Security and Privacy Preservation of Evidence in Cloud Accountability Audits

Thomas Rübsamen[1(✉)], Tobias Pulls[2], and Christoph Reich[1]

[1] Cloud Research Lab, Furtwangen University, Furtwangen, Germany
{thomas.ruebsamen,christoph.reich}@hs-furtwangen.de
[2] Department of Mathematics and Computer Science, Karlstad University, Karlstad, Sweden
tobias.pulls@kau.se

Abstract. Cloud accountability audits are promising to strengthen trust in cloud computing by providing reassurance about the processing data in the cloud according to data handling and privacy policies. To effectively automate cloud accountability audits, various distributed evidence sources need to be considered during evaluation. The types of information range from authentication and data access logging to location information, information on security controls and incident detection. Securing that information quickly becomes a challenge in the system design, when the evidence that is needed for the audit is deemed sensitive or confidential information. This means that securing the evidence at-rest as well as in-transit is of utmost importance. In this paper, we present a system that is based on distributed software agents which enables secure evidence collection with the purpose of automated evaluation during cloud accountability audits. We thereby present the integration of Insynd as a suitable cryptographic mechanism for securing evidence. We present our reasoning for choosing Insynd by showing a comparison of Insynd properties with requirements imposed by accountability evidence collection as well as an analysis how security threats are being mitigated by Insynd. We put special emphasis on security and privacy protection in our system analysis.

1 Introduction

Cloud Computing is known for its on demand computing resource provisioning and has now become mainstream. Many businesses as well as private individuals are using cloud services on a daily basis. The nature of these services varies heavily in terms of what kind of information is being out-sourced to the cloud provider. Often, that data is sensitive, for instance when Personal Identifiable Information (PII) is being shared by an individual. Also, businesses that move (parts of) their processes to the cloud, for instance by using Customer Relationship Management Software as a Service, are actively participating in a major paradigm shift from having all data on-premise to moving data to the cloud.

However, many new challenges come along with this trend. Two of the most important issues are customer trust and compliance [14,22]. These issues are

M. Helfert et al. (Eds.): CLOSER 2015, CCIS 581, pp. 95–114, 2016.
DOI: 10.1007/978-3-319-29582-4_6

closely tied to the loss of control over data. When moving to the cloud, direct control over (i) where data is stored, (ii) who has access to it and (iii) how it is shared and processed is given up. Because of this loss of control, cloud customers have to trust cloud providers that they treat their data in an appropriate and responsible way. One way to enable that trust is by strengthening transparency and accountability [12,30] of the cloud provider and services. This includes providing information about data locality, isolation, privacy controls and data processing in general.

Cloud audits can be used to check how data has been processed in the cloud (i.e., by whom, for what purpose) and whether or not this happened in compliance with what has been defined in previously agreed-upon privacy and data handling policies. This way, a cloud customer can regain some of the information he has given up control of by moving to the cloud. A central responsibility of cloud audits is the collection of data that can be used as evidence. Depending on the data processing policies in place, various sources of evidence need to be considered. For instance, logs are a very important source of evidence, when it comes to auditing the cloud operation (e.g., access logs and error logs). However, other sources of information are also important, such as files (e.g., process documentation) or events registered in the cloud management system (e.g., access control decisions, infrastructure changes, data transfers).

To capture evidence from this variety of sources, centralized logging mechanisms are not enough. We therefore propose a system for accountability evidence collection and audit. With this system, cloud providers are enabled to demonstrate their compliance with data handling policies to their customers and to third-party auditors in an automated way.

In our previous work, we proposed a concept [28] for cloud accountability audits, that enables automated collection of evidential data in the cloud ecosystem with the goal of performing accountability audits. A key mechanism of this system is the secure and privacy-friendly collection and storage of evidence. In our previous work we also explored the use of a somewhat homomorphic encryption scheme to secure evidence collected in the evidence store [17], which has proven practical but very limited in terms of performance and functionality.

In this paper, we present a more practical alternative that imposes less restrictions on evidence collection.

The contributions of this paper are:

- An architecture for automated evidence collection for the purpose of cloud accountability audits
- A process for secure and privacy-protecting evidence collection and storage

The remainder of this paper is structured as follows: in Sect. 2 we present related work in the area of secure evidence collection and cloud auditing. The core principles of Insynd are introduced in Sect. 3. Section 4 introduces the Audit Agent System (AAS) and its architecture. Following that, we present in Sect. 5 a mapping of typical characteristics of digital evidence and secure evidence collection in the cloud to how these are addressed by integrating Insynd in our audit agent system. In Sect. 6 we describe the architectural details of the

Insynd integration. We also present a scenario-based evaluation of our system in Sect. 7 and conclude this paper in Sect. 8.

2 Related Work

Redfield and Date propose a system called Gringotts [27] that enables secure evidence collection, where evidence data is signed at the system that produces it, before it is sent to a central server for archival using the Evidence Record Syntax. It is similar to our system with respect to the automatic collection of evidential data from multiple sources. However, their focus is on the archival of evidence, whereas we propose a system that also enables automated evidence processing for audits. Additionally, our system also addresses privacy concerns of evidence collection in a multi-tenant environment such as the cloud by introducing evidence encryption, whereas Redfield and Date focus on archival and preservation of evidence integrity.

Zhang et al. [31] identify potential problems when storing massive amounts of evidential data. They specifically address possible information leaks. To solve these issues, they propose an efficient encrypted database model that is supposed to minimize potential data leaks as well as data redundancy. However, they focus solely on the storage backend and do not provide a workflow that addresses secure evidence collection as a whole.

Gupta [11] identifies privacy issues in the digital forensics process, when it comes to data storage devices that typically do not only contain investigation related data, but may also hold sensitive information that may breach privacy. He also identifies a lack of automation in the digital investigation process. To address these issues, Gupta proposes the Privacy Preserving Efficient Digital Forensic Investigation (PPEDFI) framework. PPEDFI automates the investigation process by including knowledge about previous investigation cases, and which kinds of files were relevant then. With that additional information, evidence search on data storage devices is faster. However, while Gupta acknowledges privacy issues, the PPEDFI framework is focused on classic digital forensics and may not be applicable to a cloud ecosystem, where there is typically no way of mapping specific data objects to storage devices, in full.

