Migration to governmental cloud digital forensics community: economics and methodology

Gianluigi Me[1]

CeRSI - Research Center in Information Systems
LUISS Guido Carli University Roma, Italy
gme@luiss.it

Abstract. The rapid growth of elastic computing services, together with a raising need to target savings, enable new IT scenarios. In particular, Law Enforcement Agencies (LEAs) can benefit of digital forensic services sharing, lowering CAPEX and OPEX especially in case of spiky utilization of forensic hardware and software resources. In fact, considering the workloads related to the size of a forensic organizational unit, the capital expenditure (CAPEX) and operational expenditures (OPEX) differently afflict the overall cost incurred to provide response to crime investigation. Hence, Government cloud community SaaS model, under strict security requirements, can represent a viable solution to dramatically lower costs of forensics units,. This paper evaluates the Total Cost of Ownership (TCO) of in-house forensic farm and the TCO of a SaaS deployed as a government community cloud, providing Forensic as a Service to N forensic units. The costs assessment are part of the evaluation methodology to calculate the number of forensic crime cases needed to switch to Government cloud model from multiple independent forensic labs.

Keywords: SaaS, Economic Evaluation, Total Cost of Ownership, Cost Model, Digital Forensics, CAPEX, OPEX.

1 Introduction

The development of digital forensics over the last decade brought LEAs to dramatically increase case resolution effectiveness so that digital forensics, currently, represents a mandatory step in the crime investigation activity. Moreover, the huge penetration of electronic devices in daily life rapidly increased the effort in this area: e.g. FBI doubled the received cases in 2003-2012 [1] and the digital forensic area shows a 12% growth rate in 2008-2012 compound annual [2] .

In [3] Garfinkel argues that the Golden Age of Digital Forensics is quickly coming to an end of an evolution process: the pointed problems become very relevant in case of growing size of storage devices, where time to create a forensic image of a subject device and to process all of the data once it is found is frequently insufficient. In particular, huge amounts of mobile devices with increasing memory size together with large datacenter forensics require specialized hardware equipment, huge storage capability and heterogeneous software

licenses portfolio, which can be supported with high CAPEX in order to provide a prompt and complete response to crime investigation. In this expanding scenario, LEA digital forensics capacity has been strengthen in the last years, due to the huge increase of cybercrime volumes and typologies and the increasing use of Internet and digital devices to finalize common crimes [4] : however, the technological crime evolution time remains lower than the LEA organizational/technological latency due to limited budgets and finite resources.

Hence, as stated in 2006 Moore's paper [5], technological changes threaten to undermine LEA capacity for complete data analysis and the technology investments related to Digital Forensic Organizational Units (DFOUs), e.g. LEAs task forces, typically suffer of short lifetime due to rapid obsolescence of equipment versus high CAPEX needed to setup the equipment. Furthermore, the increasing digital material received by DFOUs for analysis reflects to spiky use of forensic software licenses and workload on storage and CPU cores (i.e. massive seizing for acquisition of mobile phones versus acquisition of data center storage). This scenario represents the typical fertile background for evaluating the opportunity to migrate to elastic services, providing typical cloud benefits in terms of *efficiency* (e.g. improved asset utilization, aggregated demand), *agility* (e.g. pay per use, as-a-service), *innovation* (e.g. shift focus from asset ownership to service management). In fact, when occasional or periodic load spikes happen, cloud computing is squarely tailored to provide on-demand excess resources capacity, since poor utilization rates represent a non-negligible waste factor.

In particular, IT assets in digital forensics are not used equally or continuously: as a rule of thumb, research evidences show that as computing power has indeed grown far cheaper and more plentiful, utilization rates for IT resources have rapidly decreased. All the above-mentioned considerations, in particular efficiency, agility and innovation, led many countries/regions to start the design and implementation of national strategies for Governmental Clouds (GC), e.g. in the USA in [8] , in the EU in [9]), recognizing that public authorities stand to benefit from Cloud adoption both in terms of efficiency savings and in terms of services that are more flexible and turned to the needs of citizens and business. In particular, the GC, defined in [10] as an *environment running services compliant with governmental and EU legislations on security, privacy and resilience*, is going to be deployed (or in some cases, already exists) in many western countries, providing services under public body governance to state agencies, to citizens and to enterprises.

Based on the aforementioned considerations, this paper aims to provide a methodology to identify the drivers for evaluating the appropriate model and to evaluate the convenience to switch to governmental elastic services versus maintaining the independent DFOUs, considered as a community in the GC service model.

Finally, sections 2 and 3 will introduce the Cloud and Digital Forensics background needed to better understand economic evaluations in sections 6 and 7. The section 4 will identify the objectives and the limitations of the outcomes of this study, while the section 8 will present the conclusions and the future works.

