A NASUNI WHITE PAPER

Understanding Security in Cloud Storage

Table of Contents

nasuni

Abstract

Many businesses that could already be benefiting from cloud storage are holding back due to security concerns. The cloud offers the opportunity to significantly reduce cost while providing reliable, on-demand storage. While no environment is completely secure, cloud storage does not inherently increase security risks to your business. Much of the fear stems from a lack of understanding around cloud security and, simply, peoples' natural reaction to any new technology. Our aim is to identify the major risks and to explore the security models that can be used to mitigate them.

We explore security in the context of the different types of cloud offerings available: Software as a Service, cloud compute and cloud storage. Data encryption at rest where only the user can decrypt the data is singled out as a unique capability of cloud storage, which enables a superior security model. The OpenPGP security standard is discussed at length in the context of data encryption.

Nasuni makes the Nasuni Filer: a virtual NAS that offers secure connections to major cloud storage providers. We attempt to offer a neutral view on all of the security issues; however, when building our system we chose a specific security model that we felt was the best fit for the problems at hand. At the heart of Nasuni security model is OpenPGP. We offer an in-depth description of our implementation of OpenPGP in the Nasuni Filer.

We explore security in the context of the different types of cloud offerings available: Software as a Service, cloud compute and cloud storage. Data encryption at rest where only the user can decrypt the data is singled out as a unique capability of cloud storage, which enables a superior security model.

The Cloud

Cloud is the new black. With so much buzz around cloud, it is hard to distinguish the meaningful and relevant parts to business customers. Cloud has become synonymous with anything that runs on the Web. Generally speaking for an offering to be considered cloud it must be available over the Internet, and capable of supporting large numbers of users simultaneously without significant changes to its architecture.

The cloud promises radical simplifications and cost savings in IT. Leveraging their technical expertise and economies of scale, several technology powerhouses, including Amazon, Rackspace, Google and Microsoft, have deployed a wide

array of cloud offerings. Yet, for all the potential benefits, there are also new risks that need to be understood. In order to reach the cloud, data moves over public networks where it needs to be protected. Cloud providers achieve economies of scale by building large data centers and then sharing resources among all of their customers. Because the cloud is a shared environment, the providers need to ensure that customers are identified properly and that no customer has the ability to see or change another customer's data.

Yet, for all the potential benefits, there are also new risks that need to be understood. In order to reach the cloud, data moves over public networks where it needs to be protected.

When it comes to security, it is useful to differentiate among the different cloud systems: Software as a Service, cloud compute and cloud storage. Each system poses its own set of benefits and security issues.

Software as a service (SaaS), represented by applications like Salesforce.com[1], Google Docs, Quickbooks Online and others, involves full software applications that run as a service in the cloud. Tens of thousands of companies share the common infrastructure of Salesforce.com. These companies maintain control of sensitive customer information through a combination of secure credentials and secure connections to Salesforce.com. Companies that use Salesforce tolerate the risk of their data not being encrypted at the Salesforce.com servers. This is not as a result of lax security on behalf Salesforce.com. Because SaaS runs in the cloud, the data from customers must be visible to the applications in the cloud (either not encrypted or decryptable by the SaaS code). The main benefit of SaaS is to reduce the complexity of having to configure and maintain software in-house. The success of Salesforce.com and others demonstrates that many companies have traded security concerns for the sheer utility and cost savings of not having to run their own software in-house.

Cloud compute allows customers to run their own applications in the cloud. Amazon's Elastic Compute Cloud or EC2 represents this type of system. Customers upload their applications and data to the cloud where the vast compute resources of EC2 can be applied to the data. Virtualization provides a practical vehicle to transfer compute environments and share physical compute resources in the

[1] Data in Salesforce.com or any other SaaS may actually be encrypted at some point during its life cycle. This would prevent saleforce.com employees from accidentally or maliciously taking customer data from servers that are not currently in use. However, for SaaS applications to run, they still need to have the ability to decrypt the customer data. That is, the keys must remain with the SaaS provider. That is the critical issue that weakens the SaaS security model. We actually reviewed the Saleforce.com security policies (http://www.salesforce.com/assets/pdf/datasheets/security.pdf) but could not find out if they encrypt data when it is not in use.

