THE BCS PROFESSIONAL EXAMINATION
Professional Graduate Diploma

April 2005

EXAMINERS' REPORT

The World Wide Web – Beyond the Basics

General

This paper was set for the third time this year and was again a popular paper. Once again, the general standard ranged from barely coherent to excellent.

A recurring issue: significant groups of candidates write far too much – often taking two or more pages to make a simple point worth two or three marks. Handwriting was a particular issue, with one script bordering on illegibility. There is ample time to write clearly and carefully whilst answering all questions in full.

The general level of preparation has improved significantly on previous years, but some regurgitation of textbook theory still occurs, especially on questions on more current topics – to reiterate the point highlighted in last year's report:

> *"Important note: This examination aims to be a combination of not only knowledge and recall of theoretical concepts, but also of application of these concepts to real-life situations (including the use of relevant examples, where appropriate)"*

The pass rate has increased by 10% on last year, further emphasising the improved level of preparation candidates have undertaken.

Question 1

1. *a)* Expand each of the following acronyms and define the *meaning* of each term:
   i) TCP/IP
   ii) FTP
   iii) HTTP
   iv) IMAP
   v) HTML   **(5 x 1 marks)**

   b) The current Internet Protocol allows for approximately 4,000 million unique addresses –which is quickly becoming exhausted. Define what is meant by *IPv6* and *NAT* (sometimes known as IP masquerading). How can these help to solve the address shortage?   **(4 marks)**

   c) Explain what is meant by *webcasting*?   **(1 mark)**
   i) Compare and contrast webcasting with traditional 'over-the-air' broadcasting.   **(3 marks)**

   ii) What impact has radio webcasting had on the business practices of traditional radio broadcasters (e.g. the British Broadcasting Corporation)?   **(4 marks)**

   iii) *Making specific reference to file distribution, file sharing and video streaming technologies, discuss* (with real examples) how the web distribution of music, film and television material has impacted on producers and consumers. *(Your answer should consider the technologies involved and the economic social and ethical issues.)*   **(8 marks)**

Answer Pointers

a)

   i) **Transmission Control Protocol / Internet Protocol** – A set of two communications protocols which offer reliable delivery (c.f. UDP/IP) across diverse networks. IP defines the underlying network layer, whilst TCP is the transport layer, providing the reliable delivery over IP.

   ii) **File Transfer Protocol** – A protocol used for efficient transfer of files to/from a remote machine.

   iii) **Hypertext Transfer Protocol** – A protocol used for the transfer of web files across a network.

   iv) **Internet Mail Access Protocol** – A mail protocol that interfaces with mail stored on a server, rather than fetching and storing mail on the local host machine.

   v) **Hypertext Markup Language** – A markup language used to create web pages.

b)

- IPv4 worked on 32-bit addressing. IPv6 (Internet Protocol version 6) uses 128-bit addresses, allowing for 3.4×10^{38} addresses. This is in the order of 10^{29} times greater than IPv4.

- NAT (Network address translation) allows private networks to **share** a limited number of external IP addresses. Traffic inside the network uses internal IP addresses (in the ranges 10.x.x.x, or 192.168.x.x) which do not have to be unique across the entire Internet, only on the local network. Traffic heading outside of the network has its IP address rewritten to one of the external IP addresses.

c)

   i) Webcasting is the transmission of multimedia data across the Internet. An example would be Internet radio stations (using Real Server, Shoutcast etc.).

   ii) List of differences between webcasting (WC) and over-the-air (OTA) broadcasting:
- Physical range (limited in OTA, international in WC).
- Max. number of clients (practically unlimited in OTA).
- Bandwidth and processing power limitations of webcasting.
- Quality of reception over range.

   iii) Impacts made on traditional broadcasters include:
- Increased competition between stations (due to cheap setup and limited regulation of Internet-only broadcasters).
- Increased uptake of digital radio, for comparable quality to Internet stations
- Many traditional Radio broadcasters have <u>also</u> begun streaming their broadcasts in addition to traditional over-the-air broadcasts.
- A mark is given for the use of appropriate examples of radio stations who have now started webcasting (e.g. BBC Radio 1, XFM etc.)

   iv) An open ended question.

