Michael Karlinger, Klaus Ettmayer, Michael Schrefl

Verschlüsselung bei ausgelagerter Datenhaltung

Traditionell wird Vertraulichkeit bei ausgelagerten Daten durch das den Serviceprovidern entgegengebrachte Vertrauen, durch vertragliche Regelungen zum Datenschutz sowie Mechanismen zur Zugriffskontrolle und -protokollierung erreicht. Zunehmende Datendiebstähle und dadurch verursachte Schäden erfordern weitere technische Maßnahmen. Sichere Datenverschlüsselung und effiziente Abfragebearbeitung sind schwierig gemeinsam zu erreichen, sodass diesbezüglich oft Kompromisse eingegangen werden. Gleichzeitig bestehen Wahlmöglichkeiten hinsichtlich der Entscheidung, ob Datenverschlüsselung und Abfragebearbeitung am Server oder am Client stattfindet.

Inhaltsübersicht

1 Online-Datenspeicherung

Das größte Hemmnis bei der Verwendung sogenannter Onlinespeicher[1] stellen Zweifel potenzieller Nutzer hinsichtlich der vertraulichen Behandlung ihrer Daten dar [Chen et al. 2010, S. 1]. Gegen Angriffe von außerhalb des Onlinespeichers können Maßnahmen wie Benutzerauthentifizierung und regelbasierte Zugriffskontrolle verwendet werden, sofern ein Mindestmaß an Vertrauen der Nutzer durch den Serviceanbieter erfüllt wird und dieser besagte Schutzmaßnahmen tatsächlich umsetzt. Um dies sicherzustellen, werden vertragliche Vereinbarungen, sogenannte Service Agreements, getroffen, die in dem Fall, dass die Vertraulichkeit der Daten verletzt wird, greifen und dadurch dem Nutzer juristische Mittel zur Verfügung stellen, um sich beim Serviceanbieter schadlos zu halten.

Vertragliche Vereinbarungen helfen, den durch die Verletzung der Vertraulichkeit der Daten entstandenen Schaden auszugleichen, nicht aber die Verletzung an sich zu verhindern. Darüber hinaus schützen besagte Maßnahmen ausgelagerte Daten nur gegen Angriffe von außerhalb des Onlinespeichers, sie sind aber wirkungslos gegenüber den häufig vorkommenden Angriffen von innerhalb des Systems durch zum Beispiel Serveradministratoren, die uneingeschränkten Zugriff auf physische Speichermedien und dadurch auf die darauf gespeicherten Daten haben [Takabi et al. 2010, S. 28; Richardson 2008, S. 14 f.].

Um die Vertraulichkeit ausgelagerter Daten zu gewährleisten, d.h. unbefugte Zugriffe zu unterbinden, ist es deshalb notwendig, Daten verschlüsselt zu speichern. Bei Verwendung geeigneter Verschlüsselungsverfahren ist das Lesen der ausgelagerten Daten nur für befugte Personen, also Personen, die über den entsprechenden Schlüssel verfügen, möglich.

1. Client-Server-Anwendungen zur ausgelagerten Speicherung und Verwaltung von Daten.

Der durch Verschlüsselung erreichbare Schutz vor unbefugten Zugriffen ist in zweierlei Hinsicht grundsätzlich limitiert. Erstens bietet Verschlüsselung nur effektiven Schutz vor unbefugten, lesenden, nicht aber vor unbefugten, schreibenden Zugriffen. Die Wahrung der Integrität ausgelagerter Daten kann im Unterschied zur Gewährleistung der vertraulichen Behandlung der Daten nicht durch Verschlüsselung sichergestellt werden.

Zweitens gewährleisten in der Praxis verwendbare Verschlüsselungsalgorithmen, wie der Data Encryption Standard (DES), die Vertraulichkeit von Daten nur in dem Sinne, dass die Entzifferung von Kryptogrammen nicht in angemessener Zeit durchführbar ist. Dem Anspruch perfekter Geheimhaltung, also dem Schutz der Vertraulichkeit im informationstheoretischen Sinne, genügen gängige Verschlüsselungsverfahren nicht [Schneier 1996, S. 235].

