Relação custo/benefício de técnicas utilizadas para prover privacidade em computação nas nuvens

Vitor Hugo Galhardo Moia[1], **Marco Aurélio Amaral Henriques**[1]

[1]Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas (Unicamp)
Campinas, SP, Brasil 13083–852

{vhgmoia,marco}@dca.fee.unicamp.br

Resumo. *Com a crescente utilização da computação nas nuvens, devido a suas atraentes características, surge uma grande preocupação com a segurança e privacidade nesse meio. Possíveis ataques realizados pelos próprios provedores de serviço ou por terceiros tornam os usuários relutantes na utilização desta tecnologia. Em meio a tais problemas, este trabalho apresenta uma discussão sobre os principais problemas e soluções relativos à privacidade nas nuvens e faz uma comparação preliminar destes com base em estimativas dos custos e do grau de privacidade de cada um.*

1. Introdução

A computação nas nuvens é uma boa opção para o armazenamento de dados de forma escalável e dinâmica. Com isso, evitam-se custos com infraestrutura e manutenção locais, além de ser possível pagar conforme a utilização. Outra vantagem é a disponibilidade dos dados a partir de qualquer ponto na internet. Contudo, um fator que impede uma maior adoção dessa tecnologia é a preocupação quanto à segurança e privacidade. Um potencial acesso a informações sensíveis pelos provedores é um grande problema, pois estes têm fácil acesso aos dados. Além disso, há a possibilidade de falhas na segurança da infraestrutura que permitam a invasores se apoderar de dados armazenados.

Com o objetivo de minimizar riscos na segurança foram propostas na literatura várias técnicas. Neste trabalho são discutidos os principais problemas relativos ao sigilo de dados na nuvem e alguns métodos para resolvê-los. Em seguida são feitas estimativas dos custos e benefícios de cada método a fim de permitir uma comparação preliminar entre os mesmos.

2. Problemas relacionados a privacidade na nuvem

Geralmente, os usuários carregam seus dados na nuvem de forma clara e confiam em seus provedores de serviço quanto à segurança de suas informações. Para não precisar desse tipo de confiança e aumentar a proteção dos dados armazenados, foram propostas na literatura várias formas para se aumentar o sigilo dos mesmos. Nesta seção, será realizada uma discussão sobre os principais problemas de privacidade na nuvem, apresentando algumas técnicas utilizadas para saná-los.

2.1. Sigilo do conteúdo

Consiste no acesso controlado aos dados armazenados na nuvem, os quais só podem ser acessíveis ao(s) seu(s) dono(s). Havendo este controle, os usuários não terão que confiar

cegamente em seus provedores e suas informações ficarão mais protegidas contra atacantes externos. Várias técnicas foram desenvolvidas para essa solução, sendo as duas principais a criptografia e o armazenamento distribuído dos dados particionados em larga escala (fragmentação). A primeira garante o sigilo dos dados por meio de codificação baseada em um segredo, porém traz alguns problemas como o menor desempenho em geral (há tempos extras para cifrar e decifrar) e a dificuldade de compartilhar, buscar e indexar dados cifrados. Há várias maneiras de se utilizá-la, sendo a mais comum cifrar um dado antes de enviá-lo para a nuvem e decifrá-lo apenas após o usuário tê-lo trazido de volta para seu sistema local [Padmaja and Koduru 2013]. Já a fragmentação dos dados pode utilizar técnicas como *erasure code* e *secret sharing* para prover alguma redundância entre diversos provedores de armazenamento. Assim, para recuperar o dado original é necessário buscar apenas parte desses fragmentos, acarretando na não dependência de um determinado provedor de armazenamento e em uma maior disponibilidade dos dados [Abu-Libdeh et al. 2010]. Também é possível combinar estas duas técnicas, mas permanecendo as dificuldades de compartilhar, indexar e buscar dados armazenados na nuvem [Schnjakin et al. 2011].