cloud. This approach has been used successfully by financial institutions and the life sciences to solve heavy compute models. It is expensive to run data centers full of servers ready to run complex mathematical models. The idea of sharing a compute infrastructure with other customers makes good economic sense. In a compute cloud the data can be anonymized, however it cannot currently be encrypted. That is, it is possible to obfuscate the data in such a way that is difficult for anyone to see what the data means; however in order to have a computer in the cloud operate on a data set, with today's technology, the data set must be visible to that computer (i.e. not encrypted).

Modern encryption protects data at rest so thoroughly that with proper key management, there is no significant difference between storing sensitive data in your data center or in the cloud.

Cloud storage allows customers to move the bulk of their data to the cloud. Microsoft's Windows Azure storage services and Amazon's Simple Storage Service (S3) are good examples. The initial services involved online backups and file archiving but the most recent wave of cloud storage gateways has extended the use of the cloud to all types of storage. Data growth continues to be a major source of pain and cost to businesses and cloud storage offers unlimited, on demand, reliable storage at a fraction of the cost of traditional storage.

Cloud storage is in many ways more basic than SaaS or cloud compute because the provider's main responsibility is to store the data unchanged. This brings cloud storage closer to a traditional utility model, like electricity or the Internet, where the provider does not need visibility into the specific use of its service. This is also good for security as it allows for a stricter security model. Unlike SaaS or cloud compute, cloud storage allows the data to be encrypted at rest (i.e. at the cloud provider's servers), in such a way where the cloud provider does not have the ability to read it. Security credentials and a secure connection to the cloud are still necessary, but the addition of this form of data encryption at rest dramatically increases the level of security. Unlike SaaS and cloud compute, where customer data needs to be visible to the cloud in order to deliver functionality, storage is essentially passive. Data can be completely opaque to the provider. Modern encryption protects data at rest so thoroughly that with proper key management, there is no significant difference between storing sensitive data in your data center or in the cloud.

In order to understand better how encryption protects data in the cloud it is helpful to review the top security concerns that customers have when considering cloud storage.

Security Concerns in the Cloud

Armed with the knowledge about the different types of cloud offerings—SaaS, compute and storage—we can now examine the major concerns that are keeping businesses from putting sensitive information in the cloud.

Data Leakage

Many businesses that would benefit significantly from using the cloud are holding back because of data leakage fears. The cloud is a multi-tenant environment, where resources are shared. It is also an outside party, with the potential to access a customer's data. Sharing hardware and placing data in the hands of a vendor seem, intuitively, to be risky. Whether accidental, or due to a malicious hacker attack, data leakage would be a major security violation.

With cloud storage, all data and metadata should be encrypted at the edge, before it leaves your data center. The user of the storage system must be in control of not only the data, but also the keys used to secure that data. From a security perspective, this approach is essentially equivalent to keeping your data secured at your premises.

While data leakage remains an unsolved issue in SaaS and cloud compute, encryption offers a sensible strategy to ensure data opacity in cloud storage. Data should be encrypted from the start so that the possibility of the cloud storage provider being somehow compromised poses no additional risk to the encrypted data. With cloud storage, all data and metadata should be encrypted at the edge, before it leaves your data center. The user of the storage system must be in control of not only the data, but also the keys used to secure that data. From a security perspective, this approach is essentially equivalent to keeping your data secured at your premises. It is never acceptable to encrypt data at an intermediary site before transmission to the cloud, as this allows the intermediary site to read the data. Furthermore, any encryption scheme must not rely on secrecy (other than the actual key), obscurity, or trust.