2 Government Private/Community Cloud and SaaS/PaaS services

The National Institute of Standards and technology (NIST) defined 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. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models* [16] . In particular, the deployment models defined by NIST are Public, Hybrid, Community and Private, where, in particular, the cloud infrastructure is deployed, customized, operated and maintained mainly for an organization as client supervised by cloud service provider. In the private cloud, all the resources and applications managed by the organization are strictly closed to the public. Therefore, an organization sets up a virtualization environment on its own servers, either in its own data centers or in those of a managed services provider. This structure is useful for organizations having the total control over every aspect of their infrastructure as an hard requirement. Moreover, private cloud is typically more secure than public cloud since it is customized on features that leads to more secure of their application, as consequently only selected stakeholders within organization may access and manage on a specific private cloud. Conversely, a community cloud infrastructure is procured jointly by several LEAs sharing specific needs such as security, compliance, or jurisdiction considerations. Therefore, digital forensic service for LEAs can be considered as an ICT commodity, having a common set of requirements and customers: in this case a community cloud enables the asset combination and the computing resources, data, and capabilities sharing. By eliminating the duplication of similar systems, LEAs can save money and allocate their resources more efficiently. Both private and community models fulfill the requirements of a GC forensic service provisioning: without losing generality, this paper will focus on the community cloud as the adopted model. According to [15] , although a standard definition for GC services has not been widely accepted, we assume, for the purposes of this paper, that GC services fulfill the following requirements:

- Providing private (single tenancy) or community (agreed set of tenants) Cloud to host processes and store data, to run eGovernment services, controlled/monitored locally or in a centralized way by the public body;
- Owned and managed by central government under its responsibility (private or community deployment model);
- Ensuring compliance with the infrastructure, the platform and the services with country governments and EU legislations on privacy, security and resiliency (location of the GC: on premises or off premises Cloud).

This paper will consider the GC service providing option Government to Government, where a public institution provides cloud services LEAs. Moreover, the forensic service provided by GC can be classified as critical since it

handles sensitive information (related to crime investigation) and the impact of a potential failure or leakage affects the privacy of a number of citizens. The general critical service option is already in use in 29 % of EU cloud services, as well as community/private cloud are the most used deployment models [15] .

Table 1. Cloud service models

Service model	**Delivery model**
SaaS	Companies host applications in the cloud. The service being sold or offered is a complete end-user application.
PaaS	Developers can design, build, and test applications that run on the cloud provider infrastructure.
IaaS	System administrators obtain general processing, storage, database management and other resources.

As shown in Table 1, the SaaS and PaaS models fulfill the requirements of digital forensics applications, while IaaS provides an (typically unnecessary) increase of flexibility: in SaaS the GC offers software instances to be run in single-multi tenant mode, managing the overall system. In this model, the digital forensic operator analyses data using instances of forensic software, as it uses her/his webmail. In the PaaS, the GC delivers the necessary hardware resources where the digital forensic operators con run their own software, as Virtual Machines on a remote host. Since the forensic services provided by GC should be considered critical, in this paper the SaaS alternative will be only considered, without any loss of generality of the methodology (versus PaaS). In fact, there are substantial benefits to profit and welfare in high security-loss environments associated with introducing a SaaS alternative, according to [17].

Hence, the migration from DFOUs to GC-community introduces unseen concerns related to security, in particular data leakage: in fact, even considering the software always patched at the latest version, the magnitude of user information located in one place, together with the desirability to get the criminal investigation data, SaaS service may be more susceptible to targeted attacks (e.g. Google, Salesforce and Sony data breaches). Therefore, in order to determine the applicability of migration, as suggested in [11] , after verifying the service and marketplace characteristics, the application and government readiness, the security requirements should be carefully checked. In particular, GC bodies providing forensic services have the responsibility to ensure:

- Compliance to laws, regulations, and agency requirements;
- Data characteristics to assess protections the forensic data set requires;
- Privacy and confidentiality to protect against access to information;
- Integrity to ensure data is authorized, complete, and accurate;
- Data controls (location where can be stored) and related access policies.

These aspects are discussed in section 4.

3 The Digital Forensics

The term digital forensics refers to the scientific examination, analysis, and/or evaluation of digital evidence in legal matters [18] . Digital evidence, defined as any information of probative value that is either stored or transmitted in a digital form, once gathered, must satisfy the same legal requirements as conventional evidence, i.e. it must be:

- *Authentic*, the evidence must be original and related to the crime;
- *Reliable*, the evidence must have been collected and preserved (e.g. chain of custody) using reliable procedures that if necessary could be repeated by an independent party to achieve the same result;
- *Complete*, the evidence may be used to prove guilt as well as innocence;
- *Believable*, the evidence should be formed to convince juries and prosecutors;
- *Admissible*, the evidence was collected using procedures compliant to law.

Hence digital forensics is the collection of forensic processes applied to all (heterogeneous) digital devices, in order to transform data acquired by computers, networks, mobile devices and external memories in digital evidence. Examples of digital evidence include files stored on a computer hard drive (physically, in blocks or clusters), file fragments or data items stored in a USB memory, digital videos or audios recorded e stored in a mobile equipment, packets transmitted over a network or most recent paths in a car GPS-navigator, stored in a file. For the purpose of this paper, digital forensics is a 6-phases process :

1. *Identification*: determine items, components and data possible associated with the allegation or incident; employ triage techniques;
2. *Preservation*: ensure evidence integrity or state;
3. *Collection*: extract or harvest individual data items or groupings;
4. *Examination*: scrutinize data items and their attributes (characteristics);
5. *Analysis*: fuse and correlate material to produce reasoned conclusions;
6. *Presentation*: report facts in an organized, clear and objective manner.