Answers should address a number of the following aspects:

- The evolution of FTP "trading" sites to Napster
- Use of P2P file sharing software e.g. KaZaa, eDonkey, WinMX
- Copyright infringement of music/film using these networks
- Legal action (e.g. Metallica, RIAA)
- Intellectual Property enforcement (protected "CD"s, Digital Rights Management (DRM) schemes)
- Claims of reduced revenue by Film/Music industries
- "Bedroom artists" – provides an avenue for distribution without involving traditional labels (e.g. http://www.mp3.com) – increasing diversity and opportunity
- Quality control of home-brew music/films

- Bittorrent – an efficient file distribution software with load-balancing
- Use of Bittorrent to download large media files from multiple sources
- Internet websites cataloguing Bittorrent streams for current films
- Internet websites cataloguing Bittorrent streams for TV Programmes.
- Creative Rio, Apple iPod and other mobile MP3 solutions, selling of licensed MP3 files by iTunes etc.
- Streaming of TV by select broadcasters

Examiner's Comments

Part a) was typically answered well.

Part b) The majority did well, however there were some that did not appreciate the imminent shortage of IPV4.

Part c) i) Quite a few candidates could not adequately explain the idea of webcasting, instead confusing it with standard web hosting.

Part c) ii) Reasonably well answered.

Part c) iii) Many candidates were side-tracked into talking about the history or structure of the example given (e.g. the BBC), and missed the main point of the question, which was how traditional broadcasters in general have been affected.

Part c) iv) Whilst many candidates answered this question well, answers tended to be rather unstructured and missed out the finer points.

Question 2

2. All websites are vulnerable to attack of varying degrees and type to the client, the server, and the network itself.

a) Explain, with a suitable realistic example in each case, what is meant by:
  i) *denial of service* attack **(2 marks)**
  ii) *buffer overrun* vulnerability **(3 marks)**
  iii) *phishing* (also known as password harvesting fishing) **(3 marks)**

b) *i)* Why should sensitive information, such as credit card numbers, be encrypted before transmission over the Web? **(3 marks)**
  ii) List 3 *online* methods of paying for goods and services and outline, from a security point of view, the advantages and disadvantages of each method. **(6 marks)**

c) User profiling and tracking is commonplace on the Web. Discuss, with appropriate examples, the positive and negative aspects of these activities. **(8 marks)**

Answer Pointers

(a)

i) A **denial of service** attack causes a loss of services (typically network services) to legitimate users. E.g. Attack on Microsoft.com (2003).

ii) A **buffer overrun** (or buffer overflow) vulnerability occurs when a program writes data beyond the allocated end of a buffer in memory, potentially overwriting existing code. This may lead to the program executing arbitrary code. E.g. The Morris Worm (1988), Code Red (2001).

iii) **Phishing** is the luring of sensitive information (e.g. passwords, credit card numbers) from a user by using a fake site that appears to be from a trustworthy source. E.g. Fake Ebay messages emailed (2004),

(b)

i) To prevent:

- Disclosure of personal data
- Fraudulent use of data
- Alteration of personal data in transit

ii) Methods include:
- "Real-time" Credit Card processing
- Quick
- Instant feedback
- Risk of misuse by seller
- Credit Card sent via network for later manual processing
- Credit card details sent with order details
- Order must be delayed until payment processed
- Administration of collating offline credit card processes with online orders
- Inferior to real-time processing – all the risks, with reduced benefit
- Bank secure transaction initiatives (e.g. Halifax Secure)
- No transmission of financial details to untrusted suppliers (only to the payment services company)
- Have to remember yet another login and password
- Bank Credit transfer (BACS)
- Disclosure of bank account details
- Tracking
- Time taken to clear
- Payment services e.g. PayPal/NoChex
- No transmission of financial details to untrusted suppliers (only to the payment services company)
- Commission

(c)

An open ended question. Positive (relevant targeted adverts) and negative (privacy concerns, misuse of personal data) of the issues should be explored, and a well-justified and consistent conclusion reached. Real-life examples of spyware or profiling companies (e.g. Doubleclick, BonziBuddy, Cydoor), and of companies that use profiling to deliver customised adverts (e.g. Gmail, Opera) should be used.

Examiner's Comments

This was the most popular question, and was attempted by over 80% of the candidates, but unfortunately was typically very poorly answered, with only around a third of candidates achieving a pass mark. The key failing was attempting to regurgitate theory in sections which require not just theory, but application of the theory.

Part a) i) Generally well answered.

Part a) ii) Most candidates understood the idea of buffer overrun referring to large data, but not many seemed to understand the idea of a buffer and overwriting memory.

Part a) iii) Poorly answered, with many candidates guessing, or believing phishing to refer to "cracking" a password by brute force, or capturing details by keyloggers.

Part b) i) Reasonably well answered, but note that this question did not ask for a description of encryption processes, or public/private key encryption.

Part b) ii) On the whole, answers were acceptable, though note that the question specifically asked for online methods (so cheques or phone banking is not appropriate).

