"Let a Hundred Flowers Bloom, a Hundred Schools of Thought Contend": Web Engineering in the Chinese Context

Kieron O'Hara

Web Science Research Initiative
School of Electronics and Computer Science
University of Southampton
Highfield
Southampton SO17 1BJ
United Kingdom
kmo@ecs.soton.ac.uk

Version of a chapter from Xiaoling Zhang & Yongnian Zheng (eds.), *China's Information and Communications Technology Revolution*, London: Routledge, 2009

Abstract: The revolutionary aspect of the World Wide Web is that it is a decentralised information structure. This democratic decentralisation is a key factor in the added value that the Web provides, because it facilitates the serendipitous reuse of information in new and unanticipated contexts. However its basic principle, of free flow of information packets and a very simple set of rules and standards underpinning these complex structures, is being undermined by attempts to restrict information flow. As use of the Web has spread, illiberal regimes feel threatened, but thanks to the hands-off approach of the 1990s, there are no affirmative globally-recognised principles governing the flow of information online. Currently, China is still focusing on a censorship-based approach to information control, using methods in direct opposition to the Web's essential governing principle of decentralisation. This paper examines some of the questions for engineering and policy of that clash of principles.

Introduction

There have been a number of highly illuminating studies of the effects of the Internet or the World Wide Web on Chinese society and politics – indeed there are many fascinating ones in this book. The clash between the cutting edge of technology and a relatively conservative society, or, seen from another angle, between an American, liberal and liberalising technology on a giant nation still trying to find its post-Marxian ideological bearings, is obviously interesting, not to say dramatic, and is one of the key arenas for ideological conflict in the 21st century world.[1] But social science, like any intellectual endeavour, must make simplifying assumptions, and one that is usually, if not always, made in this field is that the technology is exogenous, imposed from without, a black box with inputs and outputs, but which is not fundamentally changed by the interaction. This is, however, an untrue assumption; the Internet and the Web are defined by complex and well-planned sets of engineering protocols –

[1] Cf. e.g. Christopher R. Hughes & Gudrun Wacker (eds.), *China and the Internet: Politics of the Digital Leap Forward*, (London: RoutledgeCurzon, 2003), Françoise Mengin (ed.), *Cyber China: Reshaping National Identities in the Age of Information*, (Basingstoke: Palgrave Macmillan, 2004), Jens Damm & Simona Thomas (eds.), *Chinese Cyberspaces: Technological Changes and Political Effects*, (London: RoutledgeCurzon, 2005), Zixue Tai, *The Internet in China: Cyberspace and Civil Society*, (New York: Routledge, 2006).

human creations – and they can be changed, or broken, by anyone coming into contact with them.

In this chapter, I wish to look at the problem the other way around, and consider the potential effects of Chinese policy towards the World Wide Web. I want to think about the Web endogenously, as a piece of engineering, what assumptions underlie it, how it has developed and how it works (avoiding technicalities as far as possible), and then I will speculate about how Chinese interaction with it could create change in it. In the opening section, I will briefly describe how the Web, and the Internet on top of which it sits, developed, and what it is *for*. Then, I will consider the assumptions and principles on which the Web is founded. In that context, I will then move onto thinking about China's Web presence, and the effect that that could make, before concluding with the fundamental dilemma that faces not only the Chinese government as it considers how to interface with the Web, but also the standards bodies which administer the Web as they export their technologies to China and other illiberal polities.

The history and architecture of the Web

So completely has the Web entered the lives of people in Western democracies (particularly those on the right side of the various perceived digital divides) in the last few years that it is easy to forget how recent a development it is. The present author got his first glimpse of the Web – downloading a paper in the UK from a server in Stanford – in 1993. Ho hum, thought he. By 2000, the Web was an indispensable tool for scientific research. This section will rehearse some of the recent history of information technology, showing how the Web has built on the earlier development of the Internet.

The development of the Internet

The Internet is basically a network of computer networks. Small networks of computers, linked by telephone line or other communication routes to allow information to be passed between them, are themselves linked together, allowing information from any machine on any of the networks to be passed to any other. The key to the success of the Internet (as opposed to other methods of computer networking) is the general-purpose and universal nature of the links. Before the Internet, smaller computer networks could be and were supported by those with sufficient resources (often for military, academic, governmental or commercial applications), and those networks could be connected together. But the connections between networks tended to be special purpose, individually-tailored for the anticipated transfer of specific items of information in a specific form. The beauty of the Internet – often unappreciated – is that its protocols, which we shall describe briefly below, make few assumptions about the information that will be passed between networks, making it a uniquely general purpose tool. Information held in a wide variety of electronic representations, most unanticipated by the Internet's developers, can be passed around, pretty efficiently too, which is the root of its success. The key developments of the Internet happened during the 1960s.

