QinetiQ White Papers

Digital Investigations in the Cloud

Emma Webb-Hobson,
Senior Digital Forensic Investigator
Abstract

With the increase in cloud computing services available to home and enterprise users, digital forensic investigators need to consider what the challenges are for collecting and analysing data. This paper will examine five possible areas of data capture when collecting cloud service data which are: client side analysis (post mortem and live), server side analysis, network traffic analysis and application programming interface analysis.

Introduction

Cloud computing is the model where infrastructure, platforms, applications and services are offered over a network. Consider a webpage where data storage and interactions are all offered via a browser, such as web mail or online backups. The main storage and processing is completed on the cloud service provider (CSP) side, with cheap, low specification machines accessing services via a browser.

Increasingly, users are moving away from traditional ways of creating and storing documents and personal data. Ultraportable devices such as iPhones and netbooks are based on some local storage, and lots of online cloud storage. This makes them perfectly enabled for cloud services, where photos, documents, email and so on are stored online.

Cloud computing services can be considered as an extension of online services which have been available to home and enterprise users for the best part of 20 years. In fact, Hotmail – one of the first web mail services on the Internet – was commercially launched on American Independence Day in 1996, symbolizing “freedom” from ISP-based e-mail with the ability to access a user’s inbox from anywhere in the world.[1]

This rhetoric can be heard in the marketing commentary of today’s cloud services, the idea that clients’ data can be accessed from anywhere in the world, via a web browser.

The National Institute of Standards and Technology (NIST) defines cloud computing as “…a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” [2]

The definition also covers four different deployment models which are: private cloud; community cloud; public cloud or hybrid cloud. A private cloud is where the cloud infrastructure is operated for the purposes of a single organisation. In a community cloud the cloud is shared by several organisations with shared concerns (e.g., mission, security requirements, policy, and compliance considerations). A public cloud is an infrastructure that is made available to the general public or a large industry group and is owned by a CSP. Finally, a hybrid cloud is a composition of two or more clouds types (private, community, or public).

A big market driver for cloud services adoption is the much advertised conversion of capital expenditure to operating costs. This is a result of the pay-per-use charging model. The CSP provides scalable infrastructure, so it is the CSP that swallows the capital expenditure costs. Users are able to scale up their resources, but almost more importantly, scale back down. For example, imagine that all the investigators in a given team have a 2TB network storage device to store their evidence files on. If a server came in for analysis that is identified as 2.5TB, it is not possible to store all the relevant data on one device. There may be sufficient storage spread amongst his colleagues devices, but this is inefficient and risks cross contaminations between cases. Ignoring the arguments about the value of taking a complete forensic image of such a device, the enterprise is going to have to invest in a new, larger device, to be able to complete the analysis. This is inefficient for several reasons; there is enough storage,
it just isn’t in the right place, the enterprise has had to spend money upfront to complete the order, and the device will still remain once the requirement for it has passed.

Further market drivers for cloud computing are that in the post recession environment, enterprises can concentrate on their core offerings, and outsource enterprise functions (such as sales, marketing, human resources and finance). The availability of these services on a scalable, pay-per-use basis allows the enterprise to react quickly to change, without the upfront capital investment often required.

So what is the impact of a mass migration to cloud based services (either in an enterprise, or by home users) on digital investigations? There is apprehension from digital investigators and security experts that the use of cloud services will make obtaining digital evidence more difficult.

The best digital evidence is in user created files – documents, spreadsheets, photographs, emails, internet history and so on. With cloud services, those files are potentially no longer stored on the local machine. The key challenges are identifying whether cloud services have been used, what artefacts remain from that usage, where the relevant data is actually stored, and how it can be obtained and analysed in a forensically sound manner. This paper looks at the impact of cloud computing on digital forensics investigations and considers the possible mitigations.

1. Client Side Analysis – Post Mortem

Post mortem analysis is the traditional digital forensic methodology, where the power to the machine is cut. This “freezes” persistent data in its current state, thereby maintaining the integrity of the evidence. A forensic copy of the data is then created and analysed. When dealing with examinations in the past that involved online services such as web mail, the examination would include Internet analysis, focusing on browser artefacts, history, temporary Internet cache files, cookies and so on.

The problem is that analysis of browser artefacts can be scarce, they are generally overwritten over time, and with the increasing use of dynamically created web pages (or other cloud services), reconstructing and interpreting internet artefacts is getting more difficult. That doesn’t mean there is nothing of evidential value on the client machine. Internet history, bookmarks, cookies, installed applications, keyword searches and other fragments can at least point investigators towards the CSP(s) that have been used, which opens up a new lead in the case. It will be up to the officer in charge of the case to follow up those leads with data requests to the appropriate contacts.