Neben dem Verschlüsselungsverfahren hat insbesondere auch das zugrunde liegende Verarbeitungsprinzip entscheidenden Einfluss auf den von einem Onlinespeicher effektiv gewährleisteten Schutz der Vertraulichkeit. Werden zum Beispiel Daten am Server entschlüsselt, um Anfragen zu beantworten, werden unbefugte Datenzugriffe von innerhalb des Systems ermöglicht und der durch die Verschlüsselung erzielte Schutz der Daten aufgeweicht.

Die Gewährleistung der Vertraulichkeit ausgelagerter Daten durch Verschlüsselung hat direkten Einfluss auf die *Leistungsfähigkeit* und *Funktionalität* von Onlinespeichern.

Die Verschlüsselung von Daten stellt einen Mehraufwand bei der Anfrageverarbeitung dar. Aus diesem Grund ist die Leistungsfähigkeit eines verschlüsselten Onlinespeichers grundsätzlich geringer als jene eines unverschlüsselten. Dank immer leistungsfähigerer Hardware stellt die prinzipiell schlechtere Leistungsfähigkeit von verschlüsselten Onlinespeichern in vielen praktischen Anwendungsfällen allerdings keinen ausschlaggebenden Hinderungsgrund für deren Verwendung dar.

Gravierender als die Verschlüsselung an sich wirkt sich die bei verschlüsselten Daten aufwendigere Abfrageverarbeitung auf die Leistungsfähigkeit von verschlüsselten Onlinespeichern aus. Legt man ein zu gewährleistendes Niveau an Vertraulichkeit fest, so muss ein Kompromiss zwischen der Abfragemächtigkeit und der Leistungsfähigkeit eingegangen werden. Um einen hohen Grad an Vertraulichkeit der ausgelagerten Daten bei gleichzeitig hoher Leistungsfähigkeit erreichen zu können, bieten verschlüsselte Onlinespeicher häufig weniger Abfragemöglichkeiten an als ihre unverschlüsselten Pendants.

Hinsichtlich der Funktionalität stellt neben der Abfragemächtigkeit insbesondere auch die Mehrbenutzerfähigkeit einen kritischen Aspekt dar. Um verteilten Zugriff auf verschlüsselte Daten zu ermöglichen, ist es notwendig, einen autorisierten Benutzer in die Lage zu versetzen, einen Teil der Daten eines anderen Benutzers entschlüsseln zu können. Die Schwierigkeit dabei ist es, sicherzustellen, dass der autorisierte Benutzer tatsächlich nur den Teil der Daten des anderen Benutzers entschlüsseln kann, für den er autorisiert ist. Zusätzlich verschärft wird diese Problematik dadurch, dass sich Zugriffsrechte im Laufe der Zeit ändern. Somit muss autorisierten Benutzern die Möglichkeit der Entschlüsselung von Daten auch wieder entzogen werden können.

Um verschlüsselte Onlinespeicher zielführend in der Praxis einsetzen zu können, gilt es, Herausforderungen in den folgenden Bereichen zu begegnen:

- Vertraulichkeit
  Der verwendete Verschlüsselungsmechanismus und das zugrunde liegende Verarbeitungsprinzip müssen ausgelagerte Daten umfassend vor unbefugten, lesenden Zugriffen schützen.
- Leistungsfähigkeit
  Der Mehraufwand bei der Anfragebeantwortung soll auf den durch die Verschlüsselung

immanent entstehenden Mehraufwand reduziert werden.

- Funktionalität
  Die Abfragemächtigkeit und Mehrbenutzerfähigkeit eines verschlüsselten Onlinespeichers sollen jenen eines unverschlüsselten Onlinespeichers gleichen.

Ein gemeinsames Erreichen aller genannten Ziele in vollem Ausmaß ist aufgrund der zueinander konfliktären Beziehungen nicht möglich. Die in Forschung und Praxis entwickelten Ansätze zur verschlüsselten Online-Datenspeicherung meistern diese Herausforderungen in unterschiedlichem Maße.

2 Rahmenarchitekturen

Verschlüsselte Onlinespeicher haben die beiden charakterisierenden Gemeinsamkeiten, dass Daten am Server verschlüsselt gespeichert werden und dass der Server den lesenden und schreibenden Zugriff auf die physischen Speichermedien durchführt. Maßgebliche Unterschiede zwischen verschlüsselten Onlinespeichern ergeben sich je nachdem, ob die Verschlüsselung von Daten einerseits und die Verarbeitung von Daten andererseits vom Server oder aber vom Client durchgeführt werden.