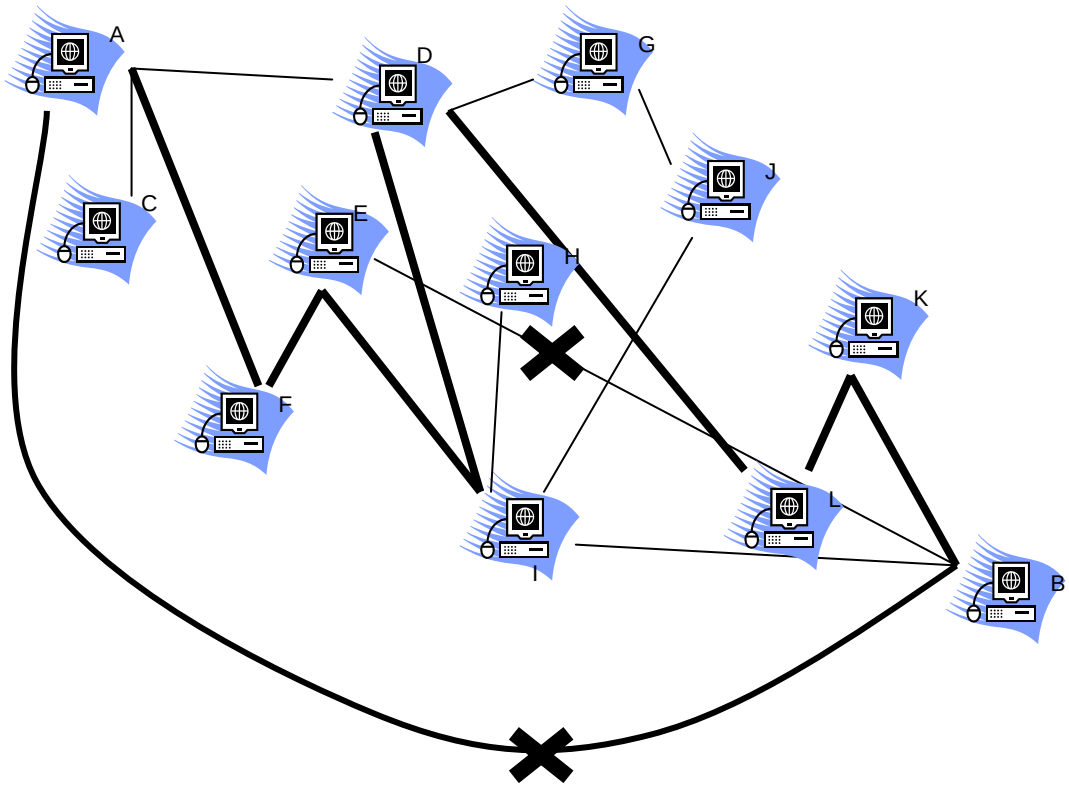
Defence

The original Internet was largely driven by the needs of the US defence establishment, and its main precursor (ARPANET) was named after the US Dept of Defense's Advanced Research Projects Agency (ARPA), and made its appearance on 21st

November, 1969. The aim was to provide large-scale communications that were efficient yet resistant to combat disruption. The system devised for routing information was a method called *packet switching*, in which a message was, in effect, chopped into little pieces (packets), which were then sent to the destination computer by separate routes. The destination computer assembled the packets into a complete message when it had received them all.[2]

The packets would be sent by a method called *dynamic routing*, in which the path by which they were sent would not be determined in advance, but worked out at the time. The pathway would be along the connections between individual computers in the network; a packet would be received by a computer, which would then try to send it to another computer which it reckoned would be closer to the ultimate destination if possible. So, for instance, in Figure 1, the aim is to get information from computer A to computer B. There is a direct route, but this can easily be knocked out. If it is the only route, then we have a bottleneck, and the connection between A and B is highly vulnerable. But if A and B are connected in a network, then there are several alternative routes. For instance, there is a route A-F-E-B. The network structure is much more robust; if this more complex route is compromised by another attack (say, the link from E to B is knocked out), then other routes, such as A-F-E-I-D-L-K-B (pictured) can be used. This is certainly not an exact science, but electronic transactions happened quickly enough that the information would eventually reach the destination in a reasonable time (in today's high-bandwidth world, downloads happen almost instantaneously, and a delay of even ten seconds is regarded as unacceptable). The dynamic routing algorithms would send each packet by a different route to the destination.

Figure 1: Getting Information from A to B in a Compromised Network



[2] Janet Abbate, *Inventing the Internet*, (Cambridge, MA: MIT Press, 1999), pp.7-42.