Part c) The answers for this part were rather vague, and lacked any contextualisation to real life examples – very few candidates successfully discussed real life companies or cases. Note that this question did not ask for a catalogue or description of profiling methods!

Question 3

3.   **Refer to the Appendix at the back of the question paper for the Figures quoted in this question.**

Cascading Style Sheets (CSS) are used to separate content from presentation.

a)   *i)*   Give THREE reasons why it is important to separate content from presentation.   **(3 marks)**

   ii)   What are the advantages of defining and linking to an external CSS file, as opposed to simply defining CSS in the header of an HTML document?   **(3 marks)**

b)   Different cultures may interpret non-content issues in particular ways (for example, in China red is associated with good fortune but in nearby Japan it is associated with anger). Briefly explain (using a different example) how CSS could be used to adapt a web page to other cultures by altering these presentational issues.   **(3 marks)**

c)   The following section exercises your ability to manipulate CSS and change the presentation of a webpage. The HTML source is given in **Figure 1**, and a visual rendering of the page (without any stylesheet) is given in **Figure 2.**

   i)   The stylesheet bcs.css (given in **Figure 3**) has now been attached to the web page. Using the template into your answer booklet), draw a diagram to illustrate what the page will now look like when it is loaded in a browser window. (You should state what browser you are assuming use of and annotate any areas of your diagram that may be unclear, such as fonts or colours).   **(9 marks)**

   ii)   A designer has given you a mock-up of how they wish the page to appear (see **Figure 5**). Modify the stylesheet bcs.css to reflect these alterations. You should list only the changes required (and line number from bcs.css) in your answer.   **(7 marks)**

Answer Pointers

(a)

i)

- Allows content to be easily updated and maintained, whilst retaining the same presentation.
- Separation allows the content to be laid out in different at any time to provide alternative representations e.g. screen, print, voice.
- More **accessible** to a wider variety of devices.
- Allows users to **customise** appearance/presentation to their own needs - style switchers.

ii)

- Easier to make **site-wide changes** - one css file rather than all pages.
- **Reducing redundancy -** less code on the page leads to faster downloads and caching.
- **Less code** on the page - easier to read and maintain.
- More **control** over code - can deliver code in preferred order for screen readers.

(b) A general discussion question; answers should mention the following points:

- Localisation can effect presentational decisions as well as language on screen
- Reference to sites such as CSS Zen Garden, where drastically different styles can be applied to the same content
- Server/client side detection of date/country/time zone/language, and auto-selection of appropriate css
- A mark is given for an appropriate example (different to the one given in the question).

(c)

i)
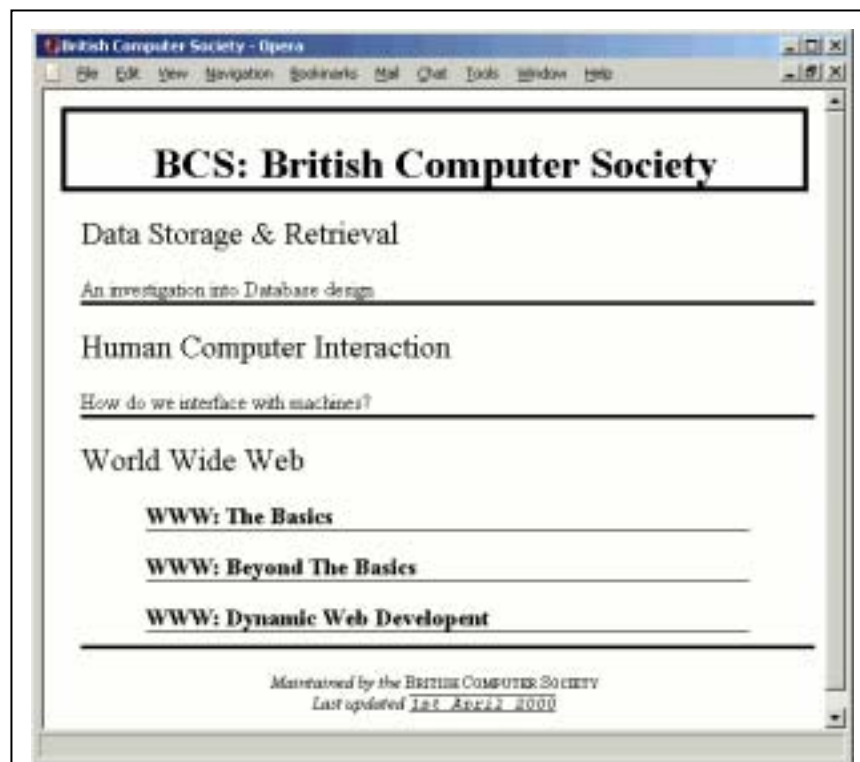
ii)

```
1.  h1 {
2.      font-family: "Courier New", Courier, monospace;
3.      font-style: italic;
4.  }
5.
6.  h2 {
7.      text-transform: lowercase;
8.      font-weight: bold;
9.  }
10.
11. .title {
12.     border: 4px dotted black;
13.     text-align: center;
14.     margin: 0px;
15. }
16.
17. .section {
18.     margin-left:20px;
19.     border-bottom: 1px solid black;
20. }
21.
22. .subsection {
23.     margin-left:0px;
24.     margin-right:0px;
25.     margin-top:10px;
26.     margin-bottom:10px;
27.     border: 1px solid black;
28.     border-bottom: 1px solid black;
29. }
30.
31. .address {
32.     font-size: medium;
33.     font-style: normal;
34.     text-align: right;
35. }
36.
37. .author {
38.     font-variant: normal;
39.     font-style: italic;
40.     font-weight: bold;
41. }
42.
43. .date {
44.     text-decoration: underline overline;
45.     font-family: "Courier New", Courier, monospace;
46.     text-transform: uppercase;
47. }
```

