2. Cloud computing and cloud storage

2.1. Cloud computing definition Cloud computing arises from the combination of the traditional computer technology and network technology, such as grid computing, distributed computing, parallel computing, utility computing, virtualization. One of the core concept of cloud computing is reducing the processing burden on user’s terminals through continuously enhancing the clouds’ handling capacity. Eventually user’s terminals are simplified into a simple input and output devices. Users can use the powerful computing and processing function on clouds and they can order their service from the cloud according to their own needs.

2.2. Cloud storage definition and it’s architecture Cloud storage is a system that provides functions such as data storage and business access. It assembles a large number of different types of storage devices through the application software which are based on the functions of the cluster applications, grid techniques, distributed file systems, etc. Cloud storage can be simply understood as the storage in cloud computing, and also can be considered to be a cloud computing system equipped with large capacity storage. Cloud storage system architecture mainly includes storage layer, basic management layer, application interface layer and access layer .

How it works

Firstly, clients compute chunk index from files structure and chunk size, then send file name and chunk index to master (mark①). Secondly, master sends chunk handle and chunk locations to clients (mark②). Thirdly, clients send chunk handle and byte range to the nearest chunkserver (mark⑤). Finally chunkserver sends data to client (mark⑥).Once clients get chunk locations from master, clients do not interact with master any more. Master does not permanently save the mapping from chunkserver to chunk. Instead, it asks each chunkserver about its chunks at master startup or whenever a chunkserver joins the cluster (③④). The master periodically communicates with each chunkserver in HeartBeat message to give it instructions and collect its state (③④).

In <http://www.sciencedirect.com/science/article/pii/S1877705811065192>

<https://www.clickssl.net/blog/top-8-cloud-computing-threats-and-its-security-solutions>

<https://www.isaca.org/Journal/archives/2015/Volume-4/Pages/security-and-privacy-challenges-of-iot-enabled-solutions.aspx?utm_referrer>=

Definição de Cloud:

Cloud computing nasce da combinação da computação tradicional com a rede. A base deste modelo é a intenção de reduzir o processamento nos terminais tradicionais, passando este a ser feito na Cloud. Deste modo, os utilizadores têm à sua disposição grande poder computacional sem a necessidade de terem que comprar e manter um sistema poderoso.

Naturalmente, da Cloud Computing deriva a Cloud Storage, que permitirá o acesso e utilização dos dados, aplicando a mesma lógica base da Cloud Computing. Assim, um grande número de dispositivos de armazenamento poderão estar conectados entre si, funcionando como um cluster, em que os dados estarão distribuídos.

Principais vantagens de uma Cloud:

On-demand self-service: cada utilizador tem à sua disposição a quantidade exata de poder computacional (uso de servidores ou de armazenamento) que necessita, sem ter que dar algo de volta ao cloud provider.

Broad network access: os utilizadores não têm que se preocupar com as plataformas e dispositivos com que estão a transmitir os dados, uma vez que o “canal de comunicação” será a Internet com um mecanismo standard definido pela cloud. Esta capacidade permite o uso de vários dispositivos IoT que, não partilhando os mesmos protocolos e interfaces, conseguem comunicar.

Resource pooling: capacidade de partilhar recursos e atribuir outros a cada um dos vários clientes, dependendo da sua necessidade atual, dinamicamente, como memória, largura de banda, etc.

Rapid elasticity: facilidade de escalar a rede de processamento e armazenamento, o que se traduz numa virtual infinitude de capacidade, para o utilizador.

Measured servisse: controlo de uso do serviço, o que permite um pagamento adequado às necessidades do utilizador mas também a recolha de dados de uso para o prestador de serviço.

Principais desafios à construção de uma Cloud confiável:

Outsourcing: Perda de controlo da posse dos dados. A Cloud fornecedora de serviço deve ser confiável e providenciar toda a segurança possível, no processamento e armazenamento de dados. A confidencialidade, integridade e não violação dos dados deve ser garantida.

Multi-tenancy: Como plataforma partilhada, a plataforma será partilhada com vários utilizadores, o que pode levar à colocação de dados no mesmo local físico que outros clientes, incorrendo no risco de um programa malicioso tentar manipular esses dados. É necessário, por isso, criar estratégias de proteção de partilha de dados.