Dynamic routing met the criteria for a system of information passing with defensive capabilities. The network was robust against attack, largely because it avoided communications *hubs*. A hub is an area with many connections, often crucial for a network. If the hub is knocked out by attacks, then the connectivity of the network is reduced dramatically – indeed, it can fall into separate disjoint networks. Similarly, if two halves of the network were connected only by a small number of communications lines, then the network can be subverted and maybe split by knocking out those lines. The packet switching system means that vulnerabilities stemming from hubs and/or crucial lines can be avoided by increasing connections between all parts of the network. Knocking out some telephone lines or computers will be damaging, inevitably, but not fatal to the network because there should always be alternative ways around. Even so, where there are bottlenecks there can still be problems. In February 2008, damage to the two undersea cables that carry 90% of the data traffic through the Suez Canal caused severe disruption to connectivity in the Middle East.[3]

Packet switching is also efficient, as communications hubs can cause bottlenecks if information traffic grows too quickly for their infrastructure. If all messages have to pass through a central exchange, then every time an extra computer is added to the network, the number of potential sender-receiver pairs doubles. As the network grows arithmetically, the information traffic tends to grow geometrically, and if the traffic management system in the hub cannot scale, the growth of the system will be restricted by congestion. By allowing packets to be sent opportunistically from computer to computer along any line that is available and plausible, messages reach their destination more efficiently. If congestion starts to build up somewhere in the system, packets can be routed around it (and, with dynamic routing, these decisions can be taken instantly while the packet is *en route*).

Science and academe

But the ARPANET, and the Internet which followed (the term 'Internet' was first used in this context in 1974), obviously had advantages beyond defence – after all, their four desiderata, avoiding congestion, efficiency, allowing general purpose communications and being scalable, are appealing to any organisation with large-scale information needs.[4] In particular, the scientific and academic communities, whose day-to-day activities concern the transfer of information to fellow researchers, students, the media and the general public, were obviously going to be interested. The Internet became, over the 1970s and 1980s, an important academic tool. When one was working on a computer connected to the Internet or its precursors, one suddenly had several informational and communicational options that were not available before. One could send messages (*email*) to people working on other computers using a system that imposed a low overhead for both sender and receiver (the sender did not need to leave his or her chair, while the receiver could access the message at a convenient time). One could get remote access to data or papers held on other computers. Files could be shared between communities. And two or more people, in remote locations, could collaborate in writing or editing files. Science and academic research became the driving ethos behind the Internet, and as a result Internet research tended to take the form of methods to *remove the friction* out of the journey of a piece of information, rather than methods to *restrict* the flow. One might wish to restrict the flow of information to promote security of intellectual property or state secrets, for

[3] "Of cables and conspiracies," *The Economist*, 9th February, 2008.
[4] Abbate, *Inventing the Internet*.

instance, or the privacy of people referred to by that information. But since the user base of the Internet at this time was relatively small, and relatively well-behaved, these were not priorities.

Standardisation and protocols

For all its complexity, the Internet is basically a very simple device. It is made up of hardware and software; the hardware, computers, telephone lines etc, is not our concern. The software consists of a series of *protocols* that define how computers talk to each other. There are two levels of protocols.

The first level is the *Internet Protocol* (IP) itself. This is the heart of the Internet – a better definition of the Internet than the one given above is a network of computer networks linked using IP. This defines the packets that are routed about the system. Above that sits the *Transmission Control Protocol* (TCP), which defines the routing system for the information packets. In effect, TCP makes a virtual connection between the sender and the receiver, even if they are not directly connected. It also treats each packet equally – although each packet needs to indicate its relationship to the rest of the message and its destination, the content of the packet need not be read *en route*, and so the whole system treats the greatest works of philosophy and the least significant pieces of tittle-tattle with equal reverence and respect. This is a good thing from the network point of view, because any method of information transfer which involved reading the content of information packets would slow the system down and create bottlenecks.

There are alternatives to TCP and IP, but these have been easily the most successful. The two protocols are often referred to together, as the TCP/IP suite.[5] The TCP/IP suite provides a basic platform upon which further applications can be defined to use the Internet, exploiting the very clean method of transferring information across a network. The freedom of movement of information whatever form it takes means that TCP/IP is an extremely versatile platform which can be used by a wide variety of applications. One such application, sitting atop the Internet, has been very successful indeed – the *World Wide Web*.[6]

The development of the World Wide Web

It is fair to say that the World Wide Web (WWW, or Web – I will use these terms interchangeably) is the main application that nowadays brings people to the Internet, so that recent users come to other applications (such as email) afterwards. The essence of the WWW is to link hypertext documents using the Internet. A browser is used to access material that may be in any one of a number of media – text, pictures, video, sound – and then can follow hyperlinks from document to document. Key technologies include the Hypertext Transfer Protocol (http, and its secure variant https), which is a protocol telling computers how to access and retrieve hypertext pages, and the Hypertext Markup Language (HTML), which enables the description of the structure of a document to be seen online (e.g. in terms of *headings*, *lists* and so on), and its supplementation with other objects, such as embedded images, interactive forms, or links to other documents. There are other protocols, languages and technical