Customer Identification

Cloud credentials identify customers to the cloud providers. This identification is a key line of defense for the SaaS and cloud compute offerings. In cloud storage, even the encrypted data can be vulnerable if your files are pooled in with those of another customer. Access to a given pool of storage is based on credentials, and if you are lumped together with another set of customers and share the same credentials, there is a risk that one of them could obtain those credentials and access your data. They would not be able to decipher it, assuming it is encrypted, but they could delete, add to, or otherwise tamper with the files. By securing your own unique credentials, however, your files will be separate. No one else will be able to log into your account and delete your data.

Key management should be so simple that users are not even aware of it: Encryption should be automatic. There should be no way to turn it off.

Data Snooping

All commercial cloud offerings rely on SSL to provide secure connections to their services. This is less of an issue when the data is already encrypted, but when data must remain visible to the provider's servers, it is imperative that a secure connection be used in order to avoid network snooping or the ability of a malicious attacker to intercept the data in transit over the public network. Strictly speaking, encrypted files do not need to be sent over a secure line – this amounts to double encryption. But it is best to assume the worst and guard against any measure of snooping of even cloud-specific metadata by only sending and retrieving data over a secure line. Both data and metadata should be completely opaque on the wire and in the cloud. Nothing – not even filenames or timestamps – should be decipherable once it leaves your premises.

… and a few more considerations that make the system usable:

Key Management

Cryptography is a double-edged sword. Strong encryption will prevent anyone from being able to see your data, including yourself, if your keys are in any way lost or corrupted. Proper key management is critical. If you botch it, there is a risk that users will not want to activate the cryptography, which then compromises security. Key management should be so simple that users are not even aware of it: Encryption should be automatic. There should be no way to turn it off. If there is no insecure mode, then there is no chance of someone accidentally sending unencrypted, vulnerable data to the cloud. Keys should also be securely escrowed, so that no one can obtain that key to access your data. Ideally, you would escrow this key yourself, but Nasuni also offers customers secure key escrow.

Performance

A strong security strategy is a necessity, but it should not seriously impact performance. Encryption of data being sent to the cloud, and decryption of files called back from the cloud, should happen with little or no impact on the user experience. Ideally, it should all happen without the user noticing a thing.

Data encryption at rest where only the user has the ability to decrypt makes cloud storage significantly more secure than other types of cloud offerings. A full implementation of the OpenPGP standard allows for maximum practical security of data through a well-understood combination of encryption programs and key management.

Inside OpenPGP

While the cloud is a new idea, security is hardly a new thing and it has made the commercial Internet possible. The security community has been working for decades on the encryption schemes that are in use today. Encryption is as ancient as war or mathematics and billions of bank transactions are performed daily over cryptographic protocols like the Secure Socket Layer (SSL).

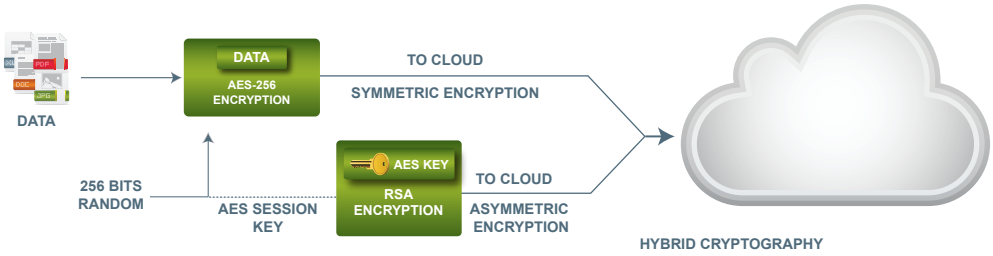
From the onset of the Internet, the security community understood that a public network would require a rethinking of previous security models in order to thrive as a civic and commercial entity. In 1991 Philip Zimmermann created Pretty Good Privacy[2] (PGP) a computer program designed to ensure cryptographic privacy and authentication. The program gained wide adoption during the early Internet because it offered the closest thing to military-grade security and the source code was available to anyone that wanted to understand and improve it. The security community concentrated its energy around improving this open-book security framework and in 1998 the Internet Engineering Task Force, the top technical body that establishes the standards that run the Internet, created the OpenPGP standard.