2.2. Sigilo do nome dos dados

Um problema pode ocorrer caso o nome do dado ou sua extensão contenha alguma informação relativa ao seu conteúdo. Para minimizar tal problema, o nome e a extensão do arquivo devem ser alterados, tomando cuidado ao renomear arquivos fragmentados, pois estes podem ser renomeados contendo nomes como dados-parte1, dados-parte2 etc., revelando não só o nome do arquivo, mas também que foi dividido. Uma das possíveis soluções para resolver esse problema é através de técnicas criptográficas. Assim, o nome do arquivo seria cifrado com uma chave secreta e isso evitaria que o nome revelasse algo sobre o arquivo e até mesmo o problema de se saber que o arquivo foi particionado. A chave secreta pode ou não ser a mesma usada na criptografia dos dados, dependendo dos procedimentos de gerência de chaves adotados.

2.3. Sigilo de localização durante acesso aos dados

Outro problema diz respeito à proteção do usuário quanto à revelação de sua localização durante o acesso aos seus dados. Há vários recursos para esse fim, como o Tor e o VPN-Proxy. Essas tecnologias têm como finalidade proteger o usuário contra vigilância e ataques. O TOR realiza o roteamento de uma mensagem através de várias máquinas aleatórias e utiliza várias camadas de encriptação para proteger o usuário e sua comunicação [Murdoch and Danezis 2005]. Já o VPN combinado com proxy cria um canal de comunicação criptografado entre dois pontos: o computador do usuário e um serviço de VPN-Proxy (como o VPNBook.com) que oculta do provedor da nuvem a localização do usuário [Duffield et al. 2005]. A utilização dessas duas tecnologias, ou outras que possuam características semelhantes, é necessária para a minimização deste problema de privacidade e obtenção do anonimato.

2.4. Sigilo sobre a posse de um dado

Se houver facilidade de um provedor ligar os dados armazenados na nuvem a seus respectivos donos, teremos outro problema de falta de privacidade. Uma das maneiras de se resolver este problema é com a adoção de identidades federadas, onde se tem provedores

de identidade separados de provedores de serviços. Os provedores de identidade, além de autenticar o usuário, têm a função de guardar os atributos dos usuários e repassá-los para os provedores de serviços de forma controlada. Há situações em que o provedor de identidade não precisa repassar nenhum atributo do usuário que possibilite ao provedor de serviços descobrir a identidade verdadeira do usuário. Além disso, para situações em que se quer uma maior privacidade, o provedor de identidade pode ser configurado para não armazenar informações sobre os acessos de seus usuários.

3. Soluções para a privacidade

Nesta seção são discutidas e comparadas técnicas para se prover algum tipo de privacidade na nuvem. Para cada uma delas é estabelecido pelos autores um Custo Relativo (CR) de acordo com uma estimativa subjetiva dos custos de tempo e espaço em relação às outras (Tabela 1). É considerado que a infraestrutura necessária para cada uma já esteja instalada e pronta para ser utilizada não sendo necessário contabilizar seu custo. Também é estabelecido um grau subjetivo de Privacidade Relativa (PR) de acordo com o nível estimado de privacidade obtido com cada técnica em relação às outras. A relação custo/benefício (R) é obtida pela divisão de CR por PR.

Tabela 1. Custo e privacidade relativos das técnicas utilizadas para prover privacidade na nuvem

Sigla	Técnica	CR	PR
T_5	Criptografar conteúdo	04	3,5
T_4	Fragmentar com redundância	10	2,5
T_3	Ocultar acesso via TOR	20	1,5
T_2	Ocultar acesso via VPN-Proxy	08	1,0
T_1	Ocultar posse dos dados (identidade federada)	02	1,5
T_0	Ocultar nome dos dados (criptografia de nomes)	01	1,0

Algumas técnicas como T_2, T_3, T_4 e T_5 dependem do tamanho do arquivo para se calcular o tempo e espaço requeridos e, por essa razão, foi considerado um arquivo de mesmo tamanho para a estimativa do peso de todas elas. A técnica de ocultar acesso está sendo considerada de duas maneiras, de acordo com a tecnologia utilizada.