[5] W. Richard Stevens, *TCP/IP Illustrated Volume 1: The Protocols*, (New York: Addison-Wesley, 1994).
[6] Tim Berners-Lee, *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*, (London: Texere Publishing, 1999).

requirements (in particular, more flexible markup languages, such as XML), but http and HTML are the basics. The China Policy Institute at Nottingham University has a home page http://www.nottingham.ac.uk/china-policy-institute/, which is written in HTML, and – as its address tells you – is transferred to your computer when you download it using http. This application – the WWW – sits neatly on top of the TCP/IP suite of protocols that defines the Internet. It is the neutrality and careful design of TCP/IP that has allowed the WWW to be created; if TCP/IP were designed less elegantly (if it made more assumptions about the information it was used to transfer), it would not be so straightforward to design something like the WWW that can sit on top of the platform. And, similarly, the WWW is designed to be as neutral as possible about applications that sit on top of it.

The end result is the huge hypermedia/multimedia space that is the WWW, which in effect allows you to treat every Web page as potentially sitting as a file on the user's computer. Markup languages such as HTML tell the individual machine how to display (or *render*, to use the technical term) the information. The way it is rendered will vary from computer to computer, and indeed window to window on a computer (open two windows with your Web browser, make them different dimensions, and then download the China Policy Institute's page onto each of them – it will appear differently, yet coherently, in both, thanks to HTML).

Navigating through the WWW is non-trivial, but made possible by various technologies. HTML allows links to be made democratically, from a page to any other you like; the reader, if she wishes, clicks on a link to go to the next page. http cleverly defines names of webpages in such a way that the computer can use the name to find the location of the resource that the name refers to. So, for instance, http://www.nottingham.ac.uk/china-policy-institute/ tells the computer to look for the contents of a directory called china-policy-institute on a computer server which is called www.nottingham.ac.uk. The name of the page is designed to be exactly the same as a conventional representation of its location (address). The simple idea of conflating names and addresses has proven to be very powerful.

A third method of navigation became required once the Web grew beyond a certain size, and people wanted access to Web pages with which they were not necessarily acquainted, and so the search engine was born (Google being the most successful at the time of writing). Search engines generally have three aspects: they must sample and store as much of the Web as they can (a huge engineering task), they must be able to search through their sample quickly and efficiently for a key word or other method of identifying a potentially interesting page, and they must then present what may be a very large quantity of information to the user. This last is the most controversial; Google uses a variation of an algorithm called *PageRank* which orders the retrieved pages in order of presumed importance. The problem with search engines is to keep their performance acceptable as the size of the WWW increases – the search engines popular in the 1990s have not been able to cope, by and large, with the scale of the current WWW, and Google has claimed top spot.[7] To see how difficult a task that is, and what feats of engineering good search engines are, consider a Google search for the three key words 'china' 'policy' and 'institute' performed on 4th January 2008. That search retrieved, from Google's stored model of the WWW, an astonishing 51,600,000 pages, all of which used the three key words somewhere in their text.

[7] John Battelle, *The Search: How Google and its Rivals Rewrote the Rules of Business and Transformed Our Culture*, (Boston: Nicholas Brearley Publishing, 2005).

What is even more astonishing is that the retrieval took a mere 0.29 seconds. And the brilliance of PageRank is shown by the fact that, once it had ordered those 51.6 million pages on the basis of its assumptions about the searcher's requirements, the home page of the China Policy Institute at Nottingham (for which I was indeed looking) was *first on the list*. Second on the list was the Wikipedia page describing the CPI (and including a picture of its lovely headquarters at China House). Any human attempt to perform this task on this scale would not only take weeks rather than fractions of a second, but would also be massively more error-prone.[8]

Properties of the WWW

This architecture means that the WWW has certain important properties, which are intended to be invariant aspects of the Web experience. In other words, if these properties were subverted, then the Web experience would be lessened, and the WWW would be reduced to something closer to the average information dispersal technology.