The Security Audit as a Service (SAaaS) system proposed by Dölitzscher et al. [9,10] is used to monitor cloud environments and to detect security incidents. SAaaS is specifically designed to detect incidents in the cloud and thereby consider the dynamic nature of such ecosystems, where resources are rapidly provisioned and removed. However, the main focus of SAaaS is not to provide auditors with a comprehensive way of auditing the cloud provider's compliance with accountability policies, which requires additional security and privacy measures to be considered in the data collection process.

3 Insynd

Insynd is a cryptographic scheme where a forward-secure *author* sends messages intended for *clients* through an untrusted *server* [23,24]. The author is

forward-secure in the sense that the author is initially trusted but assumed to turn into an active adversary at some point in time [5]. Insynd protects messages sent prior to author compromise. The server is untrusted, which is possible thanks to the use of Balloon, a forward-secure append-only persistent authenticated data structure [23]. This means that the server storing all messages can safely be outsourced, e.g., to traditional cloud services. Clients are assumed trusted to read messages sent to them by authors. Insynd contains support for clients to also be in the forward-security model, by discarding key-material as messages are read. For sake of ease of implementation, Insynd is designed around the use of NaCl [6], an easy-to-use high-speed cryptography software library.

Insynd provides the following properties:

Forward Integrity and Deletion Detection. Nobody can modify or delete messages sent prior to author compromise, as defined by Pulls et al. [25]. This property holds independently for Balloon (the data structure) and the Insynd scheme. For Balloon, anyone can verify the consistency of the data structure, i.e., it is publicly verifiable [23].

Secrecy. Insynd provides authenticated encryption [2].

Forward Unlinkability of Events. For each run by the author of the protocol to send new messages, all the events sent in that run are unlinkable. This implies that, e.g., an attacker (or the server) cannot tell which events belong to which client [24]. When clients receive their events by querying the server, if they take appropriate actions including but not limited to accessing the server over an anonymity network like Tor [8], their events remain unlinkable.

Publicly Verifiable Proofs. Both the author and client receiving a message can create publicly verifiable proofs of the message sender (the author), the receiving client (by registered identity), and the time the message was sent relative to e.g. a time-stamping authority [24]. The proof-of-concept implementation of Insynd uses Bitcoin transactions [20] as a distributed time-stamping server.

Distributed Settings. Insynd supports distributed authors, where one author can enable other authors to send messages to clients it knows of without requiring any interaction with clients. Client identifiers (public keys) are blinded in the protocol, ensuring forward-unlinkable client identifiers between different authors [24].

Pulls and Peters show that Insynd provide the above cryptographic properties under the assumptions of the decisional Diffie-Hellman (DDH) assumption on Curve25519, an unforgeable signature algorithm, an unforgeable one-time MAC, a collision and pre-image resistant hash function, a IND-CCA2 secure public-key encryption scheme, and the security of the time-stamping mechanism (in our case, the Bitcoin block-chain) [24]. The prototype implementation of Insynd shows performance comparable to state-of-the-art secure logging schemes, like PillarBox [7], securing syslog-sized messages (max 1 KiB) in the order of hundreds of microseconds on average on a commodity laptop. We stress that Insynd is subject to its own review and evaluation; in this paper, we use Insynd as a building block to facilitate secure evidence collection and storage for cloud accountability audits.

4 Audit Agent System

In the following, the main actors, components and the general flow of information from the evidence-producing source to the audit report in our Audit Agent System (AAS) are described.

4.1 Privacy and Accountability Cloud Audit System Actors

The main actor using the AAS is the *Cloud Auditor*. According to NIST, a cloud auditor is a "A party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation." [16] In general, a cloud consumer, cloud provider or an independent third-party can act as a cloud auditor. Depending on the actual stakeholder that assumes the role of the auditor, isolation issues can arise:

- A *data protection authority (DPA)* typically acts in good faith as a third-party and assesses privacy policies. Therefore, they typically have broad access to a provider's internal documentation, infrastructure and potentially customer's data.
- A *commercial third-party auditor* is usually a specialized service provider (e.g., a penetration or security testing specialist) acting on behalf of the cloud provider. Their access to information is similar to that available to the DPA.
- A *customer* can also assume the role of an auditor, however with a much more limited scope of available information. We consider two major sub-types, businesses as customers and individuals as customers.

In our proposed system, we consider business customers (e.g., companies using cloud services to replace their IT) to be potential auditors but exclude private individuals. Additionally, providers use the AAS internally for self-auditing to regularly and continuously assess their policy compliance and detect potential violations in a timely manner. Depending on the view on an organization (i.e., depending on who assumes the role of cloud auditor), data protection is an issue to consider, when potential confidential information is processed during an audit. This means data confidentiality, integrity and isolation have to be preserved during an automated audit.

4.2 Architectural Components Audit Agent System

The architecture of the Audit Agent System (see Fig. 1) is based on the use of software agents. This allows for improved flexibility by allowing to rapidly react on infrastructure changes, and improved extensibility especially with respect to data collectors that are used to gather information that is evaluated during an audit. The collectors are adapters for the various heterogeneous sources of evidence in a cloud environment. In Sect. 6.1, we describe more details of how the collectors work. The architecture of the AAS comprises of the following components: *Audit Policy Module (APM)*, *Audit Agent Controller (AAC)*, *Evidence*

