that their peers were intentionally narrowing the scope of engagements to minimize risk for any issues found due to the punitive nature of the scheme. Two CHECK providers stated that they had heard of such issues themselves. One client insisted that they were interested in having an expansive scope for the CHECK scheme. Such an approach provides financial benefits over having one approach for aspects critical to PSN compliance, and another non-mandatory test for other services. However, the punitive nature discourages such an approach, with the additional complexity that any issues found lead to poor reflection of security capabilities compared to peer organizations. The Cabinet Office does provide a four-page document (two pages of content) on scoping CHECK; however, for the clients the call was clear: greater guidance is needed to ensure consistency within the scheme.

Questions of authorization arose when engagements involved third party services. All providers stated confirmation was sought before engagements began; however, the methods used varied. The preferred method involved the client signing a document to state their legal authority to test systems within scope, with the provider requiring no further proof. One provider stated this was because it was 'too time-consuming to check it all'. A minority of providers required explicit confirmation from the third party. Cloud services were an exception, with providers often demanding email or written confirmation from the cloud service. Such authorization was found to be obtained with relative ease, except for smaller or 'non-Western' cloud services. Providers stated that undisclosed third party systems were often uncovered during the initial reconnaissance stage of the practical assessment, notably with mail servers, and that the lack of third party authorization was a common reason for delayed testing.

Providers were questioned on their understanding of the legality for conducting engagements outside UK borders. Providers were largely unaware, with the bigger providers stating that such engagements would be cleared by their legal department. The general approach was to offset risk onto the client on the assumption they would have greater knowledge of local laws, and to ensure to never stray from the scope set out for the engagement. Legal cover would then be provided by authorization from the client. One provider felt there was not enough legal guidance around the Computer Misuse Act (1990), even within the UK. 'Where does the Misuse Act stop and a new law begin?' This proved to be a bigger issue for smaller provider organizations. One such provider stated that they have had multiple enquiries about work from the USA; however, they have not taken on business because they do not understand how 'protected they are'. This provider felt that UK Trade & Investment should be making more effort within this domain. 'They have to understand that process and advise if they're wanting to push an export strategy'.

Practical assessment

> '*We're aligned to OWASP.*'
>
> '*Whatever's available.*'
>
> '*A combination of everybody's.*'
>
> '*Our internal methodology is based on all that stuff.*'
>
> Providers on methodologies

Known hostilities towards standardization of the technical aspects of penetration testing led to a strategic choice in study design to focus on other aspects of engagements. The hostility towards uniformity in testing procedures was evident through the description of the methodologies used by providers. All providers stated that the methodology was their own, but influenced by other community standards. Out of the 32 providers, 10 mentioned Open Source Security Testing Methodology Manual (OSSTMM) and 16 mentioned Open Web Application Security Project (OWASP) in general terms, with three providers more specifically mentioning the OWASP Testing Guide. Other methodologies in use during the interviews were the PTES by six providers and those adopted by the National Institute of Standards and Technology (NIST) (again in general terms, without reference to a specific standard) by three providers. One provider noted that NIST 800–115 has gained prominence in the past year, due to the new PCI DSS version 3 standard, but described it as 'old and invalid'. Providers stated that no public methodologies were followed for high

assurance and/or safety-critical environments requiring specialist approaches, such as industrial control systems, as none have been created. Despite the lack of peer reviewed methodologies, it is worth stating that both CHECK and CREST organizations need to have and provide a defined methodology before acceptance onto the schemes.

The potential legal and ethical perils of social engineering were also discussed heavily with clients, who expressed a consensus that while the industry is largely self-responsible, there is room for legal and ethical guidelines 'for the protection of those doing the testing, as much as anything else'. However, it was unanimously agreed that this was not for the domain of standardization. For UK providers, CREST was frequently cited as the desired source as it could align with their existing code of ethics and complaints processes.