This six stage process model and several other similar models [19] form the basis of the majority of digital forensic investigations. In particular, the former 3 steps are related to the acquisition (optionally performed off LEA premises), the latter to the analysis and presentation, which represent the core activities where this paper focuses for migration to GC services. In the current scenario (*as is*), the activities performed by DFOU forensic operators (basically from a LEA) are basically acquiring digital device content on a not volatile, not-rewritable memory support (optionally removable), with appropriate software/hardware tools (with data integrity as main hard requirement). The image of the seized device is, then, analyzed on premises, in a LEA forensic farm equipped with ad-hoc hardware/software: all the installed equipment has CAPEX and OPEX charged on the LEA budget. In the scenario based on FaaS (*to be*), LEA forensic operator acquires digital device content with appropriate software/hardware tools and uploads it to GC storage, to be further analyzed via digital forensic software

clients through GC, typically on a VPN (Virtual Private Network) between LEA labs and GC. Consequently, all the needed equipment has CAPEX and OPEX charged on GC budget: the GC-community service dimensioning and SLA definition to set the acceptable requests loss and the allocation policy LEA are needed to appropriately share the resources. Nonetheless, this is out of the scope of this paper, which, instead, focuses on the overall costs sustained by Government to support the GC versus individual DFOUs choice. In phases 4-5-6, CAPEX components related to digital forensics have twofold features: the former is related to hardware and software components suffering the rapid obsolescence with continuous need of updates. In fact, e.g., the operating systems life cycle, new hardware plugs for mobile devices, new or updated data formats drive the need to have always updated equipment in order to guarantee the effectiveness of the digital investigation. The latter is related to forensic software vendors, offering a large set of common capabilities and some specialized (unique) capabilities: hence, the forensic analysis quality and completeness is strictly related to the availability of all the different tools. Therefore, the ideal forensic labs should have all the forensics equipment on the market with latest updates installed. This scenario represents a classical prerequisite for cloud adoption: in fact, as multiple software programs have to be used, many licenses can remain idle for long time (depending on how a single forensic tool is able to respond to criminal investigation requests), leading to a not-optimal utilization rate per license. For these reasons, forensic applications fall in the general classification presented in [20] , where applications receiving cost benefits by running in the cloud include:

- applications with huge range of loads, occasional or periodic load spikes (crime evidences arrival process can be modeled as the sum of Poisson processes);
- capacity is hard to predict (crime evidences hardware and software resources requirements are not predictable);
- equipment purchasing is expensive and idle most of the time;
- new applications requiring additional data center space or infrastructure investment, such as new storage, cooling or power systems.

4 Objectives and Remarks

The main objective of this paper is to provide a methodology to evaluate the economic convenience, by central government perspective, to switch to GC community SaaS for LEAs and other governmental bodies. In particular, the enabling conditions when the transition is convenient will be highlighted together with the costs structure. The working scenario depicted in this paper foresees the invariance of forensic activities 1-3, while 4 to 6 are evaluated for switching to GC services. Before presenting, the following assumptions represent the baseline of the work:

Assumption 1: Although private cloud is the most acceptable model to store forensic images, GC community model for critical services can also fulfill the requirements at lower cost. In fact, forensic images are for restricted access only

(investigators, judges, lawyers dealing with the case) and can contain confidential information. Hence it is assumed a security impact level evaluation (e.g. Impact Level, IL5, [21] , [22]) to establish the requirement for forensic GC services, since IL5, e.g., protects from major, long-term impairment to the ability to investigate serious crime (as defined in legislation) and protects from a number of criminal convictions to be declared unsafe or referred to appeal (e.g. through persistent and undetected compromise of an evidence-handling system);
Assumption 2: Laws establish evidence preserved on community cloud (with explicit rules) are admissible in court. Currently, the cloud model, due to its off-DFOU-premises feature represents a new stage in the chain of custody. If no rules (e.g. ACPO-like) apply at this stage, judge can assume the evidence as invalid;
Assumption 3: The network connection cost between DFOU and GC is assumed to be zero (e.g. using an already established, flat fee, available network);
Assumption 4: A common flawed assumption in designing distributed systems states that latency is zero, which is not true, in particular for cloud services. Using SaaS through the Internet can incur a substantial cost in terms of I/O latency. However, the presented model does not account for this latency;
Assumption 5:The resources are available on demand, with no loss. This assumption can be removed in presence of the queuing model and allocation policy (out of the scope of this paper, as mentioned in previous section).