Only being able to identify the CSP is the worst case scenario, it is likely that there will be much more data of value. Fragments of web pages might be recoverable; files that have been uploaded and downloaded may be recovered from temporary folders or unallocated space, and case intelligence may provide keywords to find other areas for examination. As cloud services become increasingly prevalent and begin to emerge in real cases, examination techniques will evolve and tools will be developed to further enhance the investigation process.

2. Client Side Analysis – Live

A potentially rich source of data of particular relevance to cloud computing is likely to come from a live analysis of the client machine. Live analysis is a developing trend of analysing a machine in situ, without cutting the power.[3] Live analysis allows the capture of many areas of data that are lost in a post mortem analysis including RAM, access to mounted encrypted files and volumes, instant messenger sessions, running services and network traffic. It may now be the case that we add cloud services to that list; however more research is needed in this area.

There is a downside to live analysis however. The forensic capture is only a snapshot of the data at a single point in time. Although the integrity of data can be preserved from the point of capture onwards, it cannot be independently verified after capture, as the exact state will never be replicated.[4]

The installation and use of tools for capturing and analysing information changes the data. It is however the only way to capture temporary data. [5] It is catered for in the ACPO Guidelines, [6] so long as the examiner is technically competent, understands the “impact” of their actions, maintains comprehensive contemporaneous notes and documentary evidence (for example photos and video of the capture process) and ensures that the continuity and integrity of the evidence is maintained from the point of capture onwards. Live analysis is an exciting new area of data capture, but the big barrier for law enforcement and other agencies is that it requires the availability of technical specialists to attend the scene to perform the data capture. With post mortem examinations, non-technical “first responders” can attend the scene to seize evidence for further examination in a forensic lab. This is a more cost efficient method which suits bulk post mortem examinations. Sending specialist investigators to the scene will greatly reduce the number of machines they are able to examine and analyse, resulting in the need for more specialists. Specialist attendance may be a bitter pill that has to be swallowed as the case for live onsite analysis mounts.

Data in RAM, access to mounted encrypted files and volumes, instant messenger sessions, running services and network traffic and now, open links to clouds and volatile data relating to access of cloud services could provide essential evidential value. Investigators have to go where the evidence is, and if the best evidence is to be had from live analysis, then law enforcement and other investigatory bodies will have to evolve.

3. Server (CSP) Side Analysis

The third point of data capture is from the cloud service provider (CSP) end – the “server side” in the traditional sense. At this stage, it is not known what records and data are held by the CSPs, but it is fair to assume that extracting such information from CSPs is going to be difficult and protracted. Some situations may be more positive than others though, depending on the relationship between the client and the CSP.

For an internal investigation, where the subject is the employee of an organisation using cloud services, it is likely that the cloud CSP will be a willing partner, and will co-operate fully with the production of data. That being said, there is a lack of security or forensic readiness standards, so that important logs and records may not be available. This may be because the CSP does not want, or know how, to provide them. They may be stored, but in an inappropriate format (for example unreadable or unorganised). Finally, it may be identified that the logs were not being recorded in the first place – often due to the cost implications of storing detailed logs for any length of time. It will be up to the contractual agreements in place to ensure that logs, records and other evidential data are stored and made available to the client in a mutually agreed manner, and is likely to be a chargeable service.
4. Network Traffic Analysis

Another potential source of data capture is that of communications between the cloud client and the CSP. Monitoring requires careful legal consideration as doing so could be classed as an interception of communications. For law enforcement this would require the appropriate authorisation under the Regulation of Investigatory Powers Act 2000 (RIPA)[7] Part I and a warrant issued by the Home Secretary. It is likely only to be used in the most extreme of cases. Business monitoring however can be conducted providing the appropriate legal framework is in place. RIPA s1(6) provides a basic power in respect of private telecommunication services with the Telecommunications (Lawful Business Practice) (Interception of Communications) Regulations 2000[8] providing the basis for business monitoring. The Data Protection Act 1998 (DPA) [11] will also apply to such practices, see the Information Commissioner’s Office “Quick Guide to Employment Protection Act 1998 (DPA) [11] which covers the release of computer data for the investigation and prevention of crime. When considering data for cloud services this distinction between communication data and digital files begins to blur.

A new “Government Requests Tool” has recently been released by Google [12] which identifies the number of requests made by governments from around the world (with certain limitations) to remove or disclose user data. The tool shows that 59 requests were made by the UK to Google and YouTube between 1 July 2009 and 31 December 2009. The number could be misleading, as one request may seek data for many users, or there could be multiple requests for one user. The existence of these statistics at least goes to show that a process to obtain user data from Google exists; however the number of requests would appear to be extraordinarily low, when compared with the number of digital examinations in the UK in which the subject has used Google or YouTube.