Post-engagement

> *'I've never seen any "wow" reports, but a lot of bad ones.'*
>
> *'Shocking.'*
>
> *'Generally very hit and miss.'*
>
> *'Appalling.'*
>
> Providers on the reports of other providers

'Underwhelming' was the overarching theme in the perceptions of reporting from providers and clients. Providers expressed satisfaction with the quality of their own reports, but had largely disapproving opinions of the reports produced by other providers. A small number of providers felt there was some consistency between their direct competitors, with one large provider arguing that a level of consistency had been achieved through the movements of individuals between provider organizations within the industry.

> *'The quality varies immensely... the quality can be atrocious.'*
>
> *'Often basically a Nessus output in PDF format.'*
>
> *'Very impressed.'*
>
> *'... great deal of variability.'*
>
> *'Some are atrocious; others well thought out.'*
>
> *'The quality of the document was high.'*
>
> *'No significant quality variation.'*
>
> *'Some are so shocking, it's hilarious.'*
>
> Clients on reporting quality

Client interviews highlighted a significant perceived variability in the quality of reports from providers. The above quotes were extracted from the views of eight clients in eight organizations. One interesting finding was that the smaller clients had the best opinions on the quality of reporting. Generally, the larger the client, the greater the perceived variability.

Two widely cited issues that will not be discussed here are the packaging of vulnerability assessment as penetration tests, and the 'quality' of report content. The former is a systemic issue that stems from pre-engagement negotiations. The latter is about individual capability, which stakeholders strongly felt should be the responsibility of technical bodies.

All providers were found to follow a similar high-level reporting structure. At its most basic, all reports were described to have a managerial and technical section. Managerial sections typically contained the executive summary and engagement details (e.g. scope). Clients were moderately satisfied with provider efforts, but many felt managerial sections were still too technical, and often needed rewriting for internal communications. The technical section broadly contained the lists of discovered vulnerabilities and recommendations. The provider's implementations for both were varied. Some best practices for reporting structure that were noted include the use of document histories, information on providers involved in testing (e.g. qualifications), engagement narratives, root cause recommendations and appendices of test data (e.g. logs of tool outputs and systems 'touched' during testing). Problems with reporting began to surface when looking beyond their high-level structure. Issues can be split over three areas.

The first is the diverse use of default metrics for scoring vulnerabilities. The Common Vulnerability Scoring System (CVSS) version 2.0 was mentioned frequently, and was often mandated by some clients; however, providers were critical of it in its current form, arguing that it was only suitable for certain technologies, its scores often did not reflect real-world risks, and that it failed to account for vulnerability chains (e.g. multiple medium risk vulnerabilities created one of high risk) or the presence of mitigating security controls. Instead, providers frequently described the use of alternative metrics, such as: qualitative scores (usually high, medium and low); impact to Confidentiality, Integrity or Availability (CIA); ease of exploitation; proprietary CVSS derivatives; or a combination of multiple metrics in a matrix. For clients, the variety of scoring mechanisms was found to be problematic for tracking performance over time and comparing results between providers. Furthermore, issues were felt to be compounded due to the subjectivity in arriving at a particular score, such as when providers tried to adapt CVSS to account for its aforementioned limitations, or address one or more aspects using their own metric system. The survey highlighted a strong opposition to the potential for mandating a specific metrics system, as this is where providers felt their value was generated; however, some providers did feel that clients mandating the inclusion of unmodified CVSS scores regardless of the use of another 'main' metric system would provide a 'quick win' for consistency within the industry. Version 3.0 of CVSS is currently in development, and some providers expressed the hope that this would lead to a natural resolution and improvement of this issue.

