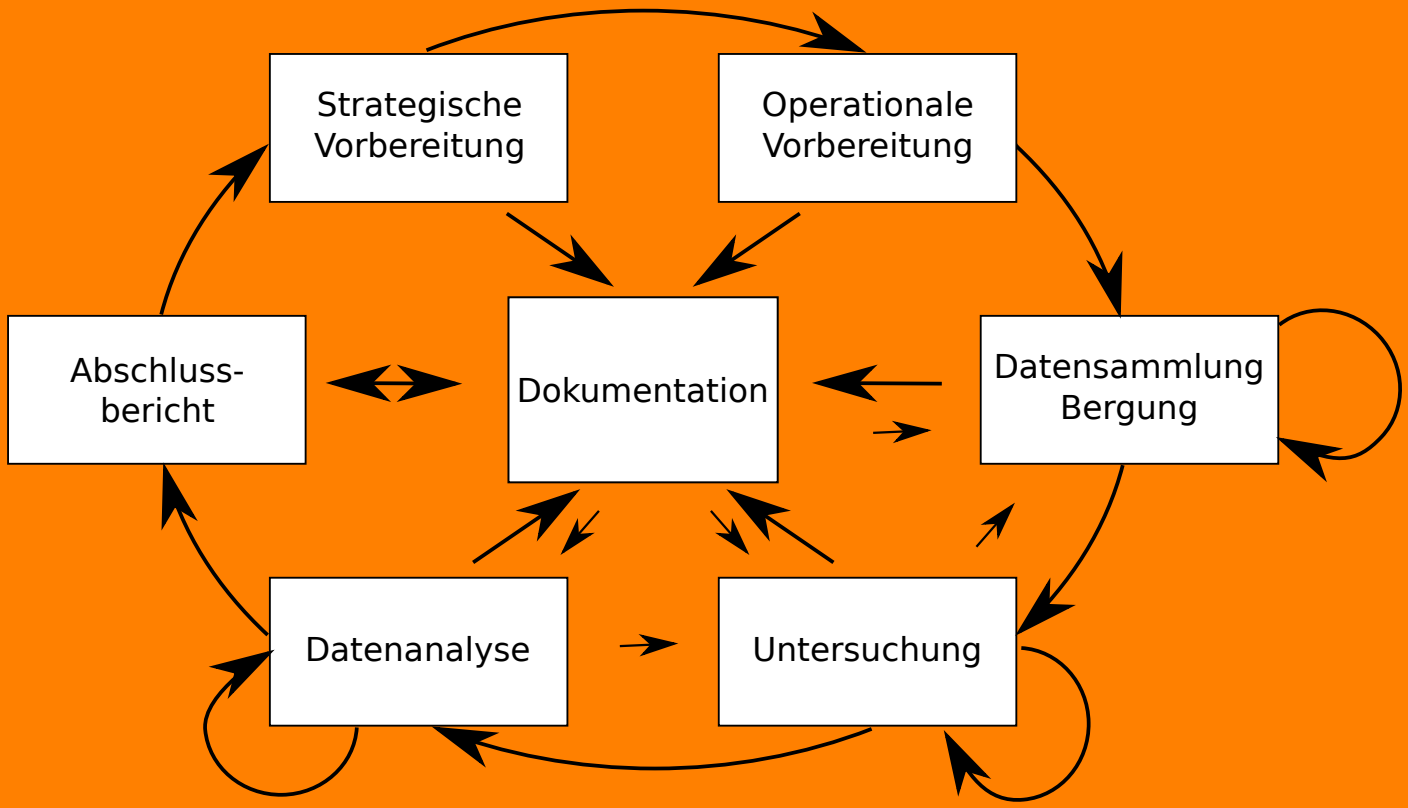
Open C3S
Open Competence Center for Cyber Security

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik



Mikromodul 8003: Grundlagen von Cloud-Forensik

Autoren:
Prof. Dr. Hans P. Reiser
Noëlle Rakotondravony
Johannes Köstler

Universität Passau

Mikromodul 8003: Grundlagen von Cloud-Forensik

Autoren:

Prof. Dr. Hans P. Reiser

Noëlle Rakotondravony

Johannes Köstler

---

1. Auflage

Universität Passau

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

AUFSTIEG DURCH BILDUNG >>
OFFENE HOCHSCHULEN

Universität Passau
Fakultät für Informatik und Mathematik
Innstraße 43
94034 Passau

1. Auflage (4. Mai 2017)

Um die Lesbarkeit zu vereinfachen, wird auf die zusätzliche Formulierung der weiblichen Form bei Personenbezeichnungen verzichtet. Wir weisen deshalb darauf hin, dass die Verwendung der männlichen Form explizit als geschlechtsunabhängig verstanden werden soll.

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 16OH12025 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

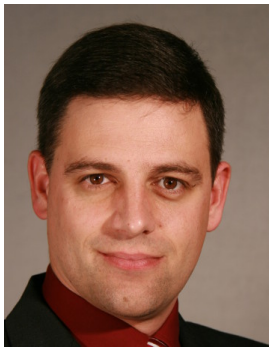
Inhaltsverzeichnis

Einleitung

I. Abkürzungen der Randsymbole und Farbkodierungen

Definition	D
Übung	Ü

II. Zu den Autoren



Hans P. Reiser ist Juniorprofessor für Sicherheit in Informationssystemen an der Universität Passau. Schwerpunkte seiner Arbeitsgruppe sind die Weiterentwicklung von Konzepten und Systemen aus dem Bereich der fehler- und einbruchstoleranten Replikation, die frühzeitliche und umfassende Erkennung von Sicherheitsproblemen und Sicherheitsvorfällen in Cloud-Umgebungen sowie die Erforschung neuartiger Sicherheitskonzepte auf Hypervisorebene.

Noëlle Rakotondravony ist seit September 2015 wissenschaftliche Mitarbeiterin in der Arbeitsgruppe von Prof. Hans P. Reiser an der Universität Passau.

Johannes Köstler ist seit Mai 2015 wissenschaftlicher Mitarbeiter in der Arbeitsgruppe von Prof. Hans P. Reiser an der Universität Passau.

Mikromodul: Grundlagen von Cloud-Forensik

1 Lernziele

Nach Bearbeitung dieses Mikromoduls sind Sie mit den grundlegenden Modellen und Vorgehensweisen der IT-Forensik vertraut. Sie haben ein tieferes Verständnis für die Herausforderungen hinsichtlich IT-Forensik in Cloud-Umgebungen entwickelt. Neben organisatorischen und juristischen Aspekten sind Sie hierbei insbesondere mit den technischen Problemstellungen und dem Stand der Forschung zu Lösungsansätzen vertraut. Sie haben einen Einblick in aktuelle Entwicklungen hinsichtlich Forensics as a Service und Forensic-Readiness-Modellen.

2 Grundlagen von IT-Forensik

2.1 Historische Entwicklung

Die unmittelbare Aufgabe von Forensik im Allgemeinen besteht in der Beantwortung der Fragen nach dem *Was*, *Wo*, *Wann* und *Wie* eines konkreten Vorfalls. Grundsätzlich verfolgt werden dabei mehrere Ziele: Neben der eigentlichen detaillierten Analyse eines Vorfalls (welche Vorgänge haben während des Vorfalls stattgefunden?) geht es einerseits um die Identifikation des Verursachers bzw. Täters und die Sammlung von verwertbaren Beweismitteln, andererseits um die Prävention gleichartiger Vorfälle in der Zukunft.

Erste Fälle zu IT-basierten kriminellen Aktivitäten gehen zurück bis in die 60er-Jahre. Nennenswert ist ein Bericht aus dem Buch „Fighting Computer Crime“ von Parker (1998), das einen Vorfall aus dem Jahr 1966 beschreibt. Dabei manipulierte ein für die Programmierung und Wartung der IT-Systeme einer Bank Beauftragter diese so, dass er sein eigenes Bankkonto unbemerkt überziehen und sich so einen finanziellen Vorteil verschaffen konnte.

Die Ursprünge der systematischen *IT-Forensik* gehen zurück auf die früher 80er-Jahre und die Anfangszeit der Personal Computer. Erst durch die massiv wachsende Verbreitung von PCs wurde die Herausforderung, digitale Spuren auf einem IT-System als Beweismittel zu sichern, zu einem häufigen, praxisrelevanten Problem. Als „Vater der IT-Forensik“ gilt hier Michael Anderson, der als Mitarbeiter des Internal Revenue Service – Criminal Investigation Division (IRS-CID) in den USA eine tragende Rolle bei der Entwicklung von Ausbildungskursen und von Software-Werkzeugen zur Beweissicherung von IT-Hardware innehatte. Im Jahr 1988 wurde in den USA die IACIS (International Association of Computer Investigative Specialists) gegründet. Aus dieser Organisation ging bald darauf ein Ausbildungskonzept für *SCERS* (Seized Computer Evidence Recovery Specialists) hervor. Bereits im Jahr 1993 fand die erste wissenschaftliche Konferenz zu diesem Thema statt.

In Deutschland wurde später das Thema IT-Forensik vom Bundesamt für Sicherheit in der Informationstechnik aufgegriffen. 2010 erschien erstmals der *Leitfaden „IT-Forensik“* des BSI, der das Ziel verfolgt, ein praxistaugliches Modell für die IT-Forensik zu definieren. Auf dieses und andere Modelle gehen wir detaillierter in den nächsten Unterkapiteln ein.

2.2 Grundlegende Definition der IT-Forensik

Eine oft zitierte Definition zu digitaler Forensik wurde 2001 in Zusammenarbeit zahlreicher Experten im Rahmen des First Digital Forensic Research Workshop (DFRWS) aufgestellt (siehe Definition 1):

D

Definition 1: Digital Forensic Science (DFRWS, 2001)

„The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."

Hinter dieser Definition steckt die Motivation, digitale Forensik als wissenschaftliche Disziplin zu etablieren und Forschung zu diesem Thema voranzubringen. Ziel ist es also, mit wissenschaftlichen Methoden Vorgehensweisen zu entwickeln und zu etablieren, die dann geeignet sind, kriminelle Handlungen zu rekonstruieren und auch unerlaubte Handlungen vorherzusehen und damit entgegenwirken zu können.

