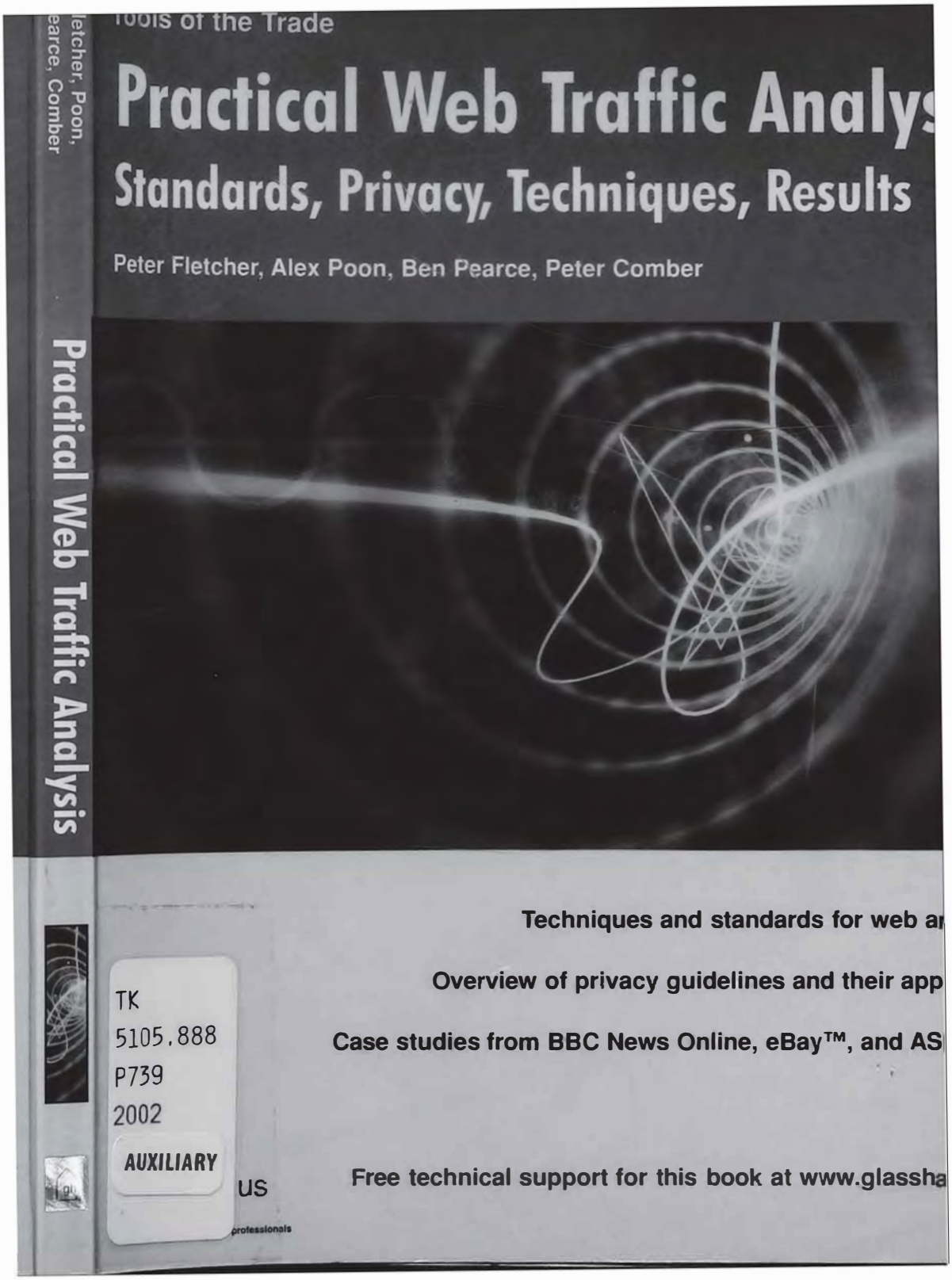


Amazon/AlmondNet
Exhibit 1011
Page 1

Summary of Contents

Introduction		1
Chapter 1:	Analytics Techniques	9
Chapter 2:	Database Log Analysis	39
Chapter 3:	Privacy	61
Chapter 4:	BBC News Online	87
Chapter 5:	eBay	111
Chapter 6:	ASPToday	139
Index		163



labor-saving devices for web professionals

Visit www.glasshaus.com





ISBN 1-90415-118-3
52999>

Practical Web Traffic Analysis:

Standards, Privacy, Techniques, Results

Peter Fletcher

Alex Poon

Ben Pearce

Peter Comber

© 2002 glasshaus

Published by glasshaus Ltd,
Arden House,
1102 Warwick Road,
Acocks Green,
Birmingham,
B27 6BH, UK

Printed in the United States
ISBN 1-904151-18-3

Practical Web Traffic Analysis:

Standards, Privacy, Techniques, Results

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical articles or reviews.

The authors and publisher have made every effort in the preparation of this book to ensure the accuracy of the information. However, the information contained in this book is sold without warranty, either express or implied. Neither the authors, glasshaus nor its dealers or distributors will be held liable for any damages caused or alleged to be caused either directly or indirectly by this book.

Cover Image

James Chew is currently studying Multimedia Computing in Melbourne, Australia. Graphic Design started out just as a hobby, but quickly grew to become somewhat of an obsession in extending his knowledge and skills. Apart from design, James plays guitar and draws inspiration from music. James also enjoys designing, and redesigning his website *http://chewman.plastiqueweb.com* which is an online folio of work he has done.

James can be contacted via his website or through e-mail, *chewman@studiowhiz.com*.



labor-saving devices for web professionals

© 2002 glasshaus

Trademark Acknowledgements

glasshaus has endeavored to provide trademark information about all the companies and products mentioned in this book by the appropriate use of capitals. However, glasshaus cannot guarantee the accuracy of this information.

eBay™ screenshots used by permission of eBay™ Inc.

