theguardian

Over to the dark side of the web

Photograph: Kim Gilmour / Alamy/Alamy
Freenet means controversial information does not need to be stored in physical data havens such as this one, Sealand. Photograph: Kim Gilmour/Alamy

Andy Beckett
The Guardian Features Thu 26 Nov 2009 09:25 GMT

In the 'deep web', Freenet software allows users complete anonymity as they share viruses, criminal contacts and child pornography

Fourteen years ago, a pasty Irish teenager with a flair for inventions arrived at Edinburgh University to study artificial intelligence and computer science. For his thesis project, Ian Clarke created "a Distributed, Decentralised Information Storage and Retrieval System", or, as a less precise person might put it, a revolutionary new way for people to use the internet without detection. By downloading Clarke's software, which he intended to distribute for free, anyone could chat online, or read or set up a website, or share files, with almost complete anonymity.

"It seemed so obvious that that was what the net was supposed to be about – freedom to communicate," Clarke says now. "But [back then] in the late 90s that simply wasn't the case. The internet could be monitored more quickly, more comprehensively, more cheaply than more old-fashioned communications systems like the mail." His pioneering software was intended to change that.

His tutors were not bowled over. "I would say the response was a bit lukewarm. They gave me a B. They thought the project was a bit wacky … they said, 'You didn't cite enough prior work.'"

Undaunted, in 2000 Clarke publicly released his software, now more appealingly called Freenet. Nine years on, he has lost count of how many people are using it: "At least 2m copies have been downloaded from the website, primarily in Europe and the US. The website is blocked in [authoritarian] countries like China so there, people tend to get Freenet from friends." Last year Clarke produced an improved version: it hides not only the identities of Freenet users but also, in any online environment, the fact that someone is using Freenet at all.

Installing the software takes barely a couple of minutes and requires minimal computer skills. You find the Freenet website, read a few terse instructions, and answer a few questions ("How much security do you need?" … "NORMAL: I live in a relatively free country" or "MAXIMUM: I intend to access information that could get me arrested, imprisoned, or worse"). Then you enter a previously hidden online world. In utilitarian type and bald capsule descriptions, an official Freenet index lists the hundreds of "freesites" available: "Iran News", "Horny Kate", "The Terrorist's Handbook: A practical guide to explosives and other things of interests to terrorists", "How To Spot A Pedophile [sic]", "Freenet Warez Portal: The source for pirate copies of books, games, movies, music, software, TV series and more", "Arson Around With Auntie: A how-to guide on arson attacks for animal rights activists". There is material written in Russian, Spanish, Dutch, Polish and Italian. There is English-language material from America and Thailand, from Argentina and Japan. There are disconcerting blogs ("Welcome to my first Freenet site. I'm not here because of kiddie porn … [but] I might post some images of naked women") and legally dubious political revelations. There is all the teeming life of the everyday internet, but rendered a little stranger and more intense. One of the Freenet bloggers sums up the difference: "If you're reading this now, then you're on the darkweb."

The modern internet is often thought of as a miracle of openness – its global reach, its outflanking of censors, its

seemingly all-seeing search engines. "Many many users think that when they search on Google they're getting all the web pages," says Anand Rajaraman, co-founder of Kosmix, one of a new generation of post-Google search engine companies. But Rajaraman knows different. "I think it's a very small fraction of the deep web which search engines are bringing to the surface. I don't know, to be honest, what fraction. No one has a really good estimate of how big the deep web is. Five hundred times as big as the surface web is the only estimate I know."

Unfathomable and mysterious

"The darkweb"; "the deep web"; beneath "the surface web" – the metaphors alone make the internet feel suddenly more unfathomable and mysterious. Other terms circulate among those in the know: "darknet", "invisible web", "dark address space", "murky address space", "dirty address space". Not all these phrases mean the same thing. While a "darknet" is an online network such as Freenet that is concealed from non-users, with all the potential for transgressive behaviour that implies, much of "the deep web", spooky as it sounds, consists of unremarkable consumer and research data that is beyond the reach of search engines. "Dark address space" often refers to internet addresses that, for purely technical reasons, have simply stopped working.

And yet, in a sense, they are all part of the same picture: beyond the confines of most people's online lives, there is a vast other internet out there, used by millions but largely ignored by the media and properly understood by only a few computer scientists. How was it created? What exactly happens in it? And does it represent the future of life online or the past?

Michael K Bergman, an American academic and entrepreneur, is one of the foremost authorities on this other internet. In the late 90s he undertook research to try to gauge its scale. "I remember saying to my staff, 'It's probably two or three times bigger than the regular web,'" he remembers. "But the vastness of the deep web . . . completely took my breath away. We kept turning over rocks and discovering things."