Auch kommt in dieser Definition zum Ausdruck, dass für forensische Untersuchungen eine systematische Vorgehensweise erforderlich ist. In der Literatur wurden von verschiedenen Autoren inzwischen viele zwar ähnliche, sich aber in Details unterscheidende Vorgehensmodelle entwickelt. In diesem Kapitel werden wir einzelne Beispiele hierzu aufgreifen. Die in obiger Definition genannten Prozessschritte der Bewahrung, Sammlung, Validierung, Identifikation, Analyse, Interpretation, Dokumentation und Präsentation von Beweismitteln lassen sich aber in unterschiedlichem Detailgrad in allen Modellen wiederfinden.

Hinsichtlich des Zwecks forensischer Untersuchungen lassen sich zwei sich überlappende Gebiete unterscheiden: Zum einen spielt Forensik eine zentrale Rolle im Bereich der Strafverfolgungsbehörden. Als Ergebnis stehen hier gerichtsverwertbare Beweise im Vordergrund. Zum anderen ist Forensik auch für interne Untersuchungen in Organisationen von großer Bedeutung. Hier gewinnt dann auch der Aspekt des Schutzes der eigenen Infrastruktur in Form von *Incidence Response* und Prävention eine größere Bedeutung. Einzelne Prozessmodelle sind daher auch vom Zweck der Untersuchungen beeinflusst. In diesem Studienbrief betrachten wir exemplarisch in Abschnitt 2.3 das Modell von Casey (2004), das auf den ersten Aspekt fokussiert ist, sowie in Abschnitt 2.4 das BSI-Modell (BSI, 2011), das stärker auf den zweiten Aspekt zugeschnitten ist.

Die grundlegenden Prinzipien hinter digitaler Forensik lassen sich nach Casey (2004) wie folgt charakterisieren:

- *Eignung (Soundness):* Dieses Prinzip wird in der Praxis oft dadurch beschrieben, dass durch die forensische Untersuchung Datenspuren nicht verändert werden dürfen. Casey (2004) weist aber zurecht darauf hin, dass der absolute Ausschluss von Veränderung zu kurz greift. Selbst bei klassischer Forensik ist dies keinesfalls immer gewährleistet. Beispielsweise werden bei der DNA-Analyse Proben entnommen und damit das Ursprungsmaterial verändert. Wichtig ist vielmehr eine exakte Dokumentation der Herkunft und des Umgangs mit Beweismitteln. Eine Veränderung von Datenspuren ist zu minimieren, muss aber nicht grundsätzlich ausgeschlossen werden, wenn diese nachvollziehbar ist.

- *Authentisierung (Authentication):* Forensische Spuren müssen angelehnt an Casey (2004) drei Eigenschaften erfüllen, um als authentisch zu gelten: Sie müssen tatsächlich von der Quelle stammen, von der sie vorgeben herzustammen, sie dürfen nicht verändert worden sein und alle zusätzlichen Informationen, die nicht direkt von der Quelle stammen (wie der Zeitpunkt der Datensammlung), müssen akkurat sein.
- *Beweismittelkette (Chain of Custody):* Im Zusammenhang mit der Authentizität ist die Beweismittelkette ein zentraler Aspekt. Zum Nachweis, dass keine Veränderungen erfolgt sind, ist durchgehend zu dokumentieren, woher die Beweismittel stammen, wo und wie sie aufbewahrt wurden und wer wann und warum Zugriff darauf hatte.
- *Integrität (Integrity):* Auch die Integrität kann als Teilaspekt der Authentizität gesehen werden. Technisch finden wir hierzu in der Regel kryptographische Hashes über Daten, mit denen nachgewiesen werden kann, dass Daten seit der Erhebung nicht verändert wurden.
- *Objektivität (Objectivity):* Ein zentraler Aspekt von forensischen Untersuchungen ist auch deren Objektivität. Die Interpretation und Präsentation von Ergebnissen sollte möglichst unvoreingenommen erfolgen und nur klare Tatsachen ohne subjektive Interpretation darstellen.
- *Wiederholbarkeit (Repeatability):* Ein weiterer Aspekt einer wissenschaftlichen Herangehensweise ist die Wiederholbarkeit. Aus originären Daten abgeleitete Analyseergebnisse sollen sich durch Dritte reproduzieren und dadurch verifizieren lassen.

Eine forensische Untersuchung, welche diesen Prinzipien gerecht wird, erfordert eine systematische Vorgehensweise. Wie bereits erwähnt, finden sich in der Literatur verschiedene Modelle für eine derartige Vorgehensweise. In der Praxis ist es weniger wichtig, welches Modell nun genau verwendet wird. Viel wichtiger ist es, dass es ein Modell gibt, nach dem die Vorgehensweise systematisch gestaltet wird.

Der Schwerpunkt dieses Studienbriefes liegt auf der Vermittlung der technischen Seite der Gewinnung von forensischen Daten, insbesondere bei cloudbasierten Systemen. Dabei ist stets zu beachten, dass diese technischen Vorgänge in einen übergeordneten forensischen Prozess eingeordnet sind. Dies ist vor allem dann der Fall, wenn Datenspuren gesammelt werden, die später z. B. vor Gericht als Beweismittel anerkannt werden sollen.

2.3 Casey-Modell

Eines der am meisten zitierten Vorgehensmodelle der digitalen Forensik stammt von Casey (2004). Dieses sehr feingranular untergliederte Modell ist in Abbildung 1 abgebildet. Während die Darstellung als Treppe suggeriert, dass es sich hierbei um eine lineare Abfolge von Prozessschritten handelt, kann es bei einer Untersuchung notwendig sein, einzelne Schritte abweichend von dieser Ordnung auch mehrfach durchzuführen, wenn beispielsweise neue Erkenntnisse gewonnen werden und daher zusätzlichen Spuren nachgegangen werden muss.

Der Auslöser für eine forensische Untersuchung ist in diesem Modell als *Alarm* oder *Anschuldigung* explizit aufgeführt. Bei einer ersten Einschätzung der Situation kann *bewertet* werden, ob der zu erwartende Nutzen überhaupt den Aufwand einer weitergehenden Untersuchung wert ist, oder der Vorgang bereits an dieser Stelle beendet wird. Der Schritt der *Incident/Crime scene protocols* dient dem „Einfrieren“ der vorliegenden Situation, um die Grundlage für unverfälschte Beweissicherung

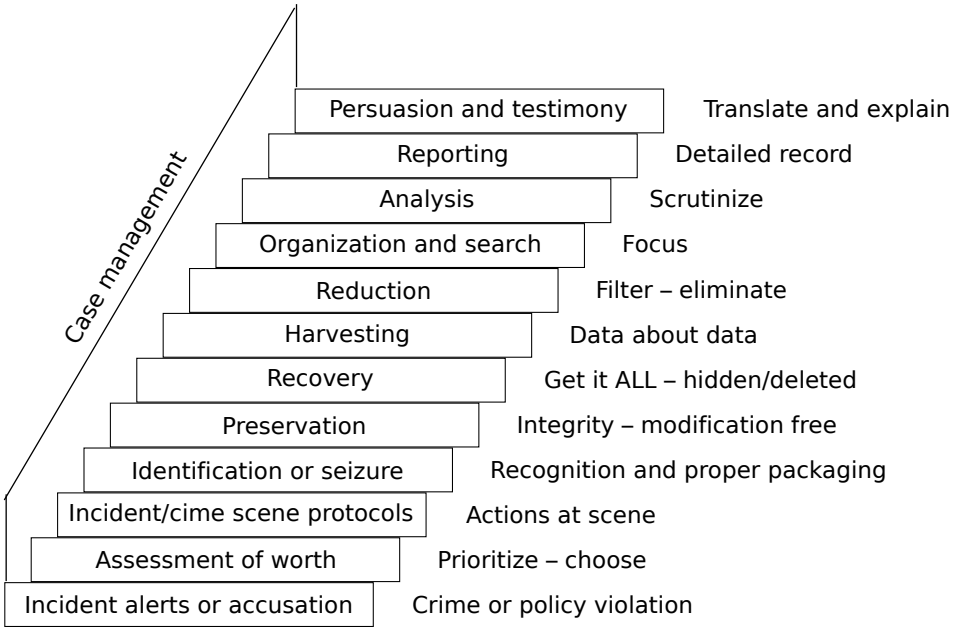


Abb. 1: Vorgehensmodell nach Casey (2004)

zu schaffen. Bei traditionellen Szenarien erfolgt auf dieser Basis im Schritt *identification or seizure* sehr oft die (physische) Beschlagnahmung von Geräten u.ä. Hier dürfte bereits offensichtlich sein, dass in global verteilten virtuellen Systemen (beispielsweise bei Datenspuren, die weltweit in mehreren Rechenzentren von Cloud-Anbietern verteilt sind) der Bedarf nach anderen Vorgehensweisen besteht.

Nach diesen ersten Schritten, die man auch als taktische Vorbereitung umschreiben kann, beginnt der technische Aspekt der Spurensammlung. Es gilt, Datenspuren zu erhalten (*Preservation*), also z. B. durch Kopien oder durch kryptographische Verfahren vor Veränderung zu schützen. Bei der *Recovery* wird angestrebt, möglichst auch alle versteckten oder gelöschten Spuren wiederherzustellen. Dies bildet die Grundlage, um zunächst möglichst vollständig alle potentiell interessanten Daten zu erfassen (*Harvesting*).

Nach diesen Schritten der Datensammlung folgen Vorgehenschritte, die der Auswertung der Daten dienen. Hierbei werden aus einer großen Datenmenge die relevanten Teile herausgefiltert (*Reduction*) und die resultierenden Ergebnisse strukturiert und aufbereitet (*Organization and search*). Auf dieser Basis erfolgt eine vertiefende Analyse der gewonnenen Daten (*Analysis*).

Das Resultat der Untersuchungen führt zur abschließenden Dokumentation der Ergebnisse (*Reporting*). Diese Dokumentation kann dann herangezogen werden, um Dritte von Ergebnissen zu überzeugen (*Persuasion and testimony*).