Such desirable properties include:

- **Decentralisation**. As a webpage author, you can link to any other page, resource or repository of data that has been published online, and indeed the creator of any other Web resource can link to you. There is no central hub to monitor linking – which helps efficiency of course, but also promotes the freedom to cite or reproduce the content of whomsoever one feels significant. Similarly, there is no hierarchical information structure, but rather a decentralised network, so information can be gathered direct from the appropriate source. The lack of a hierarchy rules out direct and pervasive censorship.
- **Open standards**. The growth of the WWW is down to the fact that several heterogeneous representation formats, languages and file types can peacefully coexist on it. This is because the WWW, which is defined by standards for the formalisms that make it up, uses open standards that can be accessed, for free, by anyone. The WWW is not a restricted piece of intellectual property. Someone writing an application for the Web can write their application, and connect it to the Web using these openly available standards. No-one has to write in a particular computer language, or to pay a licence fee, in order to build an application on the WWW, which dramatically reduces the barriers to entry.
- **Living texts**. HTML is designed to allow texts to be commented on, added to, included in other texts, referred to and linked to. Digital technology keeps texts alive and open in a way that paper does not. Think, for instance, of the comments that follow an article in an online newspaper, compared to the closure that the paper version of the newspaper can rely on; any commenting would have to be done in subsequent editions of the paper, perhaps on the letters page, and so the newspaper retains editorial control. In an online paper, the original article is often a seed for a long debate (which is admittedly often garbage!), and so the boundary between writer and reader is blurred. The

[8] For more on the architecture of the Web, the technologies it supports and its likely development over the near future, see Kieron O'Hara & David Stevens, *inequality.com: Power, Poverty and the Digital Divide*, (Oxford: Oneworld, 2006), pp.33-67.

author tends to lose control of his or her creation – ironically an effect that Plato, through the mouth of Socrates, predicted would happen (in the *Phaedrus*) with the invention of writing!

- **Non-linearity**. WWW texts can of course be linear, but equally they can exploit the hypertextual aspects of the WWW and adopt a non-linear structure, allowing the reader to follow links through a site as they interest him or her. So, for instance, a highly allusive text such as Eliot's *The Waste Land* can be rendered in such a way that one can follow the allusions around the WWW, or through the poem.[9]
- **Serendipitous reuse**. Because linking is so democratic, material can end up being reused serendipitously in contexts undreamt of by the author. In the paper era, the value of knowledge was created by *scarcity* – methods for allowing the collection of rents on intellectual property, such as copyright or patents, meant that knowledge was valuable by virtue of the restrictions placed on it. A user would have to pay (buy a book, or pay royalties or licence fees) to get hold of it. In the digital age, knowledge adds value through abundance – knowledge of little intrinsic value in an isolated context can have value added to it because it can easily be placed in a new context. So, for example, the diary of an ordinary person of the 18th century might be, in isolation, of interest mainly to the specialist historian. But in the context of a large number of such documents, it could be a great value for a range of social historians, philosophers and scientists– for instance, meteorologists could gain a lot of useful confirmatory evidence about weather patterns.

Use of the WWW

The WWW, then, is a decentralised information technology designed to have the properties outlined above. It has always, as we have seen, been defined by rules, and the rules that have determined its features have always depended on it context, and the perceived threats to it. How, then, has the WWW generally been used?

Science and the WWW

Perhaps the most important early adopters of the WWW were scientists. The performance of science, which not by coincidence has become much more data-heavy over the last few years, is now highly dependent of the Web and the Internet, serious disruption to which would produce something of a crisis. For the scientist, the WWW provides speedy access to documents, and data (and to the computing power to process those data). Various bottlenecks that typically slow down the dissemination of scientific research, such as the long period of time which publication of a book or journal article, and the limited amount of space and purchasing power available to university libraries, can be side-stepped by simply putting one's papers on the Web. This time factor is very important, as in many disciplines knowledge is advancing faster than it can be written about. The WWW also facilitates collaborative work – virtually all empirical work in the sciences and social sciences is now performed jointly by a number of researchers who are very often at different institutions.

[9] Cf. e.g. http://eliotswasteland.tripod.com/.

Commerce and the WWW

During the mid-90s, the use of the WWW moved from the universities to the private sector, with companies latching onto the commercial possibilities, and consumers relishing the opportunities to shop and bank online. These developments did require a change of culture of the WWW.[10] Conceived as a knowledge sharing technology for the use of scientists and researchers, most WWW applications assumed a benign user base acting in good faith with the good of a community in mind as well as self-interest, and there was a clear instrumental value in the freedom to use knowledge created by others. These assumptions obviously do not hold in the commercial world.

In the first place, when money is being passed around during a commercial transaction, identity becomes important. It was famously pointed out in a *New Yorker* cartoon that "on the Internet no-one knows you're a dog", and this fluidity of identity was found attractive by many of its early adopters. In the world of science, hoaxes and forgeries are rare, so if a scientific paper appeared apparently written by, say, Kieron O'Hara, one could assume with some certainty that Kieron O'Hara was the author. However, the same is not true when presented with credit card details. The flip side of the *New Yorker* joke became apparent – sometimes it's useful to be able to prove that you are *not* a dog. Hence authentication procedures needed to be developed for online use. This, notoriously, created the concomitant problem that once an online identity was established, that in itself became a commodity of value, and the hitherto undeveloped crime of identity theft became the bane of many people's lives.

