Limited use cryptographic tokens in securing ephemeral cloud servers

Gautam Kumar
University of Washington Bothell
Computing and Software Systems
gautamk@uw.edu

Dr. Brent Lagesse
University of Washington Bothell
Computing and Software Systems
lagesse@uw.edu

ABSTRACT

Many enterprises and consumers today are dependent on services which deploy their computational resources on Infrastructure as a Service (IaaS) cloud providers such as Amazon's AWS, Google's Cloud platform and Windows Azure. Cloud deployments at scale can have hundreds of virtual servers running and the same time sharing sensitive configuration information such as DB passwords and API Keys. This paper aims to propose a risk limiting architecture to safeguard against a potential information leak of sensitive configuration data using a Central Trusted Authority with hash chains as an authentication mechanism for ephemeral servers in the cloud to provide moving target defense against attackers.

CCS Concepts

•Security and privacy → Domain-specific security and privacy architectures; *Security protocols;* **•Software and its engineering → Software architectures;**

Introduction

Cloud infrastructure today is characterised by three difference classes of products. Infrastructure as a Service or IaaS, these are providers such as Amazon's AWS, Google's Cloud platform, Microsoft's Azure who offer on-demand compute, storage and network resource at the click of a button. Platform as a Service or PaaS, these are providers such as Heroku and Google's AppEngine who host a developer's code and take on the responsibility of platform maintenance, i.e maintaining the compute, storage and network resource and scaling them as needed. Software as a Service or SaaS these refer to any web based application which runs on cloud infrastructure, examples include Google Apps and Microsoft's Office 365.

According to a white paper published by Microsoft [10], more than 85% work done by enterprise IT departments is towards infrastructure maintenance. IaaS providers offer enterprises a unique advantage by minimizing the effort needed to purchase and maintain physical hardware. So it stands to reason that enterprises now consider hosting their infrastructure with IaaS providers as a necessity rather than merely a competitive advantage[18].

Hosting infrastructure on the cloud comes with its own set of unique challenges and one of the primary concerns is Security. IaaS providers offer computational and storage resource on virtualized shared hardware to maximize utilization. So the essence of securing cloud systems is using multiple layers [21] of security to increase an attacker's cost for taking over the system. One of the possible layer of security is using moving target defenses [6].

One particular problem that is common to all cloud deployments is secure credential storage. Credentials such as Encryption keys, API keys, database passwords need to be distributed to all servers that may have a connection to the open internet. These sensitive credentials are required by the servers to perform their tasks such as reading retrieving customer information, performing tasks on behalf of the customer and so on. Distributing these sensitive credentials risks sensitive data exposure which is one of OWASP's Top 10 security threats of 2013 [28].

Sensitive data exposure simply refers to unintended exposure of sensitive information such as passwords, social security numbers, date of birth and so on. In the context of our discussion these credentials are usually stored as part of a configuration file. According to a report by Risk Based Security [24] [26] the number of data leaks has dramatically increased from 2012 to 2013, to the tune of $812 million. Though sensitive data exposure in the context of cloud configurations would only constitute a small part of these leaks, leaking of such credentials can potentially lead to other massive data leaks and potential vulnerabilities being exposed because an attacker is able to gain access to sensitive resources which were previously protected by these credentials.

Sensitive data exposure can potentially be a consequence of other threats such as cross site scripting (XSS) [16], Injection is the most critical threat while XSS is the 3^{rd} most critical threat as classified by OWASP in 2013 [28].

Securing hundreds of servers against data leaks and potential vulnerabilities is a hard problem and so in this paper we propose a solution architecture which moves the responsibility secure credential storage away from individual client facing servers and migrating all credentials to a central trusted authority which acts on behalf of the client facing servers. We introduce the element of moving target defense by limit-

DOI: 10.1145/1235

ing the lifetime of client facing servers and using hash chains as an authentication mechanism.