2.4 BSI-Modell

Auch das BSI definiert ein Vorgehensmodell für den forensischen Prozess, das sich im Wesentlichen in drei Bausteine gliedert:

- Zeitlicher Ablauf: Dieser Baustein beschreibt die zeitliche Abfolge und ist somit das Äquivalent zum zeitlichen Ablauf im Casey-Modell.
- Grundlegende Methoden: Dieser Baustein beschreibt grundlegende Methoden, wie auf unterschiedlichen Ebenen des Systems (wie Betriebssystem,

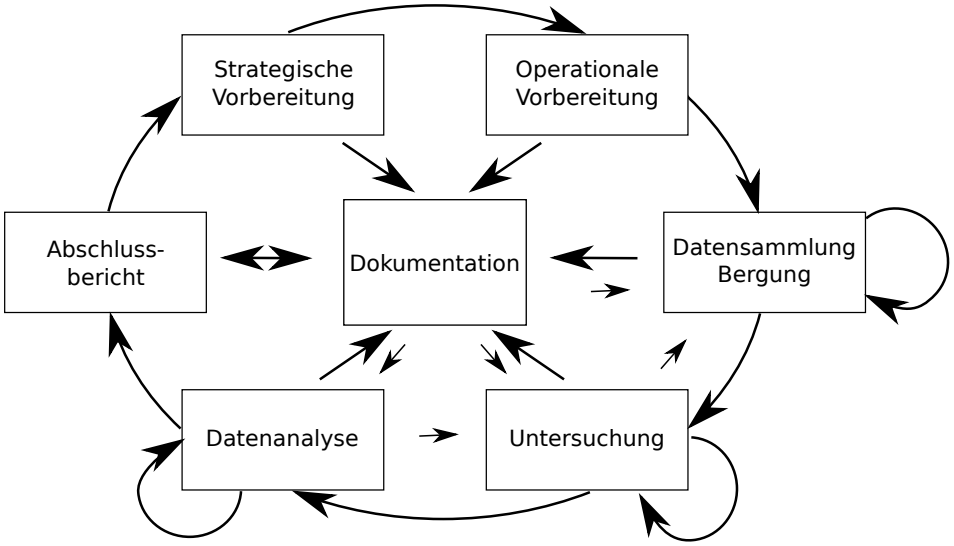


Abb. 2: Vorgehensmodell für den forensischen Prozess nach BSI (2011)

Dateisystem, Anwendungsebene) forensische Daten gewonnen werden können und ordnet diese den einzelnen Phasen des zeitlichen Ablaufs zu.

- Forensische Datenarten: Mit dem Ziel, forensisch wertvolle Daten strukturiert zu modellieren, werden in diesem Baustein zu allen grundlegenden Methoden forensisch bedeutende Datenarten beschrieben und klassifiziert.

Der zeitliche Ablauf nach dem BSI-Modell ist in Abbildung 2 dargestellt. Gegenüber dem Treppenmodell von Casey bevorzugt dabei das BSI die Darstellung als zyklischen Prozess. Auch hier gilt wieder, dass keine strikt sequentielle Abfolge an Schritten vorliegt, sondern dass diese bei Bedarf sich auch wechselseitig beeinflussen können.

Was das BSI-Modell von einigen anderen Modellen unterscheidet, ist die explizite Nennung der *strategischen Vorbereitung*. Darunter versteht man alle Maßnahmen, die vor einem Vorfall getroffen werden, um auf diesen besser vorbereitet zu sein. Hierin kann man erkennen, dass das BSI-Modell auch insbesondere auf die interne Aufklärung von Vorfällen bzw. den Schutz der eigenen Infrastruktur zielt. Beispielsweise zählt zur strategischen Vorbereitung die Dokumentation der vorhandenen IT-Infrastruktur und die Aktivierung von Logging-Mechanismen der eingesetzten Software. Im Gebiet der Strafverfolgung dagegen sind die Möglichkeiten der strategischen Vorbereitung oft stärker begrenzt.

Zur *operationalen Vorbereitung* gehören dagegen alle Maßnahmen, die nach einem Vorfall, aber vor der eigentlichen Datensammlung geschehen, wie die Planung der Vorgehensweise und die Identifikation von Datenquellen.

Daran schließt sich die *Datensammlung* an, bei der Daten möglichst unverfälscht und integritätsgeschützt erfasst werden.

Die Auswertung der Daten wird im BSI-Modell in zwei Schritte unterteilt. Während man bei der *Untersuchung* der Daten zunächst forensisch relevante Informationen aus den Datenquellen extrahiert, findet eine detaillierte Untersuchung dieser Informationen (beispielsweise Korrelation zwischen mehreren Datensätzen) im Schritt der *Datenanalyse* statt.

Die Maßnahmen zur *Dokumentation* werden in *prozessbegleitende* und *abschließende Dokumentation* unterteilt. Die prozessbegleitende Dokumentation besteht in der

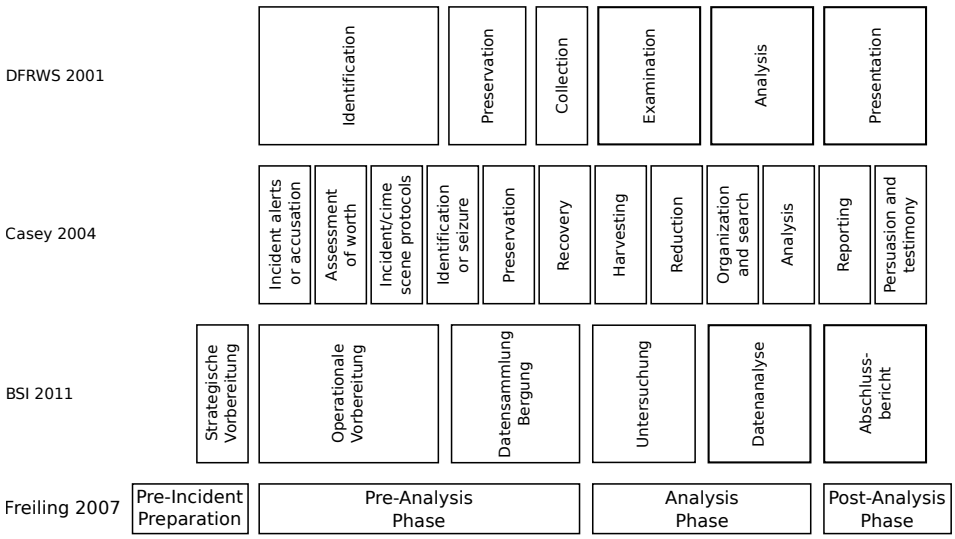


Abb. 3: Gegenüberstellung der Terminologie der betrachteten Modelle

kontinuierlichen Dokumentation parallel zu allen anderen ausgeführten Schritten. In der abschließenden Dokumentation werden die gesammelten Ergebnisse für Dritte aufbereitet, sodass aus den einzelnen Teilergebnissen ein Gesamtbild entsteht.

Für Details zu den anderen beiden Dimensionen wird an dieser Stelle auf den Leitfaden „IT-Forensik“ (BSI, 2011) verwiesen. Bei den grundlegenden Methoden wird vom BSI für unterschiedliche Schichten des Systems ausführlich beschrieben, welche Informationen dort gewonnen werden können und welche Relevanz diese Ebenen in den einzelnen Phasen des Vorgehensmodells haben. Bei den forensischen Datenarten erfolgt eine systematische Einteilung der vorhandenen, forensisch relevanten Datenquellen.

2.5 Fazit

Es besteht Konsens in der Forschung, dass die Untersuchung von Vorfällen einer systematischen Herangehensweise bedarf. Hierzu wurden von unterschiedlichen Autoren unterschiedliche Vorgehensmodelle entworfen. Abbildung 3 stellt nochmal die anfangs kurz beschriebene Vorgehensweise nach DFRWS (2001) sowie Casey (2004) und BSI (2011) gegenüber. Zusätzlich ist das verallgemeinernde und weniger feingranulare „*Common Process Model*“ nach Freiling und Schwittay (2007) dargestellt.

In der Praxis ist es weniger wichtig, welches Modell zum Einsatz kommt. Viel wichtiger ist die Erfordernis, dass auf jeden Fall ein systematisches Prozessmodell verwendet wird.

3 Herausforderungen der Cloud-Forensik

Als eine der ersten Arbeiten hat Beebe (2009) in einem Artikel auf das Problem hingewiesen, dass Cloud Computing im Bereich IT-Forensik bis dahin keine Beachtung gefunden hatte. Seitdem sind zwar einige Forschungsarbeiten zu dieser Problematik entstanden, aber noch immer gilt, dass Forensik in Cloud-Umgebungen aus juristischer, aus organisatorischer sowie auch aus technischer Sicht noch offene Probleme aufweist.

Auf die komplexen Herausforderungen aus *juristischer Sicht* kann in diesem Studienbrief nicht eingegangen werden. Es dürfte offensichtlich sein, dass sich in

der Praxis viele Probleme dadurch ergeben, dass Cloud-Dienste oft länderübergreifend zum Einsatz kommen und somit unterschiedlichsten, länderspezifischen juristischen Rahmenbedingungen unterliegen. Auch sind gerade in öffentlichen Cloud-Umgebungen relevante forensische Daten vermischt mit Daten unbeteiligter Dritter gespeichert, was weitere Herausforderungen mit sich bringt. Exemplarisch verweisen wir für eine detaillierte Betrachtung auf Heinson (2015).

Wenn man inhouse betriebene private Cloud-Infrastrukturen beiseite lässt, ergibt sich die erste große *organisatorische Herausforderung* daraus, dass stets mindestens zwei Parteien, der Cloud-Kunde und der Cloud-Betreiber involviert sind. Etablierte Vorgehensweisen forensischer Untersuchungen sind auf die lokale Sammlung relevanter Informationen ausgelegt und ungeeignet, Daten bei einem entfernten, oft global operierenden Cloud-Anbieter zu sammeln. Selbst wenn ein physischer Zugriff auf die Infrastruktur eines Cloud-Anbieters möglich sein sollte, ist eine Untersuchung ohne die aktive Mithilfe des Cloud-Anbieters kaum möglich.