BBC News is a trademark of the British Broadcasting Corporation and is used under license. Screenshots and logos used by permission of the BBC.

ASPToday screenshots used by permission of ASPToday.

Credits

Authors
Peter Fletcher
Alex Poon
Ben Pearce
Peter Comber

Additional Material
SmartGirl.org
The Onion
NYTimes.com

Technical Reviewers
Dick Bennett
Jon Duckett
Richard Foan
Mark Horner
Tim Luoma
David Wertheimer

Proof Reader
Agnes Wiggers

Commissioning Editors
Peter Fletcher
Amanda Kay

Technical Editors
Mark Waterhouse
Alessandro Ansa

Managing Editor
Liz Toy

Publisher
Viv Emery

Project Manager
Sophie Edwards

Production Coordinators
Pip Wonson
Rachel Taylor

Cover
Dawn Chellingworth

Cover Image
James Chew

Indexer
Bill Joncocks

About the Authors

Peter Comber



It is nearly seven years since Pete Comber started working with the Internet. Having just started working in direct marketing, like most junior execs he quickly became frustrated with the fragile, incomplete, and temperamental customer databases upon which the ambitious DM strategies tended to exist. With the Internet boom in the late 1990s, he became convinced that there was massive potential for the Internet to provide large volumes of data not just about what people purchased, but about what they looked at, what they discarded, and potentially granting some insight into how people came to make their buying decisions. Having analyzed Internet data for four years for the UK's biggest motoring web site, he joined the BBC at the end of 2001 in the hope that would be able to provide some useful insights into how people use the BBC News Online service.

Peter Fletcher



Peter Fletcher has been working in web development since 1997, via a degree in Philosophy and Theatre Studies and an MSc in Cognitive Science. Interests, professional and recreational, include using the Web as a distributed communications medium, analyzing web traffic data, and working with experimental performance group Stan's Cafe. His personal projects are documented at *www.joyfeed.com*. He is now a freelance web consultant and writer, dividing his time between Birmingham and Barcelona.

Ben Pearce



There's plenty about me later on, so I'd like to just say some thank-yous...

I would firstly like to thank Pete Fletcher for giving me the opportunity to talk about my work in this book. I hope it is useful and interesting reading for you.

I would also like to thank my wonderful wife Ange for all her support (not to mention the cakes!), and of course Jesus Christ who is my inspiration in everything.

Alex Poon



Alex Poon, originally from Baton Rouge, Louisiana, now lives with his wife Buffy and son Tyler in Northern California. He has a PhD from Stanford University. He was one of the first engineers at eBay, and started off as the de facto "UI guy" back in 1997. He started eBay's first user-interface group at eBay in 1999, then later ran the Advanced Technologies Group, during which his team implemented eBay's first web analytics system. Although Alex recently left eBay after five years, the company was happy to have him describe its analytics process in his own words.

To Buffy, who means everything to me.