Similarly, other security measures were needed. Attempts to develop a purely online currency never really took off, but protocols allowing credit cards to be used online without giving away too much personal information were obviously required. One did not want fraudulent transactions performed using one's card, but equally one did want authorised transactions to go through relatively smoothly and easily. As with all aspects of commerce, complex security mechanisms (passwords, security codes) put off casual users and prevent sectoral growth.

Thirdly, as users were putting information online, privacy protection was also needed. Avoiding identity theft is an aspect of this, but all online activity leaves some kind of trace, and those traces can be embarrassing. This is obviously true in industries such as pornography, but is also a factor in people seeking particular types of health care, buying subversive books or trading in shares, for example. It may be that people acting perfectly properly still do not want their details shared with others, perhaps because they do not want to be bombarded with marketing material.[11]

Fourthly, there is also the important issue of intellectual property. Sharing data is of basic importance in science, but in popular music or literature, the artists' incomes are affected by royalties paid as a result of measurable purchases of their material. There are certainly good arguments that this sort of restriction of intellectual property is not the best thing for the arts, or even for the artists, but it remains the case that someone downloading a piece of music from a peer-to-peer file-sharing site is not contributing to the artist's income in the same way that a purchaser of a CD will be. Similar arguments surrounded the early recording industry (it was thought that records would destroy the sheet music industry, which they did, but of course no-one minds that

[10] Cf. Laurence Lessig, *Code and Other Laws of Cyberspace,* (New York: Basic Books, 1999).
[11] Kieron O'Hara & Nigel Shadbolt, *The Spy in the Coffee Machine: The End of Privacy As We Know It*, (Oxford: Oneworld, 2008), O'Hara & Stevens, *inequality.com*, pp.243-271.

now), and the business model of the large record companies will change, but that innovating work still needs to be done.[12]

Finally, methods of communication that assume good faith can easily be undermined by those of bad faith. Several methods of swindling or pestering have been developed by people exploiting straightforward methods of transferring information, perhaps most notably spam and phishing, each of which is a serious problem in the online world which developed only as the WWW moved away from its academic roots.

The imperatives of the 1990s

The result of these developments in the WWW was that in the 1990s, the main imperative was to ensure that regulation was light, and governments were kept away from the Web as far as possible. It was felt that governments would interfere with the free flow of information, and free markets, and instead would try to impose restrictive security cordons on the WWW that would impede information flow, and the WWW's growth.

For example, Ira Magaziner, Bill Clinton's chief Internet policy advisor, argued that:

> Commerce on the Internet could total tens of billions of dollars by the turn of the century. For this potential to be realized fully, governments must adopt a non-regulatory, market-oriented approach to electronic commerce, one that facilitates the emergence of a transparent and predictable legal environment to support global business and commerce.
>
> Official decision makers must respect the unique nature of the medium and recognize that widespread competition and increased consumer choice should be the defining features of the new digital marketplace.[13]

Principles underlying the WWW

The result of this line of development of the WWW, and indeed the Internet, was a decentralised system designed to promote the free flow of information. In particular, the WWW became a system governed, explicitly or implicitly, on a series of principles or norms.

- **Liberalism**. The WWW is a space engineered on liberal principles, where freedom of speech is a privileged value. This is a particularly interesting point, as it has dual aspects. The WWW is a liberal space, where interactions are conceived of in a liberal manner. But also the success of the WWW, in terms of its growth (both of users, and of information held on it) depends on a liberal attitude to all the information packets moving around it.
- **Conversation and dialogue**. The WWW is intended to promote free, uninhibited self-expression, dialogue and compromise where possible.
- **The free flow of information**. Issues such as intellectual property and copyright intrude, but basically the more information flowing around the WWW, the better.

---

[12] Cf. Laurence Lessig. *The Future of Ideas: The Fate of the Commons in a Connected World*, (New York: Random House, 2001).

[13] Ira C. Magaziner, "The framework for global economic commerce: a policy perspective," *Journal of International Affairs*, 51 (1997).

- **Adding value through abundance and reuse**. If the WWW had a single purpose, it would be to add value to information through its serendipitous reuse in novel situations, as noted above.
- **Promotion of science and human progress**. In all this, the WWW embodies a whiggish point of view that society can progress, and does so through the development of intellectual capacities, among other things. A WWW which promoted pseudo-science, conspiracy theories and pornography while failing to serve science would be seen as something of a failure.
- **A platform upon which undreamt of applications can be built**. As with TCP/IP, the WWW is intended to be a clean technological platform, making as few assumptions as possible about how it will be developed, upon which new applications can be built even though they are as yet unplanned. In this way, new technologies can appear almost overnight on the WWW, because once they are developed and made available, there is nothing impeding the rapid development of a user base. Examples of some of the technologies which have grown up on top of the WWW and the Internet include: the Semantic Web; Web services; Web 2.0 and social software; peer to peer systems; software agents; pervasive or ubiquitous computing; the mobile Web; grid computing; personalised computing services; e-commerce systems; revolutionary media such as Google; and massive multiplayer online role-playing games. All of these have high numbers of users, and none of them could work if it were not for the clean infrastructure below them.

