IWS –The Information Warfare Site
Infocon Magazine Issue One, October 2003
http://www.iwar.org.uk/infocon/

Corporate Open Source Information Leakage

A self-assessment approach using Internet sources

Raphael Rues
University of Leicester, Scarman Centre
Risk, Crisis and Disaster Management
E-Mail: raphael.rues@swissonline.ch

1 Introduction

Nowadays organisations and industries have become more and more inter-connected. It is a fact that almost every major company maintains an Internet presence and relies heavily upon global and converged infrastructures. Only few companies have a web site consisting of just one page; the majority have widely and extensively integrated the Internet into their business models and supply chain (Netcraft, 2003). The result of such an integration is the availability of a considerable amount of corporate open source information on the Internet (Schneier, 2000; Murray, 2001). This information, which can be collected by simple means, poses many risks (Rues & Kunz, 2003). It can potentially be grouped into two categories:

1) Technical information necessary to connect to the Internet.
Essentially all pieces of information necessary for an organisation to maintain a presence on the Internet. Technical details such as web-server platform, operating system, firewall, IP-addresses, etc.. Unfortunately, this information can pose a serious risk, allowing a potential aggressor to gain knowledge of the corporation's IT architecture. It is a undisputable that knowledge of software vendor and version information provides a huge advantage to attackers in penetrating a system (Lee *et al.*, 2002). The website Netcraft (www.netcraft.com) with the feature 'What's that site running' offers an example of the kind of available information (Netcraft, 2003)

2) Information related to a specific organisation.
This category relates to the amount of corporate information originating from employees, managers of departments such as marketing, partners and consultants, as well as former employees – basically every kind of corporate activity for which the user has left an 'electronic mark' on the Internet (Nielsen, 2002). Unlimited amounts of information are easily available to the information-seeking individual by using Internet search engines which are becoming more and more efficient (Calishain, 2003). As an example, the Amazon website (www.amazon.com) through its 'purchase circles' service offers the possibility to ascertain the top ten books bought by employees of major corporations.

The particular self-assessment approach used in this paper offers a demonstration of corporate open source information gathering. Our aim is to demonstrate how selected indicators can provide a wealth of sensitive information (Scheer *et al.*, 2000). For each indicator the paper provides a summary of risks and recommendations concerning the issue of corporate information leakage. The paper concludes with a discussion of the possible measures available to organisations in order to reduce the threat of corporate information leakage.

2 Methodology

This paper attempts to demonstrate the state of transparency as well as the amount of 'open source information' available on selected anonymous European organisations by using a particular empirical 'corporate information leakage inventory' based on ten indicators.

The ten indicators are derived from a wide variety of security guidelines, such as COBIT (ISACF, 2003) and ISO 17799 (ISO, 2003). The authors attempted to integrate indicators derived from these two guidelines with indicators taken from the German IT Baseline Protection Manual (BSI, 2003). These ten indicators represent potential weak points from an information warfare as well as from an information security perspective.

	Indicator
1	Domain information
2	Administrator information
3	Pages referring to the organization
4	Pages referring to the organization firewalls
5	Pages referring to organisation projects
6	Pages containing a 'reference' to the company
7	Pages referring to former employees (curriculum vitae)
8	Total amount of Usenet postings with organisation's domain / address
9	Usenet postings referring to problems within the organization
10	Usenet postings referring to discussion of specific software (Microsoft) and referencing the organization

Table 1: Indicators used for the 'corporate information leakage inventory'

The methodology proposed in this paper does not intend to exhaustively cover the problem of information leakage but rather to offer the possibility to evaluate what types of sensitive corporate information are easily available on the Internet.

3 Findings

This chapter presents the key assessment's findings and recommendations for containing the 'corporate open source information leakage'.