The second area of concern is the quality and content of recommendations. Smaller clients were the most satisfied, with larger clients having more qualms. Frequent criticisms included the lack of prioritization (beyond the implicit prioritization based on the vulnerability score), categorization, and root cause analysis. Root cause analysis featured heavily in client demands, but providers were largely seen to be failing to deliver on this. One client was particularly critical of CHECK reports for their lack of root cause analysis, stating that it 'rarely happens in their CHECK reports' and that there is 'no interpretation of results'. Only seven providers (six organizations) stated that they included any root cause analysis in any form of penetration test report, although more did state that recommendations were prioritized. One of the largest clients in the project went further to argue that providers need to include scenarios in their root cause analysis to enable a greater understanding of vulnerability chains and their impact. Interestingly, a criticism of clients that arose at multiple points within the study was the claim that they often only spot fix, rather than address root causes, and that issues continue to appear in subsequent engagements (e.g. their yearly audit). While the ultimate responsibility to address systemic issues lies with the client, based on the findings in this study, it would be difficult to claim that many providers are going to great lengths to facilitate this.

The third issue is once a client has implemented remediation to vulnerabilities, they then have two options for validating its efficacy. They could either obtain a retest of the vulnerabilities (usually at an additional cost) or test

for the vulnerability themselves. Most clients of penetration testing engagements do not have the skills or training to understand and recreate the vulnerabilities themselves, which therefore means they must be empowered to do so. Only nine providers (seven organizations) stated that proof-of-concepts were included within their reports (e.g. a single command or script that can be queried or executed to demonstrate the issue), with clients describing their presence as rare. The majority of providers offered retests instead, although some providers stated that some information was provided, such as 'what tools were used'. The majority of clients expressed an interest in proof-of-concepts being made available, with some clients stating that they would also like to see attack narratives. One client stated that this was because remediation is often undone, and attack narratives would facilitate better understanding of cyber threats and empower them to implement more effective mitigating security controls. A provider argued that not including narratives or proof-of-concepts in reports was a 'business decision' by provider organizations, with another suggesting the same issues, arguing that this information will typically be produced to enable them to conduct a retest anyway. The provision of such information aids in educating the client to improve their security, but doing so would not be financially beneficial for providers.

The difficulty in achieving greater quality of reporting is balancing the need for consistency with the resistance to standardization and the providers' desire to maintain flexibility in the reporting process. Auditing and setting guidelines were two methods suggested by providers that could help to achieve this.

CHECK reports are reviewed by CESG for quality and metrics. Any issues that are found are raised with the customer and/or the CHECK company. This is supplemented with an annual audit where CHECK companies are requested to send two examples of work that best demonstrate their technical ability. CREST reports are not audited. Two providers (one CREST organization) argued that 'they should be doing them.' However, without regulatory or other external support (e.g. as with the CHECK scheme) such an approach could see opposition from providers, in addition to the practical challenges of implementing this in the private sector (e.g. due to issues around client confidentiality).

Authoritative guidelines on reporting best practices were suggested in the belief that if clients had access to such guidelines, their expectations would raise the reporting standards by providers. PTES does contain reporting standards; however, PTES has failed so far to achieve widespread awareness amongst the buying community. The provider community is aware of issues around reporting, and some providers are taking steps to address this. One example that the authors were made aware of during the study was a community project that aims to create a baseline, minimum standard for reporting. The output will be a series of guidelines outlining best practices, and an example report that will be made available to the public (i.e. both providers and clients). The example report will be produced based on the findings of a real engagement undertaken by the project's group. This project involves some providers within the PTES group; however, this project will be independent from PTES.

Conclusions

The CHECK scheme, along with technical bodies such as CREST and Tigerscheme, have successfully defined the technical capabilities of individuals who perform penetration tests, and can be seen to be making great efforts to encourage evolution within the industry. In addition, both CHECK and CREST have laid the foundations for the assessment of organizational processes that support engagements. Despite this, there are a number of issues at the start and end of the penetration testing engagement process that the industry has currently failed to address. This is not through a lack of awareness of these issues; this study has highlighted that both providers and clients are dissatisfied by the lack of transparency and consistency in industry offerings. There are issues uncovered in this research which call for further discussion and analysis. A separate publication will be released by Lancaster University in 2015 exploring various aspects of the penetration testing process in more depth.