It can be seen that many of these principles, though commonly accepted in the Western and Western-style democracies, are not universally accepted across the globe. There are many states of varying levels of illiberalism which would wish to oppose some or all of these principles, and it is to one of these, China, to which I will now turn.

China and the WWW

Chinese public policy in the 21st century can be characterised as continuing economic reform in advance of, or instead of, political reform. The government wishes to ensure the welfare and prosperity of its people, without accepting strong threats to its own position, or to a loosely-defined ‘social harmony’. There are a number of potential threats to that harmony of course, including increasing inequalities of wealth, clear divides between urban and rural interests, major migrations both internally and externally, serious environmental issues looming and restive ethnic minorities in the West and South. But one major factor, as perceived (probably correctly) by the Chinese authorities is the WWW, which favours various social groups on the right side of the digital divide, and has also been used by subversive organisations. Indeed, the Internet and the Web have actually improved Chinese governance in some respects, as their provision of timely information undermined the attempts of regional and party officials to suppress bad news (perhaps most notably about the SARS outbreak and the rising number of riots across China).

However, as has also been pointed out, the WWW is an important part of economic reform, helping free up information in markets, and putting consumers and small businesses in touch with other businesses. The Chinese e-commerce site alibaba.com, and the associated auction site taobao.com, both founded and at the time of writing run by Jack Ma (Ma Yun), are among the most successful, and most heavily used,

Internet sites in the world, for example. Furthermore, given that China has a number of ambitions in science and technology, the role of the Web is scientific research is, as has been noted, vital. The Chinese government, therefore, has in general tried to pursue a strategy of exploiting the WWW for commercial and scientific purposes, while trying to prevent it undermining social harmony.

The principle of abundance of information is replaced by restriction, notably on political commentary and political speech, as well as on "inappropriate" material such as pornography, and on technical information about how to get around security and censorship systems.[14] Control of the Web is shifted as far as possible from users to the government, and carriers of information, whose liability is in most polities limited, are turned into agents of the state.[15] And in part the hands-off strategy developed during the 1990s, to support the use of the WWW in science, research and commerce, which restricted the regulation of the Web, has allowed China (and other illiberal regimes) to flout liberal principles in this way. China now regularly uses Internet filtering technologies, looking for specific keywords or IP addresses, while regulating bloggers and Internet cafés, suborning big Western Internet firms such as Yahoo! and Google, all the while employing 30,000 secret policemen to trawl the Web for content of which it disapproves.[16]

There are arguments for censorship of course, and in many Eastern nations and cultures, social harmony is valued above individual freedom to some extent at least. But what is particularly worrying about the Chinese model in the context of the WWW is that it is particularly pernicious. For instance, in Saudi Arabia, censorship is transparent, and documented by the Internet Services Unit. The filtering is supported by a highly conservative society, to the extent that many Saudis report sites that should be made unavailable within Saudi Arabia to the authorities. There is a notification given when access is denied to a sensitive website, and there is a possibility of appeal; one can apply to the authorities that access be restored. In Singapore, there is relatively little direct filtering of the WWW by the Media Development Authority, and instead the authorities rely on muscular use of defamation laws and existing legal formulae, which are already used to suppress political comment in print and TV media. In North Korea, the entire country is pretty well cut off from the WWW, and a small number of individuals are licensed to trawl for content that can be made available to users of that nation's Intranet.[17]

In contrast, the Chinese approach is centralised, and more importantly, opaque. When, in China, one tries to download www.bbc.co.uk, the Web browser hangs. As a user, one is unaware whether there is a problem with the site, or whether it has been blocked (it will have been blocked). If one does a Web search in China, then pages are fairly likely to be blocked from one's search; even if one has used an unobjectionable keyword, sites which contain banned keywords (such as 'Falun Gong') will not appear. But again one is not notified that that has happened. As we have seen, the notions of transparency and equality of information are essential for the WWW to work on a number of levels, whether information is conceived in engineering terms as

[14] Jonathan L. Zittrain & John G. Palfrey Jr, *Internet Filtering in China in 2004-2005: A Country Study*, Open Net Initiative, April 2005, http://www.opennetinitiative.net/studies/china/ONI_China_Country_Study.pdf.
[15] Daniel J. Weitzner, *China: A Broken Link on the Web?* Keynote speech, 2006 World Wide Web Conference, Edinburgh, 2006, http://www.w3.org/2006/Talks/0424-chinabrokelink/#(1).
[16] Zittrain & Palfrey, *Internet Filtering in China in 2004-2005*.
[17] O'Hara & Shadbolt, *The Spy in the Coffee Machine*, pp.168-175.

a packet of meaningful data encoded digitally, or in social terms as a series of informative propositions. The Chinese attitude is to replace transparency with opacity, making the operation of the Web unclear. Censorship is one thing – indeed, many users' computers censor content using spam filters or filters such as Net Nanny or CYBERsitter – but it is possible to censor transparently, making it clear what has been removed.