Para a identificação das possíveis abordagens, é proposta uma forma para se criar nomes baseados nas combinações das técnicas. Há 48 possibilidades incluindo o uso de nenhuma técnica. Para ocultar a origem do acesso, as duas formas, TOR e VPN-Proxy, não podem estar na mesma combinação, e cada combinação usa um modelo de código binário como nome, sendo que 0 indica a não utilização de determinada técnica e 1 indica a sua utilização. Cada técnica ocupa uma posição em uma palavra de 6 bits, cuja ordem é T_5, T_4, T_3, T_2, T_1, T_0 e o valor decimal resultante é a representação da combinação. Considera-se que a PR_i e a CR_i de uma combinação i de técnicas são dadas pelas somas das respectivas PR e CR de cada técnica.

A Fig. 1 mostra todas as soluções possíveis com seus respectivos custos e privacidades relativas ($CR_i \times PR_i$), e também apresenta o valor de $R_i = CR_i/PR_i$ de cada combinação i na forma do tamanho dos círculos na figura. Nota-se que as combinações mais interessantes são as localizadas no quadrante inferior direito (os números indicam as combinações).

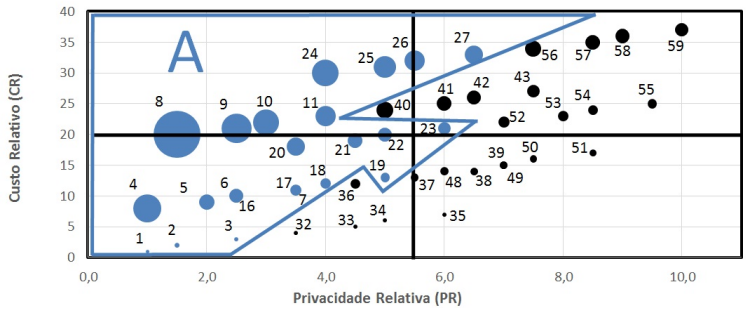

Figura 1. Custo Relativo e Privacidade Relativa para diferentes combinações

As técnicas T_0, T_1, T_3, T_4, e T_5 causam individualmente uma variação em R_i (i = 59) de +8%, +11%, -46%, -3% e +37%, respectivamente, quando são removidas desta configuração completa que usa todas as técnicas (exceto T_2). Logo, conclui-se que a adoção de técnicas como T_3 e T_4 deve ser feita com muito cuidado, pois os custos podem não compensar os benefícios. O mesmo padrão pode ser visto quando se usa T_2 no lugar de T_3, alertando para o alto custo/benefício destas duas técnicas em relação às demais.

A Fig. 1 também mostra o impacto da adoção ou não de T_5 em R. A Técnica T_5 está inativa na zona A e ativa fora dela (o bit relativo a T_5 é 0 nas combinações de 1 a 31). A inclusão de T_5 contribui significativamente para redução da razão R. Este tipo de comportamento deve ser considerado na escolha desta e de outras técnicas.

4. Conclusão

Neste trabalho foram discutidos alguns problemas relativos à privacidade nas nuvens e apresentadas algumas técnicas para amenizá-los. Uma análise dos níveis de privacidade e de seus custos relativos mostrou que algumas combinações de técnicas são bem mais eficazes em termos da relação custo/benefício que outras. Como trabalho futuro é preciso tentar obter custos absolutos mais precisos de forma a se ter uma visão mais completa que permita uma melhor tomada de decisão sobre que técnicas utilizar. Além disso, é preciso investigar formas eficientes de se combinar 2 ou mais das técnicas aqui descritas para se obter um maior grau de privacidade.

Referências

Abu-Libdeh, H., Princehouse, L., and Weatherspoon, H. (2010). Racs: A case for cloud storage diversity. In *Proc. of ACM*, SoCC '10, pages 229–240, New York, NY, USA. ACM.

Duffield, N., Greenberg, A., Goyal, P., Mishra, P., Ramakrishnan, K., and Van der Merwe, J. (2005). Virtual private network. US Patent 6,912,232.

Murdoch, S. and Danezis, G. (2005). Low-cost traffic analysis of tor. In *Security and Privacy, 2005 IEEE Symposium on*, pages 183–195.

Padmaja, N. and Koduru, P. (2013). Providing data security in cloud computing using public key cryptography. *IJESR*, 4(01).

Schnjakin, M., Alnemr, R., and Meinel, C. (2011). A security and high-availability layer for cloud storage. In *Proc. of*, WISS'10, pages 449–462, Berlin, Heidelberg. Springer-Verlag.