5 Cost Evaluation

In section 1 we introduced the cloud advantages related to supply-side savings. (e.g. large-scale data centers lower costs per equipment), demand-side aggregation (e.g. increasing server utilization rates), multi-tenancy efficiency (e.g. as the number of tenants increases, the application management and server cost per tenant decrease). The cost evaluation of SaaS migration of a Forensic Farm is based on the evaluation of the Net Present Value (NPV) over a time interval of the TCO of both alternatives, SaaS and N DFOUs. As a rule of thumb, organizations will adopt new technology system with the minimum NPV of the TCO over a time interval. The TCO of the GC is represented by the NPV of the sum of the OPEX and CAPEX (assumed on a seven years period). The costs related to internally based IT infrastructure are capital CAPEX, defined as *the amount of money spent for investments carried out from a long-term perspective to setup assets prior to their entering into operations*. In particular, in a forensic farm CAPEX are related to:
a. Hardware, as Compute nodes (split into racks), RAM (GB), Switches, connecting the racks, Controller/technology, security appliances (when the forensic lab is connected to an external network);
b. Storage (HDs): in this paper it is assumed that the Organization buys storage when needed and the organization has to replace disks periodically due to failures, with an Annualized Replacement Rate (ARR), estimated at 3%;
c. Software Licenses (with annual updates), computed as cost per single tenant;

d. Security software equipment, as firewall, anti malware, IPS/IDS etc software, as a cost related only to GC, assuming that the DFOU labs are typically disconnected from the Internet/network, while GC is online by definition.
The Operating Costs (OPEX), defined as *recurring monthly costs of actually running the equipment*, are related to
e. Power (CPU+controller + Hard disks);
f. Human resources tasked on system/network management/administration.

Furthermore, the model does not consider invariant costs related to activities 1-3 (e.g. write blocker, forensic operator FTE, whose cost is the same on both models) as well as any costs related to new legal prescriptions (variable by countries) and CAPEX depreciation. Finally, the hardware salvage value is considered as nil, due to security prescriptions and/or legal sell-ban for hardware owned by public administrations.

6 Capex and Opex

In equation form, the simplified standard capital budgeting format for calculating a purchased assets NPV, DFOU (TCO^{DFOU}) is computed as follows:

$$TCO_{(DFOU)} = \sum_{T=0}^{N-1} \frac{C_T^{DFOU}}{(1+I_R)^T} + P^{DFOU(T)} \tag{1}$$

Where is the operating cost at year T, P_{DFOU} represents the asset purchase (capital) cost and IR is the organizations cost of capital[1]. In particular, the DFOU TCO, in terms of hardware and software related to point a), b), c), e) and f) is:

$$TCO_{(DFOU)} = \sum_{T=0}^{N-1} \frac{C_T^{DFOU}}{(1+I_R)^T} + \sum_{T=0}^{N-1} \frac{L_T^{DFOU}}{(1+I_R)^T} + P_{HW}^{DFOU} \tag{2}$$

Where L_T is the lease payment at year T for software licenses, whose updates are assumed to be billed annually. At year T, the DFOU CAPEX TCO can be further split, in particular, into 4 cost contributions:

- CPU,NAS/SAN, Storage architecture cost, P_H;
- P_R is the cost of the hardware storage upgrades/failure repairing;
- Set of forensic software licenses, L_S;
- Cost of updates of forensic software licenses, L_U.

$$P^{DFOU(T)} = \frac{P_R^{DFOU(T)}}{(1+I_R)^T} + P_H^{DFOU(T)} + L_S^{DFOU(T)} + \frac{L_U^{DFOU(T)}}{(1+I_R)^T} \tag{3}$$

In particular, the extra hardware needed and the repairing costs through the years, $P_H^{DFOU(T)}$ can be resumed in (4), where cost drivers are shown in Table 2.

[1] The interest rate of its outstanding debt used to finance the purchase.

$$P^{DFOU(T)} = ((\lceil V_T \rceil_{\Omega} - \lceil V_{(T-1)} \rceil_{\Omega}) \cdot \Omega + ARR_T) \cdot M_T \quad (4)$$

Table 2. Cost drivers due to upgrades and failures

Parameter	**Description**
Ω	Size of purchased hardware (e.g.disk drives, GBytes)
ARR_T	Annual Replacement Ratio (typically $\in$ [0,5%,13,5%], assumed 3%)
P_H^{DFOU}	Hardware cost (EUR)
$\lceil V_T \rceil_{\Omega}$	returns the minimum number of Ω-sized hw units to manage V_T
K	Current per-hardware unit price (e.g. GByte storage, EUR/GByte)
V_T	Expected hardware (e.g. storage) requirement in year T (GBytes)
M_T	Predicts the cost per unit (e.g. GByte of SATA disk storage at time T)

I.e., for storage

- considering an overall hardware 3% ARR (factor 0,03 in formula (5)), as shown for hard disks in [23] for data centers;
- according to [24] , the trend line can be approximated, using regression analysis of SATA storage prices based on Pricewatch.com, as $M_t = 1.2984 \cdot e^{0.012 \cdot T}$, where T represents the time (in days) from a fixed observation data (in this case, 2003, April 23rd). Therefore, the future disk price trend conforms to the equation $K \cdot e^{C \cdot T}$, where K represents the lowest storage price per GByte available to the consumer at $T = 0$; and T represents the number of years in the future. We therefore derive M_T as $M_T = K \cdot e^{-0.0012 \cdot 365 \cdot T}$ and, finally, $M_T = K \cdot e^{-0.438 \cdot T}$.