Conclusion: a dilemma

Does this matter? The short answer is 'yes' – perhaps most obviously in terms of sheer numbers. China's Internet usage is large and increasing. According to both the China Internet Network Information Centre[18] and Internet World Stats,[19] the Chinese online presence at the time of writing is second only to the USA, more than their next two rivals put together; it is widely expected that (particularly after the spectacular rates of growth recorded in 2007), China will overtake the USA in 2008. More to the point, Internet penetration (the number of people in the population using the Internet) is a very low 12.3%, compared to relatively mature markets such as the United States (penetration 69.7%), the United Kingdom (62.3%), South Korea (66.5%) and Japan (67.1%). Only India and Indonesia, of the major Internet-using nations, have comparable 'room to grow'.[20]

The WWW has grown because of its support of the free and democratic flow of information; indeed, it is the free flow where value is added to information. The Chinese government, as part and parcel of its phased reform strategy, is much more interested in the value than the freedom – while for their part, those developing the WWW tend to work within a liberal paradigm, and are keener on the freedom than the regulation that must follow. This is a serious issue for those steering the development of the WWW.

Daniel J. Weitzner, at the time of writing Director of the World Wide Web Consortium's technology and society activities, has made the following suggestions for a way forward[21]:

- National sovereignty: accept the inevitable. China will clearly remain in charge of its own laws, and is entitled as a sovereign nation to regulate behaviour within its borders. This cannot be wished away. It may be that a 'principled' approach to China could result in a vast reduction of digital contact, as urged by a number of NGOs including, for instance, Reporters Without Borders.[22] But this does ignore the possibility of using the WWW to ease China into a more liberal stance, as well as – given the size of the Chinese online presence – risking the integrity and connectivity of the WWW itself. A hardline approach to China could split the WWW into two more or less separate networks; currently it is not clear how damaging this might be.

[18] China Internet Network Information Centre, *Statistical Survey Report on the Internet Development in China*, CNNIC, January 2008, http://www.cnnic.net.cn/uploadfiles/pdf/2008/2/29/104126.pdf. The figures are from December 2007.
[19] http://www.internetworldstats.com/index.html, figures from June 2007.
[20] *Ibid*.
[21] Weitzner, *China: A Broken Link on the Web?* The comments following each suggestion are my own.
[22] Reporters Without Borders, *Do Internet Companies Need to be Regulated to Ensure they Respect Free Expression?* Press release, 6th Jan, 2006, http://www.rsf.org/article.php3?id_article=16110

- Support free expression. Having said that, it is clear that free expression is generally taken to be a human right by most international organisations, and that support – and criticism of countries that do not respect it – should be forthcoming in the context of the WWW.
- Global transparency. National laws and government actions should be transparent – if there is censorship of the WWW, then the extent of that censorship, and the principles underlying it should be made clear. Neither of these is true in the Chinese case; indeed, the uncertainty in the Chinese system leads, and is no doubt intended to lead, to self-censorship, as commentators are literally unaware of how much what they say will be tolerated.[23]

The very name 'World Wide Web' is a statement of ambition. The WWW itself was founded by liberals, and has advanced on impeccably liberal principles.[24] It has become a space in which politics can take place,[25] and a key tool to support development, democracy and transparent governance.[26] As long as most users of the WWW were based in Western-style democracies, this was not a problem, but now that these markets are maturing, the Web's expansion in the next few years will inevitably be in illiberal communities. This is a problem for illiberal countries such as China, as has often been pointed out, but actually it is a problem for the Web too. The WWW is the most complex piece of technology ever built, yet also based on simple principles, and the idea of a democratic, decentralised space of linked information has proved to be incredibly powerful. There are serious issues about how to export that vision into illiberal polities, while preserving the essential invariants of the Web experience, to which can be traced its power.

[23] Zittrain & Palfrey, *Internet Filtering in China in 2004-2005*.
[24] Tim Berners-Lee, Wendy Hall, James A. Hendler, Kieron O'Hara, Nigel Shadbolt & Daniel J. Weitzner, "A framework for Web Science," *Foundations and Trends in Web Science*, 1(1), (2006), pp.1-134, at pp.107-109.
[25] O'Hara & Stevens, *inequality.com*, pp.1-31.
[26] *Ibid*, pp.119-242.