Neben den organisatorischen Herausforderungen lassen sich in den einzelnen Service- und Organisationsmodellen von Cloud Computing auch zahlreiche technische Schwierigkeiten identifizieren. Einen guten Überblick über diese *technischen Herausforderungen* liefert ein aktueller Forschungsartikel von Alqahtany et al. (2015), an dem sich die nachfolgende Diskussion von Herausforderungen orientiert.

3.1 Identifikationsphase

Die ersten Herausforderungen bestehen nach Alqahtany et al. (2015) in der *Identifikationphase (identification stage)*, die man im BSI-Modell der Phase der *operationalen Vorbereitung* zuordnen kann.

In dieser ersten Phase stellt sich zunächst der *fehlende Zugriff* auf die einzelnen Schichten des Systems als Herausforderung. Während im IaaS-Modell aus Perspektive des Cloud-Kunden noch Kontrolle über einen großen Teil der Infrastruktur vorhanden ist, reduziert sich diese Kontrolle im PaaS-Modell und noch mehr im SaaS-Modell. Auf technischer Ebene sind in der Praxis kaum Schnittstellen vorhanden, mit denen Cloud-Anbieter Zugriff für forensische Untersuchungen ermöglichen können.

Eine wichtige Quelle von forensisch relevanten Informationen sind Log-Dateien. Der *Zugriff auf Log-Dateien* kann allein schon dadurch erschwert werden, dass nicht bekannt ist, wo in einer über mehrere Standorte und Länder verteilten Cloud Daten gespeichert und Dienste ausgeführt werden. Zwar können im IaaS-Modell dem Cloud-Kunden innerhalb seiner virtuellen Maschinen erstellte Logs zugänglich sein, im PaaS- und insbesondere SaaS-Modell ist dies aber generell nicht der Fall. Als mögliche Datenquellen kommen in aktuellen Cloud-Infrastrukturen vom Cloud-Anbieter bereitgestellte Monitoring-Dienste in Betracht, wie wir sie in Abschnitt 5.2 näher betrachten.

Eine große Herausforderung ist auch die *Volatilität* von Diensten und Daten in Cloud-Umgebungen. Während die bedarfsgerechte Bereitstellung von Cloud-Ressourcen und deren Elastizität oft als großer Vorteil von Cloud Computing genannt werden, so sind diese aus Sicht der digitalen Forensik ein großes Problem: Ressourcen jeglicher Art (wie Speicherplatz, virtuelle Maschinen und Dienste) können genauso einfach gelöscht werden, wie sie für einen Cloud-Kunden bereitgestellt werden. Daten, die nur im Hauptspeicher einer virtuellen Maschine und nicht auf einem persistentem Medium vorhanden waren, sind nach Löschung der virtuellen Maschine verloren. Auch bei persistenten Medien ist die Wahrscheinlichkeit hoch, dass nicht mehr benötigte Ressourcen für Daten anderer Cloud-Kunden genutzt

werden und es damit selbst auf den physischen Datenträgern keine verwertbaren Spuren mehr gibt.

3.2 Datensammlung

Selbst wenn forensisch relevante Datenspuren in einer Cloud vorhanden sind, ist eine Sammlung dieser Daten ohne *Mitwirkung des Cloud-Providers* kaum durchführbar. Der konkrete Ort, wo Daten und Dienste in einer Cloud vorhanden sind, wird von der Cloud-Infrastruktur dynamisch verwaltet und lässt sich, wenn überhaupt, nur mittels der Verwaltungssysteme des Providers nachvollziehen. Die erforderliche Interaktion mit dem Provider kann zu zeitlichen Verzögerungen führen.

Zudem stellt sich die Frage nach der *Integrität* von Daten, die ggf. von einem Provider auf Anfrage bereitgestellt werden. Selbst wenn mit Hilfe des Providers Daten gewonnen werden können, ist fraglich, ob dabei interne Vorgänge nach den Anforderungen einer forensischen Untersuchung durchgeführt und dokumentiert werden. Der Beweiswert so gewonnener Datenspuren kann also fragwürdig sein.

Eine zentrale Eigenschaft von Cloud Computing ist *resource pooling*, also die gemeinsame Nutzung von Ressourcen durch mehrere Cloud-Kunden. Dies erschwert eine *Isolation* des untersuchten Systems bei der Datensammlung. Eine Isolation ist zum einen erforderlich, um eine Veränderung von Spuren durch andere Cloud-Nutzer zu verhindern. Zum anderen kann es je nach Rahmenbedingungen der Untersuchung auch erforderlich sein, keine Daten von unbeteiligten Dritten zu sammeln.

Eine weitere Herausforderung bei der Datensammlung über mehrere Standorte hinweg ist die Synchronisation von Zeitstempeln. Zeitstempel sind eine bedeutende Information bei forensischen Datenspuren. Zwar gibt es u. a. mit dem *Network Time Protocol (NTP)* ein etabliertes Verfahren, mit dem sich Uhren mit einer Genauigkeit von einigen Millisekunden synchronisieren lassen, aber es ist meist nur schwer nachzuvollziehen, ob Zeitstempel von einer derart synchronisierten Uhr erstellt wurden. Zudem können unterschiedliche Zeitzonen zu einer Fehlinterpretation führen.

In der Praxis kann auch die unzureichende Ausbildung von Forensikern hinsichtlich cloudspezifischer Untersuchungsmethoden ein Problem darstellen. Cloud-Forensik ist erst seit Kurzem in den Fokus der wissenschaftlichen Forschung geraten und geeignete Ausbildungsprogramme sind noch nicht ausreichend etabliert.

3.3 Datenanalyse

Können die Schwierigkeiten bei der Datengewinnung überwunden werden, so steht am Ende möglicherweise eine große Datenmenge, zusammengetragen aus einer großen Anzahl unterschiedlicher Quellen, zur Verfügung. Die Diversität der Datenarten sowie der Datenquellen stellt hier grundsätzlich eine Herausforderung für die Analyse dar, welche die Komplexität der Analyse von lokalen Datenspuren übersteigt.

Das Ziel der Analyse ist es auch, die genauen Abläufe und Interaktionen während eines Vorfalls möglichst genau zu rekonstruieren. Diese *Rekonstruktion* wird erschwert durch die bereits erwähnte hohe Volatilität in Cloud-Umgebungen, wodurch Daten oder virtuelle Maschinen sehr leicht unwiderruflich verloren gehen können.

Demgegenüber steht das Problem, dass *Werkzeuge und Vorgehensmodelle* für forensische Untersuchungen in der Cloud noch nicht in dem Maß etabliert sind, wie dies für traditionelle IT-Systeme der Fall ist. In Kapitel 5 werden wir auf ausgewählte Beispiele zu existierenden Werkzeugen und aktuellen Forschungsarbeiten eingehen.

3.4 Abschlussbericht / Präsentation

Am Ende einer forensischen Untersuchung steht der *Abschlussbericht*, in dem die Ergebnisse nicht nur aus technischer Sicht dokumentiert, sondern auch für externe Personen verständlich aufbereitet werden müssen. Bei komplexen Systemen sind auf eine Vielzahl von Diensten, weltweit verteilt in unterschiedlichsten Jurisdiktionen, sowie tausende von Nutzern, die gleichzeitig damit interagieren, involviert.

Im Vergleich zu einem herkömmlichen IT-System ist hier die Herausforderung erheblich größer, insbesondere für fachfremde Personen, das Gesamtbild nachvollziehbar zu präsentieren.

3.5 Fazit

Durch Cloud-Computing ergeben sich für forensische Untersuchung zahlreiche neue Herausforderungen. Neben juristischen und organisatorischen Aspekten lassen sich hierzu insbesondere technische Probleme identifizieren.

Auf Grundlage der allgemeinen Betrachtungen in diesem Kapitel betrachten wir im Folgenden detaillierter, welche Herangehensweisen in forensischen Untersuchungen bei cloudbasierten Diensten möglich sind. Abschnitt 4 geht zunächst auf den Teilaspekt der *cloudbasierten Datenspeicherung* ein. In Abschnitt 5 richtet sich der Fokus dann über die reine Datenspeicherung hinaus auf cloudbasierte Dienste.

4 Cloudbasierte Datenspeicherung

Daten werden heutzutage auf vielfältige Weise in der Cloud gespeichert. In diesem Abschnitt betrachten wir, welche cloudspezifischen Herausforderungen es bei der forensischen Untersuchung persistent gespeicherter Daten gibt. Bei der Datensammlung unterscheiden wir zwei unterschiedliche Perspektiven. Zum einem betrachten wir persistente Storage-Systeme, wie sie üblicherweise bei Cloud-Infrastrukturen verwendet werden, und die Möglichkeiten forensische Informationen aus diesen System zu extrahieren. Zum anderen gehen wir auf Dienste zur Datenspeicherung, die von Nutzern entfernt über das Netz verwendet werden und somit auch verwertbare Spuren auf Nutzerseite hinterlassen, ein. Bei der Datenanalyse diskutieren wir die Herausforderungen, die sich bei der Analyse gespeicherter Daten aus einer Systemarchitektur ergeben, die aus mehreren Abstraktionsebenen bestehen.

4.1 Cloudseitige Sammlung persistenter Daten

Bei der forensischen Untersuchung bei cloudbasierter Datenspeicherung betrachten wir zunächst die direkte Akquise relevanter Daten auf der Seite des Cloudproviders. Neben der in der digitalen Forensik etablierten Beweissicherung auf physischer Ebene der Datenträger ist es hier erforderlich, auch die Möglichkeiten der Datenerhebung auf höheren Abstraktionsebenen in Erwägung zu ziehen.

Physischer Datenzugriff

Etablierte Forensik-Techniken bestehen darin, von physichen Datenträgern zur Beweissicherung 1:1-Kopien anzulegen. Ein derartiges Vorgehen ist aber nur durchführbar, wenn ein direkter Zugang zu den physischen Datenträgern möglich ist. In einer Cloud-Umgebung kann dies bereits daran scheitern, dass es möglicherweise nicht ohne Weiteres ersichtlich ist, auf welchem Datenträger, ja sogar in welchem Data Center oder in welchem Land Daten tatsächlich gespeichert sind.