Making requests to CSPs, particularly those overseas, has been bureaucratic and time consuming in the past, and generally reserved for serious, organised or international crime. Authorised requests can be made under UK legislation – RIPA for communication data, or the Data Protection Act 1998 (DPA) [11] which covers the release of computer data for the investigation and prevention of crime. When considering data for cloud services this distinction between communication data and digital files begins to blur.

This is already a common scenario, and the frequency of this type of scenario is likely to increase as users get used to the idea of storing data online at work, and adopt the same practices at home. Additionally, there may be a perception amongst perpetrators of crime that “the cloud” is a good place to hide data. Methods exist to handle this sort of situation – law enforcement has a single-point-of-contact (SPOC) for most of the main CSPs. Requests can be made to immediately preserve data held in online accounts, and then provide copies of the data once a more in-depth request has been served.

5. Application Programming Interface (API) Analysis

Analysis The fifth and final point of capture for data in the cloud is the API itself. An API is the set of commands used by end clients to interact with the cloud service. For example, if the end user is accessing virtual machines, then the API will include commands to start, stop and restart the machine. The APIs are created by the CSP and currently there are no standard APIs. The dominant interface is Amazon’s; however its recent patent registration has caused concerns that Amazon will take legal action against others using it in the future.[10] As a result, a number of competing APIs are being developed, some open source, some not.

Non-standard APIs can result in portability conflicts, resulting in cloud users being stranded with a single CSP, and not being able to port their data to a different client. They may wish to do this following poor service from CSPs, or simply because their needs and requirements have changed. It suits the CSPs to use their own APIs, but the consensus in the community is that there will have to be standardisation at some point. Due to the immature nature of the market though, CSPs are still able to dictate their own rules. As larger enterprises move to cloud services, and require more secure, scrutinised and portable interfaces, it is inevitable that standardisation will come, as it has in other areas.

Capturing and interpreting API commands could provide valuable records of the actions taken by users and applications. The importance of this data is that it can be captured from the client side, enabling the cloud service user to have a degree of control. Even without the data files to which it refers, this information could be valuable evidence in a case, and it may be possible to recover the files themselves from alternative sources. There are not currently any known tools which are able to capture this information or present it in a format which could be used as part of an investigation.

In the following section, five scenarios are used to give practical examples of investigations with cloud service elements, and how the use of the cloud service has affected the investigation.
**Scenario B: Intellectual Property (IP) Theft**

The subject of the investigation is accessing enterprise data, held on a public cloud, via her home computer, and is downloading files for sale to a competitor. Here the employer holds the account with the CSP, so they should have no problems accessing logs and records that are of relevance to the investigation. Whether the logs and records exist, and are in a suitable format, may depend on contractual agreements, and whether the CSP and or the client have forensic readiness procedures in place. Forensic readiness procedures ensure that the organisation has the ability to gather and preserve digital evidence before an event occurs. Forensic readiness maximises the potential to use digital evidence whilst minimising the cost of the investigation. It addresses a number of key business risks, proving evidence to detect and deter crime, and preparing an organisation for the use of digital evidence in its own defence. Preparing to use digital evidence may involve enhanced system and staff monitoring. It also requires technical, physical and procedural means to secure data to evidential standards of admissibility. Processes and procedures must be introduced to ensure that staff recognise the importance and legal sensitivities of evidence. Finally, forensic readiness procedures assist with obtaining appropriate legal advice and interfacing with law enforcement in a timely manner. [15]

Forensic readiness procedures in CSPs is currently an area that is completely lacking. The recording, storing and reviewing of logs is an overhead cost, and potentially has legal risks such as privacy and disclosure issues. As cloud computing services are a relatively immature market, many of the security issues are still being discovered, and developments will be driven by market forces. Forensic readiness and other security-as-a-service offerings will only be made available to customers that demand it and are willing to pay for it. However in line with the cloud pricing model of pay-per-use, the costs to the cloud user are likely to be relatively small. Certainly it would offer value for money when compared to the cost of an investigation after a serious event. Forensic readiness should be a contractual requirement, as part of service level agreements (SLAs) between the CSP and the cloud user. For any enterprise looking to utilise such services, forensic readiness compliance should be considered essential.

This scenario identifies that some established security processes may not be present when transitioning to cloud services. The lack of standards and precedence means that the emphasis will be on the contracts and how and whether they are enforced. CSPs with dependable, well supported services will emerge as market leaders, and lead to convention and best practice in this area. Until that time, each user must be responsible for conducting their own risk assessments and due diligence. They must ensure that the CSP meets their needs not only in good times, but also provision for when it all goes wrong.