In the storage case, we derive the following formulas:

$$P_{H(Storage)}^{DFOU(T)} = ((\lceil V_T \rceil_{\Omega} - \lceil V_{(T-1)} \rceil_{\Omega}) \cdot \Omega + 0.03 \cdot \Omega \cdot \lceil V_T \rceil_{\Omega}) \cdot K \cdot e^{-0.438 \cdot T} \quad (5)$$

$$P_{H(Storage)}^{DFOU(T)} = (1.03 \cdot \lceil V_T \rceil_{\Omega} - \lceil V_{(T-1)} \rceil_{\Omega}) \cdot \Omega \cdot K \cdot e^{-0.438 \cdot T} \quad (6)$$

Analogously to the abovementioned example, the formula (4)) can be extended to the costs related to all the m hardware components, obtaining

$$P_H^{DFOU(T)} = \sum_{j=1}^{m} ((\lceil V_T \rceil_{\Omega} - \lceil V_{(T-1)} \rceil_{\Omega}) \cdot \Omega + ARR_T^j) \cdot M_T^j \quad (7)$$

as the overall formula taking into account the CAPEX of hardware upgrades and repairing. Hence, the formula in (3) can be upgraded to the GC case, taking into account a supplementary cost weights on all the CAPEX components related to the GC, adding the security hardware/software licenses costs needed to implement perimeter/host protection (8)).

$$P^{GC(T)} = \frac{P_R^{GC(T)}}{(1+I_R)^T} + P_H^{GC(T)} + L_S^{GC(T)} + \frac{L_U^{GC(T)}}{(1+I_R)^T} \tag{8}$$

Finally, resuming the TCO for GC

$$TCO_{(GC)} = \sum_{T=0}^{N-1} \frac{C_T^{GC}}{(1+I_R)^T} + P^{GC(T)} \tag{9}$$

6.1 Opex

The OPEX are calculated on the basis of the electric utility cost (ϵ) associated to the power consumed by hardware components (CPU cores (W_{Cores}), disk controllers (W_C), the disk units (W_D)) and the cost of a human operator to manage the system/data, as the salary quota for data administration, β and his salary H_T. These values are related to the annual increasing hardware/software need ($\lceil V_T \rceil_\Omega$) rate. Considering the parameters in Table 3, the OPEX can be modeled by the equation (10), as suggested in [24] , resuming the costs in C_T:

Table 3. A possible set of values for calculating the OPEX

Parameter	Description	Mock value
$\delta(EUR/KW)$	Cost of electric ultility	EUR 0.18
$W_i(KW)$	Hardware device power consumption	0.5
$P_D(KW)$	Hard disk power consumption	0.01
$\lceil V_T \rceil_\Omega$	Minimum number of hard disks	1
α	Workload ratio	0,1090
H_t	Annual salary (data management)	EUR 35000

$$C_T = (365 \cdot 24) \cdot \epsilon \cdot (\sum_{j=1}^{n-1} W_j + W_D \cdot \lceil V_T \rceil_\Omega) + \beta \cdot H_T \tag{10}$$

Derived by the formula in (10), the lab accessing the GC, the OPEX of DFOU have the following upper bounded OPEX (l is the number of hardware devices needed to access the GC)

$$C_T{}^{DFOU(GC)} = (365 \cdot 24) \cdot \epsilon \cdot (\sum_{j=1}^{n-1} W_j) \tag{11}$$

Hence, considering DFOU using some workstations to access GC SaaS (without incurring in further hw CAPEX), the formula (9) now is completed as

$$TCO_{GC} = \sum_{T=0}^{N-1} \frac{C_T^{GC}}{(1+I_R)^T} + P^{GC(T)} + \sum_N C^{DFOU(GC)} \tag{12}$$

With assumptions in Table 3 mock values, the NPV on 7 years, the OPEX account for 33%, while CAPEX account for 77%, as shown in Figure 1 .

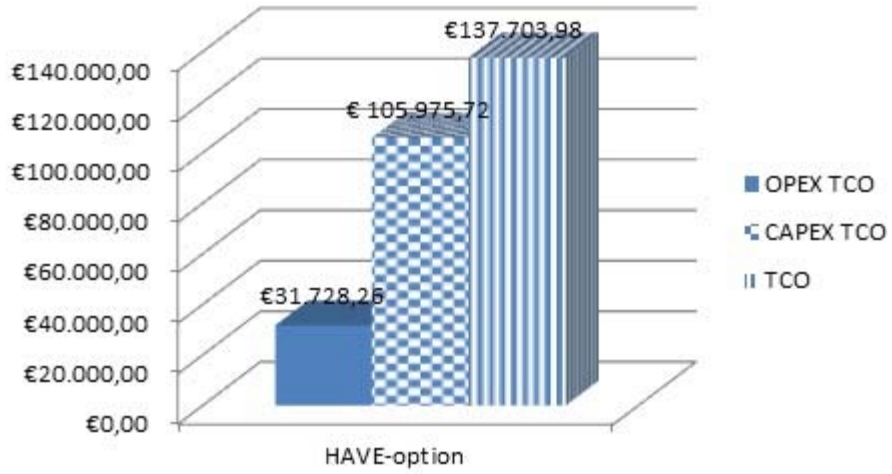

Fig. 1. TCO for a single DFOU, computed with mock values.