Massive data and intense computation: Os mecanismos de segurança normais não são suficientes quando tamanha quantidade de dados e computação está disponível. É preciso, por isso, criar novos protocolos e mecanismos de proteção.

Ecossistema de segurança e privacidade de uma Cloud:

Confidencialidade: Dados estão disponíveis para o Cloud provider e só um sentido de responsabilidade protege o consumidor.

Integridade

Disponibilidade

Responsabilidade

Preservação de privacidade – privacidade está diferenciado da segurança porque este é um grande tema ao redor da Cloud e um dos seus maiores problemas

Vulnerabilidades de uma Cloud:

Co-residence: Partilha de uma mesma estrutura física por diferentes aplicações. Pode haver Cross-VM attack e Malicious SysAdmin.

Loss of Physical Control: Os dados e serviços do cliente já não mais estão ao seu alcance, não podendo ser diretamente evitadas manipulação de dados ou software.

Bandwidth Under-provisioning: Ataques DOS também se realizam em Clouds, sendo que a capacidade real da rede é bastante inferior à capacidade de todos os serviços da mesma subrede agregados.

Cloud Pricing Model: Ataques que promovem a modificação do modelo de imposição de preços, de modo a que um utilizador/serviço pague realmente menos do que era suposto.

Cloud Integrity (foco de abordagem)

Integridade na Cloud implica que não deve acontecer manipulação de dados ou programas, quer por malware ou utilizadores nocivos e, caso esta aconteça, deve ser detetada.

**Ameaças:**

Data loss / Manipulation: No armazenamento em Cloud são armazenados dados e, como é natural aos servidores existe um risco de perda ou modificação de dados, quer intencionalmente, quer acidentalmente. Erros de administração (restores, data migration, etc) ou ataques propositados para causar a perda de controlo de dados.

Dishonest computation in remote servers: Podem acontecer de na computação em Cloud, como os utilizadores não vêem o processamento, serem devolvidos valores/dados errados para um programa ou pedido.

**Defesas:**

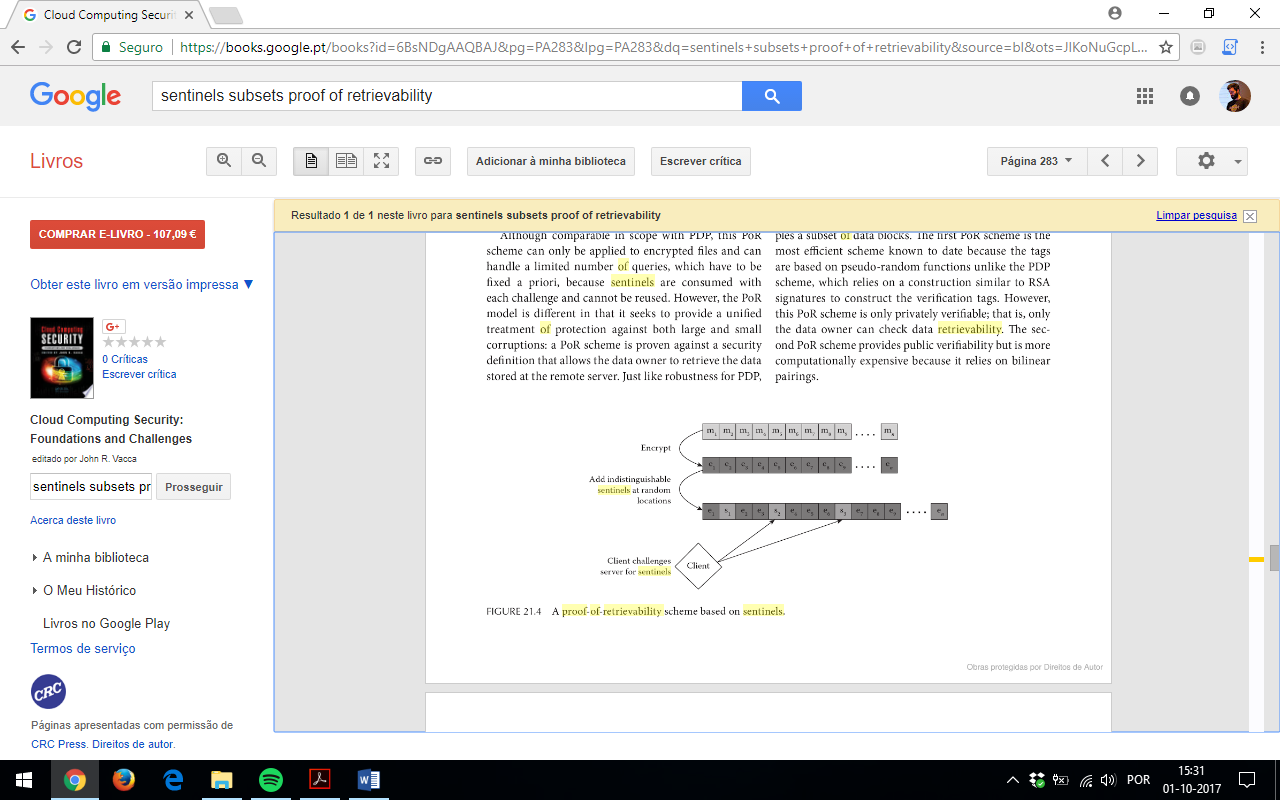
Provable Data Possession: Num ambiente de Cloud não é possível fazer o mesmo tipo de verificação que num tradicional, pois o custo de operações tradicionais seria muito grande em termos de computação (pois os dados armazenados são também eles muito grandes) e largura de banda. Deste modo, torna-se obrigatório falar de Naive Method, Original Provable Data Possession e Proof of Retrievability.