3.1 Domain information (Indicator 1)

Query: www.checkdomain.com or national Network Information Centre's (NIC)

The domain of the given organisation is established by querying the meta-domain search engine Checkdomain (www.checkdomain.com). Another possibility, although more time consuming, is to query the respective national network information center's (NIC). Assuming

that the given organisation owns several domains, the following URL of Netcraft website, http://www.netcraft.com/?host, allows to search by domain. By simply entering the name of the organisation, all registered domains will be provided (eg., organisation.de, organisation.net etc.). The result of such a domain query presents a wide array of information pertaining to the organisation.

```
Registrant:              Company Name
Address:                 Street, ZIP, Place, Country
First Registered:        September 17, 2002
Administrative Contact:  John Doe (john.doe@organisation.domain)
                         Street, ZIP, Place, Country
                         Phone / Fax
Technical Contact:       John Doe (john.doe@organisation.domain)
                         Street, ZIP, Place, Country
                         Phone / Fax
Name Servers:            DNS.Organisation.Domain IP-Address
                         NS1.Organisation.Domain IP-Address
Information Source:      Network Solutions
```

Figure 1: Example of organisation domain registration

Not only the administrative contacts but also the details for the technical contact can be found. The person listed under technical contact (see Figure 1 – John Doe) is usually also the administrator managing the networking assets. In some cases the technical contact lists also the specific function/department of the person (eg., security, communication, etc.) Therefore, it is possible to quickly gain the identity of the people managing the organisation's most sensitive assets with a simple query, unless the data is anonymized (e.g. helpdesk contact).

Summary of risks & recommendations

- Risk: Too many administrators and organisations register their domain with complete personal data. In several cases even mobile telephone numbers are also registered.
- Risk: In various cases among the full identities it is also possible to relate the exact job/duty of the administrator (technical contact).
- Recommendation: Organisations should register their domain (technical contact) anonymously (eg., dns-admin@organisation.domain), or even better using the helpdesk contact (eg., administration domain).

3.2 Administrator Information (Indicator 2)

Query www.google.com with: administrator generalities

As illustrated above it is not unusual for administrators to register the organisation domain with their proper full name. By querying Google with the generalities provided by the administrator it is possible to gain further personal background information on the specific person. For one administrator, it was possible to retrieve web documents which contained his holiday pictures, a full curriculum, as well as his personal interests. From the five financial institutions being part of the information pool used for this survey, it was possible based on

the administrator's name, to relocate technical postings as well as hardware and software information specific to the employer.

Summary of risks & recommendations

- Risk: Administrators and engineers are to be made aware of the risk of publishing information related to their organisation's infrastructure.
- Recommendation: specific awareness and guidelines to limit the amount of information heading to administrators.

3.3 Pages referring to the organisation (Indicator 3)

Query www.google.com with: +organisation.domain +e-mail

A simple search on Google based on the organisation domain would return too many documents, most of them not necessarily relevant to the organisation. The caveat of such an indicator is therefore the considerable amount of data available in Internet. The idea behind this specific query is therefore to gain documents where employees have left their 'fingerprints'. The combination of 'organisation domain' and 'e-mail' does indeed return a specific amount of open source information relevant to the organisation. To be noted, that in most cases where an e-mail address is given, additional personal information (e.g. phone numbers, pictures, postings in discussion forums, hobbies details) are available.

Summary of risks & recommendations

- Risk: Several documents have no relation to the organisation mission statement, and are of private nature.
- Risk: Employees use corporate e-mail addresses for a variety of purposes have nothing to do with their business activity. Examples of such usage are: membership in political parties, leisure activities as well as ads in personal contacts websites.
- Risk: Increase of targeted e-mail spam.
- Recommendation: Internet usage guidelines should specify that e-mail addresses are to be used only for corporate activities.
- Recommendation: Specific awareness for employees and management on issues such as privacy and spam should be realized.

3.4 Pages referring to the organization firewalls (Indicator 4)

Query www.google.com with: +organisation.domain +firewall

This indicator has manifold utility. The most obvious information that can be gained is the specific firewall vendor and version information used by the organisation (Lee *et al.*, 2002). In several cases this indicator also reveals the identity of the firewall administrator (see Indicator 2). An example of information collected using this indicator is the following posting:

```
Hi , I have a complex problem. I have fw-1 4.0 sp7 on a risc6000(
aix4.3.3).en0 is the administration interface( a 10/100) and en3 is a
Gigabit Interface.fw-1 is controlled by the management (win nt, fw-1
4.1 sp2 pathech for backward compatibility to 4.0) over a VLAN 1 on a
Cisco switch 6500. My problem is the following: i have a workstation
on the VLAN2 [IP-Address….] The rule is : User@any   [IP-Address….]
When fw-1 is down, everything works well: for this reason , the cause,
in my opinion can't be Gigabit interface of aix. Somebody can help me?
John Doe (john.doe@xxx.xxx)
```

Figure 2: Posting excerpt on a firewall vendor-specific forum

A search on Google for this person (as discussed above with Indicator 2) revealed further documents referring to technical problems of this specific organisation. For a total of three companies, one of them a financial institution, information on installation and operating problems that affected their firewalls in the past two years are available.

