Cloud Computing and its Ethical Challenges

Matteo Turilli1,2

Luciano Floridi2, 3, 4

1Oxford e-Research Centre, University of Oxford; 2Information Ethics Group, University of

Oxford; 3Research Chair in Philosophy of Information, University of Hertfordshire; 4UNESCO

Chair in Information and Computer Ethics.

Short abstract (205 ws)

The paper analyses six ethical challenges posed by cloud computing, concerning ownership, safety, fairness, responsibility, accountability and privacy. The first part defines cloud computing on the basis of a resource-oriented approach, and outlines the main features that characterise such technology. Following these clarifications, the second part argues that cloud computing reshapes some classic problems often debated in information and computer ethics. To begin with, cloud computing makes possible a complete decoupling of ownership, possession and use of data and this helps to explain the problems occurring when different providers of cloud computing retain or relinquish the right to use or own users' data. The problem of safety in cloud computing is coupled to that of reliability, insofar as users have to trust providers to preserve their data, applications and content in a reliable manner. It is argued that, in this context, data insurance could play an important role. Regarding fairness, the paper argues that cloud computing is already reshaping the nature of the Digital. Responsibility, accountability and privacy close the ethical analysis of cloud computing. In this case, the thesis is that the necessity to account for the actions of cloud computing users imposes delicate trade-offs between users' privacy and the traceability of their operations.

Long abstract (1818 ws)

According to a resource-oriented interpretation, cloud computing refers to software-related activities performed by users thanks to pools of computing resources, which are accessible through a network, where they are made available by some providers. Storage space, Virtual Machines (VMs), software and application frameworks, software for product delivery, and software applications are all typical examples of computing resources that can be made available as a service through a so called 'cloud'. These computing resources are used to implement and deploy a vast array of technologies and corresponding services, ranging from online storage and back-up solutions to web sites for e-commerce and e-mail services.

The advantage of such a resource-oriented interpretation is that it helps to highlight four characteristics of cloud computing:

1. the shift towards a utility-based conception of computing resources;

2. the virtualization of hardware resources into software resources;

3. the fundamental role played by networking; and

4. the provision of computing resources as a service.

1. A utility-based conception of soft resources

Cloud computing is built on the idea that providers offer to users the opportunity to consume just the amount of computing resources that they need, when they need them. It is pay-as-you-go computing. A utility-based conception of computing resources is of particular importance for business and research users. While for a domestic user owning a predetermined amount of hardware and software usually implies acceptable trade-offs, for computationally intensive business or research endeavours acquiring computing resources exceeding present demands can be either too expensive or unfeasible.

2. The virtualization of hardware resources.

Computing resources are usually provided by hardware, which then represents the major constrain for their flexible deployment. "Virtualisation" refers to any process whereby one can deliver computing resources usually built in hardware – like a specific CPU, a storage facility or a network infrastructure – by means of software. Clearly, such "softening" of hardware resources plays a crucial role in the implementation of a utility-based approach to their provision. The difference between deploying a virtual or a physical machine is dramatic. Once the virtualisation infrastructure is in place, the provider of virtualised hardware resources can satisfy users' requests in a matter of minutes and, potentially, to a very large scale. Likewise, terminating or halting such a provision is equally immediate.

3. The fundamental role played by networking

Cloud computing is a form of remote computing and thus networking is how it is offered, delivered and consumed. Users rely on cloud computing through a network, usually the Internet, and the products of such activity are either downloaded or built so that they can be accessed online, through the Internet. This trend is well exemplified by the development of cloud-oriented operating systems, like Google Android, Chrome OS or Apple iOS.

4. The provision of computing resources as a service

From the point of view of an infrastructure provider, cloud computing is a way to offer computing resources through VMs, and VMs as a service. Users need a computational infrastructure and providers offer a virtualised version of such an infrastructure that is functionally analogous to the one implemented in hardware. This type of cloud computing is usually known as Infrastructure as a Service (IaaS).

The same service-oriented approach to the delivery and fruition of computing resources can also be applied to platforms for software deployment. As in the case of a computing infrastructure, a software deployment platform can be provided on-demand, as a service to the software developers. This type of cloud computing is called Platform as a Service (PaaS).

Finally, the evolution of web-related technologies has prompted the development of a third type of cloud computing: the so-called Software as a Service (SaaS). In many situations, complex software applications that can be executed through freely available web browsers are a desirable alternative to applications that have to be installed directly in the operating systems of users' computers.

IaaS, PaaS and SaaS are radically changing the way in which millions of users access and consume computing resources in order to develop or exploit new types of applications, products, services and enjoy their online experiences. Obviously, such radical changes have far reaching consequences. Some of them are ethical, and the second part of this paper is dedicated to their analysis.