O MEU TEMA

POR – Proof of Retrievability:

(inserir imagem de pag 6 do T10-062…)

Atua de modo similar ao PDP. No caso, o POR tenta diminuir a quantidade de processamento e dados envolvidos, pela geração de uma key que codificará um determinado ficheiro F. De seguida, este ficheiro já encodificado será enviado para guardar na Cloud. São também adicionadas sentinelas, que são blocos indistinguíveis que, mais tarde, serão usadas para o controlo. O protocolo POR é ainda capaz de recuperar ficheiros pouco corrompidos. Só aplicável a ficheiros estáveis.



-- fim do uso do T10-062..

-- inicio do uso do literature\_survey…

O POR não obriga à manutenção da totalidade do ficheiro mas requere que, quando necessário, seja devolvida a informação que é pedida. Pode, para além da codificação do ficheiro original, ser mantida parte do ficheiro na máquina local. A grande vantagem do POR em relação ao PDP é que não necessita da devolução do ficheiro por inteiro, o que no caso de ficheiros muitos grandes é significativo.

Como não é a totalidade do ficheiro verificada, esta verificação é probabilística.

Quando se deseja efetuar a verificação, o utilizador pede à Cloud um conjunto de posições, que sabe serem sentinelas, e compara os valores devolvidos com aqueles que conhece. Caso tenham havido alterações, a sentinela terá sido também manipulada. Como pode acontecer de haverem mudanças que não tenham afetado as sentinelas, o protocolo também possui códigos de correção de erros, corrigindo ficheiros corrompidos.

Ao POR, para além da redução do esforço computacional da Cloud, também se garante encriptação de dados. O downside desta encriptação é que é exigido ao computador do utilizador um esforço computacional que não é exigido no PDP.

<http://www.arijuels.com/wp-content/uploads/2013/09/BJO09b.pdf>

http://www.arijuels.com/wp-content/uploads/2013/09/JK07.pdf

Explicando um pouco melhor o sistema de sentinelas:

Existe uma fase de precomputação em que são computados M e Q valores. M são os valores das sentinelas e Q o valor encriptado do correspondente M.

De seguida, existem dois cenários de perda de dados. No primeiro, um pequeno número de modificações foi feita e pode ser corrigida com os mecanismos de correção abordados anteriormente. No segundo, demasiadas alterações foram efetuadas e já não é possível corrigir, mas são detetadas essas alterações por falta de compatibilidade com as sentinelas especificadas. Um dos problemas do POR é que, como é probabilístico, não consegue determinar quantos bits são necessários serem modificados para que seja corrigido um erro. No extremo, pode toda a informação estar errada e apenas as sentinelas estarem de acordo com a especificação inicial, o que leva a uma insegurança no método. Este método tem também a particularidade de só poder usar as sentinelas uma vez, pois quando pede os resultados Q à Cloud, então é determinado que estes blocos são sentinelas e não o ficheiro em si.