Entsprechend den beiden Wahlmöglichkeiten der Lokation der Verschlüsselung und der Verarbeitung von Daten ergeben sich die in Abbildung 1 zusammenfassend dargestellten Rahmenarchitekturen für verschlüsselte Onlinespeicher. DieWahl der Rahmenarchitektur ist ausschlaggebend für die Güte eines Onlinespeichers hinsichtlich der identifizierten Herausforderungen und bietet sich aus diesem Grund für die Kategorisierung an.

		Verschlüsselung	
		Server	***Client***
Verarbeitung	***Server***	(I) Datenbank-systeme	(II) Forschungs-prototypen
	Client	_____	(III) Backup-Systeme + EPOD

Abb. 1: Rahmenarchitekturen für verschlüsselte Onlinespeicher

Für die clientseitige Verschlüsselung der Daten spricht der erhöhte Schutz der Vertraulichkeit, da die Daten in diesem Fall nie im Klartext am Server verfügbar sind. Für die serverseitige Verarbeitung der Daten spricht die tendenziell hohe Leistungsfähigkeit. Diese beiden Vorzüge würden offensichtlich negiert, falls Daten am Server verschlüsselt und dann am Client verarbeitet werden. Deshalb stellt diese Rahmenarchitektur nur eine theoretische Möglichkeit dar, die weder in der Praxis noch in der Forschung verfolgt wird.

2.1 Serverseitige Verschlüsselung und Verarbeitung

Datenspeicherung: Bei serverseitiger Verschlüsselung und Verarbeitung werden zu speichernde Daten im Klartext vom Client an den Server geschickt und dort mit dem für den Benutzer hinterlegten Schlüssel verschlüsselt und gespeichert. Um den einem Benutzer zugeordneten Schlüssel eruieren zu können, ist eine Benutzerauthentifizierung zwingend erforderlich.

Abfrageverarbeitung: Um Daten abzurufen, schickt der Client eine Abfrage im Klartext an den Server. Die Berechnung des Abfrageergebnisses erfolgt so wie bei unverschlüsselter Datenspeicherung. Einziger Unterschied dabei ist, dass Daten, die für die Berechnung des Abfrageergebnisses benötigt werden, nach dem Lesen vom physischen Speichermedium zuerst entschlüsselt werden, bevor sie verwendet werden können. Das berechnete Abfrageergebnis wird abschließend im Klartext an den Client zurückgegeben. Da der Server vollen Zugriff auf die Daten eines Benutzers hat, ist die Abfragemächtigkeit im Vergleich zu unverschlüsselten Onlinespeichern nicht eingeschränkt.

Da Abfragen gänzlich am Server abgearbeitet werden, ist der entstehende Mehraufwand bei der Anfragebeantwortung im Vergleich zur unverschlüsselten Online-Datenspeicherung auf ein Minimum, nämlich den durch die Verschlüsselung der Daten immanent entstehenden Mehraufwand, reduziert. Der Kommunikationsaufwand bei serverseitiger Verschlüsselung und Verarbeitung ist ebenfalls nicht höher als bei unverschlüsselter Online-Datenspeicherung, da nur die Anfrage an sich vom Client an den Server geschickt und nur das tatsächliche Anfrageergebnis vom Server an den Client zurückgegeben wird. Online-Datenspeicher mit serverseitiger Verschlüsselung und Verarbeitung zeichnen sich deshalb durch hohe Leistungsfähigkeit aus.

Mehrbenutzerfähigkeit: Der gemeinsame Zugriff auf Daten durch mehrere Benutzer wird bei serverseitiger Verschlüsselung und Verarbeitung durch konventionelle Authentifizierungs- und Autorisierungsmechanismen ermöglicht. Stellt ein authentifizierter Benutzer eine Abfrage, wird vom Server überprüft, ob dieser Benutzer über entsprechende Rechte verfügt, um die für die Beantwortung der Anfrage notwendigen Schreib- und Leseoperationen auf den Daten durchzuführen. Das Ergebnis der Anfrage wird vom Server nur dann berechnet, wenn der Benutzer über die notwendigen Rechte verfügt.

