Internet (content) Filtering

Definition

BCS understands IT is a key enabler of the digital economy, which describes the exchange of information, goods and services through an information technology platform. BCS seeks to ensure everyone benefits from IT and we work to promote wider social and economic progress through the advancement of information technology.

Background/Context

Internet (content) filtering is the process whereby some agent (generally the internet service provider — ISP) refuses to return a requested piece of internet content. It is often confused with search engine filtering such as ‘Safe Search’, whereby a search engine (e.g. Google) does not return answers even though it knows about them. This paper does not discuss search engine filtering, which is in a state of flux given the recent European Court of Justice Judgment over the ‘right to be forgotten’. Internet filtering is often described as ‘parental control’, even though until recently the ‘parent’ in fact only had a take-it-or-leave-it control, and there has been no research on the effective use of the configuration options.

Internet filtering is not a perfect technology, and has two risks: false positive (censoring something that ‘ought not’ to be censored) and false negative (failing to block that which ‘ought’ to be blocked). There are two basic approaches to internet filtering (which are also at the basis of ‘Safe Search’ technology).

One is black-listing: analogous to ‘I’m not going to read you that book because it’s called “Lady Chatterley’s Lover”, and I’ve been told not to read that book’ or ‘I’m not going to show you that film because it’s classified 18’. The major problem with this approach is that a vanishingly small proportion of the internet has been examined for banning/classification. Hence false negatives are common here.

The other is content-based filtering: analogous to ‘I’m going to stop reading that book now because I’ve come across this banned word in it’, or ‘I’m going to stop showing this film because the last frame was more than 40 per cent pink’. There are two problems with this approach. The first is that, under the Regulation of Investigatory Powers Act, it is probably illegal for ISPs to do, since it involves inspecting the content. The second is that of false positives: many books contain occasional `banned' words, and a frame may well be pink because of a sunset, or a swimming gala.

There has been an explosion in the availability and publicity for internet filtering since the Prime Minister’s 22 July 2013 speech. BCS also understands that ISPs have sought, and been given assurances, on the relationship of internet filtering to the Regulation of Investigatory Powers Act, as they were being actively encouraged to do things that were hitherto probably illegal.

BCS position on the key issues:

BCS welcomes the broader availability of and increased publicity for, internet filtering. However, BCS is worried that internet filtering is being described as ‘the solution’, whereas it is merely part of the solution, and the more zealously it is applied, the less effective it becomes. Two statements from
the UK Safer Internet Centre (BCS emphasis) are worth recalling: ‘filters can be a helpful tool in reducing the chances of coming across something upsetting’ and ‘remember that filtering is only part of the solution’.

**What should be censored?**

There are actually two questions here: deciding which categories of materials should be defined, and either placed on a banned list, or offered as options if the filtering is configurable; and deciding which materials belong in which category.

1. **Which categories?** The one category on which there is widespread societal agreement is area of child abuse (technically speaking, indecent images of children, which are illegal to possess under the Protection of Children Act 1978) by use of blacklisting technology. More support could be given to the people, largely volunteers, who do the initial reporting, and to the blacklisting process, generally under the auspices of the Internet Watch Foundation (IWF). **BCS commends the extent to which ISPs and mobile operators cooperate with the IWF to manage and apply these blacklists.**

2. It should be noted that this sort of blacklisting has both financial and non-financial costs
   a. The ISPs need to install and operate substantially more powerful equipment to do filtering than is needed to pass through requests unexamined;
   b. The ISPs, very largely, fund the IWF;
   c. There is a risk of error and ‘false positives’: for example one such prevented edits to the whole of Wikipedia, and YouTube “Safety Mode” has blocked collections of landscapes;
   d. It is difficult to get right in terms of what is filtered: the [Irish study of filtering in schools](#); showed that 50 per cent of schools reported that filtering occasionally blocked valid educational sites, with a further 20 per cent reporting that it regularly did so;
   e. It is difficult to get right technically. Filters, particularly those implemented transparently within the network, can create unexpected technical interactions with other internet services. The Wikipedia example above, and the accidental global blocking of YouTube by a state agent illustrate the problem. Filters may also, by interfering with the end-to-end design principle, hinder or prevent the development and adoption of new internet services. **BCS recommends that those designing and implementing filters use approaches that minimise these risks, and allow them to be detected and rectified as quickly as possible**
   f. Filtering encourages the use of bypasses, and BCS has anecdotal evidence that children are learning how to use (largely US-based) virtual private networks to bypass filtering.

3. The CMS Select Committee asked about ‘filtering out extremist material, including … and material intended to promote terrorism or other acts of violence’. There has been much resistance to the Internet Watch Foundation’s widening its remit to the other material in the Select Committee’s question, and **BCS does not believe that this is the way forward.** A move into filtering such content would be overtly political and would not carry the same degree of international consensus, as one person’s/state’s terrorism is another person’s/state’s jihad.

4. **Which content should be categorised?** Currently this is largely done by human-controlled blacklisting: an image (typically) is reported and suitable changes are made to the blacklisting databases.