Noch am einfachsten stellt sich die Situation bei einer Untersuchung im Private-Cloud-Modell dar. Hier ist ein direkter Zugriff auf die physischen Datenträger und der Einsatz von traditionellen forensischen Vorgehensweisen im Allgemeinen grundsätzlich möglich. Auch kommt keine „multi tenancy“ durch mehrere unterschiedliche Kunden vor, sodass auch dies kein Hindernis für forensische Untersuchungen darstellt. Nichtsdestotrotz stellt sich die Situation schwieriger dar, als bei einem einfachen Desktop-PC mit einem persistenten Datenträger (Festplatte, Solid-State-Drive und ähnliches). Alleine schon die Storage-Größe eines lokalen Rechenzentrums wird in einer Untersuchung ein Hindernis darstellen, eine forensische 1:1-Kopie anzulegen. Auf die Schwierigkeiten bei der Analyse gehen wir in Abschnitt 4.2 näher ein.

Schwieriger ist die Lage in öffentlichen Cloud-Infrastrukturen dar. Gerade große Cloud-Anbieter betreiben Data Center an mehreren Orten und mehreren Ländern. Auch selbst wenn der Speicherort von Daten bekannt und ein physischer Zugang möglich sein könnte, enthalten die physischen Datenträger Daten von potentiell einer Vielzahl von Cloud-Nutzern, und eine Beschlagnahmung aller Datenträger zum Zweck der forensischen Untersuchung der Nutzung durch einen einzelnen Nutzer hat eine störende Auswirkung auf andere, nicht beteiligte Nutzer, die in der Regel kaum zu rechtfertigen ist. Insbesondere bei einer organisationsinternen privaten Untersuchung eines Vorfalls wird bei einem öffentlichen Cloud-Anbieter ein Zugriff auf physische Medien, die auch Daten anderer Nutzer enthalten können, nicht durchsetzbar sein.

Virtueller Datenzugriff

Besser durchführbar ist der Praxis ein virtueller Zugriff auf physisch gespeicherte Daten. Einen Snapshot von *persistentem Speicher* zu erzeugen ist dabei die einfachste Variante. Technisch ist dieser Vorgang in einer Cloudumgebung sogar meist einfacher durchzuführen als auf einem dedizierten System. Beispielsweise bietet Amazon's Infrastrukur die direkte Unterstützung, um jederzeit von einem EBS-Volumne (Elastic Block Storage) ein Copy-on-Write-Snapshot anlegen zu können[1].

Ein derartiger Copy-on-Write-Snapshot stellt eine 1:1-Kopie des virtuellen Dateisystems dar. Insofern enthält es exakt die gleichen Informationen wie das Original-Dateisystems, wie sie dem untersuchten Cloud-System aus Benutzersicht zur Verfügung stehen. Dennoch birgt dieses Vorgehen zwei relevante Nachteile gegenüber direkter Datenakquise auf physischer Ebene.

Zum einen stellt ein ein solcher Schnappschuss lediglich eine identische Sicht „von oben“ (aus Perspektive des Datennutzers) dar. Intern wird in der Regel nur eine virtuelle Kopie der Verzeichnisstruktur des untersuchten Datensystems angelegt, Für die zugeordneten Datei-Inhalte verweisen Original und Kopie physisch auf die gleichen Inhalte. Erst bei einer Modifikation der Daten wird eine echte Kopie dieser Daten angelegt, so dass die Änderungen nicht im Schnappschuss sichtbar werden.

[1] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-creating-snapshot.html [abgerufen am 2017-03-01]

Bei einer forensischen Untersuchung eines Dateisystems auf physischer Ebene lassen sich oft auch Spuren finden, die aus der Perspektive „von oben“ nicht sichtbar sind. Hierzu zählt insbesondere die Rekonstruktion von gelöschten Daten. Bei einer Copy-on-Write-Kopie eines virtuellen Datensystems ist im Allgemeinen davon auszugehen, dass nur aktiv genutzte Inhalte kopiert werden und gelöschte Daten in der Kopie nicht vorhanden sind. Eine oft bedeutende forensische Datenquelle steht dadurch bei virtuellem Datenzugriff nicht zur Verfügung.

Zum anderen erfolgt die Erzeugung eines virtuellen Schnappschusses durch Mechanismen, welche durch den Cloud-Anbieter bereitgestellt werden. Es erfolgen also bei der Datengewinnung auf virtueller Ebene Vorgänge, auf die der Cloud-Anbieter direkten Einfluss hat, bzw. deren korrekte Durchführung vom Cloud-Anbieter abhängen. Hier stellt sich die Frage, ob oder inwieweit die Authentizität und Integrität der so gewonnenen Daten sichergestellt werden kann. In der Praxis vorhandene Mechanismen sind im Allgemeinen nicht speziell für die Anforderungen an forensische Prozesse ausgelegt. Es lässt sich argumentieren, dass auch bei physischem Datenzugriff Abhängigkeiten von Dritten bestehen. Beispielsweise verlassen sich oft eingesetzte forensische Geräte zum Anlegen einer 1:1-Kopie einer Festplatte auf die korrekte Funktionsweise der Firmware dieser Festplatten. Nichtsdestotrotz erfordert die Informationsgewinnung durch virtuellen Datenzugriff wesentlich umfangreichere Operationen unter Kontrolle des Cloud-Anbieters.

4.2 Einfluss von Abstraktionsschichten auf die forensische Analyse

Verallgemeinert lässt sich das Problem der Unterschiede bei der Datengewinnung auf physischer und virtueller Ebene als Einfluss von Abstraktionsebenen auf die forensische Untersuchung beschreiben.

Bei einer genaueren Betrachtung lässt sich erkennen, dass in der Praxis nicht nur zwei unterschiedliche Ebenen (physisch und virtuell) zu finden sind. Vielmehr besteht ein Storage-Systemen heutzutage oft aus einer Vielzahl übereinander geschichteter Ebenen. Eine vertiefte Diskussion der Unterschiede zwischen Datengewinnung auf unterschiedlichen Ebenen bedarf daher einer individuellen, detaillierten Betrachtung der jeweils vorliegenden geschichteten Architektur.

Architekturen mit mehreren verschiedenen Abstraktionsebenen finden sich nicht nur innerhalb von öffentlichen Cloud-Infrastrukturen. Vielmehr können auch bei lokalen Storage-Systemen mehrere Ebenen unterschieden werden, wie z. B. physische Festlatten, virtuelle Festplatten eines Hardware-RAID, logische Partitionen innerhalb solcher Festplatten sowie virtuelle, Festplattengrenzen überschreitende „Volumes“ eines Storage-Systems. Auch bei einer lokalen Inhouse-Infrastruktur einer privaten Cloud führt die Vielzahl von Abstraktionsebenen dazu, dass die Rekonstruktion der Sicht von „oben“ und die Extraktion von versteckten oder gelöschten Platten zunehmend komplexer wird. Eine geeignete Unterstützung in forensischen Werkzeugen fehlt oft. Dadurch kann eine Datengewinnung und -analyse selbst bei vorhandenem physischen Zugriff eine große Herausforderung darstellen.

Praxis-Beispiel

Ein Praxis-Beispiel für eine derartige mehrschichtige Architektur zeigt Abbildung 4. Auf unterster Ebene befinden sich die physischen Festplatten (im Beispiel vier Stück, in der Praxis können dies leicht mehrere hundert sein). Einzelne Teile der Festplatten können als RAID-Set organisiert sein, die in der nächsten Ebene zu logischen Festplatten gefasst werden. Vor allem aus Performanzgründen (I/O-Lastverteilung)

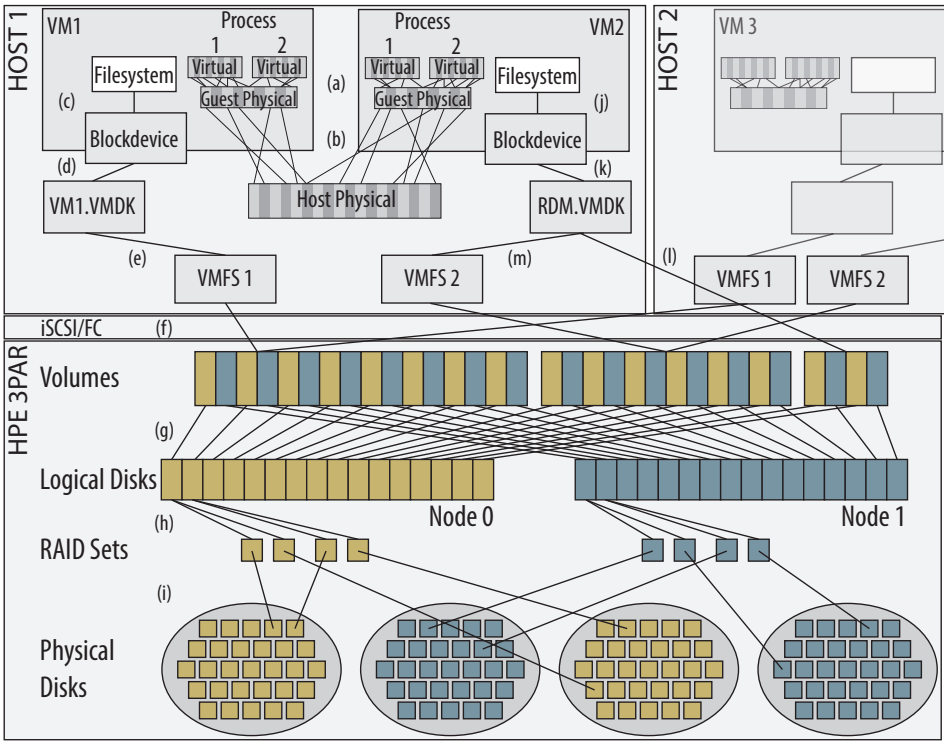


Abb. 4: Exemplarische Darstellung einer Storage-Architektur im privaten Cloud-Rechenzentrum nach Freiling et al. (2017)

ist es nicht unüblich, dass diese Volumes per „Striping“ zusammengefasst und im Beispiel per VMFS den einzelnen Hosts zur Verfügung gestellt werden.