Standards, of course, need to be well formed to avoid the potential to suffocate and hinder rapidly evolving industries, such as the one we find with penetration testing. Self-regulation in such an environment is an ideal solution; however, as one provider stated, when it comes to current industry offerings, it can be a 'Wild West'. This is not due to a lack of capability within the industry, as the technological bar has been set and maintained by the technical bodies. One provider argued that the 'UK security industry can provide anything the market asks for'; the problem is that 'it [the market] does not ask the right questions'. Standards bodies can facilitate this process of enabling and educating the market to do so. Based on this study's findings, three recommendations are proposed. Given the importance (and rapid growth) of penetration testing, resolving these needs for best practice quickly would aid providers and buyers.

Terminology standardization

Providers offer a diverse mix of qualities and depth of penetration tests. The buying market in the words of one provider is in need of something to compare 'like-for-like'. Standardization of terminology is recommended to enable clients to make more informed procurement decisions. Stakeholders argued such an approach would aid in addressing the commoditization of penetration testing, whilst not being tied to a specific region, and thus open to internationalization.

The PTES is one notable community project, which is working towards a similar goal. Standards bodies should look towards developing relationships with community efforts to achieve similar terminology models. Such an approach leverages existing work by subject matter experts. Furthermore, in the opinions of providers, the reputation of standards bodies can aid in bringing these concepts to the mass market.

Reporting guidelines

Although the high-level structure of reports was found to be relatively standardized between providers, at a lower level, stakeholders were found to be dissatisfied with various characteristics of reports. For providers, it was predominantly their experiences of seeing competitors offering vulnerability assessments badged as penetration tests. The most effective resolution to this arguably comes not from a focus on the report itself, but around the standardization of terminology. For clients, it was the inconsistency between providers, and a lack of depth in the provision of metrics and recommendations (e.g. root cause analysis) to empower the client to understand more about the security issues within their environment.

The survey results recommend such a standard to be a guideline rather than a specification. Rigid standardization would likely see significant opposition; providers see their reports as a means to differentiate themselves and add value to their offering to the client. However, guidelines describing best practices could be produced, and have the potential to provide an effective alternative. Through standards bodies, such guidelines would likely gain significant exposure within the buying community, which may be otherwise difficult to achieve. Exposure facilitates education, which empowers clients to make informed decisions when interacting with providers, and gives them a clear conception of what they should expect. Such guidelines could address all of the aforementioned issues within this study, while describing best practices around the processes that support report production (e.g. quality assurance). As with the recommendations on terminology, standards bodies should look to leverage existing efforts within the community to raise reporting standards. This should include working with technical bodies in the UK, such as CREST, while remaining aware that guidelines should not be region-specific.

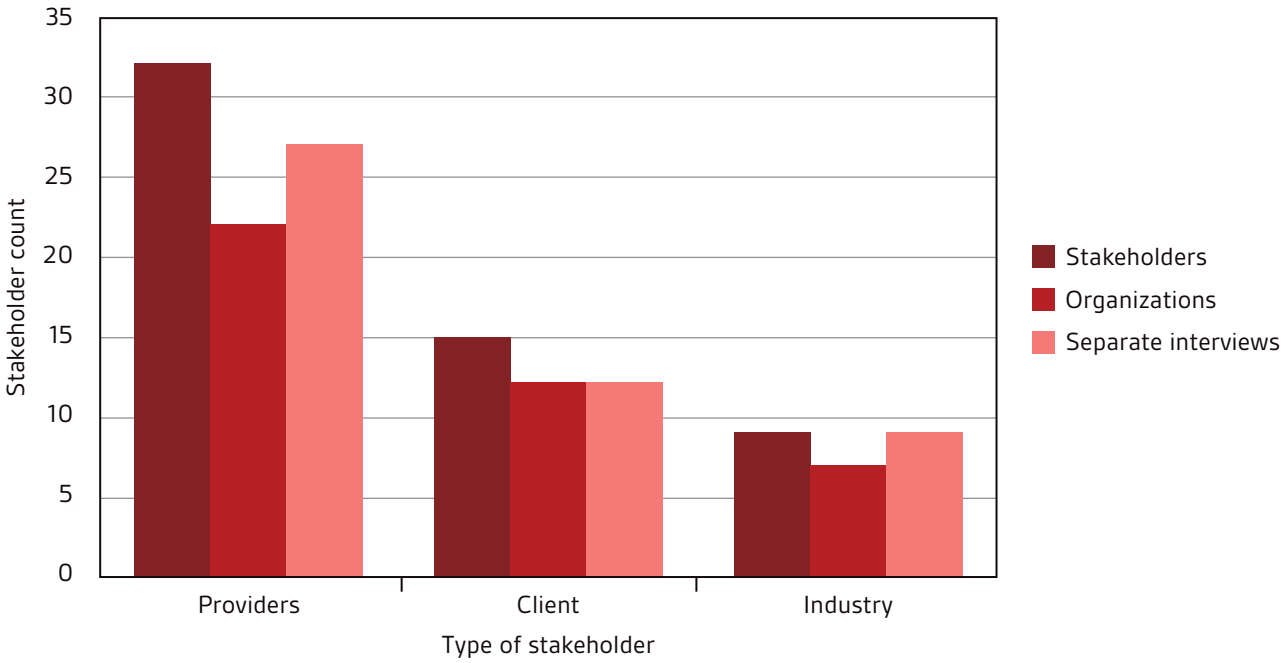
Penetration testing auditing guidelines (primarily for ISO/IEC 27001)