OpenPGP has stood the test of time while being challenged by the best minds in cryptography.

OpenPGP is not a cipher or a specific encryption algorithm; rather it establishes a framework for how to combine widely available algorithms into a secure system. The current OpenPGP document published by IETF is in the form of Request for Comment or RFC-4880[3]. An RFC document stands to an extended period of peer review. OpenPGP has stood the test of time while being challenged by the best minds in cryptography. There are many implementations of the OpenPGP standard. The Free Software Foundation developed its own implementation under the auspices of the GNU project, GnuPG. It is hard to overstress the importance of not only having the security standard be open but also having access to the source code of the software implementation. This allows for an extensive and thorough review process. When it comes to security, the security you don't know is always worse than the security you know. A standard also means that data encrypted with one implementation of the standard can be decrypted with another implementation thereby protecting the longevity of data access.

[2] Pretty Good Privacy is a reference to "Ralph's Pretty Good Grocery" in Garrison Keillor's Lake Wobegon. Source: http://en.wikipedia.org/wiki/Pretty_Good_Privacy

[3] http://tools.ietf.org/html/rfc4880 Full disclosure: David Shaw who is a co-author of RFC4880 is an employee of Nasuni. He is responsible for the security architecture of the Nasuni products and services.

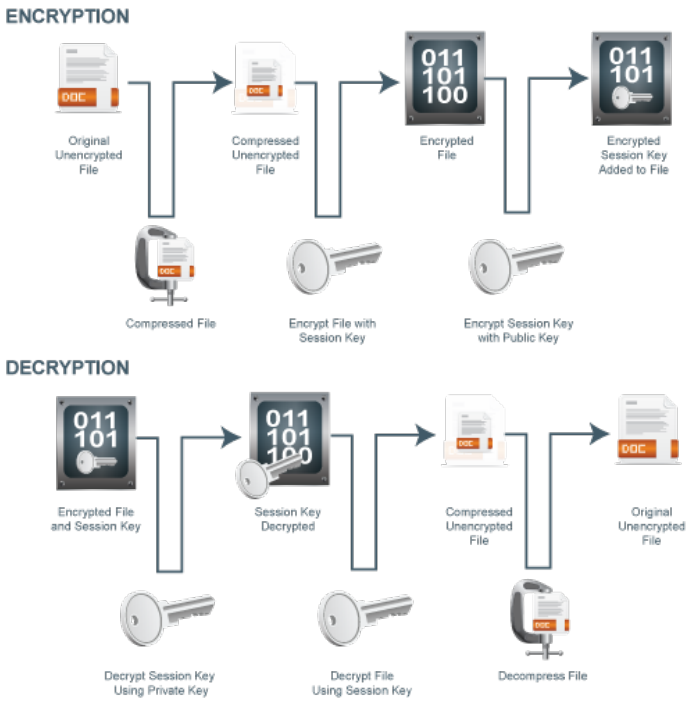


OpenPGP hybrid encryption to the cloud

OpenPGP combines symmetric and asymmetric encryption schemes to form a security model that not only protects the data but does so in a way that is practical and does not compromise the performance of the system. Symmetric encryption, where the same key is used to encrypt and decrypt, tends to be fast at encrypting lots of data. The Advanced Encryption Standard AES-256[4] is a symmetric encryption scheme used by the U.S. government. The 256 indicates the size of the key in bits. OpenPGP uses symmetric encryption like AES-256 to encrypt data and asymmetric encryption like RSA (Rivest-Shamir-Adleman) to encrypt the keys used by AES-256. Asymmetric encryption simplifies key management, but is generally much slower than symmetric encryption. This hybrid approach—using the fast symmetric encryption to encrypt data and the slower asymmetric encryption only to encrypt the (comparatively small) keys—allows data to be encrypted efficiently and a very high level of granularity. Every data packet in the cloud can be protected separately with its own symmetric key and those keys can be managed together through their combined asymmetric key. This allows a practical level of control and granularity of keys and encrypted objects. The asymmetric keys can be maintained by the user in a key ring that becomes the single point of access control to the whole system.