**Scenario C: Breach of Acceptable Use Policy**

The subject of the investigation is an enterprise user who has been looking at adult pornography websites during breaks. The employer uses a private cloud infrastructure, where employees have access to virtual machines through a browser.

Whilst not illegal, this would certainly constitute a disciplinary issue and warrant an internal investigation. The success of this investigation depends on a number of factors – does the employer have suitable disciplinary procedures outlining breaches of acceptable use policies as a disciplinary offence? Is there an acceptable use policy in place which prohibits access to the material described? Is the investigation being conducted in a free and fair manner, by an independent party?

Presuming these aspects are all in place, then the evidence will reside in access logs, Internet history and temporary Internet files. Where will this information be stored? Very little information of relevance will reside on the end user’s machine. There may be artefacts relating to the date and times the user accessed the cloud, but as an employee they would have had legitimate access, so this data might be irrelevant.

The useful Internet history, temporary Internet files and other artefacts for Internet analysis will reside within the virtual machine (VM) assigned to the user in the cloud. This could go either way. On the positive side, digital forensic analysis of virtual machines is relatively well developed and researched, with material and tools available. Analysis of the virtual machine disk is very straightforward – for example a VMWare virtual machine disk (.vmdk) file can be dragged and dropped into most forensic analysis software, and hash values generated to ensure integrity. VM disks have unallocated clusters in the same way that physical disks do, which can be data carved for deleted files and fragments. Additionally, virtual machine files contain an exact copy of physical memory in use by the virtual machine. Analysis of physical memory is an exciting area in digital forensics, with the increasing size of memory available, and improved techniques and research into the data traces that can be recovered from it.

The negative side is that new virtual machines can be created in a few clicks, and old ones destroyed just as easily. Once a virtual machine has been destroyed, recovering the virtual disk or virtual memory files from the cloud would be at best, non-trivial, and at worst nigh impossible! Apart from virtual machine analysis, investigators can use other records such as firewall logs and so on, to determine user activity. These methods are documented in numerous other papers – the intention of this scenario is to highlight the strong possibility that analysis of virtual machines will become commonplace as cloud based virtual workstations are adopted. Furthermore, the above emphasis on forensic readiness as a defence for employers is also applicable in this case. For example, how long are the firewall and Internet access logs retained for? Who has access to them? Are they recorded per user so that the relevant data can be easily obtained, and the privacy of other innocent users is maintained? Are they recorded in a format that can be easily exported and reviewed? Are they stored in such a way that prevents modification which would compromise their integrity? How often are the virtual machines refreshed? Is data relating to user actions recorded separately? How long is that information held? Are there backups of the virtual machines? Can they be rolled back to a previous state?

Even though in this scenario the employer should have full access to the relevant data, it is not straightforward to capture it for use in an investigation. Again this scenario highlights that thought is required, and systems and procedures need to be established prior to an event occurring. This is to ensure that when the event occurs, the right people are in the right place, with the right access, to be able to obtain the data needed for a digital investigation. This scenario also highlights the importance of acceptable use policies, disciplinary procedures, and a framework to conduct fair internal investigations.
Scenario D: Contractual Dispute with CSP

In this scenario the CSP is the subject of the investigation and is therefore potentially hostile. It is in the best interests of CSPs to ensure confidentiality, availability and integrity of data, to preserve its reputation and future business. But what happens when a CSP, on whom the end user is dependant, has a disaster, malfunction or mishap, which results in a loss of revenue to the customer, and constitutes a breach of the CSP’s contractual and / or service level agreements.

The investigator, working for, or on behalf of the aggrieved customer, is going to be seeking evidence to substantiate the claim as described above. Consider scenarios where the customer has lost access to their data temporarily – is this the fault of the CSP or the Internet service provider (ISP)? How about if files have been corrupted? How can the client prove that they were not already corrupted when they were uploaded, or that some action taken by an employee has caused the error? What if some of the cloud functionality is reduced, for example email services become unavailable for a day. How can the client quantify that as a loss and where does the evidence reside to show that the CSP is at fault (if that is the case)? If it can be determined where the relevant data is held, is it possible for investigators to get access?

In this scenario it is likely that the investigation would take a legal route, being based on contracts and civil litigation. It is worth considering what would happen in the event of a contractual dispute with a supplier, and how the client would get access to relevant data under such circumstances. Monitoring and recording of the service received should be conducted by the client, rather than relying on the CSP to provide them with details of any service outages or problems.