Background

Cryptographic hash function [25]

A cryptographic hash function is any one way function which meets the following requirements

- Preimage resistance
- Collision resistance
- Second Preimage resistance

A hash function has preimage resistance if given a hash value h it is computationally infeasible to find any message m such that $h = hash(k, m)$ where k is the hash key.

A hash function is collision resistant if, given two messages m_1 and m_2 it is hard to find a hash h such that $h = hash(k, m_1) = hash(k, m_2)$ where k is the hash key.

A hash function has second pre-image resistance if given a message m_1 it is computationally infeasible to find a different message m_2 such that $hash(k, m_1) = hash(k, m_2)$ where k is the hash key. The second pre-image resistance is a much harder property to achieve for hash functions. This property is closely related to the birthday problem [14].

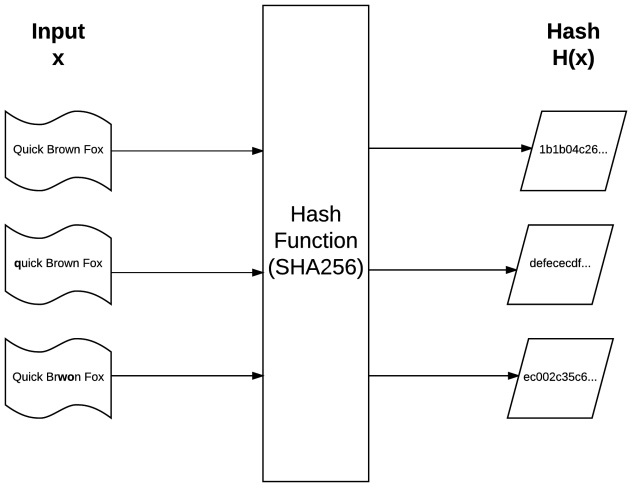

Figure 1: A simplified view of a hash function which represents its input and potential result. The length of the hash sum always remains the same regardless of the input size. Any small change in the input drastically changes the output.

Hash Chains [11]

Leslie Lamport [12] first proposed the use of hash chains in his paper on a method for secure password authentication over an insecure medium.

"A hash chain is a sequence of values derived via consecutive applications of a cryptographic hash function to an initial input. Due to the properties of the hash function, it is relatively easy to calculate successive values in the chain but given a particular value,it is infeasible to determine the previous value"

A hash chain in essence is merely the successive computation of a Cryptographic hash function on a given value.

As an example, Let x be the initial password and H be the cryptographic hash function. A hash chain of length 2 would be $H^2(x) = H(H(x))$. A hash chain of n values is denoted as $H^n(x)$ and the i^{th} value in the chain would be computed as $x_i = H(x_{i-1})$.

For a given value in the chain x_i its computationally infeasible to determine the previous value in the chain x_{i-1}.

Proposed solution architecture

The proposed architecture to defend against the threat of sensitive data exposure is to use a Central Trusted Authority (CTA) responsible for storing sensitive information. The CTA would act as a proxy and would make requests on behalf of client facing servers, refer Fig. 2.

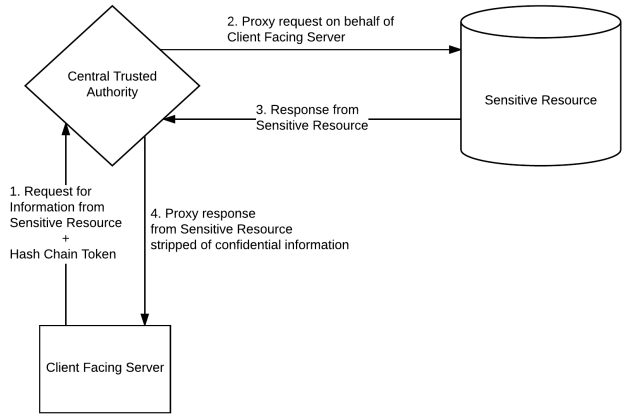

Figure 2: Architectural overview of the system. This figure describes the three primary modules involved. The client facing server, The central trusted authority and a sensitive resource.