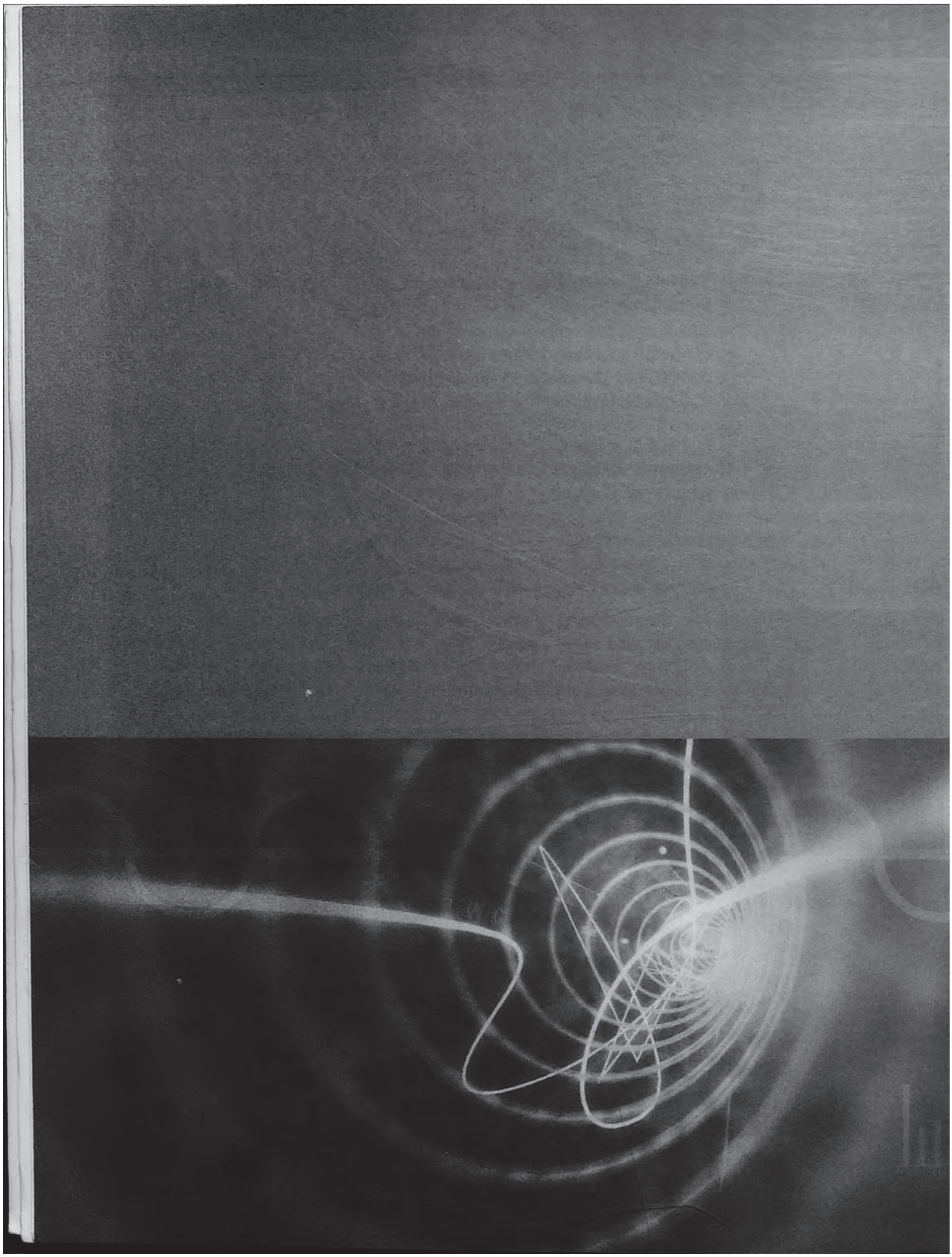


Table of Contents

Introduction **1**

What's it all About? **1**

Who is this Book for? **3**

What's Inside? **3**

Support and Feedback **6**

Web Support 6

Analytics Techniques **9**

1. Comparative Overview of Analysis Methods **9**

Server Logs 9

Advantages 10

Disadvantages 10

Conclusion 11

Panel Surveys 11

Advantages 11

Disadvantages 12

Conclusion 13

UNIVERSITY OF WASHINGTON LIBRARIES

Browser Analysis: Tagging, or Web Bugs....13
Advantages....13
Disadvantages....14
Conclusion....15
The Skeptical Analyst....15
Log Files....15
What Does the Log Not Tell Us?....17
What Can We Say for Certain?....20
Analog....20
Cookies....29
Standards for Web Traffic Analysis....31
Metrics....32
Summary....36

2. Database Log Analysis....39
Dealing with the Token....40
Token Creation....40
Cookie Management....41
The Database....43
Admin Tables....44
Data Tables....44
The Import Logs Module....44
The Import Queries....46

Amazon/AlmondNet
Exhibit 1011
Page 10

Starting the Import Process (frmRunImport) 48
Specifying the Format (spcLogFile) 48
Adding the Cookie Data (qryAppendCookies) 49
Adding the Page View Data (qryAppendPageViews) 50
Adding the Referrer Data (qryAppendReferers) 52
Adding the Click-Through Data (qryAppendLinkClickThroughs) 53
Clearing the tblLogs (qryDeleteLogsContent) 53
Finding Our Largest Referrer (qryRefererByDate) 54
Page Views by Date 55
Unique Cookie Visitors by Day 58
Total Clicks on a Particular Tracking Link 58
Summary 59

3. Privacy 61
Concerns 61
Cookies 61
Data Privacy 63
Legislation 64
The EU Directive and "Safe Harbor" 64
COPPA 66
Other Organizations 67
The Platform for Privacy Preferences Project (P3P) 68
APPEL 70

Amazon/AlmondNet
Exhibit 1011
Page 11

Case Studies: Real Life Privacy 72
SmartGirl: www.SmartGirl.org 72
Personal Information 73
User Tracking and Cookies 74
Opting In, Opting Out 74
Additional Measures and Security 74
The Onion: www.theonion.com 75
Personal Information 76
User Tracking and Cookies 77
Data Analysis 77
Opting In, Opting Out 79
The New York Times: www.nytimes.com 79
Personal Information 80
User Tracking and Cookies 80
Data Analysis 81
Opting In, Opting Out 82
Seal Programs 82
Summary 83

4. BBC News Online 87
Background and Overview 87
The Traffic Information We Gather 90
The Importance of Page Impressions 91
Overview of Technical Architecture 94

Amazon/AlmondNet
Exhibit 1011
Page 12

Three Requirements of Our Traffic Analysis 94
League Table Generator.....95
Audited Log Analysis 99
Use of Cookies 101
Detailed Visitor Analysis.....101
Drilling Down into User Behavior: The One-Hit Wonder Phenomenon 103
Analyzing Extreme Demand: September 11, 2001.....106
News Service or Archive Service? 107
Summary 109

5. eBay 111
What is eBay? 111
How eBay Makes Money 112
Why eBay Needs Web Traffic Analysis 112
Measuring the Completion Rate of Key Processes.....113
Measuring the Effectiveness of Marketing Initiatives.....113
Measuring the Effectiveness of Searching and Browsing 114
Gathering Technographic Data 114
Gathering Anonymous Visitor Statistics.....114
eBay's Web Traffic Analysis System 115
Requirements.....115
Page View Reports 115
Unique Visitor Reports.....116
Page Flow Analysis 117
Technographic Reports 118
Page Property Reports 119

Amazon/AlmondNet
Exhibit 1011
Page 13

Client-Side JavaScript 120
How the System Works 121
JavaScript, Cookies, and Web Bugs 121
Random Sampling 122
Respecting Users' Privacy 123
"Long Distance" Tracking 123
How Page Names and Page Properties Are Set 124
Putting It to Real Use 126
Selling Flow 127
The Old Flow Versus the New Flow 127
The Analysis 129
Bidding Analysis 131
The Four Finding Methods 132
The Goal 134
Our Analysis 135
Summary 136