Summary of risks & recommendations

- Risk: Availability of information on specific weaknesses of the organisation published predominantly by administrators and engineers, allowing to potentially prepare an intrusion.
- Recommendation: Public postings (e.g. on vendor discussion forum) of sensitive information should not be permitted. Alternatively, administrators and engineers should anonymise their postings, so that assets and identity of the organisation remain concealed.
- Recommendation: Specific awareness for administrators should be realized.
- Recommendation: Constant monitoring of the published information

3.5 Pages referring to projects (Indicator 5)

Query www.google.com with: +organisation.domain +(project OR projects)

This indicator allows to gain a detailed view of projects originated by the specific organisations. By using the Boole'an operator OR various variations of the keyword project can be searched. The collected data indicates a variety of projects information realized by the organisation. The amount of information 'left behind' is considerable: for two insurance companies classified data (contained in Powerpoint presentations) related to their IT network could be retrieved.

Summary of risks & recommendations

- Risk: Various organisations publish several details regarding their projects. It is not only possible to ascertain the budget and/or the project team (with phone numbers and even pictures), but also to collect project reports, as well as hardware and software details.
- Risk: In several cases minutes and projects reports, to be deemed in some cases confidential, are encountered.
- Recommendation: Project managers should be aware of information about their project that could potentially leak out. Awareness of the people working on the project (e.g. freelancers) should specifically target this risk.

3.6 Pages referring to organisation references (Indicator 6)

Query www.google.com with: +organisation.domain +(reference OR references)

Indicator 6 "references" provides information about an organisation's business networks. Most of the data gained through this indicator are pertinent to assets and services (for extent software and hardware details) sold to the organisation by a business partner.

Summary of risks & recommendations

- Risk: Merging of references reveal details about the organisation software and hardware assets.
- Recommendation: Companies should specify in their contracts with external partners that information about projects, even in form of best cases or references should not be disclosed or published.

3.7 Pages referring to former employees (Indicator 7)

Query www.google.com with: +organisation.domain +(curriculum OR resume OR lebenslauf)

The interesting aspect of this indicator is related to the wide range of information that can be gained regarding the present and past activities of the specific organisation. A project manager provided a complete list of projects that were performed at a company in the energy market. The projects covered a span of three years, including their outcome, the total investment and project return on investment (ROI).

Summary of risks & recommendations

- Risk: In search for a job, confidential information about past activities can be published in resumes/curriculums by former employees.
- Recommendation: Constant Internet monitoring on specific information contained in curriculums and affecting the organisation.
- Recommendation: A legal clause in the working contract should prohibit to post on the Internet resumes/curriculums containing financial and technical details of past activities.

3.8 Usenet postings with organisation's domain / address (Indicator 8)

Query groups.google.com with:: +author:@organisation.domain

This indicator is used to measure the total amount of Usenet postings indexed by Google Newsgroups (groups.google.com) for a given organisation. The analysis is based solely on the organisation domain. Such an indicator allow to quickly ascertain how widespread the usage of Internet within the organisation is. To be noted that most of the published information is originated by people working in the IT department.

Summary of risks & recommendations

- Risk: The absence of specific Internet guidelines creates the risk that all sorts of information, including sensitive technical information, is published in newsgroups.
- Risk: Productivity of employees can be analysed and measured. For one employer, an insurance institution, more than 340 posting generated by a single person were found in a timeframe of 12 months. 90% of postings have nothing to do with technical problems but with scuba diving, parachuting and formula one races. By clicking on the specific posting, under view original format, technical details of the organisation as well as the timestamp of the posting can be gained.
- Recommendation: Awareness of employees, especially about the fact that postings once published can not be easily removed (see Chapter 4 Measures).
- Recommendation: Internet usage guidelines should clearly address the usage of Usenet.

3.9 Usenet technical postings (Indicator 9)

Query groups.google.com with: +author:@organisation.domain +problem

The goal in using this indicator is to measure the quantity of specific technology information related to problems affecting the organisation hardware and software assets.