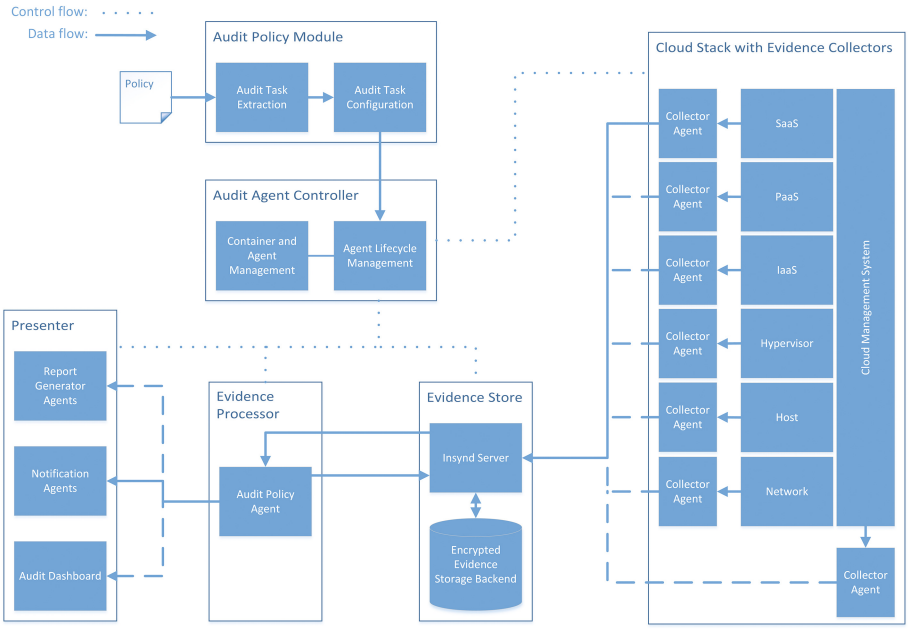

Fig. 1. Privacy and accountability cloud audit system architecture.

Processor and Presenter (EPP) and *Evidence Store (ES)*. Especially the Evidence Store and the aforementioned collection agents make heavy use of Insynd to assure that data protection requirements are being met. To a lesser extent, the AAC and EPP also utilize Insynd for securely transporting evidence.

All components are implemented as software agents based on the Java Agent DEveleopment framework (JADE) [13] and make heavy use of the JADE Agent Communication Language (ACL) for agent interaction. In the following, we describe the architecture components:

Audit Policy Module. The main input to the AAS are machine-readable policies that describe data handling obligations (e.g., access control), security controls (e.g., service configuration) and data protection mechanisms (e.g., encryption). From such policies, tasks for collecting evidence, and rules for evaluation of the evidence with the goal of producing a compliance statement are extracted. Additional input to the APM is provided by the auditor. We assume, that there is always a need for at least some manual input for defining an automated audit because the input policy might not be complete with respect to all the parameters that are required for an automated audit. Such parameters include the audit type (periodic or event-driven), the frequency (e.g., daily, monthly...) but also more task-specific information that is not provided by the input policy. Depending on the actual audit task, the input comprises of policies and auditor-supplied information:

1. Policies, which define obligations that have to be fulfilled by the cloud provider, such as data access restrictions and usage policies, requirements for the implementation of privacy controls, data retention requirements and general security requirements. The A4Cloud [1] research project develops a machine-readable policy language based on the Primelife Policy Language [3] called Accountability PPL [4]. The A-PPL is capable of describing obligations providers have to adhere to, such access control rules and data handling (e.g., data location, purpose etc.). A-PPL serves as the main input to the Audit Agent System and for defining audits.
2. It is possible that an input policy does not necessarily include all information required for mapping policy requirements to specific evidence sources, collectors (e.g., evidence source specific REST client or log parser) and evaluators (e.g., API endpoints, access credentials). That information is provided by the auditor.

With the above mentioned data, the APM builds audit tasks - a combination of evidence collector, processor and presenter agents - and passes that task on to the Audit Agent Controller for instantiation.

Audit Agent Controller. The AAC is the core component of the Audit Agent System. Its main responsibility is the management (i.e., instantiation, configuration, deployment) of any type of agent in the AAS. The main input comes from the APM, which effectively instructs the AAC on how to setup specific audit tasks. A typical audit task deployment in AAS is called an audit workflow. The typical audit workflow (depicted in Fig. 2) is as follows:

1. *Preparation*: The APM extracts audit task configuration from the policy, combines it with input provided by the auditor and passes it on to the AAC.
2. *Configuration*: According to the input provided by the APM, the AAC configures audit policies, its tasks and corresponding collection and evaluation agents.
3. *Instantiation*: the AAC instantiates the previously configured agents as well as the associated evidence store.
4. *Migration*: Agents are migrated from the core platform where the AAC is running to the target platforms (agent runtime environments as close as possible to the evidence source).
5. *Monitoring*: During the agents' lifetime, the AAC monitors registered platforms and registered agents, handles exceptions, and manages the creation, archival and deletion of evidence stores
6. *Termination*: The AAC disposes of the collector and evaluation agents when they are not needed anymore. It also handles archival and / or deletion of the corresponding evidence store in that case.

Evidence Processor and Presenter. After the collector agents have gathered evidence data and stored it in the evidence store, the evaluation agent(s) of an audit task retrieve that data and analyze it according to the rules that have been

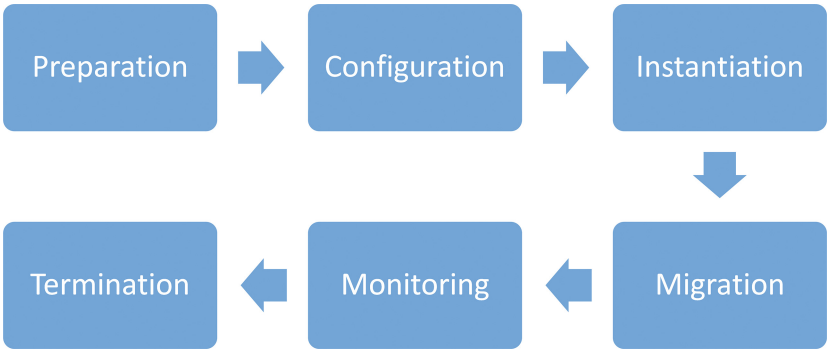

Fig. 2. Audit Agent System Architecture - Audit Workflow.

extracted from the policies in the preparation phase by the APM. The results that are produced by the evaluation agents are written back to the evidence store. A result can either positive (e.g., a message of proven compliance or the absence of a violation) or negative (e.g., a violation that is detected by the evaluation agent). Additionally the result is passed on to presenter agents that inform the auditor about the audit results. Currently the presenter agents can either display the audit result in a web-based dashboard or pass on the violation in a machine-readable format to other tools or services via a REST API. The whole of processor and presenter agents logically forms the EPP component. It is thereby irrelevant, where these agents are running as long as they are able communicate via a network, which helps in balancing the load that can be introduced with complex analysis mechanisms or the sheer amount of evidence data that needs to be analyzed. According to the complexity of task, due to the amount of obligations, or the volume of evidence to analyse, different verification processes may need to be considered for the evaluation agents, ranging from log mining, checking for predefined tokens or patterns, to automated analysers and automated reasoning upon the audit trail.