Scenario E: Civil Dispute between Corporations

The subject of the investigation is an enterprise engaged in a civil dispute with a competitor. Electronic discovery (ediscovery) is an established method for obtaining documents from corporations and organisations. Some of the issues regarding disclosure of digital files include the differing formats in which files are stored, identifying where they are stored, identifying relevant documents while maintaining confidentiality and corporate sensitivities and extracting documents from the sheer volume of data that is often held by large companies. Also, digital files often have associated metadata (file properties such as dates and times for creation and modification, author name and version numbers). Files need to be identified based on metadata as well as content, and the process must ensure the integrity of those files and the associated metadata.

What are the implications for electronic discovery investigations when one or both parties is storing their data in the cloud? There are the obvious jurisdictional issues. If an enterprise is using the public cloud to store data then that data could be stored anywhere in the world. How does an investigator begin to identify where that data is held, and how can it be physically accessed?

The answer to these issues is that the enterprise itself should know where its own data is held, as per various legal requirements and in order to comply with requests for that data (including investigative requests).

The Sarbanes-Oxley Act 2002 (SOX) is US legislation that has had implications in the UK and internationally. It is described as "an act to protect investors by improving the accuracy and reliability of corporate disclosures made pursuant to the securities laws, and for other purposes" [16] and requires companies to have audit controls and disclosure rules in place.

E-discovery experts have been dealing with data stored in disparate locations for some time and many developed processes exist to assist in accessing that data.

Part 31 of the Civil Procedure Rules [17] governs the responsibility of companies to disclose data to an aggrieved party, and to support any authorised party to which the data is relevant. The definition of documents is broad, and covers computer databases, emails, documents, scanned documents and metadata. It also covers electronically recorded communications and activities such as instant messaging and online systems. An appropriate court order issued under this legislation allows for electronic discovery to take place in civil disputes.

This scenario emphasizes that enterprises must carefully scrutinise their arrangements with any third party CSP to ensure that they can effectively accommodate electronic discovery in these circumstances. This includes being aware of the legal ramifications of storing data internationally.
Conclusions

This paper has considered a number of impacts of cloud computing on digital forensic investigations. As with any development in technology, the paradigm shift towards online services will create new complexities relating to the access and acquisition of evidential data.

Some of the issues raised include jurisdictional problems, as many of the cloud services available are based in the US or other non-UK countries. Ways to resolve this include using and improving existing methods for international co-operation on data requests.

Previous experience with more traditional online services, such as webmail, may indicate a way forward when conducting live or post mortem analysis of end client machines.

The analysis of virtual machines will certainly become more important as cloud models are incorporated into corporate architectures.

Existing legal requirements may assist investigators, forcing some types of data to be stored in the UK, thus reducing some of the jurisdictional issues. Other existing legislation includes the scope to cover online services, and the requirement for enterprises to comply with digital investigations.

Forensic readiness in enterprise continues to be an essential requirement to ensure that companies are in a position to comply with digital investigations both when they are the target of an investigation, and when they are the victim. Businesses must ensure that they have contracts in place with their CSPs that are tailored to their needs, and ensure compliance and due diligence.

Best practice standards amongst CSPs will assist in providing a consistent approach to data capture and assistance with digital investigations. The demand for forensic readiness and self-regulation in CSPs will only be driven by a demand from customers, who need to ensure that alongside the cost savings, they consider the implications of somebody else storing and managing their data.

The UK Government, law enforcement and its international partners must work together to improve international agreements and processes. As the cloud grows, so the world shrinks. Digital investigators need to find new ways to work with the developments in technology, just as they have done with each change that has gone before.

References


About QinetiQ

A FTSE 250 company, QinetiQ is a leading international provider of technology-based services and solutions to government departments as well as commercial and industrial companies around the world. We employ 14,000 people in the UK, US and Australia – the majority of whom are scientists and engineers, who use their in-depth knowledge to support our customers.

For more information see: www.QinetiQ.com/cyber

About the author

Emma Webb-Hobson is a Senior Cyber Security Consultant based at Cody Technology Park in Farnborough. She has been working in digital forensics since 2004 and has undertaken forensic analysis in criminal and civil cases for both public and private sector clients.

In April 2009 she joined QinetiQ helping to establish the new Digital Investigation Services site at Farnborough. More recently Emma has been part of a “Securing the Cloud” taskforce developing tools and services to provide cloud assurance and security.

For further details or to discuss any of the topics in this paper, please contact:

Emma Webb-Hobson
Senior Cyber Security Consultant
QinetiQ
Email: Cyber@QinetiQ.com
Phone: +44 (0)8700 100 942