[4] http://en.wikipedia.org/wiki/Aes-256

Nasuni addresses security by relying on the Open-PGP encryption framework, acquiring unique cloud credentials for each of its customers, and ensuring that all data is completely opaque on the wire and in the cloud.



Encryption and decryption of files

OpenPGP also specifies several important details, including proper salting, cipher modes, and other fine points. OpenPGP calls for a variant of cipher feedback (CFB) mode that avoids the drawbacks of less secure techniques, such as ECB. With CFB, as each new block of data is encrypted, part of the previous block is worked into it. This way, there is feedback from the very first block through to the last. If you change a byte early on, that change propagates through the entire system. Patterns in the data are completely hidden.

Security in the Nasuni Filer

Nasuni addresses security by relying on the OpenPGP encryption framework, acquiring unique cloud credentials for each of its customers, and ensuring that all data is completely opaque on the wire and in the cloud. Using OpenPGP, the Nasuni Filer encrypts all data and metadata before sending it to the cloud, and transmits it all over an encrypted line. The Filer also masks filenames, file sizes, timestamps, and more – data and metadata are completely opaque. Keys are securely escrowed and encryption is automatic: The Filer can only send

encrypted data to the cloud. Finally, by using the AES-256 cipher, the Filer is able to encrypt and decrypt data quickly, ensuring that users enjoy strong security without sacrificing performance.

Features

AES-256

OpenPGP offers a small number of carefully selected ciphers. Nasuni uses the AES-256 cipher, which is arguably the strongest in the civilian world. Jon Callas, the primary author of the OpenPGP standard, once used this story to frame the strength of 128-bit keys in layman's terms:

If a customer's files have been modified, the Filer will detect this change when it accesses the relevant data or metadata.

"Imagine a computer that is the size of a grain of sand that can test keys against some encrypted data. Also imagine that it can test a key in the amount of time it takes light to cross it. Then consider a cluster of these computers, so many that if you covered the earth with them, they would cover the whole planet to the height of 1 meter. The cluster of computers would crack a 128-bit key on average in 1,000 years."

"If you want to brute-force a key, it literally takes a planet-ful of computers. And of course, there are always 256-bit keys, if you worry about the possibility that government has a spare planet that they want to devote to key-cracking." [5]

The Filer uses these 256-bit keys.

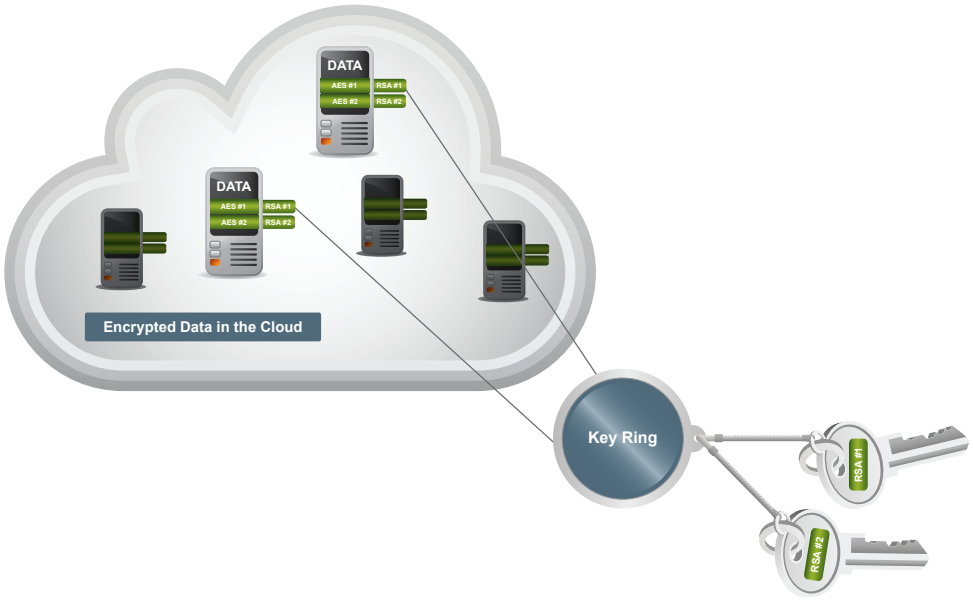
Modification Detection

A problem at a given cloud vendor—an accident or a malicious break-in—could lead to modification of a customer's files. Since Nasuni does not stand in the data path, we cannot prevent tampering, but we can detect it.