7 Economic Model

In 2006 Moore argued in [5] that the incentives of technology companies, LEAs and society do not always align. In particular, he cast the problem of recovering digital evidence in economic terms and how technology choices can impact the costs imposed on law enforcement, in terms of choice of proprietary and open standard formats. As a consequence of problems faced by LEAs in recovery data, the highest heterogeneous license software portfolio represents the best response. Hence, GC can represent the solution to benefit of this heterogeneity lowering the costs. In this scenario, we figure out to have N DFOUs accessing to the forensic services, namely forensic software (e.g. Encase, FTK, XRY, Cellbrite) provided by the GC. In particular, every single forensic service is bound to a maximum service capacity, represented by the related total number of software licenses available in the GC portfolio. Although the dimensioning of the GC is the target of a further paper, for the sake of methodology completeness overview, the related model can be represented as the sum of N independent Poisson processes and the related queuing model is represented by an Engset model [25] , with an allocation policy following, e.g., the non cooperative resource allocation game. The well-established related literature shows that cloud services can be modeled with learning curves, in order to determine the reduction in costs per unit as

experience grows while providing services with increasing maturity. The gained economies of scale are modeled by (13), where the marginal cost of producing the x-th unit, denote by c(x), is given by

$$c(x) = K \cdot x^{log_2 \alpha} \tag{13}$$

Where

- K is the cost to produce the first unit, As a result of the previous sections, K is represented by the TCO of the GC;
- $C(x)$, is the cost to provide the xth forensic service (digital investigation case);
- x is the number of digital forensic cases;
- α is the learning percentage (expressed as a decimal), with $\alpha \in [0, 1]$. The higher the value of α and K, the higher the production cost for the GC provider I.e., it is estimated that Amazon when running 30,000 EC2 instances, incurred in cost per instance of \$0.08/hr while their original cost per EC2 instance was estimated to be \$2.20/hr, representing an 80% learning factor. As reported in [26] , a typical cloud provider, as the total number of servers in house doubles, the marginal cost of deploying and maintaining each server decreases 10-25%, thus the learning factors are typically within the range (0.75, 0.9).

The Figure 2 shows, in general, the cost comparison between the two choices, as reported in [27], for a company, where the concave line represents the in-house cost function based on the learning curve model while the straight line corresponds to the linear cost using cloud service. The sketch shows that cloud

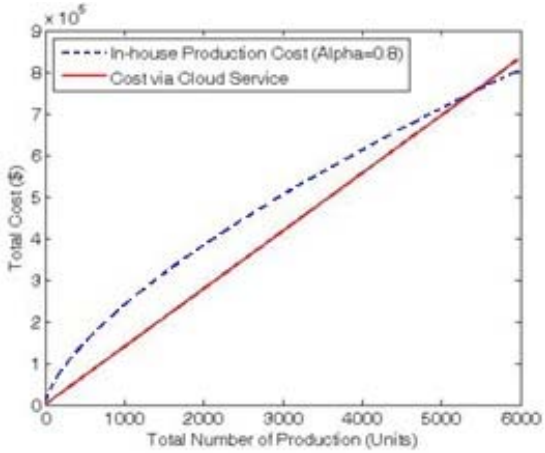

Fig. 2. Cost comparison:In-house vs Cloud Service.

service is generally attractive for small to medium businesses. The TCO shown

in the previous paragraphs applies in the learning curve costs, considering K as the TCO. The total cost can be computed by integrating the above formula, obtaining

$$\frac{K \cdot x^{1+log_2\alpha_S}}{1+log_2\alpha} \tag{14}$$

Conversely to well established literature [28] , [27] where the adoption of cloud services is related to the profits and costs of the cloud provider, it is assumed that, since the choice is between a GC and multiple governmental DFOUs, no profit should be considered for the central cloud provider. The convenience is lowering the costs less for the overall digital forensic service. Therefore, the migration to the aggregate (governmental) community cloud of the n DFOUs happens when the aggregate costs of DFOUs overcome the cost of the community GC service. This is explained by the following formula:

$$\sum_{j=1}^{n} \frac{K_j \cdot x^{1+log_2\alpha_j}}{1+log_2\alpha_j} > \frac{K_S \cdot x^{1+log_2\alpha_S}}{1+log_2\alpha_S} \tag{15}$$

Assuming, for the sake of simplicity, that $\alpha_j = \alpha_F$ for all the DFOUs, the formula (15) is

$$\sum_{j=1}^{n} K_j \cdot (\frac{x^{1+log_2\alpha_F}}{1+log_2\alpha_F}) > \frac{K_S \cdot x^{1+log_2\alpha_S}}{1+log_2\alpha_S} \tag{16}$$

which is solved by

$$x > (\frac{(1+log_2\alpha_F) \cdot K_S}{(1+log_2\alpha_S) \cdot \sum_{j=1}^{n} K_j})^{log\frac{\alpha_F}{\alpha_S{}^2}} \tag{17}$$