Examiner's Comments

Part a) was generally answered well in both subsections.

Part b) Very few candidates provided alternative examples or amplified issues of localisation, instead using the examples stated in the question itself.

Part c) Most attempts were highly proficient with a very few that found re-rendering the page difficult.

Question 4

4. *a)*   In the context of digital images, what is meant by *watermarking*?   **(2 marks)**

  i)   *So*me websites attempt to protect images using methods such as blocking the right mouse click, removing the menu bar or "encrypting" the source code with JavaScript. Discuss, with reasons, the effectiveness of each of these techniques in preventing the user from saving a copy of the image.   **(6 marks)**

  b)   How is it possible on the server-side to restrict access to a web page to certain people or networks?   **(2 marks)**

  i)   With reference to real-life initiatives (such as Internet Content Rating Association and the PICS standard) and products (such as Internet filters), describe the different approaches to restricting access to Web resources on the client-side.   **(4 marks)**

  c)   The Web has had a positive influence in many areas of society. However, as with all technology, it can be used for purposes which may be viewed as immoral, unethical and, in some cases, illegal.

  i)   List, with a brief rationale for the reason, an example of a website that may be viewed as:
Immoral/Unethical
Illegal
(You should provide one distinct example for each category.)   **(4 marks)**

  ii)   Identify and evaluate possible measures that can be taken to minimise the impact of questionable sites on society.   **(7 marks)**

Answer Pointers

  (a)
  i)   (Digital) watermarking is the process of embedding identifying code into a digital file, for copyright, identification or authentication purposes.

  ii)
- **Blocking right click** – Will stop casual users of certain browsers from being able to right-click save the file, but does not stop people from finding the image source in the HTML and downloading the image file directly. Also, does not work on all browsers (e.g. Opera).
- **Removing menu bar** – Stops people from being able to view the source directly from the browser, but does not stop someone from a) opening the cached source file on disk or b) using a tool such as wget to download the HTML.
- **"Encrypting" source code** – Relies on Javascript being enabled. Will stop all but the most determined person from downloading the image – but the image can be found in the browser's internet cache.

  (b)
  i)   On the server side, two main methods to restrict content are to restrict access by hostname/IP address (.htaccess) , and restrict access by username/password (.htpasswd).

  ii)   Content filters (e.g. NetNanny) can be used to **whitelist** (allow only specified sites) or **blacklist** (prevent access to certain sites, allowing all others). Filters can evaluate the **content** of the site (by looking for e.g. keywords). Schemes such as PICS and ICRA attempt to add self-declared ratings to a page, which can then be filtered out by the client-side browser.

(c)

    i)   This question is heavily dependent on culture and laws of a country. With regards to immoral and unethical sites, any strong example is acceptable, so long as it is <u>consistently justified</u> in the answer. Distinct examples should be used for each section (despite the fact that some content might rightfully fall under more than one category). The following are some examples:

- Arguably **Immoral/Unethical** – Pornography, sites glorifying drug/alcohol abuse, suicide sites, plagiarised material
- **Illegal** – Child pornography, music/video/software piracy sites, Hate sites

    ii)   Some possible measures to minimise the impact of sites on society:

- **Legal action** – difficult to enforce since the WWW is international, and sites illegal in one region may not be in others.
- **Web Censorship** – (as seen in China) makes access to material more difficult, but does not prevent people connecting remotely to a machine outside of the censored region. May lead to "false positives" – harmless sites being blocked. May lead to negative public opinion in certain countries that advocate "free speech".
- **Filtering software** – Blocks questionable content from dependants. Can be switched off, or avoided by public Internet access terminals (e.g. in Cyber Cafés). Filters may not catch all dubious content (false negatives) or may catch innocent sites (false positives).
- **Age restrictions/verification** – (e.g. requiring a credit card) only adults would be able to access questionable sites, to protect children. Possibility of adults allowing children access under their account.
- **Education** – Guidance and education might prevent people from being affected by the questionable content.