The Modification Detection Code (MDC) system in OpenPGP is a built-in tamper alarm. MDC incorporates a hash of the contents in the encrypted object. If a customer's files have been modified, the Filer will detect this change when it accesses the relevant data or metadata.

[5] The email is available here: http://www.interesting-people.org/archives/interesting-people/200607/msg00058.html

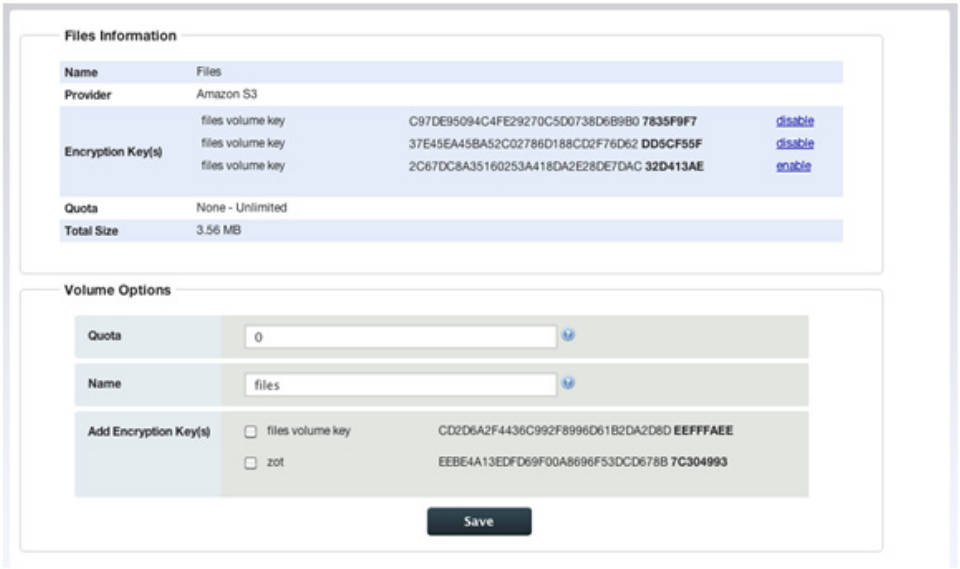
Implementation



Key management of encrypted data in the cloud

OpenPGP Key

When a customer boots the appliance for the first time, the Filer generates an RSA OpenPGP key. From this point forward, this key is associated with the newly created trial cloud volume.



Nasuni Filer key management window

Nasuni automatically escrows this initial key for the customer, but if customers moving from trial mode to the pay model do not want us to escrow their key, we suggest that they either create an additional cloud volume with their own key or generate a new key for the cloud they have been using and remove the trial key.

Creating a new key is an important security step. Nasuni can escrow keys, but for security purposes, we would rather not. We have numerous practices in place to protect the keys Nasuni escrows. Yet there is still a small but non-zero risk that a rogue Nasuni employee could access customer data if—and only if—we are storing that customer's key.

We trust our employees, but we do not ask our customers to share that trust. This is good security practice. No strong encryption scheme should rely on trust.

Key Escrow

Modern encryption offers the strongest protection to your data in the cloud; however, it is also the quickest way to lose access to your data. If the keys are lost, the data will no longer be able to be decrypted. Hence key escrow is crucial to avoid inadvertent data loss.

Currently, we offer customers the option of escrowing their keys at Nasuni. In the future we intend to offer key escrow with a third party, but in the meantime, we have established several practices to protect these keys.