Penetration tests currently contribute to ISO/IEC 27001 audits under an isolated ISO/IEC 27002 security control 'technical compliance review'. Auditing guidelines have been produced previously in ISO/IEC 27008; however, they are in need of revision. Standards bodies should look to provide a greater link between the scope of an engagement and its findings, and ISO/IEC 27002 security controls beyond the narrow categorization of a technical compliance review. Stakeholders mentioned the perception of ISO/IEC 27001 being a 'checkbox exercise', and where penetration testing is of relevance in the audit process, the audit is merely a confirmation that it has occurred, rather than a detailed analysis of its findings to determine whether the security controls that have been implemented are consistent with the objectives of the ISMS.

Auditing guidelines provide the opportunity to link the sociotechnical security controls of ISO/IEC 27002, and the sociotechnical assessment of a penetration test. Penetration testing may only be one assessment methodology, but it continues to rise in popularity and is increasingly seen as a regulatory requirement. Furthermore, it is arguably the most realistic methodology currently available for simulating cyber threats. As part of this process, it is recommended that standards bodies examine the integration of penetration testing with other standards, such as ISO 31000. A diverse array of metrics can be proposed as part of penetration testing engagement, but what is its meaning for risk management, and how does this impact the risk that is to be managed as part of ISO/IEC 27001?

Annex A: Stakeholder composition

Figure A1 – Stakeholder interview composition



In total, 54 stakeholders were interviewed across 46 separate interviews. Stakeholder categories and composition are described below. The duration of provider interviews was between 40 minutes and 2 hours, and between 25 minutes and 1 hour 15 minutes for client interviews.

- **Providers** – 32 stakeholders were interviewed across 27 separate interviews, and 22 provider organizations. This includes providers of penetration tests in its various forms and services (e.g. IT Health Checks, and various variants of specific technology-focused 'penetration tests'). Two stakeholders were employed by provider organizations based outside the UK. Out of 32 providers, 10 (across 8 organizations) were from CHECK accredited organizations, and 18 providers (across 13 organizations) were from CREST member companies. Insufficient information was collected to calculate the number of CHECK or CREST qualified individuals that were interviewed. Such individuals do not need to work for CHECK or CREST organizations, nor does one have to be CHECK or CREST qualified to work for one.
- **Clients** – 15 stakeholders were interviewed across 12 separate interviews with 12 client organizations. To achieve a broad representation of client experiences, the organizational size of clients interviewed was highly varied. Client stakeholders ranged from microenterprises (i.e. < 10 employees) to large enterprises (i.e. < 250 employees, with 3 stakeholders in organizations of > 1000 employees). Furthermore, 5 representatives from UK local government were interviewed. Included within the total client count were 2 stakeholders who worked in consultancy roles (e.g. in one case, as a CLAS consultant) to procure penetration tests and identify remediation strategies for third parties.
- **Industry bodies** – 9 stakeholders were interviewed across 9 separate interviews with 7 industry bodies. Included within this count were 2 stakeholders from provider organizations who also spoke about their roles within an industry body. This count of 9 stakeholders does not include the 3 providers who spoke about their work on the community standard, PTES. Stakeholder organizations included: BSI; CESG; the Department for Business, Innovation and Skills (BIS); Tigerscheme; IASME; and QG Management Standards.