The processing or analysis of evidence consists of two steps:

1. Retrieve the appropriate information from Evidence store.
2. A verification process, which checks the correctness of recorded events according to defined obligations and authorizations.

Evidence Store. The ES is the central repository for storing evidence. Some of the more important characteristics of evidence are that they are associated with a policy for which they were collected and contain supporting information such as log entries collected by an agent, which points out a potential policy violation or incident. For each cloud tenant, there is a separate ES to ensure basic data protection principles are being adhered to by isolating tenants and their data. This addresses some of the confidentiality and privacy issues associated with a share data pool for potentially sensitive information.

There are several approaches to harmonizing the storage format for digital evidence that can be reused in the ES such as [15,26,29]. AAS uses a custom evidence format that is based on concepts described in [26,29].

Securing the transport and storage of evidence is a considerable challenge. The remainder of this paper focusses on how this is achieved in AAS by utilizing Insynd.

5 Audit Evidence Storage Requirements

In this Section, we present a comparison of general evidence attributes, how they apply in the context of evidence collection for cloud accountability audits and how the integration of Insynd solves key issues in evidence storage.

5.1 Requirements of Digital Evidence

In [19] the core principles of any evidence are described as:

Admissibility. Evidence must conform to certain legal rules, before it can be put before a jury.
Authenticity. Evidence must be tieable to the incident and may not be manipulated.
Completeness. Evidence must be viewpoint agnostic and tell the whole story.
Reliability. There cannot be any doubts about the evidence collection process and its correctness.
Believability. Evidence must be understandable by a jury.

These principles apply to common evidence as well as digital evidence. Therefore, the evidence collection process for audits has to consider special requirements, which help in addressing these attributes and ensure best possible validity in audits and applicability in court.

In Table 1 we present a mapping of the previously described evidence attributes and how they are supported by the integration of Insynd as a means of storing evidence records. We thereby focus on the key properties of Insynd as described in Sect. 3.

Table 1. Mapping the Impact of Insynd Properties to Evidence Attributes.

		Insynd	
		Forward Integrity and Deletion Detection	Publicly Verifiable Proofs
Evidence Store	Admissibility		
	Authenticity	✓	✓
	Completeness	✓	✓
	Reliability	✓	✓
	Believability		

Admissibility of digital evidence is influenced by the transparency of the collection process and data protection regulation. Digital evidence can be any kind of data (e.g., e-mail messages, social network messages, files, logs etc.). Insynd does not have any direct influence on the admissibility of the evidence stored in it.

Authenticity of digital evidence before court is closely related to the integrity requirement put on evidence records. Evidence may not be manipulated in any way and must be protected against any kind of tampering (willingly and accidentally). Insynd ensures that data cannot be tampered with once it is stored.

Completeness is not directly ensured by Insynd, but rather needs to be ensured by the evidence collection process as a whole. Especially important are the definition of which evidence sources provide relevant evidence that need to be considered during the collection phase. Insynd can complement the evidence collection process by providing assurance of that all data stored in the evidence store are made available as evidence, and not cherry-picked.

Reliability is indirectly supported by integrating necessary mechanisms into the evidence collection process, such as Insynd.

Believability of the collected evidence is not influenced by implemented mechanisms, but rather by the interpretation and presentation by an expert in court. This is due to judges and juries usually being non-technical, which requires an abstracted presentation of evidence. Insynd does not influence the believability in that sense.

5.2 Privacy Requirements

Not all requirements that a secure evidence storage has to fulfill can be captured by analyzing the attributes of digital evidence. Other aspects have to be taken into account to address privacy concerns. Protecting privacy in the process of evidence collection is utmost importance, since the collected data is likely to contain personal data. For cloud computing, one limiting factor may be whether or not the cloud provider is willing to provide deep insight into its infrastructure. Table 2 presents a mapping of privacy principles and properties of our evidence process.

Below we summarise some key privacy principles:

Confidentiality. of data evolves around mechanisms for the protection from unwanted and unauthorized access. Typically, cryptographic concepts, such as encryption, are use to ensure confidentiality of data.

Table 2. Mapping of Insynd properties to Evidence Collection Requirements.

		Insynd		
		Secrecy	Forward Unlinkability of Events	Forward Unlinkability of Recipients
Evidence Store	Confidentiality	✓	✓	✓
	Data Minimisation		✓	✓
	Purpose Binding			
	Data Retention	✓		

Data Minimization. states that the collection of personal data should be minimized and limited to only what is strictly necessary.

Purpose Binding. of personal data entails that personal data should only be used for the purposes it was collected for.

Retention Time. is concerned with how long personal data may be stored and used, before it needs to be deleted. These periods are usually defined by legal and business requirements.

Insynd and our evidence process provides various mechanisms that support these privacy principles.

Confidentiality. A central property of Insynd is that it is always encrypting data using public-key cryptography. By encrypting the evidence store, compromising the privacy of cloud customer data that has been collected in the evidence collection processes becomes almost impossible by attacking the evidence store directly. This goes as far as being able to safely outsource the evidence store to an untrusted third-party, a key property of Insynd [24].

Data Minimisation. Furthermore, Insynd provides forward unlinkability of events and client identifiers, as described in Sect. 3, which helps prevent several types of information leaks related to storing and accessing data. Collection agents are always configured for a specific audit task, which is very limited in scope of what needs to be collected. Agents are never configured to arbitrarily collect data, but are alway limited to a specific source (e.g., a server log) and data objects (e.g., a type of log events).