5. Some people say that more should be done, and imply, without saying so, that content-based filtering should be used, so that more such material could be blocked. This would require a major change in society’s attitude to censorship, as well as, probably, primary legislation to enact fundamental changes to the Regulation of Investigatory Powers Act. **BCS does not believe that this is either feasible or desirable.** The technology is nowhere near mature enough (a major foreign government’s project to block content on the basis of an inappropriate ratio of skin tones was abandoned), and the risk of false positives, e.g. medical textbooks, is too great.
The role of education

1. The report of the Byron Review made a powerful analogy: ‘At a public swimming pool we have gates, put up signs, have lifeguards and shallow ends, but we also teach children how to swim’. To this she could well have added ‘and we help parents to teach children to swim, and we teach parents to be lifeguards;’

2. The sort of education necessary here for children is not technology education, it is societal education. For this reason BCS believes that it belongs in the general Personal, Learning and Thinking Skills (PLTS) category, rather than in ICT- or computing-specific classes. There is excellent advice at the Get Safe Online website, and class templates such as https://www.isc2cares.org/safe-and-secure/ are available;

3. The IWF’s comment on the home-produced material points again in this direction. ‘These findings provide evidence to support the importance of the education work delivered by child protection agencies to raise awareness of the permanence of information on the internet and the risks inherent to young people in creating and distributing this type of content.’ Recent ‘right to be forgotten’ cases refer to search engines, and do not help in this context.

4. A challenging question is what and how much education is appropriate for parents. Some advice and help on ‘parental controls’, both on the configuration and on tools such as Google Safesearch and YouTube’s SafetyMode, most of which have come along since many parents first encountered the internet, is also appropriate. Similarly, parents need to become aware of the (excellent, and recommended in the Byron report) PEGI rating system for games and the advice at http://www.askaboutgames.com/, another site which has emerged since many parents learned about the internet and which recent learners will not necessarily come across. BCS believes that schools should be encouraged to facilitate and host such classes in the PTA context;

5. Such classes will need to cover technology, but should probably be wider. Parenting guidance is sometimes sought where parents would like support in terms of how to engage with their children on social topics that children might explore on the internet. These social topics are all related to what we might class as ‘growing up on the internet’ and have security facets to them. Topics might include; management of friendships mediated by the internet, trolling and other forms of internet mediated abuse, balancing internet mediated activities with offline activities, identity and projection of self via the internet etc. Again, Get Safe Online has good age-specific materials, but these need to be drawn to the attention of parents, and their attention refreshed as the children grow older. Information about the BCS/Get Safe Online joint Cyber to the Citizen initiative can be viewed here.

James Davenport, BCS Security Community of Expertise (author)
Louise Bennett, BCS Security Community of Expertise (spokesperson)
Andrew Cormack, BCS Security Community of Expertise (spokesperson)
Internet Content Filtering: technical background

There are currently three mechanisms which cause some internet content to be filtered or blocked:

1. Internet Watch Foundation Blocking
2. High Court Orders
3. Content Filtering Services

These three different ways equate to three different reasons for blocking or filtering being deployed in networks and it is important to understand the background to each in order to understand the overall content filtering position in the UK.

The Internet Watch Foundation was founded by industry to try and deal with the issue of child abuse images being posted on the internet. It is funded largely by industry but is mainly, though not exclusively, responsive in approach: people can report suspect images to the IWF for investigation. The IWF publishes (secret) lists of URLs that should be blocked to ISPs, who in turn implement blocking systems in their networks to prevent the images being accessed. This is largely effective at preventing people from inadvertently stumbling upon this type of content but it would be naïve to think that effective in dealing with hardened criminals trading in such content and who use less visible forms of communication. This form of blocking is largely uncontentious except when mistakes are made in the grey area of artistic or medical publications. A well known example of this occurred in 2008 when Wikipedia published an image of a record album, resulting in the IWF causing a block to an image that probably is not illegal under UK law and also much more of the site than was appropriate.

Since 2012, ISPs have been subject to High Court Orders requiring them to block specific IP addresses or URLs. The most well known case relates to The Pirate Bay site and around 50 other sites have been subject to similar orders since then. Recently, the major ISPs have defended a case attempting to extend the jurisdiction to include sites offering fake or counterfeit goods on the basis of trade mark infringement. These orders appear to be effective for some sites, but other sites use all kinds of techniques to evade the block. It is clear that the law in this area is out of date, struggles in the modernity of the internet age and is badly in need of an update. In particular, politicians and others involved in the legal process struggle to understand that the internet has evolved with the intent of making it accessible and open to everybody in the world. The design of the internet makes it very difficult and often technical impossible for ISPs to invoke some of the content focused ideas for filtering and blocking that people come up with.

Content Filtering Services differ from the previous two types because they are optional for customers. It is generally possible for customers to choose different categories of “protection” or opt out of filtering altogether. The provision of these services arose as a result of Government pressure for ISPs to do something to protect children from unsuitable internet content, but without any proper and detailed technical consideration of exactly how this could be done effectively. Of course, there is no magic indicator attached to internet content that can be used to identify what is or what is not suitable for any given age to consume. At best this can only be considered a broad brush approach, with most ISPs relying on well known global companies to grade internet site content into different categories. It is not feasible to deliver a 100% effective solution that cannot be bypassed with technical expertise. As with anything involving the internet, the environment is naturally open and there are many ways to achieve all kinds of objectives given sufficient time and effort deployed.

However, for the majority of consumers, the filtering products now deployed by the major ISPs provide a reasonable level of protection and much more than was available to parents previously. There were mumblings in the media at one point about whether the filtering products infringed the Regulation of Investigatory Powers Act (2000). Given that the filtering service is optional for
consumers to adopt or not (although they may be forced to make a decision), and the communications data or content is not passed on to a third party, it is unlikely to infringe that act and the UK Government has issued its own opinion to that effect.