Dieses Beispiel illustriert das bereits beschriebene Problem, dass eine genaue Lokalisierung von Daten (auf welchen physischen Festplatten finden sich für ein Filesystem einer virtuellen Maschine relevante Informationen?) eine komplexe Aufgabe darstellt. Eine vollständige 1:1-Kopie aller vorhandenen physischen Festplatten ist alleine schon aufgrund der Datenmenge eine kaum zu leistende Aufgabe. Aber auch die Analyse dieser Datenmenge bringt Probleme mit sich. Neben dem erforderlichen Aufwand (sowohl zeitlich als auch dafür notwendige Ressourcen) ist es bei Analyse von Rohdaten erforderlich, mit geeigneten Werkzeuge die Vielzahl von Abbildungen zwischen den einzelnen Schichten zu rekonstruieren. Dies wird in der Praxis oft an der fehlenden Verfügbarkeit von dafür geeigneten Werkzeugen scheitern.

Einfluss von Abstraktionsschichten

Die Komplexität der Extraktion von forensischen Informationen aus Rohdaten auf physischer Ebene kann in der Praxis dazu führen, dass man bei der Datensammlung auf höhere Ebenen zurückgreifen muss. Grundsätzlich sind alle aktiven Daten (also die Daten, die dem Nutzer in der virtuellen Maschine als im Dateisystem gespeicherte Informationen zur Verfügung stehen) auf allen Schichten in einer derartigen hierarchischen Architektur vorhanden. Bei Analyse von Daten einer

tieferen Schicht ist es erforderlich, dass Problem der Rekonstruktion der aktiven Abbildung zwischen den Schichten zu lösen (vgl. Definition 2)

Definition 2: Definition von *active mapping reconstruction* nach Freiling et al. (2017)

Beschreibt man die Relation zwischen Informationen aus Elementen einer oberen Schicht h und einer unteren Schicht l zum Zeitpunkt t als Abbildungsfunktion φ_t, so lässt sich das Problem der Rekonstruktion der aktiven Abbildung zwischen diesen Schichten wie folgt definieren:

„Given a copy e of l at time t, solving *active mappings reconstruction* means to

- find and decode φ_t from e and
- enumerate all elements of φ_t. “

D

Bei dieser Definition wird davon ausgegangen, dass Informationen über die Abbildung in der unteren Schicht gespeichert sind, wie dies in der Praxis i. d. R. der Fall ist. Beispielsweise finden sich Informationen über die Abbildung von logischen Partitionen einer Festplatte auf Blöcke der physischen Festplatte in einer auf der physischen Festplatte gespeicherten Partitionstabelle. Lässt sich eine entsprechende Abbildungsfunktion für alle Abstraktionsschritte in einer hierarchischen Architektur rekonstruieren, so ist das Ziel erreicht, die Informationen aus Anwendungssicht in der obersten Ebene aus den Rohdaten einer unteren Ebene zu gewinnen.

Aus forensischer Sicht ist es über diese Rekonstruktion hinaus von Relevanz, ob es auf unterer Schicht Informationen gibt, die nicht aus der Perspektive der oberen Schicht vorhanden sind. Hierzu zählen insbesondere Daten, die zu einem früheren Zeitpunkt dort vorhanden waren, die aber durch Löschung oder Veränderung von Daten zum Untersuchungszeitpunkt auf oberer Ebene nicht mehr vorhanden sind. Freiling et al. (2017) definieren hierzu die Problemstellung der Rekonstruktion gelöschter Abbildungen (siehe Definition 3).

Definition 3: Definition von *last deleted mappings reconstruction* nach Freiling et al. (2017)

„Given a copy e of l at time t, solving *last deleted mappings reconstruction* means to

- find and decode φ_t from e,
- enumerate all elements of φ_t,
- enumerate a non-trivial subset of all deleted mappings of φ_t, and
- to enumerate the most recently deleted mappings from that subset“

D

Ist dieses Problem in einer hierarchischen Architektur lösbar, so stehen auf der unteren Ebene forensische Informationen zur Verfügung, die bei einer Analyse von Daten auf der oberen nicht vorhanden sind. Nach Definition 3 stehen dabei gesicherte Erkenntnisse über in der Vergangenheit gespeicherte Daten zur Verfügung. In einer schwächeren Version ist es auch möglich, dass eine Rekonstruktion eine mögliche Zuordnung in der Vergangenheit aufzeigt, es aber nicht gesichert ist, dass dies die letzte vorhandene Abbildung war. Dies hat zur Folge, dass die

Daten der unteren Ebene zwar vom rekonstruierten Ursprung auf oberer Ebene stammen können, dies aber nicht gesichert ist.

Der Artikel von Freiling et al. (2017) zeigt einige Beispiele auf, wie einige praxisrelevante Beispiele auf dieses abstrakte Modell abgebildet werden. Für eine weitergehende Verwendung in der Praxis besteht aber Bedarf nach weiteren Forschungsarbeiten sowie nach der Entwicklung von entsprechenden forensischen Werkzeugen.

4.3 Nutzerseitige Sammlung forensischer Datenspuren

Nutzt ein Anwender einen Cloud-Dienst, um Daten zu speichern, so ist dies im Allgemeinen damit verbunden, dass auch auf dem lokalen Endgerät des Anwenders entsprechende Datenspuren vorhanden sind. Oft werden Dienste wie z. B. Dropbox verwendet, um Dateien zwischen verschiedenen Geräten zu synchronisieren. In diesem Fall sind die synchronisierten Dateien nicht nur auf Seite des Dienstanbieters in der Cloud vorhanden, sondern ggf. auch auf den Endgeräten des Anwenders. Eine Erfassung relevanter Datenspuren beschränkt sich daher nicht nur auf die verwendeten Clouddienste. Vielmehr kann insbesondere ein Endgerät des Nutzers eine wichtige Quelle von Informationen sein.

Die Herausforderung, die sich hier nun stellt, ist die große Vielfalt an Diensten, die ein Anwender verwenden könnte. Die verwendeten Protokolle und Datenformate sind in den meisten Fällen proprietär und nicht öffentlichen bekannt, manche Dienste sind auch nur kurzlebig oder unterliegen in ihrer internen Struktur dynamischen Veränderungen. Es fehlen dadurch in vielen Fällen wissenschaftlich fundierte Untersuchungen, welche verwertbaren Spuren bei der Verwendung einzelner Dienste entstehen können, sowie auch forensische Werkzeuge, welche diese Spuren angemessen auswerten.

Dennoch finden sich für verschiedene Dienste, die in der Praxis eine große Verbreitung gefunden haben, wissenschaftliche Untersuchungen zu verwertbaren Spuren. Im Folgenden betrachten wir hierzu das Beispiel Dropbox[2].

Beispiel Dropbox

Eine ausführliche Untersuchung der Spuren, die auf Client-Geräten bei Verwendung des Storage-Diensts Dropbox aufzufinden sind, wurde von Quick und Choo (2013) vorgenommen.

Die Autoren beschreiben darin ihre systematische Untersuchung, an welchen Stellen bei der Verwendung von Dropbox Spuren hinterlassen werden. In einer aufwendigen Reihe von Experimenten wurde untersucht, wie sich Aktionen wie das Installieren der Dropbox-Client-Software, deren Deinstallation sowie der Zugriff auf Dateien, die in Dropbox gespeichert wurden, über das Dropbox-Webfrontend auf im lokalen Dateisystem vorhandene Dateien auswirken.

Datenspuren lassen sich dabei an vielen Stellen finden, wie beispielsweise in der Windows-Registry oder im Thumbnail-Cache des Webbrowsers bei Zugriff auf Bilder über das Dropbox-Webfrontend. Wir verzichten an dieser Stelle auf eine detaillierte Wiedergabe der Ergebnisse von Quick und Choo (2013), sondern verweisen den Leser stattdessen auf den Original-Artikel.

Eine weiterführende Darstellung von aktuellen Forschungsergebnissen im Bereich der forensischen Untersuchung von Cloud Storage mit Fokus auf die nutzerseitige

[2] http://www.dropbox.com [abgerufen am 2017-04-01]

Spurensammlung mit Betrachtung von Microsoft SkyDrive, Dropbox, Google Drive und ownCloud findet sich bei Quick et al. (2014).

5 Cloud-Forensik

Während der letzte Abschnitt Daten, die in der Cloud gespeichert werden, betrachtet hat, wollen wir nun den Fokus auf Dienste in der Cloud legen, die in egal welchem Servicemodell von einem Angriff betroffen sein können. Daher stellt sich die Herausforderung, mit welchen technischen Mitteln diese Dienste einer forensischen Untersuchungen unterworfen werden können.

5.1 Offline-Analyse von System-Snapshots

Eine erste mögliche Herangehensweise besteht darin, von einem cloudbasierten System einen Snapshot zu erzeugen, der den persistenten Speicher, den volatilen Hauptspeicher oder auch beides umfassen kann.

Die Möglichkeit, einen Snapshot von *persistentem Speicher* zu erzeugen, wurde bereits im vorangegangenen Abschnitt thematisiert. Einen Snapshots vom *Hauptspeicher* zu erzeugen ist zumindest im IaaS-Modell grundsätzlich mit Tools, die auf dem Zielsystem ausgeführt werden, möglich. Diese Methode ist auch auf nichtvirtualisierten Systemen eine übliche Herangehensweise, hat aber den Nachteil, dass eine Interaktion mit dem Zielsystem notwendig ist und dieses dadurch verändert wird. Angewendet in der Cloud, bringt diese Methode aber gegenüber traditionellen IT-Systemen sogar den Vorteil mit sich, dass durch die Cloud jederzeit geeigneter persistenter Speicher zum Ablegen des Speicherabbilds zur Verfügung gestellt werden kann.

Darüber hinaus ist aber auch der Hypervisor grundsätzlich technisch in der Lage, ein komplettes Hauptspeicherabbild in einer Datei zu speichern. Beispielsweise bietet Xen mit dem Kommando `xl core-dump` genau diese Möglichkeit[3]. Hier hat ein virtualisiertes System ganz klar einen Vorteil gegenüber den Möglichkeiten auf einem nicht virtualisierten System.