Purpose Binding. Neither Insynd nor our evidence process can directly influence the purpose for which collected data is used. Indirectly, the use of an evidence process like ours, incorporating secure evidence collection and storage, may serve to differentiate data collected for auditing purposes with other data collected e.g., for marketing purposes.

Retention time poses a real challenge. In cloud computing, the precise location of a data object is usually not directly available, i.e., the actual storage medium used to store a particular block is unknown, making data deletion hard. However, if data has been encrypted before storage, a reasonably safe way to ensure "deletion" is to discarding the key material required for decryption. Insynd supports forward-secure clients, where key material to decrypt messages are discarded as messages are read.

In Sect. 7, we also describe the threat model for the system described in this paper and present an evaluation of how Insynd is used to mitigate these threats.

6 Secure Evidence Storage Architecture

In this Section, we provide an architectural overview of the integration of Insynd into a secure evidence collection and storage process. We describe the overall architecture and its components, how the components of Insynd are mapped into the Audit Agent System and which setup process is required to use Insynd for securing evidence collection and storage.

6.1 Architecture

In this Section we discuss the architectural integration of Insynd as an evidence store in our audit system. There are basically three different components required to perform secure evidence collection. Figure 3 shows an overview of these components - *Evidence Source*, *Evidence Store* and *Evidence Processing* (see Sect. 4 and Fig. 1 for reference) - as well as the flow of data between them. From the various sources of evidence in the cloud, evidence records are collected that will be stored in the evidence store on a per-tenant basis. The evidence store is thereby located on a separate server. As previously mentioned, the server may be an untrusted third-party cloud storage provider. This is important to ensure so that this approach scales well with a growing number of tenants, evidence sources and evidence records.

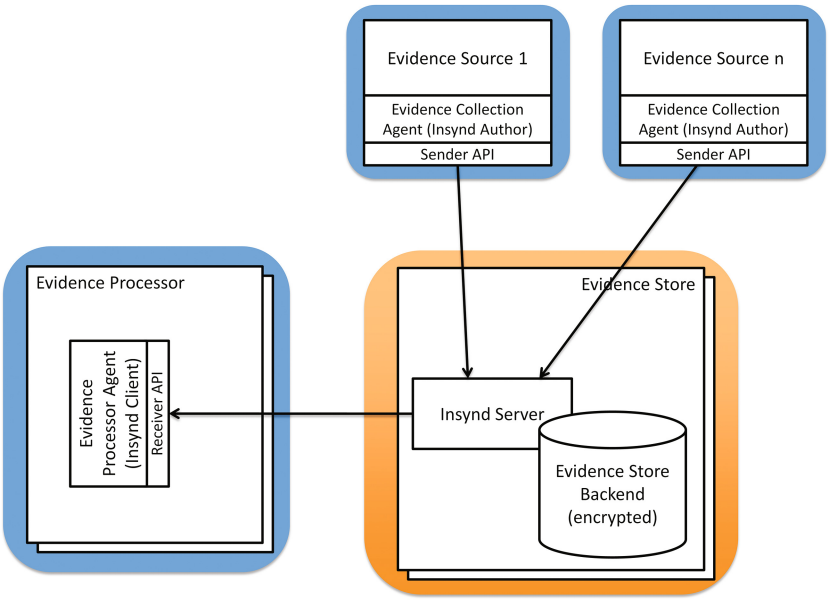

Fig. 3. Evidence Collection, Storage and Processing Workflow.

Evidence Collection. There are various evidence sources to be considered, such as logs, cryptographical proofs, documentation and many more. For each, there needs to be a suitable collection mechanism. For instance, a log parser for logs, a tool for cryptographical proofs or a file retriever for documentation. This is done by a software agent called *Evidence Collection Agent* that is specifically developed for the data collection from the corresponding evidence source. The collection agent acts as an *Insynd Author* meaning it uses the *Sender API* to store evidence into the Evidence Store. The encryption happens in the Sender

API. Typically, this agent incorporates or interfaces with a tool to collect evidential data, for instance forensic tools, such as file carvers, log parsers or simple search tools. Another type of collection agent have client APIs implemented to interface with more complex tools, such as Cloud Management Systems (CMS). Generally, these agents receive or collect information as input and translate that information into an evidence record, before storing it in the Evidence Store.

Evidence Storage. From the Evidence Collection Agent, evidence records are sent to the Evidence Store. The Evidence Store is implemented by the *Insynd Server*. Since Insynd functions as a key-value store for storing evidence records (encrypted messages identified by a key) NoSQL or RDBMS-based backend for persisting evidence records can be used. All data contained in the Evidence Store is encrypted. Each record is addressed to a specific receiver (e.g., an Evidence Processing Agent). The receiver's public key is used in the Sender API to encrypt the record on the Evidence Store. This means that only the receiver is able to access the evidence data from the Evidence Store. Isolation between tenants in a single Evidence Store is achieved by providing one container for each tenant where his evidence records are stored. However, even stronger isolation is also possible by providing a separate Evidence Store hosted on a separate VM. Additionally, Evidence records require a unique identifier in the Evidence Store to enable selective retrieval of records. In our implementation, we use a combination of a policy identifier and a rule identifier (where a rule is part of a policy) to enable the receiver to reduce the amount of records to receive to a manageable size.

Evidence Processing. Evidence Processing components are located at the receiving end of this workflow. The Receiver API is used by the processing agent (Insynd Client) to retrieve evidence records from the Evidence Store. The receiver can request multiple records from a period of time at once. The Client is also in possession of the corresponding private key to decrypt evidence records, which means records can only be decrypted at the Client.

6.2 Identity Management and Key Distribution

Since asymmetric encryption is such an important part of our system, we describe the encryption key distribution sequence next. In this software agent-based system, the automated setup of key material and registration with Insynd is particularly important. Figure 4 depicts the initialization sequence of collection and processing agents with a focus on key distribution.