1. Ownership, possession, and use

Cloud computing is part of the contemporary tendency towards the deflation of the notion of ownership and the uniqueness of what is owned. The underlying idea is that use does not imply ownership, for it might require only temporary possession, especially when the good in question is a mere clone, that is, an instance of a type, indiscernibly replaceable by any other instance of the same type. Owning and therefore maintaining large and complex hardware resources is a limiting, expensive and often unsustainable overhead for users. The issue here is that, while the ownership of the hardware supporting computing activities is not needed or wanted anymore, the ownership of the outcome of such activities remains vital. To put it simply, users want to store data in the cloud, they do not want to own such cloud, only possess its services so that they may be able to use them, but they also want their data to be and remain their own data. The difficulty is to make sure that users only possess and use services but own their data, while providers own and use their services but only possess users' data.

2. Safety, reliability and data insurance

Storing large amount of potentially sensitive data – including mobile phone usage records, personal e-mails, work-related documentation or photographs – on hardware facilities owned by private companies poses not only the problem of who retains the ownership of those data, but also of how and why the storage provider should be trusted in managing them properly. Specific definition of 'misuse' may be provided and supported by different legislations, policies and contracts, but all this would only partially mitigate the issue concerning data mining, user profiling and the possibility of data leakages or loss. The solution here does not seem to lie only in the improvement of the legal constraints that can make providers trustworthy, but also, if not mainly, in transferring the full ownership and control of the data access and usage from the provider to the user.

The need for trusting a cloud computing provider has profound consequences also for the safety of the data processing. This because safety in cloud computing involves not only the usual need for authorised-based access to user data and operating systems but also the reliability of the offered services. It should be the responsibility of the provider to inform its users about the technological and policy-based measures taken in order to guarantee data integrity, while it should be the responsibility of the user to be informed about this issue.

Finally, because cloud computing decouples the physical possession of data (by the provider) from their ownership (by the user), it also offers an unprecedented opportunity. Nowadays it is still common and easy to insure a machine (e.g. a laptop or a mobile) on which the data are stored, but not the data it stores. This because, although data may be invaluable and irreplaceable, they are also perfectly clonable at a negligible cost, contrary to physical objects, so it would be impossible for an insurer to ascertain their irrecoverable loss or corruption. On the contrary, once it is the provider that physically possesses the data and is responsible for their maintenance, the user/owner of such data should rightly expect to see them insured, for a premium of course, and to be compensated in case of damage, loss or downtime.

3. Fairness and digital divide

Cloud computing adds a new dimension to the problem of distributive justice in general and the digital divide in particular. On the one hand, it clearly contributes to a democratisation of computing resources through their potential wider distribution at a lower cost, thus further levelling the technological playing field. But the digital divide is not just a problem about the unfair availability of computing resources, it is also, and sometimes above all, a problem of accessibility and usability, and in these two respects, cloud computing may easily exacerbate it. Recall that cloud computing depends on affordable, dependable, safe and fast networks as well as on the IT skills of the empowered users. So cloud computing might reshape the borders of the digital divide, and place them between users who have access to the right kind of online services and the required skills to make the most of the computing resources made available there, and those who do not. Paradoxically, the cloud is potentially available everywhere, and in theory we may only need very basic terminals to erase the digital divide, but in fact it is shifting the discrimination between 'onlife' and 'offlife' people.

4. Control and responsibility

One of the defining characteristics of cloud computing is to shift the control of a computational infrastructure from the provider to the user. Once the user has obtained one or more VMs, he may run a large amount of different operating systems with all the required software customisations. The user retains full technical control over her VMs: not only about switching it on or off, and installing and running whatever software it is technically feasible, but also about deciding how the VMs are networked with each other and with the whole Internet. Of course, users remain legally responsible for their wrongdoing (e.g., if they breach the contract with the provider), but they are not pre-emptively incapacitated to misuse the provided infrastructure. It follows that, in an IaaS, users are assumed to be entirely responsible of their computing activities because they are fully empowered. This leads to a more complicated issue, the relationship between accountability and privacy.

5. Accountability and privacy

Accountability is used to enforce the need for full responsibility, and in this sense it may be seen as a positive factor in the management of cloud computing. However, accountability has a direct impact on the levels of privacy and anonymity of the users. In order to be accountable, users' actions need to be traceable and, as such, their physical identity must be (not necessarily known but at least) knowable to the provider, while their actions must leave meaningful traces that can be used to identify, prove and quantify the damage or offence caused by reckless behaviours. There is no space for a fully anonymous use of an IaaS or for a notion of privacy forbidding the logging of user activity. At the same time, there should be no overreaction. Arguably, a principle should be endorsed for which, among all the available implementation of accountability, the one that minimizes the erosion of the right to privacy and to anonymity is chosen. For this reason, solutions based on federated authentication and authorisation and policed logs access should be preferred to those based on proactive and invasive practices, like deep packet inspection or proactive log mining.

Conclusion

In this paper, we have seen what cloud computing is and what sort of new ethical challenges is already posing. Cloud computing is part of a macroscopic, information revolution, which is fostering new ethical sensitivities and shaping some important moral questions. It seems that tackling them successfully will require an e-nvironmental approach, which might cover all environments, natural, virtual and hybrid, and might be able to treat as authentic and genuine all forms of

existence and behaviour, even those based on synthetic and engineered artefacts. Ultimately, the challenge will be to reconcile our roles as both agents within such new infosphere, and stewards of it. The good news is that this is a challenge we can meet.