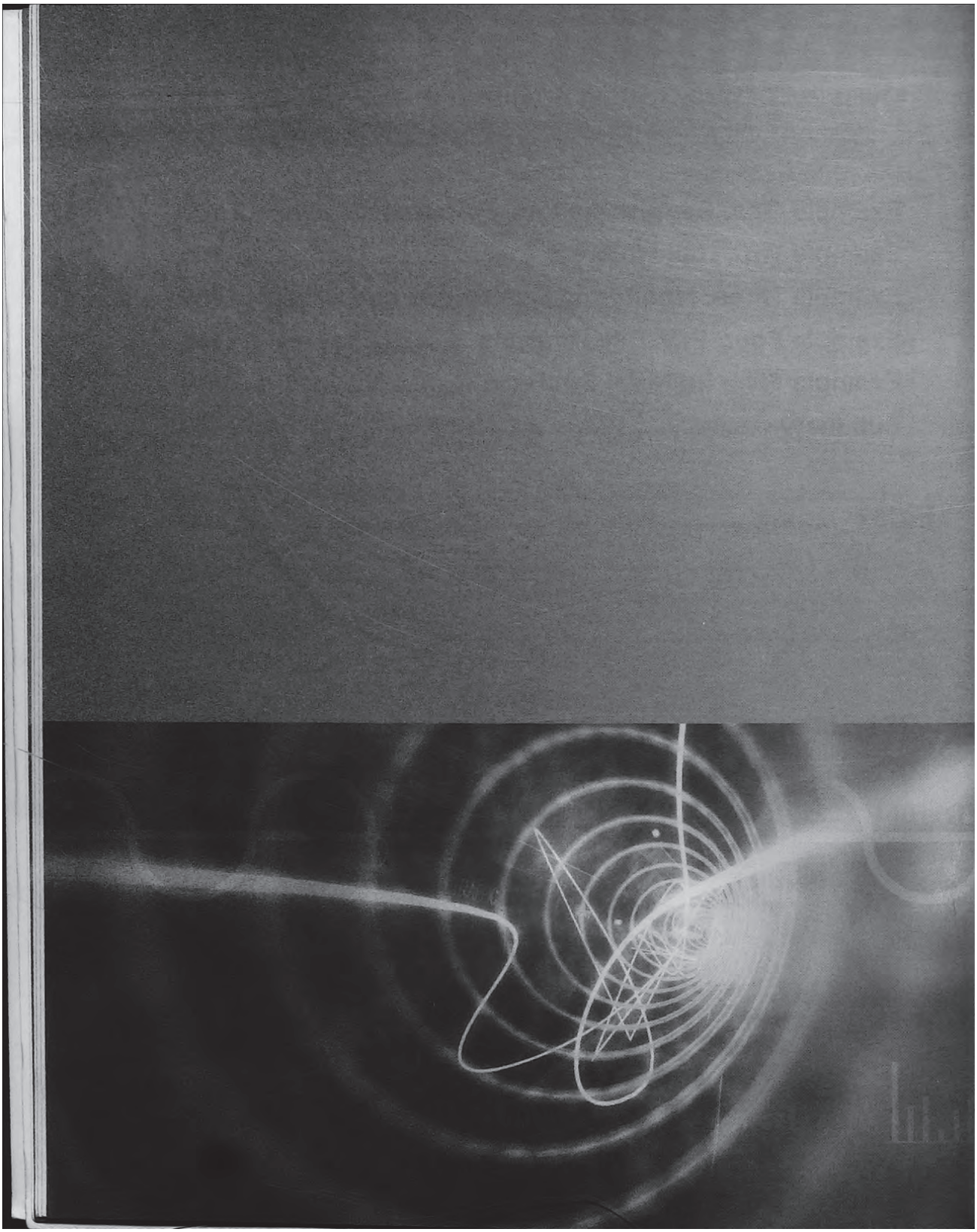
6. ASPToday 139
Content and Visitor Profile 139
History and Background 140
Role of Web Traffic Analysis 142
Problems with Early Log Analysis 143
Cookie Data Analysis 144

Amazon/AlmondNet
Exhibit 1011
Page 14

Overview of Web Hosting Architecture.....145
Overview of Business Logic Architecture.....145
Database Structure 148
Example One: Regular Visitors, Frequent Viewers 150
Example Two: Topic Interest 154
Example Three: Monitoring Campaigns and Schemes 156
Example Four: PDF Site Analysis 158
Example Five: Editorial Strategy 158
Summary 160

Index 163

Amazon/AlmondNet
Exhibit 1011
Page 15



Web Traffic Analysis Introduction

Knowing how much traffic your site gets, and which parts get more than others, can be vitally important to understanding what visitors are getting from your site. With a bit of analysis of your server logs, you could discover that people are using your site in a way you didn't expect. Once you get a better idea of usage patterns for your site, you can begin to work on ironing out any problem areas such as that page that most shoppers never get beyond.

So, if you've ever wondered how to find out which day of the week your site gets most visits, this is the book for you. In fact if you've ever wondered how much traffic your site gets, then this is definitely the book for you. In the coming pages we look at the most important aspects of analyzing web traffic, including a comprehensive discussion of the key terms and their usage.

What's it all About?

To get the full benefit from *Chapter 1*, there are things you should know before we get there. You'll have noticed that we didn't say "If you want to know how many people have visited your site, this is the book for you," and with good reason. As we'll learn in *Chapter 1*, we can't really know this mystical figure for a number of reasons. Chiefly, because the site is accessed via HTTP which only recognizes requests for information and doesn't maintain a connection with a visitor, we can at best only get an indication of how many people have visited the site in any given time period. Instead, it has become widely accepted that a different figure be used, one which we can measure accurately, and that is the number of page requests.

In fact, this is all we can really measure with any hope of accuracy – pretty much everything else is extrapolation and educated guesswork.

Don't worry that you won't be able to tell your advertisers that you don't know how many people visit your site each day, because they shouldn't be expecting you to tell them. They should be interested in how many page impressions your site generates per day, and with a little help from this book you'll be able to tell them.

Despite having said that there is only one thing we can measure with real accuracy, there are a number of other elements to analyzing your web traffic. The first of these is knowing where to find evidence of your traffic, and how to interpret it. As we'll see, initially in *Chapter 1* but also throughout the book, this information comes to you in the form of log files generated by the web server hosting your site. Every time the server receives a request for a page, be it a simple HTML file, a picture, or a dynamic server-side scripting file, it makes a record of the request. This record includes the date and time of the request, where it came from, the file requested, and the manner in which it was requested. Importantly, if there was a cookie involved in the request, the name of the cookie is recorded as well. Counting the unique cookies in each log is as close as we get to knowing how many people visited the site – not everyone accepts cookies in their browser, for a variety of reasons. Some people may visit the site from more than one computer, and some people may even use more than one browser on the same computer.