Forum	Date
adobe.acrobat.windows	15. Sep 2003
comp.unix.solaris	02. Sep 2003
comp.os.vms	01. Sep 2003
comp.lang.perl.misc	29. Aug 2003
comp.soft-sys.wxwindows	29. Aug 2003
mailing.unix.ethereal-dev	27. Aug 2003
comp.os.qnx	24. Aug 2003
comp.lang.java.security	22. Aug 2003
comp.lang.pascal.delphi.databases	22. Aug 2003
comp.unix.shell	17. Aug 2003
weblogic.developer.interest.management	15. Aug 2003

Figure 3: Selected organisation Usenet postings containing the indicator 'problem'

Most of the technical postings are generated by few people. These users also seem to work almost exclusively in IT departments.

Summary of risks & recommendations

- Risk: Indicator 9 "problem" leads to the discovery of various technical problems affecting organisations. In some cases the publication of problems appear to be originated by consultants, looking to solve the IT problem of the contractor. Recommendation: Internet usage guidelines should specify the use of Usenet.
- Recommendation: Specific awareness training for employees and management on issues such as knowledge management and communication among teams should be conducted.

3.10 Usenet postings referring to discussion of specific software (Indicator 10)

Query groups.google.com with: +author:@organisation.domain +group:*microsoft*

This last indicator is used to measure the type of postings related to a particular software vendor. For demonstration purposes, Microsoft was chosen. The basic use of this indicator is the possibility for various organisations to not only assess the employed software/hardware but also the types of problems affecting a specific product.

Summary of risks & recommendations

- Risk: Although specific technical information can be easily gained by other means, it is possible to ascertain the magnitude and time of problems and issues affecting a specific organisation.
- Risk: Various postings illustrate the use of specific software, in this case components of the Microsoft Office Professional suite.
- Recommendation: An analysis of the postings originating from the organisation could be used to offer specific training for more efficient use of the software package.

4 Measures

As mentioned initially, this survey is far from depicting the complete picture of the open source information issue as well as the problem of corporate information leakage into the public domain. Nevertheless, it provides an impressive snapshot of the amount of information available from publicly accessible sources. Specifically the survey emphasize the following recommendation:

4.1 Organisational and technical measures

- implementation of internet usage guidelines should address the disclosure of sensitive information (Indicators 1-10)
- guidelines for participating in public online forums (Indicators 4, 8-10)
- security/information disclosure guidelines should include and consider external partners (Indicators 5,6)
- setting up a strategic "weak signals" interpretation unit, working on the basis of Open Source Information indicators in order to reliably identify low-probability / high-impact events (Indicators 1-10)

4.2 Remove content from Google's index

Google allows to remove single document from its database. If the assessment contained in this paper should reveal organisation sensitive information which should be better deleted, you can access the remove content possibility under this URL: http://www.google.ch/remove.html.
By choosing remove individual pages you will access the process for deleting the sensitive content. As a matter of authentication, Google ask you to register yourself with E-Mail to which you will receive the necessary identification steps in order to delete the material.

4.3 Awareness

The organisation should develop an 'Information Security Program' intended to meet the basic goal of providing security commensurate with the risk and magnitude of the harm resulting from the loss, misuse, or unauthorized access to or modification of information used by the organisation. Following activities should be included:

- awareness programs and specific training to administrators and engineers evidencing the threat of 'corporate open source information leakage' (Indicators 1-10)
- improvement of corporate communication culture, specifically for administrators and engineers (Indicators 4,9-10)

5 Conclusions

This paper, based on a framework of ten indicators, has illustrated the following aspects:

First, the collected data illustrates and emphasizes the human contribution in disclosing sensitive data. The phenomenon does not seem to be pervasive among all employees, but limited to selected groups such as administrators and managers.

Second, an analysis of the collected data illustrates that most corporations are "open" to the outside, in spite of most likely having a security organisation and having invested heavily in security (e.g. implementing hardware such as firewalls). The data collection for a company takes only a few minutes, yet it yields a huge amount of information, which often creates more than "transparent" enterprises as it was demonstrated.

Third, the organisation should design and implement appropriate information security controls to limit the leakage of information, including that lost through industrial espionage. Given the sheer amount of information circulating in most companies, it is important to focus the effort – by assessing the overall extent of information leakage and undertake an impact assessment to identify the most important risk areas. The self-assessment approach introduced in this paper, can offer a first idea about the organisation situation.

