1

Fragmentation and Compartmentalization of Virtual Space

1.1. The nymph Carna and Internet census

One day, a computer scientist wondered how many Internet users could navigate this immense digital map, which is today's global network. Therefore, he created a small and perfectly harmless spy program named Carna Botnet, in honor of the nymph Carna, who became goddess of the *cardo*, the "hinge" or "axis", that is to say, the divinity of the gates in Roman rites, a charge that she inherited from the god Janus, who had taken her virginity in exchange.

> "The first day [of June] is consecrated to you, Carna, goddess of the hinges. She opens that which is closed, she closes that which is open; these are the attributes of her divinity" [1].

While Janus, honored on January 1, opens the first part of the year, Carna is celebrated on June 1, opening the second half of the year. The anonymous creator of Carna Botnet intimately knew her Latin letters. Carna is indeed a multifaceted goddess. If she was to be given tribute and sacrifice in June, the month in which the days were the longest, it was from her reign of the calendar that the period of the year begins, in which days begin to shorten until the end of summer and slowly turn into winter. The goddess of light, Carna is therefore also a goddess of darkness and concealment, or even a

1 Ovid, *Fasti*, VI, 107.

goddess of the underworld, with whom she is associated. This second attribute also makes her worth being considered as the goddess of organs and internal functioning of the human body, the "goddess of the human body viscera", so says the Latin author Macrobe[2].

From May to October 2012, the small program named after the goddess of entrails attempted to list all objects connected to the Internet with an IPv4[3] address. Out of a total of 4.3 billion IPv4 addresses available, Carna Botnet counted 1.3 billion active addresses in October 2012; 729 million occupied domains and 141 million addresses protected behind a firewall. In a previous "large Internet census" in 2006, 187 million visible users were counted. The latest estimates[4] put the total number of this era of smartphone users at just under four billion. Nevertheless, from his large "2012 Internet census", Carna Botnet's creator drew the conclusion that the development of IPv6[5] addresses might make any further census attempts in the future impossible. Five years on, in 2017, the growth of the connected objects industry and the ever-increasing number of users proved it right: it is impossible to know exactly how many users, devices, objects or servers are connected to the Internet today. Estimates and statistics continue to be produced, but remain doomed to be approximated.

It is therefore probably impossible to establish a precise geography of the global network today, and this can only be celebrated if we consider that the Internet must remain a virtual space in which privacy, anonymity and user freedom must be preserved. This libertarian concept contrasts with that of states and governments, striving to know and monitor cyberspace zones entrenched behind the barriers of encryption, or even more simply, lost in the ocean of data accumulated since the creation and privatization of the Internet. States are not alone in wanting to know about this *new terra incognita* of the digital universe, as private and public sector economic operators are also looking for ways to take advantage of them. In the face of various attempts at state regulation and commercial penetration of the "hidden Internet", communities and individuals are now trying to hide behind the supposed sanctity of encryption keys that allow, as way of an

2 Macrobe, *Saturnales*, I, 12, 31–33.

3 A 32-bit electronic identifier format, mainly used today on the Internet.

4 *EMC Digital Universe Infobrief*, research and synthesis accomplished by the *International Database Corporation* (IDC), April 2014.

5 A 128-bit electronic identifier format to replace IPv4.

example, two million users to surf the Internet anonymously using the Tor browser (The Onion Router), in the name of protecting privacy and freedoms, and sometimes for less admirable reasons.

The very provocative "Declaration of the Independence of Cyberspace" [6], issued in 1996 by John Perry Barlow, particularly resonates today. In Barlow's days, it was a protest against the Telecommunications Act that, according to the author of the "Declaration of Independence", supplied this virtual space of freedom (that is the Internet) to commercial appetites and the regulatory fury of companies and governments. But in 1996, the Internet was still in its infancy. Twenty years later, in 2017, we are not far off from considering that Barlow's wish came true. Cyberspace has somehow found its independence through its own extension and the phenomenon of encrypted networks. According to Campbell [CAM 07], author of a report for the European Parliament in 1997, the battle of cryptography was already lost by governments at the dawn of the 21st Century, which leaves room for the expansion of private networks and thus offers the prospect of a World Wide Web that is very difficult to control, a virtual territory largely beyond the reach of legislation; a gray zone between public and private space. The exponential growth of the global network, a vast web of networks and subnetworks numbering in the tens of thousands, in itself guarantees the relative powerlessness of States – which never had the extensive monitoring capabilities that they do today – from controlling Internet traffic. The mass of data represented by the circulation of these immaterial flows is impossible to process. Today, the Internet is simply too vast to be submitted in its entirety to the authority of regulatory bodies, or even to be comprehensively understood and apprehended.