Once we have the log file, we need to know what to do with it. Rather than have someone with a pencil and some paper read through the whole thing tallying up the page impressions and unique cookies, there's a selection of software available to do the hard work for you. Some of this software is very expensive, and some of it, like the main tool we'll be using in this book, is so cheap it's free! Specifically, we'll be looking at how a tool called Analog can be used to interpret your web server's log files. If you feel up for a challenge, you could write your own log file analysis utility, a common choice. We'll be taking a look at how one of these works, the very one in fact that is used to analyze the log files for our own site, *http://www.glasshaus.com*.

Of course, as anyone who has done any research involving statistics will know, having some numbers doesn't mean you'll know what's happening. Interpretation, whether we like it or not, has some part to play, as well as application. In the selection of case studies that make up the second half of the book, we'll see some of the benefits that can be gleaned from in-depth analysis of your site traffic. If you know where people stop trying to pay for your products, you know where you should really improve your site.

Amazon/AlmondNet
Exhibit 1011
Page 18

Who is this Book for?

This book is for you if you need to learn more about web traffic analysis, whether you've been told to present some figures to wow the board or you've been told to get the figures so someone else can use them to wow the board. It's for people who want to know what to put in traffic reports, what you can know about your visitors, what you're not allowed to know, and of course how to get the information in the first place.

We're not going to assume that you know vast amounts about statistics, or are capable of hand-coding Amazon.com. What we do assume is an intermediate level of knowledge about the Web: basic scripting experience is useful for the technical parts, as is some familiarity with Access, and you'll need to know (or be prepared to learn) the basics of HTTP and how the Web works.

What do I Need to Begin?

For this book, a web site and access to its log files are a good start. Once you have your log files, you'll want to look at them with either Analog (*http://www.analog.cx/*), or the hand-built system available as part of the code download from *http://www.glasshaus.com*, which is built in Access 2000 (so you'll need a PC for those parts). We also use Excel 2000 for further analysis of the data from the hand-built system.

What's Inside?

As we've already seen, *Chapter 1* covers all the basics of web traffic analysis; from data analysis methods and theory to tools for log file analysis. If you find terms in later chapters that aren't explained there, *Chapter 1* is the place to look.

> *Reach is a common measure used in audience research, it measures the "market share" that any given site has claimed*

> *The server is unable to log where the user goes after leaving the site*

> *Applying standards and employing auditors is not a replacement for proper intelligent interrogation of the data being presented*

Amazon/AlmondNet
Exhibit 1011
Page 19

Chapter 2 takes a look at a hand-built web traffic analysis system, using Access 2000 and Excel 2000. We introduce the database in which log files are turned into easy-to-understand numbers, and look at how the information that goes in there is gathered, including setting cookies.

> *One of the key facets of this database solution is the ability of the analyst or administrator to specify exactly which data are required.*

> *The general principles that it exemplifies can be applied to almost any database and coding solution*

Chapter 3 tackles the thorny issue of privacy – it's all very well being able to track a visitor across your site, but some of the ways of doing this invade their privacy, and can expose us to legal problems. As well as examining what laws there are to be aware of, we talk about some real sites and discuss their privacy policy and how it impacts their gathering of web traffic metrics.

Chapters 4, *5*, and *6* are case studies, in which the people responsible for web traffic analysis for three large sites explains how they gather their data, and what they do with it.

> *There is no single "correct" attitude towards cookies, the situation changes depending on the circumstances*

> *It is especially important to engender trust in the mind of the user*

In *Chapter 4*, Peter Comber tells us how the BBC monitors the traffic on their vast content site, *News Online* (*http://news.bbc.co.uk*). As well as the differences in size, we see how useful it is to have frequent access to traffic reports, and how they indicate which stories people are most interested in.

> We have learned that most "web metrics" - including page impressions - are extremely context sensitive.

> *There was an assumption that, as the archive of articles was both so rich and so large, there must be a huge number of page impressions drawn from this part of the site*

Amazon/AlmondNet
Exhibit 1011
Page 20

In *Chapter 5* Alex Poon gives us the low-down on how *eBay* (*http://www.ebay.com*) deal with their own unique problems. Since knowing when people stopped trying to buy something is vitally important to any commerce venture, eBay used detailed web traffic analysis to discover the usage patterns of their visitors.

> *For every page on the eBay site, we wanted the ability to see how users got to that page, and where users went after it.*

> *We found it cost-prohibitive to track every single page view from every single eBay user. Instead, we implemented a random sampling algorithm that tracks only one out of every 100 users to the eBay site*

Finally, in *Chapter 6*, Ben Pearce talks about the development of web traffic analysis for the *ASPToday* site (*http://www.asptoday.com*), from its inception in February 1999 to the site's latest version in August 2002. Here we gain an insight into the ways that web traffic analysis can help to drive the redevelopment of a site.

> *The site depends upon accurate intelligence to inform the key business decisions that have influenced its evolution from a simple static web site to a dynamic e-commerce application.*

> *Learning from early experiences of log analysis, the technical team and business team worked together to determine what information was required*

Amazon/AlmondNet
Exhibit 1011
Page 21

Support and Feedback

Although we aim for perfection, the sad fact of book publication is that a few errors will slip through. We would like to apologize for any errors that have reached this book despite our best efforts. If you spot an error, please let us know about it using the e-mail address *support@glasshaus.com*. If it's something that will help other readers, then we'll put it up on the errata page at *http://www.glasshaus.com*.

This e-mail address can also be used to access our support network. If you have trouble running any of the code in this book, or have a related question that you feel the book didn't answer, please mail your problem to the above address quoting the title of the book (*Practical Web Traffic Analysis*), the last 4 digits of its ISBN (**1183**), and the relevant chapter and page number.