The client facing servers would use hash chains to authenticate with the CTA. Hash chains cryptographically limit the number of times a key can be used. Limited use was intentionally selected to promote the creation of a moving target for attackers.

Assumptions

Client facing servers are the servers which are exposed outside the private cloud network environment. These client facing servers could potentially be load balancing servers, compute servers.

The client facing servers are assumed to be ephemeral. This is common in many cloud deployments [27] and contributes to the moving target nature of the security architecture. Companies such as Netflix expect this behaviour with their chaos engineering architecture [1]. This allows for higher reliability of their server infrastructure.

Client facing servers are expected to shut down after their hash chain expires. This contributes to the ephemeral nature and also to moving target defense of the overall system.

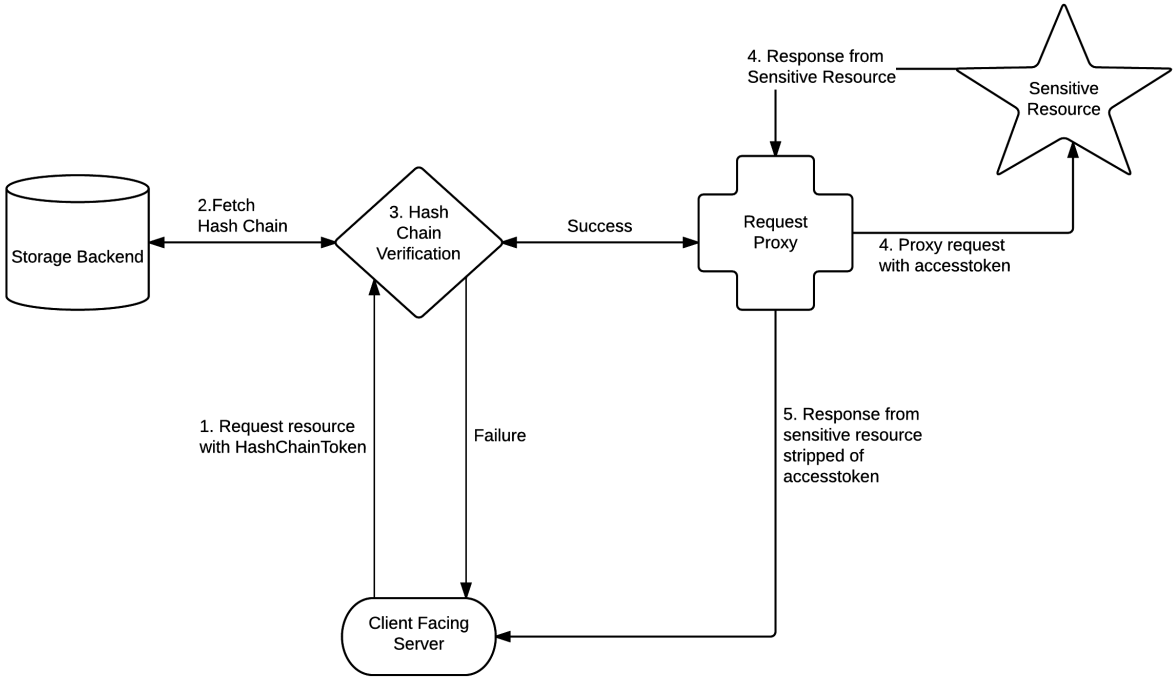

Figure 3: Architecture of the Central Trusted Authority. The CTA consists of a storage backend, a hash chain verifier and a request proxy.

CTA Architecture and Implementation

The Central Trusted Authority consists of three primary components, A hash chain verifier, A storage backend and a request proxy. The CTA generally performs three roles within the system which are

- Create new hash chain
- Verify hash chains
- Proxy requests

Creating new hash chain

Hash chains are created by iteratively hashing a secret key K, n number of times. After the hash chain is created the CTA stores $H^n(K)$ in the storage backend and returns K and n to the client facing server. The secret key K is not stored by the CTA. The client facing server can now use the the secret key n number of times. This process is detailed in algorithm 1.