1.2. Dimensions of cyberspace

There is a need for early comprehension on some definitions. The Internet is a global computer network composed of millions of public and private networks, made up of a set of sites, pages and databases accessible via the World Wide Web, invented in the early 1990s by CERN[7] computer scientists Tim Berners-Lee and Robert Cailliau. The World Wide Web is

6 See the text in the Appendix.

7 European Council for Nuclear Research, today known as the European Organization for Nuclear Research.

only one application among many (among others, the various e-mail systems and peer-to-peer file-sharing systems) that provides access to the Internet. The latter, since its official creation on January 1, 1983, and its opening to commercial exploitation in the 1990s, thus brings together an ever-increasing number of users, but also connected objects and databases, which can be accessed through browsers such as Google Chrome, Yahoo, Internet Explorer, Opera and other lesser-known browsers. According to internetlivestats.com, there were 3,611,467,000 network users on April 14, 2017 and 1,177,754,000 online sites. According to the same site, in April 2017, 2,580,768 e-mails were sent per second (including an overwhelming majority of spam sent by robots), 7,578 tweets, 776 photos uploaded to Instagram, 59,779 Google searches and 43,277 GB of data exchanged per second. In 2016, the International Communication Union estimated that less than half of the world's population had access to the global network. This leaves the Internet an impressive margin for growth.

This exponential growth seems to remove all significant issues on the size of the Internet. In July 2000, a study by Cyveillance, "Sizing the Internet" [MUR 00], estimated the size of the Internet to be more than 2 billion pages. Five years later, the strategic intelligence company DIGIMIND produced a study showing that the Internet had about 64 billion pages, while an Italian study announced in the same year that 11.5 billion pages were indexed by the main search engines. Nowadays, if we estimate the number of Websites created on the Internet at nearly 1 billion 200 million, it is very difficult to know how many active pages this can correspond to. In terms of data volume, the size of the Internet was therefore estimated at 4.4 zettabytes in 2014[8], equivalent to a number of digital tablets that would cover two-thirds of the distance from the Earth to the Moon, if they were end-to-end. The same study estimates that in 2020, the volume of data represented by the Internet will have exceeded 44 zettabytes, six times the distance from the Earth to the Moon, using the previous example. The rate of growth of the Internet now makes it possible to reach such orders of magnitude that we can now speak of a true "alchemy of multitudes", just as the researchers Francis Pisani and Dominique Piotet did [PIS 08, p. 188].

This "alchemy of multitudes", which brings together 4.4 zettabytes of data, nearly 50,000 different networks, 1 billion 200 million sites and an almost incalculable number of connected objects, constitutes the Internet,

8 Data Observatory, July 2014, op. cit., p. 8.

where the World Wide Web allows nearly 4 billion users to navigate. However, the vast majority of these users are unaware of most of these vast digital resources and will only visit a limited number of sites and pages that have been archived since the Internet was created. The most widely used search engines, such as Google and Yahoo, are supposed to reference at best only 15–20% of all content on the Internet, because of a number of restrictions that can be quickly outlined in the form of a table (see Table 1.1). These different types of content, which are indexed by search engines differently – or not at all – will determine the existence of subsets within the Internet, depending on the accessibility of online data and content. If we consider the Internet as a vast set of networks, where the application of the World Wide Web makes it possible to navigate and allows search engines to orientate themselves, from page to page or site to site, we will be able to distinguish several subsets that constitute the world web according to the different categories of contents.

Type of content	Description
Contextual content	Page content varies according to the context of access (e.g. e-mail homepage).
Dynamic content	Content hosted on a server, accessible by a request, whose display as a page is determined by a set of scripts controlled from the server. The user will access this type of page through a search engine. Dynamic pages are generated and controlled by an application from the hosting server (e.g. corporate or government Websites, or booking forms).
Static content	Page whose content is simply stored on a server database or online and made available on the World Wide Web via HTTP (HyperText Transfer Protocol).
Content with limited access	This type of content will not necessarily be encrypted, but search engine access will be limited by the use of anti-robot protocols (which will specify exclusion zones, but won't be necessarily respected by all automated search software) or CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart).

Non-HTML or non-indexable content	Specific formats not recognized by search engines. Flash content or using Javascript.
Private content	Password-protected content (typically: a private forum, e-mail, protected customer account, etc.).
Encrypted and protected content by application	Any type of content hosted on an alternative network such as Tor, I2P or Freenet. In this case, accessing it requires the installation of a specific software or search engine.
Orphan pages and archived content	The expression "orphan page" means a page that is no longer linked to another page or for which a search engine has been unable to locate the link. Archived content refers to all archived pages of a site, or even a site itself, that have become inaccessible to search engines.