Web Support

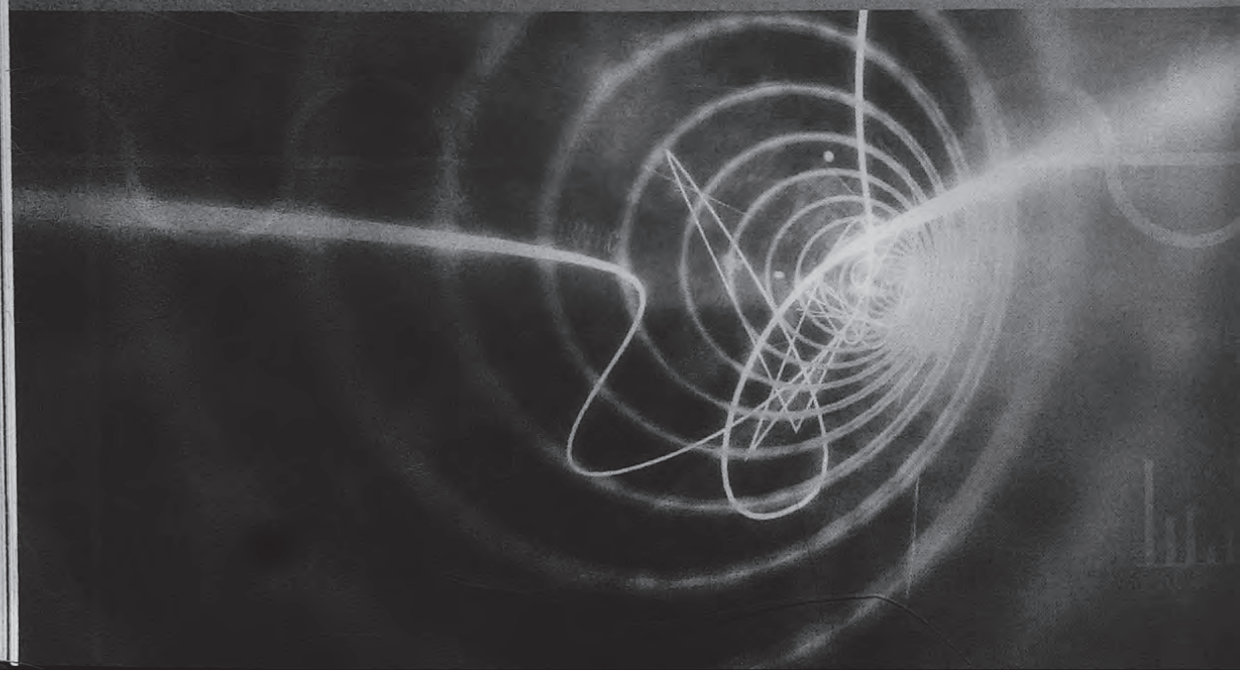
You'll want to go and visit our web site, at *http://www.glasshaus.com*. It features a freely downloadable compressed version of the code for this book, in both `.zip` and `.sit` formats. You can also find details of all our other published books, author interviews, and more.

Amazon/AlmondNet
Exhibit 1011
Page 22

1

- Theory and terminology
- Guidelines - what you can measure
- Using Analog

Author: Peter Fletcher



Amazon/AlmondNet
Exhibit 1011
Page 23

Analytics Techniques

We want to know who is coming to our web site, whether they have been here before, where they came from, what they are looking at, what route they take through the site, how long they spend on each page, and why they leave. Which, if any, of these objectives are possible?

In this chapter, we will define the landscape of Web Traffic Analysis, beginning with the various methods available to us, followed by a discussion of some of the advantages and disadvantages of these methods. Then we go back to basics, and determine exactly what it is that we can know, and what we cannot know, and the extent to which it is safe to make some assumptions. Having established these principles, we examine the emerging standards for web traffic analysis, and discuss whether there are any good "rules" that we should follow.

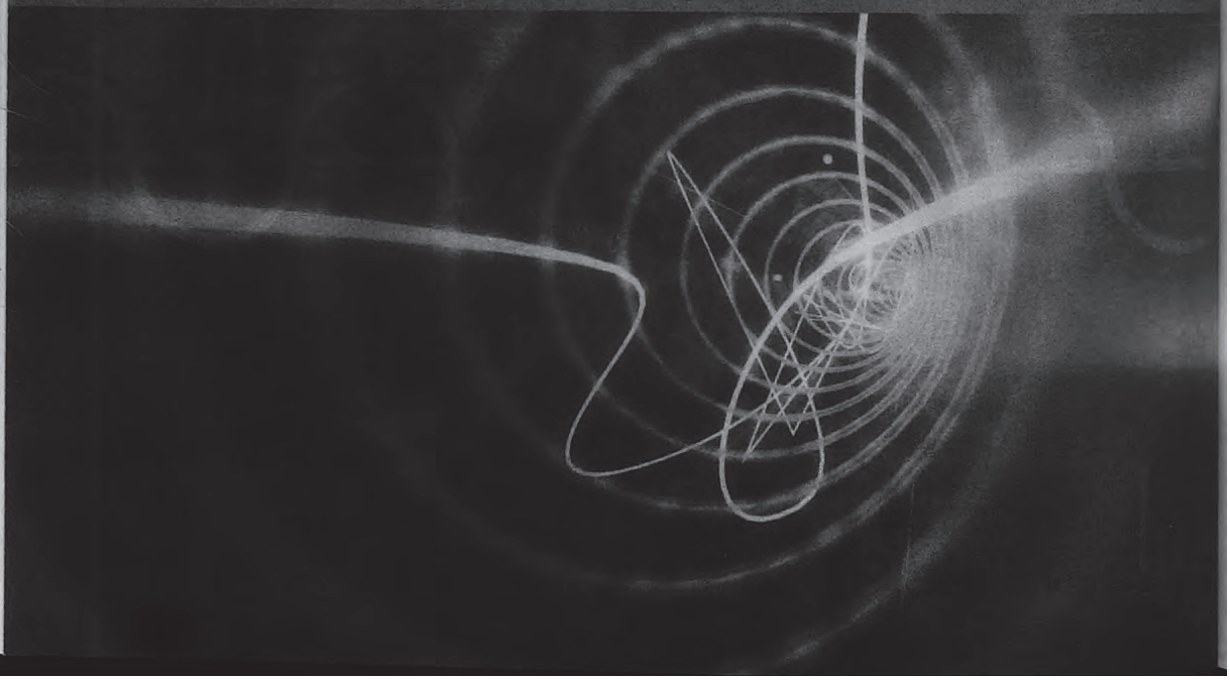
Comparative Overview of Analysis Methods

There are three broad types of web traffic analysis, defined by the source of the data that is being recorded and analyzed. These are data collected from activity on the web server, data collected from panel surveys, and data collected from the client or browser using "tags" or "web bugs". We can look at these in turn to determine their relative merits.