Hash chain verification

Hash chains are used to authenticate client facing server requests which require access to a sensitive resource. The hash chain verification is detailed in algorithm 2.

Security Threats

The proposed architecture of using a Central Trusted Authority (CTA) to proxy requests on behalf of all the clients places the CTA as a single point of failure. Thus the various components of the CTA need to be analysed and hardened for security threats.

Data: Hash Chain secret K and Hash chain length N
Result: Hash Chain $H^N(K)$

$i \leftarrow 1$;
$H^1(K) \leftarrow H(K)$;
while $i <= N$ **do**
  $i \leftarrow i + 1$;
  $H^i(K) \leftarrow H(H^{i-1}(K))$;
end
return $H^i(K)$ *where* i *equals* N ;

Algorithm 1: Generating a Hash Chain

Data: Authentication key $H^{i-1}(K)$ from the client
Result: Response from sensitive resource

Let $H^i_{client} = H(H^{i-1}(K))$;
Let $AuthenticationData = fetch(H^i_{client})$;
if *AuthenticationData exists in storage backend* **then**
  replace H^i_{client} with $H^{i-1}(K)$ in storage backend ;
  fetch and return response from sensitive resource ;
else
  return authentication failure ;
end

Algorithm 2: Verification of Hash Chain authentication

Our architecture proposal does not prescribe any particular hardware / VM co-location for any components in the CTA. Thus each component can be implemented on separate machines or the CTA can be instantiated as a whole within a single VM. In this section we detail the potential security threats against each component and possible mitigation strategies.

Malicious Request Proxy

The request proxy component acts on behalf of the client facing server to make a call to a sensitive resource such as an API endpoint. The request proxy is also responsible for attaching authentication information such as an API Key when contacting the API endpoint. Upon receiving a response from the sensitive resource the request proxy strips out sensitive information such as API Keys and Refresh Tokens before forwarding the response to the client facing server.Under this threat scenario the request proxy is assumed to be untrustworthy and potentially malicious.

Chen et al [4] propose a solution to this problem of a Malicious proxy using trusted hardware such as Trusted Platform Module (TPM) or the IBM 4758 cryptographic coprocessor [22].

Chen et al assume the proxy, the CTA in our architecture, is malicious but is incapable of modifying the underlying hardware. The CTA executable is also verified by a trusted third party to operate correctly as a proxy as described in [22].

There are three primary attacks that a malicious actor may perform on the CTA. First, the attacker may try to expose the sensitive information stored, such as API keys and DB passwords, from the CTA using vulnerabilities in the CTA executable. Second, the attacker could potentially modify the requests / responses which are being proxied by the CTA. Third the attacker could potentially launch a reboot attack to inject a malicious executable after attestation to carry out one of the above attacks.

Prior work [15, 19] on preventing reboot-attacks can be leveraged to impede the attacker's ability to inject a malicious executable. The proposed architecture also allows for the CTA to be placed behind secure corporate firewalls to further limit the risk of a malicious take over by an attacker. Further work is needed to fully secure the CTA against a malicious proxy.

Insecure Storage back-end

Our proposed architecture can accommodate various kinds of storage backends and thus potential security threats against the storage backend may vary.

Considering the scenario of an RDBMS system, The most common vulnerability is SQL Injection and a lot prior of work has been done to secure RDBMS systems against SQL Injection [8, 2, 9] and those techniques to drastically limit the risk of data leakage or system takeover.

Using a pre-hardened database as a service such as Amazon's RDS or Google's CloudSQL can further improve the security of our storage backend [5] .

Performance testing

In our testing hash chain initialization showed a linear increase in time complexity as shown in figure 5. This is in line with our expectations for hash chain implementations.