In Fig. 4, we introduce an additional component beyond those already described in the general architecture: the *Controller*. The Controller serves as an entry point that controls the agent setup and distribution process in the audit system. It is an important part of the lifecycle management of the system's agents (e.g., creating and destroying of agents or migration between platforms).

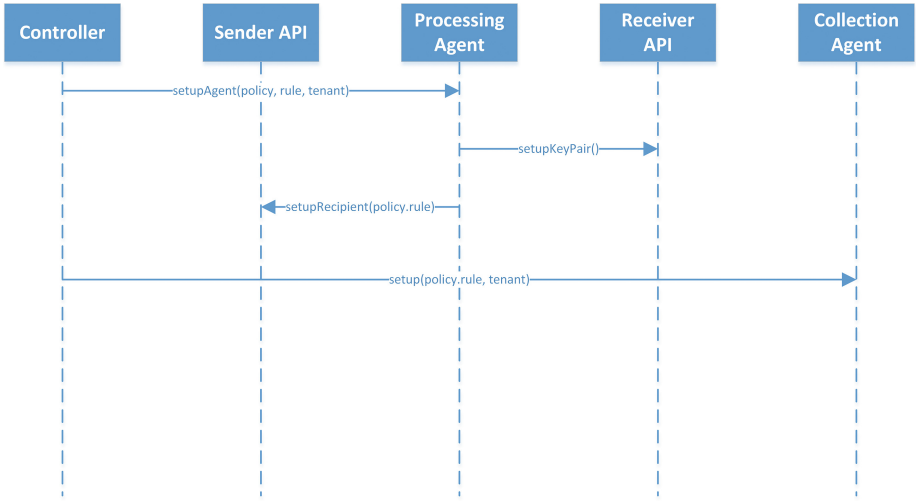

Fig. 4. Evidence Collection Setup Sequence.

In Fig. 4, we describe the initialization sequence for a simple scenario, where a particular tenant wishes to audit compliance with a policy and one rule included in that policy in particular. The following steps have to be performed to setup the evidence collection and storage process for that particular rule:

1. In the first step, a Processing Agent is created and configured according to the input policy and rule respectively for the tenant.
2. During the setup phase, the Processing Agent sets up a key pair at the Receiver API. The Receiver API is a RESTful service that holds private key material and is therefore located at the same servers hosting the Processing Agents (i.e., a trusted environment).
3. After the key material has been generated, the Processing Agent registers itself as a recipient at the Sender API. For this, it uses a unique identifier generated from the policy ID and the rule ID (i.e., *policyID.ruleID*).
4. In the last step, the Controller sets up the required Collection Agents and connects them with the corresponding Processing Agents by using the unique recipient identifier.

Now, it is possible for the Collection Agents to send evidence records to their corresponding Processing Agents. The messages will be encrypted at the Sender API service before storage, using the provided recipient's public key. The Processing Agent then pulls the evidence records from the Evidence Store using the Receiver API the records are decrypted using the receiver's private key.

7 Evaluation

In this Section we present an informal security evaluation of the system we have implemented for secure evidence collection. We describe the evidence

collection work flow using a fictitious scenario. By applying the evidence collection and storage process to the setting described in this scenario, we demonstrate how the requirements stated in Sect. 5 are addressed. Additionally, we provide a model that states threats and adversaries to the process as well as the mitigation functions introduced by Insynd.

In this scenario, the CCOMP company is a customer of the Infrastructure as a Service provider CloudIA. In particular, we analyze the security properties of the evidence collection process by looking at the data at rest as well as the data in transit protection at any time during the flow from the evidence source to its processor. We thereby assume that CloudIA is using OpenStack [21] as a its Cloud Management System (CMS), since this a widely popular open source CMS, which we use for developing our audit agent system. However, any other CMS could be used as well as long as it provides the needed monitoring interfaces.

7.1 Scenario

CloudIA is specialized in providing its customers with virtualized resources in the form of virtual machines, networks and storage. CCOMP has outsourced most of its IT services to CloudIA. Among them is a service that processes data of CCOMP's customers. For that data, CCOMP has to guarantee data retention. CCOMP has identified snapshots to be one major problem with respect to the data retention policy, since the virtual machine's storage is duplicated in the process. This means for CCOMP that in order to be compliant with the data retention policy, a snapshot of that virtual machine may have a maximum lifetime of one day, which limits its usefulness to e.g., backing up before patching. Now, we assume a trustworthy but sloppy administrator at CCOMP who creates a snapshot before patching software on the virtual machine, but then omits deleting the snapshot after he is done. However, an automated daily audit of its cloud resources was put in place by CCOMP to detect such compliance violations.

7.2 Implementation

The collection agent required for the above scenario communicates with our OpenStack CMS to gather evidence of the CMS behavior regarding virtual machine snapshots. The processing agent contains the logic for detecting snapshot violations (i.e., base virtual machine and a maximum age of the snapshot derived from the retention policy). The collection agent is deployed at the CMS controller node and has access to OpenStack's RESTful API. The processing agent is located on the same trusted host as the controller agent (see Fig. 3 for reference). The evidence store is located on a separate, untrusted virtual machine. Now, the following steps are performed:

1. The collection agent opens a connection to the OpenStack RESTful API on the same host and requests a history of snapshot events for CCOMP's virtual machine. Despite there being no communication over the network, HTTPS is

used to secure the communication between the collection agent and the CMS. Since the policy only requires information about snapshots to be collected, the CMS agent limits evidence record generation to exactly that information, nothing more.
2. The collection agent sets up the receiver of the evidence according to the process depicted in Fig. 4 and sends the collected records to the evidence store (Insynd). The communication channel is encrypted using HTTPS and the payload (evidence records) is encrypted with the receiving agent's public key.
3. The processing agent pulls records from the evidence store in regular intervals (e.g., every 24 h), analyses them and triggers a notification of a detected violation. The communication between the processing agent and the evidence store is secured using HTTPS.
4. In the last step, evidence records are deleted because their retention limit has been reached. This is done by discarding the keys required for decryption.