Festplatten- und Hauptspeicherabbilder lassen sich genauso untersuchen wie entsprechende Abbilder in einem herkömmlichen IT-System. Bei Hauptspeicherabbildern stellt sich allerdings das Problem, dass in einer öffentlichen Cloud für einen Cloud-Kunden oder einen externen Untersucher im Allgemeinen nicht die Möglichkeit besteht, auf entsprechende Schnittstellen des Hypervisors zuzugreifen. Anwendbar sind diese Methoden also vor allem in privaten Cloud-Infrastrukturen, in denen man den erforderlichen direkten Zugriff auf das System hat. Auf dieses Problem werden wir in Kapitel 5.3 genauer eingehen.

In öffentlichen Cloud-Umgebungen ergibt sich als weiteres Problem, dass diese Methode nur im Servicemodell IaaS uneingeschränkt eingesetzt werden kann, da nur dort die Ressourcen eines Cloud-Kunden einer virtuellen Maschine entsprechen. Im PaaS-Modell sind mehrere Varianten bei der Trennung von Benutzern auf einer Plattform möglich. Werden einzelne Nutzer auf separate zugrunde liegende virtuelle Maschinen abgebildet, so ist ebenso eine Verwendung von Hauptspeicherabbildern dieser virtuellen Maschine möglich. Im SaaS-Modell dagegen sind Daten unterschiedlicher Nutzer nicht auf diese Weise getrennt, und insbesondere bei internen Untersuchungen eines Kunden ist die Bereitstellung von Schnittstellen, die Abbilder erzeugen, die auch Daten anderer Kunden enthalten, nicht akzeptabel.

[3] https://xenbits.xen.org/docs/4.8-testing/man/xl.1.html [abgerufen am 2017-03-01]

5.2 Existierende Monitoring-Dienste

Eine weitere Informationsquelle insbesondere in öffentlichen Cloud-Umgebungen können Schnittstellen des Cloud-Providers sein, mit dem sich Log-Daten von Monitoring-Diensten des Providers abfragen lassen. Hierbei handelt es sich im Allgemeinen um proprietäre Schnittstellen eines Anbieters, die dafür gedacht sind, mehr Transparenz für den Nutzer zu bieten, Reaktionen auf Fehler zu ermöglichen oder die Diagnose von Problemen zu unterstützen. Anzutreffen bei aktuellen Cloud-Anbietern sind also keine Monitoring-Dienste oder -Schnittstellen, die dediziert für forensische Untersuchungen ausgelegt sind. Da es hierbei keine standardisierten Schnittstellen gibt, sind dazu bei der Analyse dieser Informationsquelle jeweils anbieterspezifische Werkzeuge notwendig sind.

Exemplarisch nennen wir hier Amazons Dienst AWS CloudTrail, „ein Service für die Überwachung von Governance, Compliance, Betrieb und Risiken in Ihrem AWS-Konto“[4]. AWS CloudTrail ist in der Lage, alle API-Aufrufe eines Kunden an der Managementschnittstelle mitzuprotokollieren. Damit lassen sich z. B. Änderungen an AWS-Ressourcen wie das Anlegen und Löschen von Storage-Volumes oder von virtuellen Maschinen nachvollziehen. Eine Nutzung von AWS CloudTrail erfordert die Aktivierung der Protokollierung, es ist also eine *Vorbereitung* notwendig, um bei späteren Vorfällen diese Informationsquelle nutzen zu können.

5.3 Forensic Readiness

Unter dem Stichwort „Forensic Readiness“ haben einzelne Forschungsarbeiten Lösungsvorschläge erarbeitet, wie man die Verfügbarkeit forensischer Datenquellen in Cloudumgebungen verbessern, indem man in der Cloud-Management-Infrastruktur geeignete Schnittstellen und Mechanismen bereitstellt.

FROST

Eine der ersten Arbeiten zu diesem Thema ist FROST (Forensic OpenStack Tools) von Dykstra und Sherman (2013). Die Autoren entwerfen drei forensische Werkzeuge, mit denen die Cloud-Management-Plattform OpenStack erweitert wird. Ziel dabei war es, auf der Basis von FROST den Einsatz herkömmlicher forensischer Tools zu ermöglichen.

Die Werkzeuge bieten Zugriff auf folgende Datenquellen:

- Abbild der virtuellen Festplatte einer virtuellen Maschine;
- Log mit allen Zugriffen eines Nutzers auf die Schnittstelle des Cloud-Providers;
- Firewall-Logs der virtuellen Maschinen des Nutzers.

Bei FROST werden Maßnahmen zum Schutz der Integrität dieser Daten getroffen. Die Festplatten-Abbilder werden durch eine kryptographische Prüfsumme geschützt und für die Log-Daten stellt die Architektur einen authentischen Log-Dienst bereit.

Die Integration von FROST in OpenStack erfolgt auf zwei Ebenen. Zum einen wird das Management-Frontend erweitert, um dem Cloud-Nutzer im Dashboard Zugriff auf die forensischen Werkzeuge zu geben. Zum anderen werden auf den Compute-Knoten, auf denen virtuelle Maschinen des Kunden ausgeführt werden, Mechanismen zum Aufzeichnen von API-Logs und Firewall-Logs ergänzt.

[4] https://aws.amazon.com/de/cloudtrail/ [abgerufen am 2017-04-01]

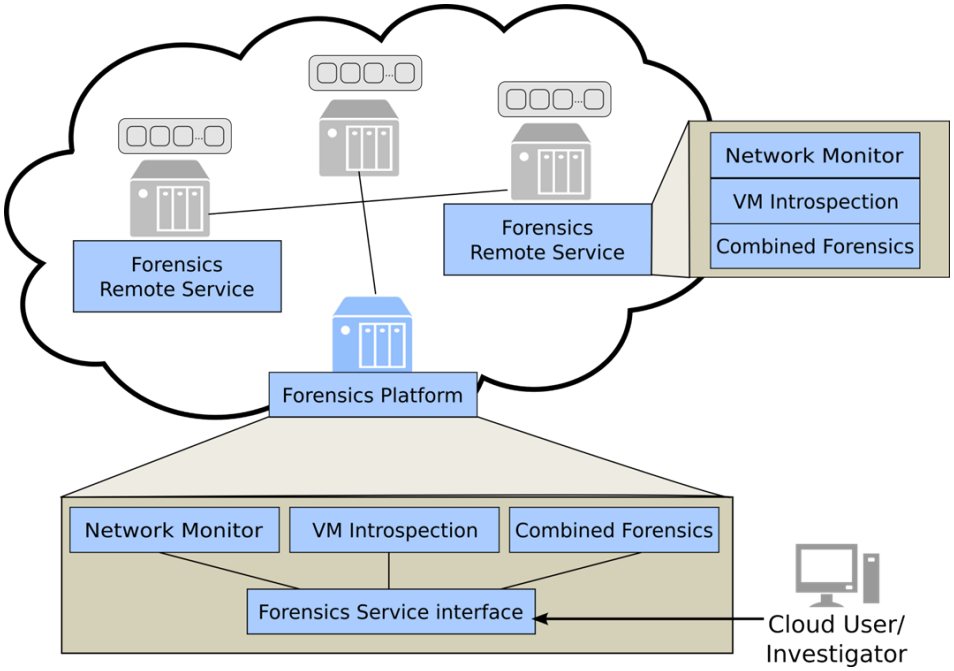


Abb. 5: Überblick über die Architektur von LiveCloudInspector (Zach und Reiser, 2015)

LiveCloudInspector

Aus architektureller Sicht verfolgt LiveCloudInspector (Zach und Reiser, 2015) einen ähnlichen Ansatz (siehe Abbildung 5). LiveCloudInspector stellt als Frontend eine Forensik-Plattform bereit, mit der ein Nutzer Information aus verschiedenen Datenquellen extrahieren kann. Der Zweck von LiveCloudInspector unterscheidet sich allerdings von dem von FROST: Während FROST auf die Extraktion von Logs und Disk-Snapshots fokussiert ist, zielt LiveCloudInspector auf die Live-Analyse eines laufenden Systems.

Für diese Live-Analyse werden die Cloud-Rechner, die virtuelle Maschinen ausführen, um einen Forensik-Dienst erweitert. Dieser Dienst stellt Mechanismen bereit, um neben dem Monitoring des Netzverkehrs auch den Hauptspeicherinhalt des Zielsystems analysieren zu können. Da der Zugriff auf die VMI-Schnittstellen des Hypervisors von einer privilegierten Domäne aus erfolgt (in der Prototyp-Implementierung von der Dom0-VM eines Xen-Hypervisors), ist es nicht möglich, kundenspezifischen Analysecode direkt auszuführen. Vielmehr beschränkt sich die Analyse auf fest vorgegebene Operationen. Aus forensischer Sicht ist hier vor allem das Anlegen eines vollständigen Abbilds des Hauptspeichers einer virtuellen Maschine von Interesse.

Die Beispiele FROST und LiveCloudInspector zeigen, dass es durchaus technisch machbar ist, über den in der Industriepraxis verbreiteten Stand (wie beispielsweise Amazons AWS CloudTrail) hinaus forensische Dienste und Schnittstellen anzubieten. Nichtsdestotrotz bieten auch diese Ansätze nicht die Möglichkeit, (virtuell) vor Ort beliebige Live-Analysen durchzuführen. Hierfür ist es notwendig, geeignete, sichere Verfahren zu entwickeln, die einen entsprechenden entfernten Zugriff auf Introspektions-Schnittstellen zur Live-Analyse des Zielsystems ermöglichen. In der Forschung gibt es hierzu erste Ansätze, auf die wir im folgenden Abschnitt eingehen werden.