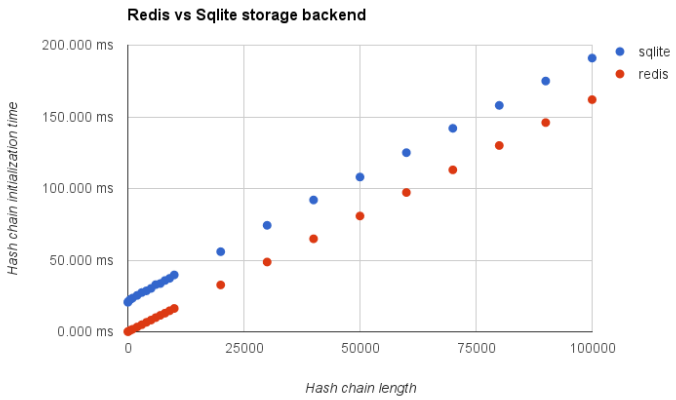

Figure 5: Linear time complexity of the hash chain initialization. Execution platform Intel i5-5200U 2.2Ghz, 12GB RAM, SHA-512 implemented in python.

A hash chain of length 100,000 takes 150 - 200 ms to initialize based on the storage backend used.

Limitations

The proposed architecture and implementation are ideally suited to cloud IaaS providers such as Amazon AWS where ephemeral servers are easily managed. Deploying the proposed system to a cloud provider without the ability to quickly provide large number of ephemeral servers could result is significant performance degradation.

Hash chains inherently posses certain vulnerabilities such as a Hash chain cycle and Hash chain length oracle attacks which can potentially reduce the effectiveness of the proposed architecture [13].

Providing moving target defense

Green et al [7] identify Unpredictability, Vastness and Revocability as some of the criteria for evaluating moving target defenses. In this section we try define how our implementation conforms to these criteria.

Unpredictability. Cryptographic hash functions [25] are designed to be collision resistant and comply with the avalanche effect and thus the generated sequence is highly unpredictable in nature. Hash chains can compound this effect by sequential application of the cryptographic hash function.

Vastness. A secure cryptographic hash function such as SHA-256 has 128 bit of security in its worst case scenario which translates to potentially 2^{128} values which can be safely computed without resistance thus offering a vast state space.

Revocability. The proposed architecture allows individual chains or a group of chains to be revoked merely by purging them from the database. Further group / hierarchical revocability can be achieved using Merkle hash tree implementations as a means of Hash chain generation.