7.3 Threat Model

To demonstrate which security threats exist for the evidence collection process and Insynd is used to mitigate them, we describe the threat model for this system categorized according to the STRIDE [18] threat categorization:

- **S**poofing Identity
- **T**ampering with Data
- **R**epudiation
- **I**nformation disclosure
- **D**enial of Service
- **E**levation of Privilege

We have identified the following major threats to the evidence collection and storage process:

- *Unauthorized access to evidence (S,I)*: the protection of evidence from being accessed by unauthorized persons. Possible adversaries are a malicious third-party evidence storage provider (cloud service provider), another tenant (isolation failure) or an external attacker. Using Insynd for evidence collection and storage addresses this threat since recipients of messages are authenticated using appropriate mechanisms such as user credentials for API authentication and public keys for encryption.
- *Data leakage (S,I)*: the protection from unintentional data leakage. This could be caused by misconfiguration (e.g., unencrypted evidence being publicly available). Using Insynd for evidence collection and storage addresses this threat by encrypting data by default.
- *Eavesdropping, (T,I)*: the protection of evidence during the collection phase, especially in transit. Possibly adversaries are another tenant (isolation failure) or external attackers in case evidence is transported to an external storage provider or auditor. Using Insynd for evidence collection and storage addresses this threat by using transport layer as well as message encryption.

- *Denial of Service (D)*: the protection of the evidence collection and storage process from being attacked directly with the goal of disabling or shutting it down completely (e.g., to cover-up simultaneous attacks on another service). Possible adversaries are external attackers. This is a very generic threat that cannot be addressed by a single tool or control but rather requires a set a measures (on the network and application layer) to enhance denial of service resilience.
- *Evidence manipulation (T,R,I)*: the protection of evidence from intentional manipulation (e.g., deletion of records, changing of contents, manipulation of timestamps). Possible adversaries are malicious insiders and external attackers. Using Insynd for evidence collection and storage addresses this threat, since Insynd provides tampering and deletion detection.

Some of these threats can be mitigated by implementing appropriate security controls (i.e., using Insynd for evidence transport and storage). It provides effective protection by employing security techniques described in Sect. 3.

7.4 Requirements Evaluation

In this section, we evaluate the integration of Insynd against the requirements described in Sect. 5. In step 1 of the fictitious scenario, the data minimization principle is being followed because the specialized agent only collects evidence on the existence of snapshots.

This workflow is secure as soon as the collection agent inserts data into the evidence store in step 2. More precisely, evidence records are tamper-evident and encrypted. This is true, even though the evidence is actually stored on an untrusted virtual machine. The only way to compromise evidence now, is to attack the availability of the server hosting the Insynd server.

When the processing agent in step 3 retrieves records for evaluation, it can be assured of the authenticity of the data and that it has been provably collected by a collection agent. Since evidence records may be subject to maximum data retention regulation, records that are not needed anymore are deleted.

As previously mentioned in Sect. 6 we use JADE as an agent runtime. To secure our system against non-authorized agents, we use the TrustedAgents add-on for the JADE platform. This ensures that only validated agents are able to join our runtime environment. This effectively prevents agent injection attacks, where malicious agents could be inserted at either the collection or processing side to compromise our system.

As can be seen, the evidence records are protected all the way from the evidence source to the processing agent using only encrypted communication channels and having an additional layer of security (message encryption) provided by Insynd. Additionally, while the evidence is being stored, it remains encrypted.

7.5 Scalability

Obviously, since there is a vast amount of evidence sources and therefore a potentially equal number of collection agents, ensuring the scalability of the process and the implementation is very important. This has been considered very early in the design process by choosing an software agent-based approach for the system architecture. Software agents are inherently distributable and allow for complex message flow modeling in an infrastructure. Therefore, the core components evidence collection, storage and processing become distributable as well. In our future work, we'll focus on the scalability aspects. We will follow a methodology where we focus on the following technical key scalability indicators:

- Data transfer volume: amount of evidence data being transferred over the network
- Message volume: amount of evidence message transmissions over the network
- Storage volume: amount of storage required for evidence
- Encryption overhead: performance impact introduced by encryption and decryption

Based on the identified performance impact of each of these indicators, in the second step, we model different message flow optimization strategies to alleviate their impact and ensure scalability.

8 Conclusion and Future Work

In this paper, we presented our system design and implementation for secure evidence collection in cloud computing. The evidence provides the general basis for performing cloud accountability audits. Accountability audits take a large variety of evidence sources and data processing requirements into account.

We showed what the requirements for a secure evidence collection process are and demonstrated how these issues are addressed by incorporating Insynd into our system. We described how the core principles of digital evidence are addressed by our system. Additionally, we considered data protection principles for the evidence collection process, how they influence our approach and how they are addressed in our system by integrating Insynd. For this, we presented the relevant architectural parts of our prototype. Additionally, we provided an overview of how the evidence collection is integrated in our system for automated cloud audits.

In our future work, we will focus on the scalability of our audit system in general and the scalability of the components involved in evidence collection in particular. For that reason, we will focus on the distribution of the audit system and evidence collection not only in the same domain (i.e., in the same infrastructure), but also taking into account outsourcing and multi-provider collection scenarios.

Acknowledgements. This work has been partly funded from the European Commission's Seventh Framework Programme (FP7/2007–2013), grant agreement 317550, Cloud Accountability Project - http://www.a4cloud.eu/ - (A4CLOUD).