3

- Legislation
- Guidelines and organizations
- Case studies from SmartGirl.org, The Onion, NYTimes.com

Author: Peter Fletcher



Privacy

So far we have been discussing the technical aspects of traffic analysis and data collection, but privacy is becoming a very important part of building and managing a modern web site. Now we review the main elements and concerns of the privacy issue, the legislative constraints that are being developed, and the various protocols, organizations, and programs that are trying to make privacy less daunting and problematic both for web professionals and users. We then put this in context, by looking at three case studies, which demonstrate how actual web sites are meeting the challenges that privacy presents to both the business of the Web, and user confidence. Finally, we wrap up with a summary of best practices and guidelines, suggesting the issues to cover in a privacy policy.

Concerns

Concerns over privacy can be categorized into several interlinked areas, which we will look at in turn. They are, briefly: the identification of individuals' behavior (often connected to the use of cookies), the collection and abuse of personal data (such as e-mail addresses), the security of the data once collected, and the ability to verify – and delete – incorrect data being stored.

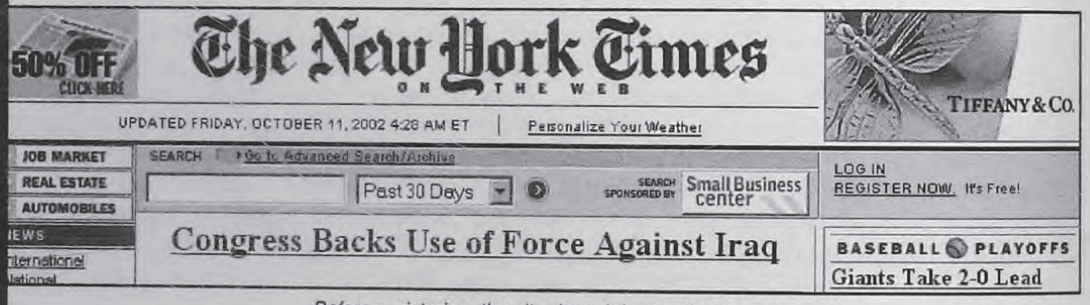
Cookies

As use of the Web has grown, users have gradually become more aware of the fact that site owners can gather and store information about their movements through a site. A fear of the implications of this has been projected onto the main technical innovation that makes it possible: the cookie. Opinion about this humble text file varies from the entirely benign view that cookies do not identify individuals and can be disabled by the user, to the more extreme belief that they are a grave threat to our personal privacy, and should be avoided at all costs. What is the correct attitude towards the privacy implications of cookies?

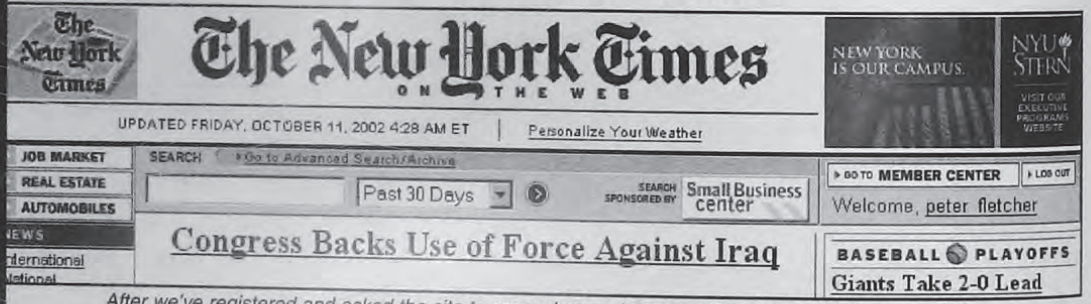
> *There is no single "correct" attitude towards cookies, the situation changes depending on the circumstances*

There is no single "correct" attitude towards cookies, and, like much else that we discuss in this book, the situation changes depending on the circumstances: the type of site, the user, what is done with the cookie and the information that it generates, whether the cookie is "permanent" or deleted at the end of the browser session, and the ability of the user to opt out of any data collection. The basic fact of the cookie, however, is that it does not by default identify a known individual to the web site; it is merely an anonymous token that can be used to help the site deliver personalized pages to the user, or count the number of unique users (as defined in *Chapter 1*) that are visiting the site. This is complemented by the fact that users can disable cookies if they so desire, thus destroying any ability on the part of the site to track them, for whatever purpose.

Cookies are often used to identify an individual if the site that is using them requires some form of registration, either during a session in which the user has logged in, or from one session to the next, removing the obligation from the user to sign in each time on the same machine. For example, The New York Times on the Web *(http://www.nytimes.com/)* uses cookies to identify you between visits:



Before registering, the site doesn't know who we are



After we've registered and asked the site to remember us, it knows who we are next time we visit – in fact it welcomes us at the right-hand side of the header.

If we disable cookies before registering, or remove the cookies after registering, *NYTimes.com* no longer recognizes us.

Amazon/AlmondNet
Exhibit 1011
Page 27

Though we have established that cookies are not innately dangerous, there is a lot of uncertainty in the browsing public, and not all of it due to paranoia or misinformation, about how cookies work and what they can or cannot reveal. Concerns were raised, for example, when ad-serving company DoubleClick™ purchased Abacus Direct, a large catalog database firm. It was believed by some that this would allow DoubleClick to cross-reference its cookie data with the profiling information contained in the Abacus database. For more information on this case, we suggest that you visit the Electronic Privacy Information Center (*http://www.epic.org/*) and DoubleClick (*http://www.doubleclick.net/*).