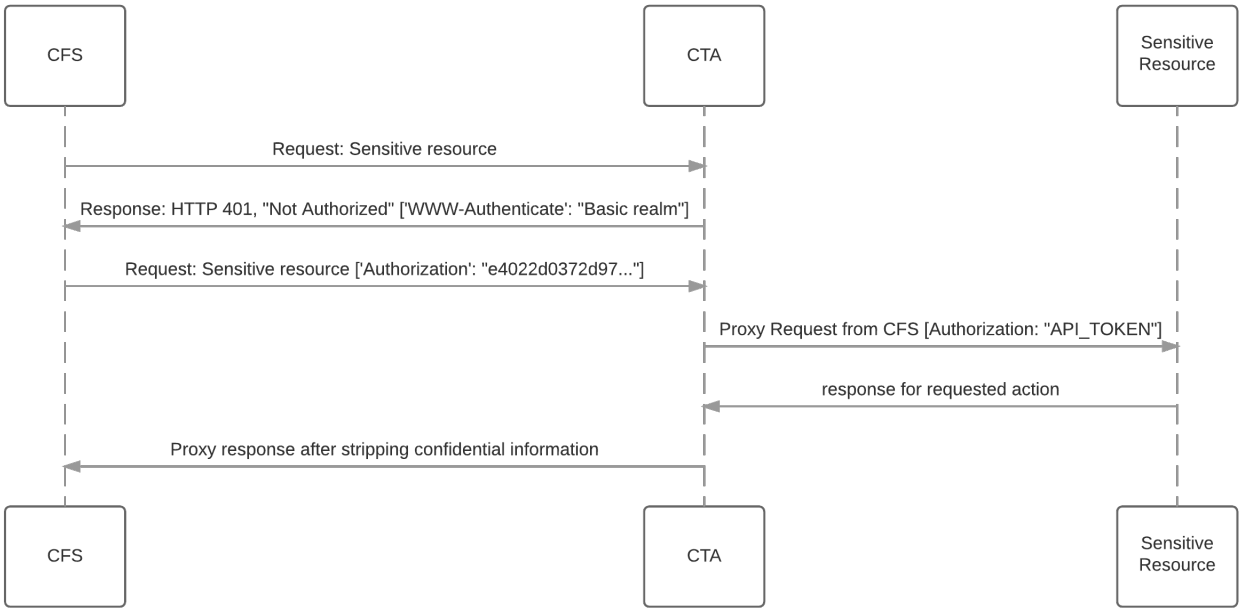

Figure 4: Sequence diagram describing the authentication and proxying capabilities of the CTA. CFS refers to Client Facing Server while CTA refers to Central Trusted Authority

Artificial diversity through DHT

The proposed architecture lends itself well to establishing artificial diversity using a Distributed Hash Table to store the hash chains as suggested by Morell et al [20], thus further contributing to the un-predictability of the system and further limiting the risk introduced by a malicious node.

Future work

Formulating an ideal balance between server lifespan and hash chain length to optimize computational resources in a cloud system, this can be derived from current work into load balancing in cloud systems [23]. Such a formulation can be used by dev ops engineers in choosing an ideal hash chain size based on the performance requirements of a cloud system.

Integrating Timed Release Cryptography [3] as a hash chain renewal mechanism to potentially increase the lifespan of a server after a certain cool off period.

A Merkle hash tree implementation middle-ware can potentially provide hierarchical authentication authorization [29] capabilities to the CTA.

Related work

Confidant [17] is a library maintained by Lyft, a transportation network company based out of San Francisco. Confidant provides an implementation of the Central trusted authority server with encryption at rest, authentication and authorization handled by AWS's Key Management Service, KMS and a storage backend of DynamoDB. This severely hampers the ability of a potential user to deploy a Confidant instance on a different cloud IaaS provider beside Amazon's AWS.

Confidant also serves as an inspiration for the CTA component of the proposed architecture.

Summary

In this paper we propose an architecture to provide moving target defense and minimize the damage that attackers can cause by centralizing the storage of sensitive configuration information and taking advantage of the limited use property of hash chains to authenticate potentially vulnerable client facing servers.

1. REFERENCES

[1] A. Basiri, N. Behnam, R. d. Rooij, L. Hochstein, L. Kosewski, J. Reynolds, and C. Rosenthal. Chaos engineering. 33(3):35–41.

[2] S. W. Boyd and A. D. Keromytis. SQLrand: Preventing SQL injection attacks. In *International Conference on Applied Cryptography and Network Security*, pages 292–302. Springer.

[3] K. Chalkias and G. Stephanides. Timed release cryptography from bilinear pairings using hash chains. In *Communications and Multimedia Security*, pages 130–140. Springer.

[4] R. Chen, A. Reznichenko, P. Francis, and J. Gehrke. Towards statistical queries over distributed private user data. In *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pages 169–182.

[5] C. Curino, E. P. Jones, R. A. Popa, N. Malviya, E. Wu, S. Madden, H. Balakrishnan, and N. Zeldovich. Relational cloud: A database-as-a-service for the cloud.

[6] D. Evans, A. Nguyen-Tuong, and J. Knight. Effectiveness of moving target defenses. In S. Jajodia, A. K. Ghosh, V. Swarup, C. Wang, and X. S. Wang, editors, *Moving Target Defense*, number 54 in Advances in Information Security, pages 29–48. Springer New York. DOI: 10.1007/978-1-4614-0977-9_2.

[7] M. Green, D. C. MacFarland, D. R. Smestad, and C. A. Shue. Characterizing network-based moving target defenses. In *Proceedings of the Second ACM Workshop on Moving Target Defense*, MTD '15, pages 31–35. ACM.