Annex B: Key points for new and modified standards

Standardization of terminology

- Clients argued the quality of service they received from different providers was highly varied, and that the level of testing offered by different providers was unclear.
- Providers were vexed at the limited opportunities for client interaction during the tender process (which is becoming increasingly common), and the potential loss of work due to the homogeneous perception of penetration testing, with clients choosing providers purely on cost. Homogeneity exists because of a lack of varied terminology and the 'widespread' 'up-branding' of offerings considered vulnerability assessments to penetration tests.
- Stakeholders expressed a desire for clearer service definition, but providers were opposed to tiered levels of *services* and certifications, due to concerns around internationalization.
- Terminology enables differentiation, mitigating penetration testing's commoditization.
- Existing subject matter experts (e.g. PTES) have identified this need and are pursuing solutions; standards bodies should look to establish community links.

Reporting guidelines

- Clients and providers were critical about the consistency of the industry's reporting.
- Individual competence is not the cause of reporting failures; it is an operational decision.
- Metric use is heterogeneous; clients had difficulty tracking performance over time and between providers.
- Providers do not empower their clients to understand the security implications of their environments (e.g. through attack narratives and root cause analysis of vulnerabilities).
- Providers felt the report is where value was added to an engagement, and were strongly opposed to strict standardization of reporting content.

Auditing guidelines for ISO/IEC 27001 where penetration testing contributes as audit evidence

- Stakeholders argued that the characterization and isolation of a penetration test as a 'technical compliance review' diminished its potential contribution to the assessment of controls. Such controls are not purely technical, but could also be social and physical.
- Penetration testing (and its derivatives) was felt to provide a demonstrable means of assessing security controls, above and beyond alternative assurance techniques.
- There was limited awareness of ISO/IEC 27008; where it existed, stakeholders considered the standard inadequate and in need of updating.
- Stakeholders felt the checklist culture for 'Plan, Do, Check, Act' of ISO/IEC 27001 restricts auditor analysis of operational risks – such guidelines would institutionalize it.

Annex C: Standards referenced

References below correct as of February 2015, please reference BSI or ISO/IEC for latest versions.

BS EN ISO 9001:2008. *Quality management systems. Requirements*

BS ISO 31000:2009. *Risk management. Principles and guidelines*

BS ISO/IEC 15408–1:2009. *Information technology. Security techniques. Evaluation criteria for IT security. Introduction and general model*

BS ISO/IEC 15408–2:2008. *Information technology. Security techniques. Evaluation criteria for IT security. Security functional components*

BS ISO/IEC 15408–3:2008. *Information technology. Security techniques. Evaluation criteria for IT security. Security assurance components*

BS ISO/IEC 18045:2008. *Information technology. Security techniques. Methodology for IT security evaluation*

BS ISO/IEC 27001:2013. *Information technology. Security techniques. Information security management systems. Requirements*

BS ISO/IEC 27002:2013. *Information technology. Security techniques. Code of practice for information security controls*

PD ISO/IEC TR 19791:2010. *Information technology. Security techniques. Security assessment of operational systems*

PD ISO/IEC TR 20004:2012. *Information technology. Security techniques. Refining software vulnerability analysis under ISO/IEC 15408 and ISO/IEC 18045*

PD ISO/IEC TR 27008:2011. *Information technology. Security techniques. Guidelines for auditors on information security controls*