References

1. A4Cloud FP7 Project (2015). http://www.a4cloud.eu/
2. An, J.H.: Authenticated encryption in the public-key setting: security notions and analyses. IACR Cryptology ePrint Archive 2001, 79 (2001). http://eprint.iacr.org/2001/079
3. Ardagna, C.A., Bussard, L., Vimercati, S.D.C.D., Neven, G., Paraboschi, S., Pedrini, E., Preiss, S., Raggett, D., Samarati, P., Trabelsi, S., Verdicchio, M.: Primelife policy language (2009). http://www.w3.org/2009/policy-ws/papers/Trabelisi.pdf
4. Azraoui, M., Elkhiyaoui, K., Önen, M., Bernsmed, K., De Oliveira, A.S., Sendor, J.: A-PPL: an accountability policy language. In: Garcia-Alfaro, J., Herrera-Joancomartí, J., Lupu, E., Posegga, J., Aldini, A., Martinelli, F., Suri, N. (eds.) DPM/SETOP/QASA 2014. LNCS, vol. 8872, pp. 319–326. Springer, Heidelberg (2015). http://www.eurecom.fr/publication/4381
5. Bellare, M., Yee, B.: Forward-security in private-key cryptography. In: Joye, M. (ed.) CT-RSA 2003. LNCS, vol. 2612, pp. 1–18. Springer, Heidelberg (2003)
6. Bernstein, D.J., Lange, T., Schwabe, P.: The security impact of a new cryptographic library. In: Hevia, A., Neven, G. (eds.) LatinCrypt 2012. LNCS, vol. 7533, pp. 159–176. Springer, Heidelberg (2012). http://dx.doi.org/10.1007/978-3-642-33481-8_9
7. Bowers, K.D., Hart, C., Juels, A., Triandopoulos, N.: PillarBox: combating next-generation malware with fast forward-secure logging. In: Stavrou, A., Bos, H., Portokalidis, G. (eds.) RAID 2014. LNCS, vol. 8688, pp. 46–67. Springer, Heidelberg (2014). http://dx.doi.org/10.1007/978-3-319-11379-1_3
8. Dingledine, R., Mathewson, N., Syverson, P.F.: Tor: The second-generation onion router. In: Blaze, M. (ed.) Proceedings of the 13th USENIX Security Symposium, 9–13 August 2004, San Diego, CA, USA, pp. 303–320. USENIX (2004), http://www.usenix.org/publications/library/proceedings/sec04/tech/dingledine.html
9. Doelitzscher, F., Reich, C., Knahl, M., Passfall, A., Clarke, N.: An agent based business aware incident detection system for cloud environments. J. Cloud Comput. Adv. Syst. Appl. **1**(1), 9 (2012)
10. Doelitzscher, F., Ruebsamen, T., Karbe, T., Reich, C., Clarke, N.: Sun behind clouds - on automatic cloud security audits and a cloud audit policy language. Int. J. Adv. Netw. Serv. **6**(1,2), 1–16 (2013)
11. Gupta, A.: Privacy preserving efficient digital forensic investigation framework. In: 2013 Sixth International Conference on Contemporary Computing (IC3), pp. 387–392, August 2013
12. Haeberlen, A.: A case for the accountable cloud. In: Proceedings of the 3rd ACM SIGOPS International Workshop on Large-Scale Distributed Systems and Middleware (LADIS 2009), October 2009
13. JADE: Java Agent Developement framework (2015). http://jade.tilab.com
14. Jansen, W., Grance, T.: Sp 800–144. guidelines on security and privacy in public cloud computing. Technical report, National Institute of Standards and Technology, Gaithersburg, MD, United States (2011)
15. Jerman Blaič, A., Klobučar, T., Jerman, B.D.: Long-term trusted preservation service using service interaction protocol and evidence records. Comput. Stand. Interfaces **29**(3), 398–412 (2007). http://dx.doi.org/10.1016/j.csi.2006.06.004
16. Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., Leaf, D.: Nist cloud computing reference architecture (2011). http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909505

17. Lopez, J., Ruebsamen, T., Westhoff, D.: Privacy-friendly cloud audits with somewhat homomorphic and searchable encryption. In: 2014 14th International Conference on Innovations for Community Services (I4CS), pp. 95–103, June 2014
18. Microsoft Developer Network: The Stride Threat Model (2015). https://msdn.microsoft.com/en-US/library/ee823878(v=cs.20).aspx
19. Mohay, G.M., Anderson, A.M., Collie, B., de Vel, O., McKemmish, R.D.: Computer and Intrusion Forensics. Artech House, Boston (2003). http://eprints.qut.edu.au/10849/. For more information about this book please refer to the publisher's website (see link) or contact the authors
20. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. Consulted **1**(2012), 28 (2008)
21. OpenStack: Openstack (2015). http://www.openstack.org/
22. Pearson, S.: Toward accountability in the cloud. IEEE Internet Comput. **15**(4), 64–69 (2011)
23. Pulls, T., Peeters, R.: Balloon: a forward-secure append-only persistent authenticated data structure. In: Pernul, G., Y A Ryan, P., Weippl, E., Torres, C.F., Jonker, H., Mauw, S., Diao, W., Liu, X., et al. (eds.) ESORICS. LNCS, vol. 9327, pp. 622–641. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24177-7_31
24. Pulls, T., Peeters, R.: Insynd: secure one-way messaging through Balloons. Cryptology ePrint Archive, Report 2015/150 (2015)
25. Pulls, T., Peeters, R., Wouters, K.: Distributed privacy-preserving transparency logging. In: Sadeghi, A.R., Foresti, S. (eds.) WPES, pp. 83–94. ACM (2013)
26. R. Brandner, U.P., Gondrom, T.: Evidence record syntax (ERS) (2014). http://tools.ietf.org/html/rfc4998
27. Redfield, C. M., Date, H.: Gringotts: securing data for digital evidence. In: 2014 IEEE Security and Privacy Workshops (SPW), pp. 10–17, May 2014
28. Ruebsamen, T., Reich, C.: Supporting cloud accountability by collecting evidence using audit agents. In: 2013 IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom), vol. 1, pp. 185–190, December 2013
29. Turner, P.: Unification of digital evidence from disparate sources (digital evidence bags). Digit. Investig. **2**(3), 223–228 (2005). http://dx.doi.org/10.1016/j.diin.2005.07.001
30. Weitzner, D.J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., Sussman, G.J.: Information accountability. Commun. ACM **51**(6), 82–87 (2008). http://doi.acm.org/10.1145/1349026.1349043
31. Zhang, R., Li, Z., Yang, Y., Li, Z.: An efficient massive evidence storage and retrieval scheme in encrypted database. In: 2013 International Conference on Information and Network Security (ICINS 2013), pp. 1–6, November 2013