[8] W. G. Halfond and A. Orso. AMNESIA: analysis and monitoring for NEutralizing SQL-injection attacks. In *Proceedings of the 20th IEEE/ACM international Conference on Automated software engineering*, pages 174–183. ACM.

[9] W. G. Halfond, J. Viegas, and A. Orso. A classification of SQL-injection attacks and countermeasures. In *Proceedings of the IEEE International Symposium on Secure Software Engineering*, volume 1, pages 13–15. IEEE.

[10] R. Harms and M. Yamartino. The economics of the cloud.

[11] D. Horne. Hash chain. In H. C. A. v. Tilborg and S. Jajodia, editors, *Encyclopedia of Cryptography and Security*, pages 542–543. Springer US. DOI: 10.1007/978-1-4419-5906-5_780.

[12] L. Lamport. Password authentication with insecure communication. 24(11):770–772.

[13] D. Lee. Hash function vulnerability index and hash chain attacks. In *2007 3rd IEEE Workshop on Secure Network Protocols*, pages 1–6.

[14] L. M. Lesser. Exploring the birthday problem with spreadsheets. 92(5):407–411.

[15] B. Libert and D. Vergnaud. Tracing malicious proxies in proxy re-encryption. In S. D. Galbraith and K. G. Paterson, editors, *Pairing-Based Cryptography âĂŞ Pairing 2008*, number 5209 in Lecture Notes in Computer Science, pages 332–353. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-85538-5_22.

[16] M. T. Louw and V. N. Venkatakrishnan. Blueprint: Robust prevention of cross-site scripting attacks for existing browsers. In *2009 30th IEEE Symposium on Security and Privacy*, pages 331–346.

[17] lyft. Confidant: Your secret keeper. A library to store and retrive senstive configuration within a central trusted authority encrypted at rest using Amazon KMS. https://github.com/lyft/confidant.

[18] A. McAfee. What every CEO needs to know about the cloud. 89(11):124–132.

[19] J. M. McCune, B. J. Parno, A. Perrig, M. K. Reiter, and H. Isozaki. Flicker: An execution infrastructure for TCB minimization. In *ACM SIGOPS Operating Systems Review*, volume 42, pages 315–328. ACM.

[20] C. Morrell, R. Moore, R. Marchany, and J. G. Tront. DHT blind rendezvous for session establishment in network layer moving target defenses. In *Proceedings of the Second ACM Workshop on Moving Target Defense*, MTD '15, pages 77–84. ACM.

[21] A. Panwar, R. Patidar, and V. Koshta. Layered security approach in cloud. In *3rd International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2011)*, pages 214–218.

[22] B. Parno, J. M. McCune, and A. Perrig. Bootstrapping trust in commodity computers. In *2010 IEEE Symposium on Security and Privacy*, pages 414–429. IEEE.

[23] M. Randles, D. Lamb, and A. Taleb-Bendiab. A comparative study into distributed load balancing algorithms for cloud computing. In *2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pages 551–556.

[24] Risk Based and Security. An executiveâĂŹs guide to 2013 data breach trends.

[25] P. Rogaway and T. Shrimpton. Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In B. Roy and W. Meier, editors, *Fast Software Encryption*, number 3017 in Lecture Notes in Computer Science, pages 371–388. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-25937-4_24.

[26] X. Shu, D. Yao, and E. Bertino. Privacy-preserving detection of sensitive data exposure. 10(5):1092–1103.

[27] L. M. Vaquero, L. Rodero-Merino, and R. Buyya. Dynamically scaling applications in the cloud. 41(1):45–52.

[28] D. Wichers. OWASP top-10 2013.

[29] X. Yi and W. Wang. The cloud access control based on dynamic feedback and merkle hash tree. In *2012 Fifth International Symposium on Computational Intelligence and Design (ISCID)*, volume 1, pages 217–221.