The research presented in this report was undertaken by:

William Knowles
William is undertaking an EPSRC Industrial Case Ph.D. that is supported by the Airbus Group (formerly EADS) where he researches Industrial Control System security metrics. This Ph.D. is being undertaken at Security Lancaster, an EPSRC and GCHQ recognized Academic Centre of Excellence in Cyber Security Research. He is also a qualified Tigerscheme Qualified Security Team Member (QSTM) and ISO/IEC 27001:2013 Lead Auditor.
Email: w.knowles@lancaster.ac.uk. Twitter: @william_knows.

Alistair Baron
Alistair is a Security Lancaster Research Fellow in the School of Computing and Communications at Lancaster University, UK. His primary research involves applying natural language processing techniques to cyber security challenges, including social engineering, extremism and other serious online crimes. Alistair also teaches penetration testing and digital forensics modules on the GCHQ-certified M.Sc. in Cyber Security at Lancaster. He has a B.Sc. (Hons) and Ph.D. in Computer Science from Lancaster University.
Email: a.baron@lancaster.ac.uk. Twitter: @al586.

Tim McGarr
Tim is the Market Development Manager for the Information Technology area within Standards Development in BSI. Tim has specific responsibility for the direction and development of newer standards areas. Tim has been working at BSI since 2009. Prior to BSI, he spent 5 years working in the legal publisher LexisNexis in the strategy department. Before this, he worked as a management consultant for CGI and an internal consultant for BT. Tim has an MBA from HEC Paris, France.
Email: tim.mcgarr@bsigroup.com. Twitter: @Tim_McGarr.

The research was supported by:

Security Lancaster
Security Lancaster is one of only four flagship Lancaster Research Centres and was amongst the first eight Academic Centres of Excellence in Cyber Security Research in the UK. It is one of the few multi-disciplinary centres to tackle human and technological challenges to cyber security by integrating computer science and communication systems researchers with expertise from sociology, international relations, behavioural science and law. Involving over 70 researchers, the centre is internationally renowned for its cyber security research on network resilience, security of communications, securing mobile and cyber-physical systems, intelligent tools for tackling cyber crime and studies of user behaviours leading to cyber security threats. This research constitutes a multimillion pound grant portfolio funded from a variety of sources, including research councils (EPSRC, Economic and Social Research Council (ESRC)), the European Commission, JANET and direct investment from industry and government organizations.
URL: http://www.lancaster.ac.uk/security-lancaster/.

The authors would like to thank the EPSRC through its financial support for Lancaster University's Impact Acceleration Account (Grant Reference EP/K50421X/1).

About BSI Group

BSI is the business standards company that equips businesses with the necessary solutions to turn standards of best practice into habits of excellence. Formed in 1901, BSI was the world's first National Standards Body and a founding member of the International Organization for Standardization (ISO). Over a century later, it continues to facilitate business improvement across the globe by helping its clients drive performance, manage risk and grow sustainably through the adoption of international management systems standards, many of which BSI originated. Renowned for its marks of excellence, including the consumer recognized BSI Kitemark™, BSI's influence spans multiple sectors including aerospace, construction, energy, engineering, finance, healthcare, IT and retail. With over 70,000 clients in 150 countries, BSI is an organization whose standards inspire excellence across the globe.
URL: bsigroup.com.

BSI Group Headquarters
389, Chiswick High Road
London W4 4AL
United Kingdom

T: +44 (0) 845 086 9001
E. cservices@bsigroup.com
bsigroup.com

BSI UK
Kitemark Court
Davy Avenue
Knowlhill
Milton Keynes MK5 8PP
United Kingdom

T: +44 (0) 845 080 9000
E: MK.customerservices@bsigroup.com
bsigroup.com

BSI Group America Inc
12950 Worldgate Drive
8th Floor Monument II
Herndon
VA 20170
USA

T: +1 800 862 4977 / 703 437 9000
E. inquiry.msamericas@bsigroup.com
bsiamerica.com