Fourth, the organisation should develop an early warning system intended to ensure that, on a continuous basis, risks and dangers - in particular those that could pose a threat to the company's existence - are identified in good time, and that the responsible decision-makers are informed so that countermeasures can be taken. This early warning system should identify typical risk areas for the organisation that are monitored continuously or on an ad hoc basis with the help of key data and indicators. Such an early identification system enables all relevant changes in industry and society that are communicated through the press and other media to be identified and monitored by specialists. By properly interpreting the weak signals, the organisation will also be able to apply the concept of 'lessons learned' therefore preparing adequately the organisation to possible future problems (Toft & Reynolds, 1997: 54-58).

Finally, the amount of corporate open source information available on the Internet, more than being a technical and organizational issue, appears to be a cultural issue. Differences in communication and corporate culture illustrate how most of the employees have taken on the Internet as a means of pervasive communication.

It is the authors´ goal to further enhance the method presented. To assist in performing a self assessment, the possibility of developing an automated tool is currently being discussed. A further issue to be researched is the feasibility of an independent benchmark authority that would assess the corporate level of transparency, based on information leakage. Such an isomorphic approach would allow the creation of best practices for protecting and restricting access to corporate open source information.

6 Annex 1

	Indicator	Specific Google Boolean Operator
1	Domain information	www.checkdomain.com and /or http://www.netcraft.com/?host
2	Administrator information	www.checkdomain.com
3	Pages referring to the organization	+organisation.domain +(e-mail OR email)
4	Pages referring to the organization firewalls	+organisation. domain +firewall
5	Pages referring to organisation projects	+organisation. domain +(project OR projekte)
6	Pages containing a "reference" to the company	+organisation. domain +reference
7	Pages referring to former employees (curriculum vitae)	+organisation. domain +(curriculum OR resume OR lebenslauf)
8	Total amount of Usenet postings with organisation's domain / address	+author:@organisation. domain
9	Usenet postings referring to problems within the organization	+author:@organisation. domain +problem (*or +problema, +probleme*)
10	Usenet postings referring to discussion of specific software (Microsoft) and referencing the organization	+author:@organisation. domain +group:*microsoft*

Table 2: Google Boolean operators used in the survey (Calishain, 2003:3-11)

7 References

BSI (2002) „Grundschutzhandbuch - IT Baseline Protection Manual“, Bundesamt für Sicherheit in der Informationstechnik, Edition May 2002, Bonn
URL: http://www.bsi.de

Calishain, T. & Dornfest, R. (2003) *Google Hacks*, Sebastopol CA: O’Reilly

ISO (2000) “ISO/IEC 17799:2000 Code of Practice for Information Security Management”
URL: http://www.iso.ch

ISACF (2003) “ISACF: Control Objectives for Information and Related Technology”
URL: http://www.isaca.org/cobit.htm

Lee, D. et al. (2002): “Detecting and Defending against Web-Server Fingerprinting”, in *Proceedings of the 18th Annual Computer Security Applications Conference*, Las Vegas 2002.
URL: http://www.acsac.org/2002/papers/96.pdf

Lyman, P. et al. (2000) “How Much Information?”, School of Information Management and Systems at University of California at Berkeley
URL: http://www.sims.berkeley.edu/how-much-info

McClure, S. et al. (1999) *Hacking Exposed: Network Security Secrets and Solutions*, Berkeley
CA: Osborne / McGraw-Hill,

Netcraft (2003) Netcraft Web Server Analysis
URL: http://www.netcraft.com

Rues R., Kunz, P. (2003) “Geschäftsrisiken von gläsernen Unternehmen” in *D.A.CH Security*, University of Erfurt, Germany

Scheer, S. et al. (2000) “A Methodology to Retrieve, to Manage, to Classify and to Query Open Source Information - Results of the OSILIA Project"
in - *JRC Technical Note No. I.01.016,* ISIS - RMDS, Anti-Fraud Information Management, Milan
URL: http://hosting.jrc.cec.eu.int

Schneier, B. (2000) *Secrets and Lies: Digital Security in a Networked World*, Wiley Computer Publishing, New York

Toft, B & Reynolds S. (1994) *Learning from disasters: a management approach*, Oxford: Butterworth-Heinemann