In 2001 he published a paper on the deep web that is still regularly cited today. "The deep web is currently 400 to 550 times larger than the commonly defined world wide web," he wrote. "The deep web is the fastest growing category of new information on the internet … The value of deep web content is immeasurable … internet searches are searching only 0.03% … of the [total web] pages available."

In the eight years since, use of the internet has been utterly transformed in many ways, but improvements in search technology by Google, Kosmix and others have only begun to plumb the deep web. "A hidden web [search] engine that's going to have everything – that's not quite practical," says Professor Juliana Freire of the University of Utah, who is leading a deep web search project called Deep Peep. "It's not actually feasible to index the whole deep web. There's just too much data."

But sheer scale is not the only problem. "When we've crawled [searched] several sites, we've gotten blocked," says Freire. "You can actually come up with ways that make it impossible for anyone [searching] to grab all your data." Sometimes the motivation is commercial – "people have spent a lot of time and money building, say, a database of used cars for sale, and don't want you to be able to copy their site"; and sometimes privacy is sought for other reasons. "There's a well-known crime syndicate called the Russian Business Network (RBN)," says Craig Labovitz, chief scientist at Arbor Networks, a leading online security firm, "and they're always jumping around the internet, grabbing bits of [disused] address space, sending out millions of spam emails from there, and then quickly disconnecting."

The RBN also rents temporary websites to other criminals for online identity theft, child pornography and releasing computer viruses. The internet has been infamous for such activities for decades; what has been less understood until recently was how the increasingly complex geography of the internet has aided them. "In 2000 dark and murky address space was a bit of a novelty," says Labovitz. "This is now an entrenched part of the daily life of the internet." Defunct online companies; technical errors and failures; disputes between internet service providers; abandoned addresses once used by the US military in the earliest days of the internet – all these have left the online landscape scattered with derelict or forgotten properties, perfect for illicit exploitation, sometimes for only a few seconds before they are returned to disuse. How easy is it to take over a dark address? "I don't think my mother could do it," says Labovitz. "But it just takes a PC and a connection. The internet has been largely built on trust."

Open or closed?

In fact, the internet has always been driven as much by a desire for secrecy as a desire for transparency. The network was the joint creation of the US defence department and the American counterculture – the WELL, one of the first and most influential online communities, was a spinoff from hippy bible the Whole Earth Catalog – and both groups had reasons to build hidden or semi-hidden online environments as well as open ones. "Strong encryption [code-writing] developed in parallel with the internet," says Danny O'Brien, an activist with the Electronic Frontier Foundation, a long-established pressure group for online privacy.

There are still secretive parts of the internet where this unlikely alliance between hairy libertarians and the cloak-and-dagger military endures. The Onion Router, or Tor, is an American volunteer-run project that offers free software to those seeking anonymous online communication, like a more respectable version of Freenet. Tor's users, according to its website, include US secret service "field agents" and "law enforcement officers . . . Tor allows officials to surf questionable websites and services without leaving tell-tale tracks," but also "activists and whistleblowers", for example "environmental groups [who] are increasingly falling under surveillance in the US under laws meant to protect against terrorism". Tor, in short, is used both by the American state and by some of its fiercest opponents. On the hidden internet, political life can be as labyrinthine as in a novel by Thomas Pynchon.

The hollow legs of Sealand

The often furtive, anarchic quality of life online struck some observers decades ago. In 1975, only half a dozen years after the internet was created, the science-fiction author John Brunner wrote of "so many worms and counter-worms loose in the data-net" in his influential novel The Shockwave Rider. By the 80s "data havens", at first physical then online locations where sensitive computerised information could be concealed, were established in discreet jurisdictions such as Caribbean tax havens. In 2000 an American internet startup called HavenCo set up a much more provocative data haven, in a former second world war sea fort just outside British territorial waters off the Suffolk coast, which since the 60s had housed an eccentric independent "principality" called Sealand. HavenCo announced that it would store any data unless it concerned terrorism or child pornography, on servers built into the hollow legs of Sealand as they extended beneath the waves. A better metaphor for the hidden depths of the internet was hard to imagine.

In 2007 the highly successful Swedish filesharing website The Pirate Bay – the downloading of music and films for free being another booming darknet enterprise – announced its intention to buy Sealand. The plan has come to nothing so far, and last year it was reported that HavenCo had ceased operation, but in truth the need for physical data havens is probably diminishing. Services such as Tor and Freenet perform the same function electronically; and in a sense, even the "open" internet, as online privacy-seekers sometimes slightly contemptuously refer to it, has increasingly become a place for concealment: people posting and blogging under pseudonyms, people walling off their online lives from prying eyes on social networking websites.