Posing $K_F = \sum_{j=1}^{n} K_j$ and $\alpha = \frac{\alpha_F}{\alpha_S}$

$$x > (\frac{(1+log_2\alpha_F) \cdot K_S}{(1+log_2\alpha_S) \cdot K_F})^{log_\alpha 2} \tag{18}$$

In the formula (18), the cost of providing the first unit of service (here, the first digital forensic investigation case) in both cases is represented by K_S and K_F. In particular, assuming $K_F = TCO^{DFOU}$ and $K_S = TCO^{GC}$ for $T = [0, \ldots, n]$ years (which generalizes the cost of the first unit) enables the simulation of the years needed to return of the investment of migration with respect to the overall number of investigation cases. The related Figure 3 depicts the behavior of convenience with respect to the ratio of TCO ($K = \frac{K_S}{K_F}$) in both cases: in fact, when the number of cases is greater than right-end member it is convenient to migrate. The Figure 3 further shows how the formula in (18) can help the discussion: assuming $\alpha_F = 0.9$ and $\alpha_S = 0.75$ in fact, as the ratio of TCOs increases, the needed cases/per year increase exponentially in order to have convenience. Since assumed α values are likely for the two matching cases, the curve in Figure 3 shows that, i.e., when the cost of the community cloud is

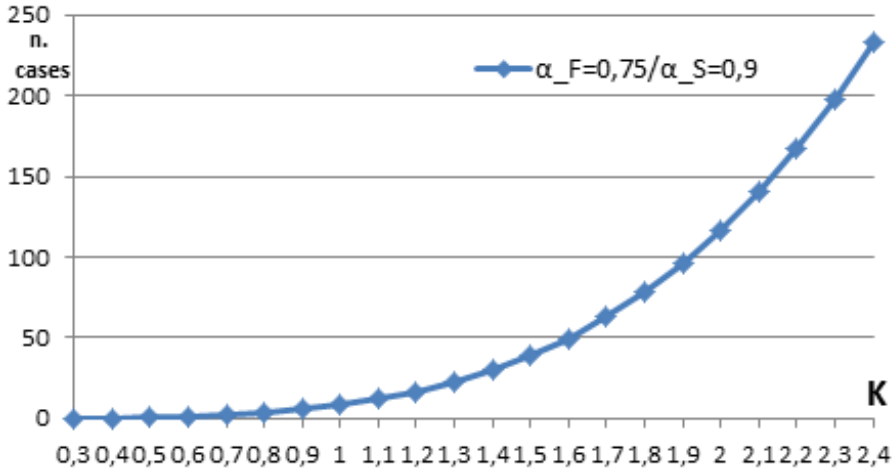

Fig. 3. Number of cases vs TCO ratio

double of the sum of all the DFOU TCO ($K = 2$), migrating to community GC is convenient for a number of cases greater than 117.

8 Conclusions and Future Works

The evolution of technological activities together with the increased volume of cases expected in the future make digital forensic activity squarely sitting on the G.Reese requirements for cloud convenience: if your usage patterns are relatively static, your savings in the cloud will be nominal (or potentially even nonexistent). On the other hand, if your consumption varies widely above or below planned capacity on an ongoing basis, your savings will be much higher. Hence, due to the availability of GC services and the need to rationalize the spending for LEAs, improving the quality, it is important to assess costs and evaluate alternatives. This paper has proposed a methodology to fulfill these requirements, depicting a scenario where the single DFOUs can benefit in terms of costs reduction and quality deriving from the migration to GC. The security requirement, due to the migration increased value for network attack, has been taken into account as a supplementary cost weighting on GC setup costs. Nonetheless, the results presented in this paper have to be completed, in future works, with the system dimensioning based on a queue model of the system and the allocation policy. In particular, the former describes the system behavior, following the dynamics of a continuous-time birth-death Markov process (Markov chain). This lossy system, where forensic operators enter the system if at least one of the software licenses is free, will be modeled assuming exponential inter-arrival time and service time, K servers, no queuing. The latter relies on the GC provider to supply their computing needs, requiring specific QoS to be maintained in order to meet the service requirements. The game theory approaches the problem where players

want to maximize their returns which depend also on actions of other players. Vast literature (e.g. [29] , [30] , [31] where has been shown that Nash equilibrium always exists if the resource allocation game has feasible solutions) proposed game-theoretic method to solve the optimization problem of resource allocation from the viewpoint of the GC. Future works will complement this paper in these above-mentioned directions.