Examiner's Comments

This question was attempted by only a third of candidates and exhibited an almost full range of answers (from 2/25 up to an excellent answer of 25/25).

Part a) The key meaning of watermarking is not about reducing contrast to create background images for web pages!

Part a) i) Generally well answered in terms of disabling menus/right click, and the ways to circumvent this protection, though very few candidates understood the process of "encrypting" source code (also known as obfuscating the source code), which is really to prevent casual viewers from understanding the code easily.

Part b) i) This part was not answered well. Answers tended to be vague, and did not reference real life initiatives such as PICS, as specifically requested. Many candidates fell back to discussing firewall technologies, which was rather irrelevant.

Part c) i) Generally well done and justified.

Part c) ii) A lack of real life examples held back most candidate's answers.

Question 5

5. *a)* The WWW is used for a diverse range of purposes.

　i) Identify FIVE distinct categories of website (in terms of intended purpose) and, for each category, identify an appropriate real-life example. **(5 marks)**

　ii) Considering any TWO of the categories identified above, discuss how one can measure the "success" of each site. For each category, you should identify at least four criteria (and a short justification) that could be sensibly used to measure "success". **(8 marks)**

b) *i)* Define the terms *load testing* and *performance testing* when applied to a website. What (if any) are the differences between them? **(5 marks)**

　ii) Given a B2C website (e.g. `http://www.amazon.com`), devise a plan for testing the site, with specific emphasis on load and performance testing. **(7 marks)**

Answer Pointers

a)

　i) Types of website include:
- Educational
- Governmental
- Support (Help, drivers etc)
- Commercial
- Entertainment
- Community/Informational

　ii) Some criteria include:
- Traffic – measured by Number of hits
- Revenue generated
- Interest, feedback or other correspondence generated – number of received emails
- Popularity– measured by number of links from external websites to your site, or google ranking

b)

　i) **Load testing** involves evaluating a website in its ability to:
- Handle volumes of data/bandwidth usage
- Handle numbers of concurrent requests/transactions

It is evaluating these aspects under the **expected** usage of the system (as opposed to stress testing).

Performance testing involves evaluating a website with respect to how fast it performs a task under a given workload.

The difference between them is that performance testing is interested in benchmarking the time actually taken to respond, whilst load testing is interested in the site staying active under the anticipated traffic.

　ii) A general, consistent and coherent framework should be drawn up, with reference to *some* (not all) of :

(1) the following testing methods:
- load testing
- exhaustive testing
- recovery management assessment
- hyperlink testing/navigation
- intrusion detection

(2) the following metrics:

- What percentage of dynamic page requests are processed within an acceptable time limit (e.g. 10 secs non-encrypted, 15 secs encrypted) when accessed by a typical client (e.g. 56K Modem, 512KB Broadband)
- Average response time
- Max number of concurrent users
- Transactions per second that the website can handle whilst still providing an acceptable level of performance (with definition of "acceptable level of performance" e.g. response <20 sec)
- System breaking point – how many requests can be concurrently handled before the system falls over (and a definition of "falls over" e.g. response time >= 30 sec, or 404 Page Not Found/"Server Busy" error messages)
- How are loads beyond the max capacity handled? (e.g. Throttling of bandwidth, Refusing new requests)
- Determine the slowest and the fastest 10% web pages downloaded
- Network bandwidth required to support estimated and maximum number of clients
- Reference to load distribution, caching, mirroring, serves offered by e.g. akamai

Examiner's Comments

Another popular question.

Part a) i) Well done, but some presented b2b, b2c, c2c as distinct categories, when they are all in the "commercial" category of site.

Part a) ii) Poorly answered – the question asked candidates to identify 4 criteria as indicators of success, but instead the response was to give a list of desirable features – to test a site you requite a means of judging how "well" a site stands in light of these features.

Part b) i) Very confused response – load testing was often interpreted as meaning the successful loading and rendering of an individual page, rather than issues of bandwidth, concurrent users etc. Some ideas on performance testing, but again this tended to focus on the functionality of individual pages with respect to HCI.

Part b) ii) Good general test plans, though lacking in structure and missing the emphasis on load and performance testing, rather than functional testing.