"The more people do everything online, the more there's going to be bits of your life that you don't want to be part of your public online persona," says O'Brien. A spokesman for the Police Central e-crime Unit [PCeU] at the Metropolitan Police points out that many internet secrets hide in plain sight: "A lot of internet criminal activity is on online forums that are not hidden, you just have to know where to find them. Like paedophile websites: people who use them might go to an innocent-looking website with a picture of flowers, click on the 18th flower, arrive on another innocent-looking website, click something there, and so on." The paedophile ring convicted this autumn and currently awaiting sentence for offences involving Little Ted's nursery in Plymouth met on Facebook. Such secret criminal networks are not purely a product of the digital age: codes and slang and pathways known only to initiates were granting access to illicit worlds long before the internet.

To libertarians such as O'Brien and Clarke the hidden internet, however you define it, is constantly under threat from restrictive governments and corporations. Its freedoms, they say, must be defended absolutely. "Child pornography does exist on Freenet," says Clarke. "But it exists all over the web, in the post . . . At Freenet we could establish a virus to destroy any child pornography on Freenet – we could implement that technically. But then whoever has the key [to that filtering software] becomes a target. Suddenly we'd start getting served copyright notices; anything suspect on Freenet, we'd get pressure to shut it down. To modify Freenet would be the end of Freenet."

Always recorded

According to the police, for criminal users of services such as Freenet, the end is coming anyway. The PCeU spokesman says, "The anonymity things, there are ways to get round them, and we do get round them. When you use the internet, something's always recorded somewhere. It's a question of identifying who is holding that information." Don't the police find their investigations obstructed by the libertarian culture of so much life online? "No, people tend to be co-operative."

The internet, for all its anarchy, is becoming steadily more commercialised; as internet service providers, for example, become larger and more profit-driven, the spokesman suggests, it is increasingly in their interests to accept a degree of policing. "There has been an increasing centralisation," Ian Clarke acknowledges regretfully.

Meanwhile the search engine companies are restlessly looking for paths into the deep web and the other sections of the internet currently denied to them. "There's a deep implication for privacy," says Anand Rajaraman of Kosmix. "Tonnes and tonnes of stuff out there on the deep web has what I call security through obscurity. But security through obscurity is actually a false security. You [the average internet user] can't find something, but the bad guys can find it if they try hard enough."

As Kosmix and other search engines improve, he says, they will make the internet truly transparent: "You will be on the same level playing field as the bad guys." The internet as a sort of electronic panopticon, everything on it unforgivingly visible and retrievable – suddenly its current murky depths seem in some ways preferable.

Ten years ago Tim Berners-Lee, the British computer scientist credited with inventing the web, wrote: "I have a dream for the web in which computers become capable of analysing all the data on the web – the content, links, and transactions between people … A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines." Yet this "semantic web" remains the stuff of knotty computer science papers rather than a reality.

"It's really been the holy grail for 30 years," says Bergman. One obstacle, he continues, is that the internet continues to expand in unpredictable and messy surges. "The boundaries of what the web is have become much more blurred. Is Twitter part of the web or part of something else? Now the web, in a sense, is just everything. In 1998, the NEC laboratory at Princeton published a paper on the size of the internet. Who could get something like that published now? You can't talk about how big the internet is. Because what is the metric?"

Gold Rush

It seems likely that the internet will remain in its Gold Rush phase for some time yet. And in the crevices and corners of its slightly thrown-together structures, darknets and other private online environments will continue to flourish. They can be inspiring places to spend time in, full of dissidents and eccentrics and the internet's original freewheeling spirit. But a darknet is not always somewhere for the squeamish.

On Freenet, there is a currently a "freesite" which makes allegations against supposed paedophiles, complete with names, photographs, extensive details of their lives online, and partial home addresses. In much smaller type underneath runs the disclaimer: "The material contained in this freesite is hearsay . . . It is not admissable in court proceedings and would certainly not reach the burden of proof requirement of a criminal trial." For the time being, when I'm wandering around online, I may stick to Google.

✉ Send to a friend

g Contact us

Go back to the web story

Browse guardian jobs

Soulmates dating
Search for a date now

Get text alerts

Sponsored features

Mobile marketing

Win a trip to Cuba

Get the latest Red Bull Air Race news and videos

Enjoy England- take part and win

[   ] [Search]