Table 1.1. *Different types of content*

1.3. *Deep web*, *darknet* and *dark web*

The best thing to do with science today is to use it to explore the present. "Earth is today's alien planet", William Gibson said in an interview with the American channel CNN in 1997. Certainly, virtual space today is one of the most fascinating subjects of this unsettling terrestrial strangeness. Cyberspace is even the most faithful technical materialization of the Freudian *Unheimliche*. A virtual bottomless pit, a Pandora's box of fantasies, the Internet arouses the most imaginative theories in order to grasp the reality and complexity of the global network. One of the most popular today postulates the existence of "Mariana's Web", which would be the last level of the Internet, accessible only after a trip into digital darkness that has nothing to envy of the descent into the hells of Orpheus, all the way from level 1, frequented by M. The entire world, up to a kind of mythical plan containing almost all the secrets of the universe, from the manufacture of quantum computers to the location of Atlantis and the secrets of the Illuminati. Of course, this ultimate level of the Internet is supposed to be controlled by a mysterious secret society.

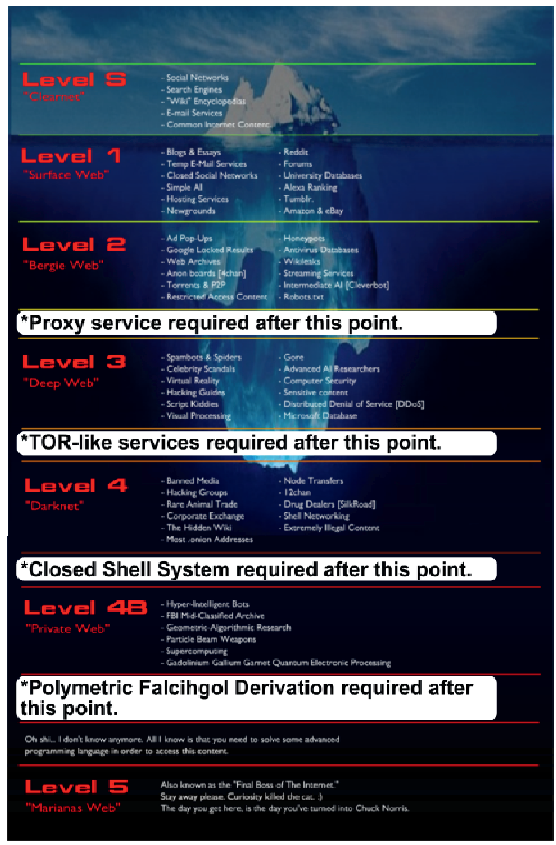

Figure 1.1. *The Internet in the form of an iceberg: a fantasy representation*

Another popular representation is that of the iceberg (Figure 1.1), which has the merit of offering a representation that is easy to grasp for the human mind: on the first level, the "surface web", and then the different levels of the "deep web" to the mysterious "darknet". The image, however convenient as it may be, is nevertheless based on a profoundly wrong apprehension of the division of virtual space. This division actually takes place according to the *accessibility* of content, as shown in Table 1.1: contextual, static, dynamic, limited and protected content. The volume of data also determines the ability of search engines to access data that is not necessarily protected, but simply poorly referenced or even impossible to reference. Therefore, in the vast entirety formed by the Internet, we can distinguish three subsets: the

surface web, which can be accessed without any problems from simple queries on the most famous search engines, the deep web, a vast ensemble made up of non-indexed, private or difficult to access content and the darknet: the "hidden Internet", made up of various darknets, and alternative networks. Just as the World Wide Web is the application that makes it possible to navigate the Internet, dark webs are the subsets formed by sites and applications that make it possible to navigate, communicate and exchange on an alternative network.

Sites whose addresses ends with Onion are accessible via Tor (The Onion Router). Applications such as Grams or OnionCity (search engine on the darknet, see Figure 1.2) or sites such as TorShops or SheepMarketPlace (the latter being an alternative to Silk Road) are also part of the dark web.

Figure 1.2. *The Grams search engine*

We will then understand what distinguishes the deep web from the dark web: the latter does not even represent 0.005% of all the platforms present on the Internet. On the "classic" Internet, in other words the unencrypted and normally accessible without the help of a specific protocol or application, a forum such as 4Chan (Figure 1.3), which was widely publicized because of

cases of hacking involved in US and French presidential campaigns, can provide a good example of data considered to belong to the deep web: it can be easily accessed by browsing through the forum, but a direct request on Google or Yahoo will not allow access to the content of the forum.

Figure 1.3. *www.4chan.org homepage*

If the Internet is a fantastic thread for all conspiracy theories, the darknet is an ideal fantasy object for the postmodern folklore of digital esotericism. The fascinating idea of a "hidden Internet" still leads to frequent confusion between darknet and deep web. The deep web, a term that could be translated as the "deep Internet", means nothing more than all the online databases, publications and archives that cannot be indexed by traditional search engines, due to the exponential growth of the global network since its creation in 1983 and privatization in 1994. According to the most recent studies, the "deep web" represents a volume of data 4,000–5,000 times bigger [CHE 17, pp. 26–38] than the surface web. The darknet, on the other hand, refers to a much more restricted part of the Internet whose access, contrary to the deep Internet, is regulated by very specific protocols and where, in contrast, legal regulations may not always apply according to the standards defined by law.