The key itself is encrypted and sent to Nasuni over an encrypted channel. Escrowed keys can only be accessed through the use of both a passphrase and a physical, hardware-based key—an OpenPGP smartcard. We use a smartcard instead of simply encrypting the keys because the smartcard has to be physically present to decrypt. It cannot be copied and used later by a rogue employee. The smartcard has safeguards against brute force attacks, too. If the smartcard is stolen and someone incorrectly guesses the passphrase three times, the card will "brick" itself and become unusable.

If a Nasuni customer's data center is destroyed, or in the event of some other disaster, they can recover their data by logging into their account from a new location, downloading and starting an entirely new copy of the Filer. This new Filer will detect that the customer is reinstalling and start the recovery process, which

can only be performed with the proper key. Customers who escrow their own keys can upload them at the required point in the process, but for a customer who escrows with Nasuni, we provide the customer with a session key that will allow them to de-escrow their keys, without exposing any other escrowed keys.

Upload encryption keys to recover Nasuni Filer

Unique Cloud Credentials

Nasuni's cloud partners provide access to a given account based on credentials. Clouds will only let customers access their accounts if they have the proper credentials. Each Nasuni customer receives a unique set of credentials that enables their copy of the Filer to establish a link with the chosen cloud. This helps ensure that no customer can interfere with another customer's data.

Data Opacity

The Nasuni Filer encrypts customer data before sending it to the cloud, but it also renders that data completely opaque. Even an encrypted file can betray certain details, such as size or filename.

Traffic analysis, a known attack against secure systems, capitalizes on this sort of information, teasing out information based on traffic patterns such as

timestamps or filenames. To help guard against this and other attacks, the Filer renders both data and metadata unreadable before sending it to the cloud.

To help guard against this and other attacks, the Filer renders both data and metadata unreadable before sending it to the cloud.

Random Filenames

If filenames are readable, information could be leaked, so the Filer encrypts them, masking these details. Cloud files have quasi-random filenames that are unrelated to actual customer filenames.

Chunking

The size of a file could give away some information about its contents. An 8 GB file, even with a randomized name, might be flagged as an object of interest. The Nasuni Filer's chunking technique—large files are broken down into smaller pieces before being dispatched to the cloud—helps to prevent this. Compression, also automatic for all objects stored in the cloud, further contributes to size differences. All objects in the cloud are smaller than a certain predetermined size.

Encrypted Metadata

The Nasuni Filer encrypts data and metadata. If this were not the case, a hacker could find an encrypted file of interest by scanning that metadata. This includes the filename, as described above, but the Filer also encrypts file size, timestamps, the location within the directory tree, etc. None of this is decipherable.

The end result of all this is that if someone were to hack into a Nasuni customer's cloud, all they would see would be a huge number of files with equally long, unrelated filenames, and no other revealing metadata. From the perspective of the clouds, or someone hacking into them, the contents would be nearly indistinguishable from each other.

Discussion

For all the new benefits that come with the cloud, there are also a new set of risks. Each of the cloud systems—SaaS, cloud compute and cloud storage—comes with a unique set of benefits and security challenges. In this paper we've explained how and why cloud storage allows for a stricter security model, in part because data can be encrypted at rest in the cloud with a proven, tested standard like OpenPGP in such a way that nobody but the user can decrypt it. We detailed the work that went into creating OpenPGP, and addressed some of the concerns, including data leakage, customer identification, and data snooping, that are keeping businesses from moving sensitive information to the cloud. Finally, we detailed how Nasuni protects customer data through OpenPGP encryption, unique customer credentials, and complete opacity of data and metadata. We do not ask our customers to trust the clouds, and we do not expect them to trust Nasuni, either. They should not have to. Most companies in the cloud space have some measure of access to customer data, even if they insist in a EULA that they will not look at it. By design, Nasuni cannot read customer data.

Concerns about security in the cloud are legitimate, but here at Nasuni we believe these risks can be mitigated with the right security strategy.

Sales: sales@nasuni.com, 800-208-3418
Support: support@nasuni.com, 888-662-7864
or 888-6NASUNI

nasuni