Third-Party Cookies

These concerns came about because of third-party ad-servers' ability to place cookies on users' machines as they browse different sites. As we know from *Chapter 1* a cookie may only be retrieved by the server it was set by, which prevents any given web site from tracking the path of an individual user across the Web. Users visiting sites that use third-party ad-serving technology, however, might also receive a cookie from that server. This cookie is commonly referred to as a "Third-Party Cookie", as it is set not by the site the user is explicitly visiting, but by a completely different server. This same third-party server therefore technically has the ability to track its cookies as each individual user visits sites that use that ad-server, potentially building up a "profile" (albeit an anonymous one) of that user's surfing habits. This profile can be used to target ads at a user based on their surfing habits. Opinion is divided on this, as some feel this is a genuine benefit to the user, increasing the relevance of the information being served, while others consider it an intrusion into their private business. The key here, perhaps, is choice. DoubleClick, a provider of third-party ad-serving solutions, has introduced an "opt-out" cookie system, details of which are contained in their privacy policy: *http://www.doubleclick.com/us/corporate/privacy/privacy/default.asp.* Specifically, the policy states **"DoubleClick does not use your name, address, e-mail address, or phone number to deliver Internet ads. DoubleClick does use information about your browser and Web surfing to determine which ads to show your browser".**

Cookies have been designed to work in such a way that they do not of themselves pose a threat to privacy, but it is possible to deploy them in ways with which many users might feel uncomfortable. For this reason, it is important to be explicit in your privacy policy about exactly what cookies you set, and what you intend to do with the data collected as a result. We will cover this in more detail later in this chapter.

Data Privacy

The other broad area of concern for privacy is the collection, and integrity, of personal or sensitive information: this might include e-mail addresses, residential addresses, credit card details, purchasing history, medical information, and so on. The potential danger here is not that sites will be collecting information without the consent of the user, as it is fairly clear what is going on when a web site asks us to enter our address, but rather what happens to the data once we have clicked "submit ".

Amazon/AlmondNet
Exhibit 1011
Page 28

The factors that we need to consider are:

- What will the web site do with my personal information?
- How can I be sure that the web site will not use this information for purposes other than that to which I have consented?
- Will the web site intentionally distribute my personal information?
- Will the web site keep my information secure, and take steps to ensure that it is not unintentionally distributed?

As before, these are all key points that should be covered in any privacy policy, but there are also third-party agencies that can help us to gain the trust of our users, and these are also the main concerns of the various legislative bodies that have intervened in the interests of citizens' privacy. We will discuss some of the organizations that exist shortly, but first we're going to examine the more formal legislative controls that have been introduced.

Legislation

This section is intended to give a helpful overview of the legal constraints that are being introduced, concentrating on the situation in the US and the EU, with some suggestions for further research. Privacy legislation, like most legislation, is legally complex and technical, and circumstances will change between businesses and jurisdictions. Clearly, when considering your site's privacy policy and its terms and conditions, you should consider taking formal legal advice.

In this overview, we will consider the implications and intentions of two key legal interventions: the EU Data Protection Directive, and the US Children's Online Privacy Protection Act (COPPA). For details of the EU legislation, visit: *http://europa.eu.int/comm/internal_market/en/dataprot/index.htm*; and for more details on COPPA visit: *http://www.ftc.gov/bcp/conline/pubs/buspubs/coppa.htm.*

The EU Directive and "Safe Harbor"

The EU Directive binds all 15 member-states of the EU to safeguard the privacy of their citizens according to a basic set of principles, the intention being to harmonize and standardize the level of data protection and privacy across all of the member states. In principle, the legislation is designed to assert the following conditions:

- It covers those who collect, hold, or transmit personal data
- It puts them under an obligation to collect data only for "specified, explicit, and legitimate purposes"

Amazon/AlmondNet
Exhibit 1011
Page 29

- It obliges them to only hold data that is "relevant, accurate, and up-to-date"
- It mandates that the collection and retention of the data should be transparent, meaning it should be clear that the data collection and storage is being conducted, and that the individual has the clear right to opt out of this process

In particular, the Directive asserts the rights:

- To access the data being held
- To know where that data originated
- To rectify inaccurate data
- To legal recourse in the event of illegal use of the data
- To withhold permission to use the data for specific purposes (for example, sending spam or junk mail)

Furthermore, in the case of sensitive information, defined as "ethnic or racial origin, political or religious beliefs, trade union membership or data concerning health or sexual life", this can only be processed with the explicit consent of the individual (they have to OPT IN to this process).

The EU Directive is a formal guideline of principles for member-states to implement. Each state is free to draft legislation covering the Directive in its own manner, so long as it conforms to the Directive.

There are various exceptions and provisos which refer to the rights and conditions, allowing a certain degree of flexibility in the drafting of this legislation by member states (for example, for legitimate "journalistic, artistic, or literary purposes"), and safeguards are required to be put in place to monitor this process. This flexibility is overseen by independent "data supervisory authorities" in each member state.

The impact of this legislation is that organizations (or, indeed, individuals) are not entitled to collect or process data about citizens of the EU without constraint. Data collectors must make clear what information is being collected, for what purpose, and provide the means for individuals to check the accuracy of the data held (and correct it if necessary), and opt out of certain uses of the data.

Whilst legislation within member-states specifically covers the use of data within the EU, and also the transfer of data between member-states, there is an assumption that EU citizens ought to be able to enjoy the same privacy and data protection rights regardless of the geographical destination of the data they provide. This therefore has an impact not only on web sites that operate from within the EU, but any web site that technically "transfers data" from inside of the EU zone. In effect, any web site hosted outside of this area is transferring data when it receives a request from an EU resident.

Amazon/AlmondNet
Exhibit 1011
Page 30