CloudVMI und Forensik-VMs

Eine der ersten Arbeiten, die das Problem der fehlenden Verfügbarkeit des Zugriffs auf VMI-Schnittstellen in öffentlichen Clouds thematisiert hat, ist CloudVMI (Baek

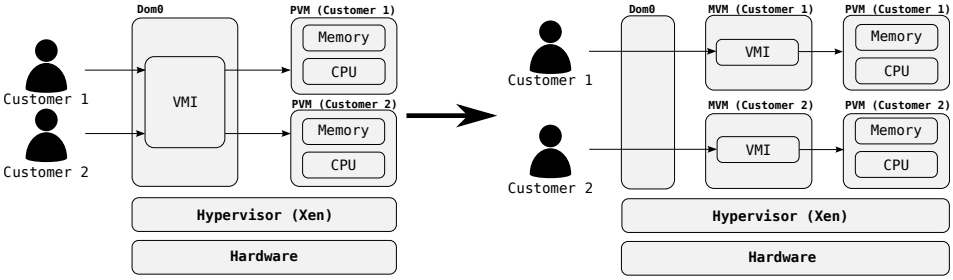


Abb. 6: Verwendung von VMI-Operationen in Dom0 im Vergleich zu dedizierten Monitoring-VMs in der CloudPhylactor-Architektur

et al., 2014). Ziel von CloudVMI ist es, die VMI-Schnittstelle zu virtualisieren und Cloud-Kunden als Dienst bereitzustellen. Dabei wird die VMI-basierte Analyse in zwei Teile gesplittet: Auf der einen Seite stellt der Cloud-Provider einen Dienst bereit, der Anfragen zu Speicher-Seiten des Zielsystems auf VMI-Operationen an der Hypervisor-Schnittstelle abbildet, und auf der anderen Seite kann der Cloud-Kunde beliebigen Analysecode verwenden, der über das Netz auf den bereitgestellten Dienst zugreift.

Die Architektur von CloudVMI löst zwei zentrale Schwierigkeiten bei der Verwendung von VMI in öffentlichen Clouds: Es ermöglicht den entfernten Zugriff auf VMI-Schnittstellen als Dienst in der Cloud und durch die Verwendung und Durchsetzung von Policies, die regeln, welche Nutzer auf welche Ziel-VMs zugreifen können, löst es auch das Problem der Isolation zwischen unterschiedlichen Cloud-Kunden. Der entfernte Zugriff auf VMI-Operationen bringt aber auch den Nachteil größerer Latenzen mit sich. Als nur ein Beispiel betrachten wir die VMI-Operation *vmi_pause_vm()* (Anhalten einer virtuellen Maschine, um einen konsistenten Schnappschuss des Hauptspeichers analysieren zu können). Im von Baek et al. (2014) untersuchten Beispiel benötigt diese Operation lokal bei direkten Zugriff auf VMI-Schnittstellen ca. $3.5\mu s$, während der Zugriff über CloudVMI ca. $60\mu s$ dauert, einem Anstieg um mehr als den Faktor 15. Je nach Anwendungszweck kann dieser Overhead in der Praxis unverhältnismäßig groß sein. Wird beispielsweise eine virtuelle Maschine während einer Hauptspeicher-Analyse angehalten, so kann die höhere Dauer dieser Analyse einen nicht akzeptablen Einfluss auf die Ausführungsgeschwindigkeit des Zielsystems haben.

Eine Lösung für dieses Problem wurde in der Forschungsarbeit CloudPhylactor (Taubmann et al., 2016) vorgeschlagen und als Prototyp implementiert (siehe Abbildung 6). Anders als bei CloudVMI, bei dem Zugriffspolicies als Teil eines entfernt ansprechbaren Diensts implementiert wurden, werden bei CloudPhylactor diese Policies in den Hypervisor verlagert. Im Prototyp-System kommt der Hypervisor Xen zum Einsatz, der intern auf Basis einer Implementierung von Flask (Flux Advanced Security Kernel) und der Verwendung von Xen Security Modules (XSM) bereits die Grundlagen für derartige Policies bereitstellt.

In der CloudPhylactor-Architektur werden nun Flask-Policies definiert, die einer *Monitoring Virtual Machine (MVM)* den Zugriff auf eine andere virtuelle Maschine auf dem selben Host mittels VMI erlauben. Der Effizienz der VMI-Operationen ist dabei nahezu identisch zu VMI-Operationen, die direkt auf der Dom0 ausgeführt werden. Gleichzeitig wird durch die Policies sichergestellt, dass Isolation zwischen einzelnen Cloud-Kunden gewahrt werden kann. Neben einer Durchsetzung von Policies im Hypervisor erfordert CloudPhylactor auch eine entsprechende Verwaltung von Policies in der Managementschicht der Cloud-Infrastruktur. Auf dieser Ebene wird geregelt, welche Nutzer eine MVM für ein bestimmtes Zielsystem anlegen dürfen, und es werden für die erzeugten virtuellen Maschinen geeignete Flask-Label generiert, die dann den gewünschten Zugriff ermöglichen.

5.4 Bewertung dieser Möglichkeiten

Alle diese Erweiterungen zeigen auf, dass es durchaus technische Möglichkeiten gibt, das Problem der Identifikation und Sammlung forensischer Informationen in Cloud-Infrastrukturen zu verbessern. Derzeit sind dem Autor aber keine nennenswerten Bestrebungen größerer Cloud-Anbieter bekannt, derartige Techniken in ihren Infrastrukturen anzubieten.

Diese Bewertung ist jedoch auch hinsichtlich der Cloud-Servicemodelle zu differenzieren. Alle beschriebenen Ansätze sind jeweils für das IaaS-Modell entworfen, und nur dort können sie uneingeschränkt als geeignet angesehen werden. Eine Erweiterung auf das PaaS ist insofern denkbar, wenn eine Platform für einen PaaS-Cloud-Kunden isoliert innerhalb einer virtuellen Maschine bereitgestellt wird. In allen anderen Fälle sind die beschriebenen Methoden ungeeignet, weil bisher keine Ansätze existieren, auf diesem Weg gezielt die relevanten Daten eines Kunden zu identifizieren oder die Zugriffsmöglichkeiten gezielt auf einen Kunden einzuschränken.

6 Übungaufgaben

Übung 1

Welche Bedeutung haben Vorgehensmodelle für eine forensische Untersuchung?

Ü

Übung 2

Erläutern Sie das Vorgehensmodell des BSI für den forensischen Prozess und vergleichen Sie dieses mit anderen bekannten Vorgehensmodellen.

Ü

Übung 3

Beschreiben Sie wesentliche Herausforderungen, die sich bei einer forensischen Untersuchung von cloudbasierten Diensten in den Phasen der Datenidentifikation, Datensammlung und Datenanalyse ergeben.

Ü

Übung 4

Ein Cloud-Nutzer einer öffentlichen Cloud ist mit einem Missbrauch seiner Dienste konfrontiert. Welche wesentlichen Unterschiede gibt es bei einer von ihm durchgeführten forensischen Untersuchung zwischen den Servicemodellen IaaS, PaaS und SaaS?

Ü

Übung 5

Eine Organisation nutzt eine private Cloud-Infrastruktur zur Speicherung von Daten, die auf Geräten von Mitarbeitern genutzt werden. Beschreiben und vergleichen Sie unterschiedliche Möglichkeiten der Sammlung forensischer Informationen.

Ü

Ü

Übung 6

Beschreiben Sie den Unterschied zwischen Rekonstruktion aktiver Abbildungen und Rekonstruktion der letzten gelöschten Abbildung im abstrakten Modell von Freiling et al. (2017).

Ü

Übung 7

Beschreiben Sie die technischen Herausforderungen und Lösungsansätze einer Live-Analyse von Hauptspeicherinhalten bei Verwendung einer privaten Cloud im IaaS-Modell.

Verzeichnisse

I. Abbildungen

II. Definitionen

III. Literatur

Saad Alqahtany, Nathan Clarke, Steven Furnell, und Christoph Reich. Cloud Forensics: A Review of Challenges, Solutions and Open Problems. In *2015 International Conference on Cloud Computing (ICCC)*, S. 1–9, April 2015.

Hyun wook Baek, Abhinav Srivastava, und Jacobus Van der Merwe. CloudVMI: Virtual Machine Introspection As a Cloud Service. In *Proceedings of the 2014 IEEE International Conference on Cloud Engineering*, IC2E '14, S. 153–158, Washington, DC, USA, 2014. IEEE Computer Society.

Nicole Beebe. Digital Forensic Research: The Good, the Bad and the Unaddressed. In *Advances in Digital Forensics V - Fifth IFIP WG 11.9 International Conference on Digital Forensics, Orlando, Florida, USA, January 26-28, 2009, Revised Selected Papers*, S. 17–36, 2009.

BSI. Leitfaden „IT-Forensik". online verfügbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Leitfaden_IT-Forensik.pdf [abgerufen am 2017-03-15], Bundesamt für Sicherheit in der Informationstechnik, 2011.

Eoghan Casey. *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet*. Academic Press, 2. Aufl., 2004.

DFRWS. A Road Map for Digital Forensic Research. In *Proceedings of the First Digital Forensic Research Workshop*. DFRWS, 2001.

Josiah Dykstra und Alan T. Sherman. Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform. *Digital Investigation*, 10, Supplement:S87 – S95, 2013.

Felix Freiling, Thomas Glanzmann, und Hans P. Reiser. Characterizing Loss of Forensic Information due to Abstraction Layers. *Digital Investigation*, 20:S107–S115, 2017.

Felix C. Freiling und Bastian Schwittay. A Common Process Model for Incident Response and Computer Forensics. In *IT-Incidents Management & IT-Forensics - IMF 2007, Conference Proceedings, September 11-13, 2007, Stuttgart, Germany*, S. 19–40, 2007.

Dennis Heinson. *IT-Forensik: Zur Erhebung und Verwertung von Beweisen aus informationstechnischen Systemen*. Mohr Siebeck, 1. Aufl., 2015. ISBN 978-3161537011.

Donn B. Parker. *Fighting Computer Crime: A New Framework for Protecting Information*. John Wiley & Sons, Inc., New York, NY, USA, 1998. ISBN 0-471-16378-3.

Darren Quick und Kim-Kwang Raymondo Choo. Dropbox analysis. Data remnants on user machines. *Digital Investigation*, 10:3 – 18, 2013.

Darren Quick, Ben Martini, und Kim-Kwang Raymond Choo. *Cloud Storage Forensics*. Syngress Publishing / Elsevier, 2014.

Benjamin Taubmann, Noelle Rakotondravony, und Hans P. Reiser. CloudPhylactor: Harnessing Mandatory Access Control for Virtual Machine Introspection in Cloud Data Centers. In *The 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-16)*, 2016.

Julian Zach und Hans P. Reiser. LiveCloudInspector: Towards Integrated IaaS forensics in the Cloud. In *Proc. of the 15th IFIP Int. Conf. on Distributed Applications and Interoperable Systems (DAIS)*, 2015.