References

1. Crane, B.: Digital Forensics: A Decade of Changes POLICE Magazine, 11 11 2013. Available: http://www.policemag.com/blog/technology/story/2013/11/digital-forensics-a-decade-of-changes.aspx. [Last access: 26 2 2014].
2. Forbes: 8 Hot Industries for Startups in 2013 Available: http://www.forbes.com/pictures/fghj45fjl/4-digital-forensic-services/. [Last access: 26 2 2014].
3. Garfinkel, S.: Digital forensics research: The next 10 years Digital Investigation , n. 7, pp. 64-73, 2010.
4. Beebe, N.: Digital forensics research: the good, the bad, and the unaddressed Proceedings of the Fifth annual IFIP WG 11.9 international conference on digital forensics, 2009.
5. Moore, T.: The Economics of Digital Forensics Proceedings of Workshop on Economics of Information Security, 2006.
6. Cramm, S.: The Truths about IT Costs Harvard Business Review, vol. 87, n. 2, p. 28, 2009.
7. Carr, N.: The end of corporate computing Sloan Management Review, vol. 46, n. 3, pp. 67-73, 2005.
8. Berstis, V.: Fundamentals of grid computing IBM Redbooks Paper, 12 November 2002. Available: http://www.redbooks.ibm.com/redpapers/pdfs/redp3613.pdf. [Last access: 25 2 2014].
9. Dharanibalan,G.: Want to move from capex to opex? Think cloud. The Economic Times, 2013 2 2013. Available: http://articles.economictimes.indiatimes.com/2013-02-27/news/37330988_1_cloud-service-providers-consumers-capital-expenditure. [Last access: 25 2 2014].
10. Choudhary, V. : Comparison of software quality under perpetual licensing and software as a service. Journal of Management Information Systems , vol. 2, n. 24, p. 141165, 2007.
11. Kundra, V. : Federal Cloud Computing Strategy, 8 2 2011. Available: https://www.dhs.gov/sites/default/files/publications/digital-strategy/federal-cloud-computing-strategy.pdf. [Last access: 25 2 2014].
12. Federal Risk and Authorization Management Program (FedRAMP) Available: www.fedramp.gov. [Last access: 25 2 2014].
13. European Commission: Digital Agenda for Europe Available: http://ec.europa.eu/digital-agenda/ [Last access: 25 2 2014].
14. European Commission: Unleashing the Potential of Cloud Computing in Europe, 27/9/2012). Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0529:FIN:EN:PDF. [Last access: 24 2 2014].

15. ENISA: Good Practice Guide for securely deploying Governmental Clouds ,2013. Available: http://www.enisa.europa.eu/activities/Resilience-and-CIIP/cloud-computing/good-practice-guide-for-securely-deploying-governmental-clouds. [Last access: 24 2 2014].
16. Mell, P., Grance, T.: The NIST Definition of Cloud Computing NIST-Special Publication 800-45, 2011.
17. August, T.,Niculescu, M.F., Shin, H.: Cloud Implications on Software Network Structure and Security Risks Proceeding of WEIS 2013, September 6, 2013..
18. Scientific Working Group on Digital Evidence (SWGDE): Digital evidence standards, principles and best practices Available: https://www.swgde.org/documents/Current
19. Carrier, B., Spafford, E.: Getting physical with the digital investigation process International Journal of Digital Evidence, 2003.
20. Reese, G.: Cloud Application Architectures: Building Applications and Infrastructure in the Cloud, O'Reilly Media; 1 edition , 2009.
21. The National Technical Authority for Information Assurance: Extract from HMG IA Standard No.1, 10 2009. Available: http://www.cesg.gov.uk/publications/Documents/business_impact_tables.pdf. [Last access: 27 2 2014].
22. Gawen, E.: So what is IL3? A short guide to business impact levels HM Government-G-Cloud, 9 3 2012. Available: http://gcloud.civilservice.gov.uk/2012/03/09/so-what-is-il3-a-short-guide-to-business-impact-levels/. [Last access: 27 2 2014].
23. Schroeder, B., Gibson, G.: Disk Failures in the Real World:What Does an MTTF of 1,000,000 Hours Mean to You? Proceedings of 5th Usenix Conf. File and Storage Technologies (FAST 07), 2007.
24. Walker, E., Brisken, W., Romney, J.: To lease or not to lease from storage clouds IEEE Computer, vol. 43, n. 14, pp. 44-50, 2010.
25. Brandwacht, L.: Master thesis:dimensioning the Cloud, 11/1/2013. Available: www.math.leidenuniv.nl/scripties/MasterBrandwacht.pdf. [Last access: 26 2 2014].
26. Amit, G., Xia, C. H.: Learning curves and stochastic models for pricing and provisioning cloud computing Service Science, vol. 3, p. 99109, 2011.
27. Phillips, R. L. : Pricing and Revenue Optimization Stanford University Press, 2005.
28. B., Vasilakos, A., Lesser,V.: Evolutionary stable resource pricing strategies Proceeedings of ACM SIGCOMM, August 1721, 2009, Barcelona, Spain, 2009.
29. Doulamis, N., Doulamis, A., Litke, A., Panagakis, A.: Adjusted fair scheduling and non-linear workload prediction for QoS guarantees in grid computing Computer Communications, vol. 30, n. 3, pp. 499-515, 2007.
30. Wei, G., Vasilakos, A. V., Zheng, Y. ,Xiong, N.: A game-theoretic method of fair resource allocation Journal of Supercomputing, vol. 54, p. 252269, 2010.
31. Mendel, T., Takahashi, S.: Enterprise IT budget outlook: Europe. Forrester Research, 2007. Available: http://www.forrester.com/Research/Document/Excerpt/0,7211,41668,00.html.