Die von Onlinespeichern mit serverseitiger Verschlüsselung und Verarbeitung angebotene Funktionalität entspricht somit sowohl im Hinblick auf die Abfragemächtigkeit als auch auf die Mehrbenutzerfähigkeit der eines unverschlüsselten Onlinespeichers.

Datenschutz: Die hohe Leistungsfähigkeit und umfangreiche Funktionalität gehen klar zulasten des Schutzes der Vertraulichkeit der Daten. Da die Schlüssel am Server verfügbar sind, können Insider die ausgelagerten Daten grundsätzlich entschlüsseln. Auch falls einem Angreifer die Schlüssel nicht zur Verfügung stehen, so kann er dennoch alle für die Beantwortung von Anfragen benötigten Daten ausspionieren, da diese temporär am Server entschlüsselt werden. Die Daten sind somit lediglich vor Angriffen geschützt, bei denen der Angreifer nur Zugang zum physischen Speichermedium, nicht aber zu den verwendeten Schlüsseln hat.

Um unbefugte Zugriffe während der Übertragung von Anfragen und Anfrageergebnissen zwischen Client und Server zu unterbinden, werden bei serverseitiger Verschlüsselung und Verarbeitung üblicherweise die Kommunikationskanäle durch Verschlüsselung, zum Beispiel SSL, gesichert.

Marktsituation: Verschlüsselte Onlinespeicher mit serverseitiger Verschlüsselung und Verarbeitung werden bereits in Form von verschlüsselten Datenbanksystemen von führenden Datenbankherstellern, wie zum Beispiel Oracle [Oracle 2011], angeboten. Im Unterschied dazu sind derzeit noch keine verschlüsselten Online-Dateiverwaltungssysteme mit serverseitiger Verschlüsselung und Verarbeitung auf dem Markt.

2.2 Clientseitige Verschlüsselung und serverseitige Verarbeitung

Datenspeicherung: Bei clientseitiger Verschlüsselung und serverseitiger Verarbeitung werden Daten am Client verschlüsselt und an den Server zur Speicherung auf dem physischen Speichermedium geschickt. Der für die Verschlüsselung notwendige Schlüssel muss vom Benutzer am Client bereitgestellt werden.

Abfrageverarbeitung: Um Daten abzufragen, schickt der Client eine Abfrage an den Server. Der Server berechnet das Abfrageergebnis direkt auf den verschlüsselten Daten, d.h. ohne die Daten temporär zu entschlüsseln, und schickt das Abfrageergebnis in verschlüsselter Form an den Client zurück, der abschließend das Abfrageergebnis entschlüsselt.

Um die Berechnung des Abfrageergebnisses direkt auf den verschlüsselten Daten zu ermöglichen, werden sogenannte homomorphe

Verschlüsselungsalgorithmen für die clientseitige Verschlüsselung der Daten verwendet. Die spezielle Eigenschaft homomorpher Verschlüsselungsalgorithmen ist, dass bestimmte algebraische Operationen das gleiche Ergebnis bei der Ausführung auf den eigentlichen Daten und bei der Ausführung auf den aus diesen Daten erzeugten Kryptogrammen liefern. Die verfügbaren algebraischen Operationen hängen dabei vom konkret verwendeten homomorphen Verschlüsselungsalgorithmus ab. Zum Beispiel erlauben sogenannte ordnungserhaltende Verschlüsselungsalgorithmen Größer/Kleiner-Vergleiche auf Kryptogrammen durchzuführen [Agrawal et al. 2004].

Bei clientseitiger Verschlüsselung und serverseitiger Verarbeitung zeigt sich der notwendige Kompromiss zwischen Leistungsfähigkeit und Abfragemächtigkeit besonders stark. Homomorphe Verschlüsselungsalgorithmen, die nur einzelne algebraische Operationen ermöglichen, sind zwar effizient, bieten aber eben nur sehr geringe Abfragemöglichkeiten. Sogenannte vollständig homomorphe Verschlüsselungsverfahren, wie zum Beispiel das von Gentry kürzlich entwickelte [Gentry 2009], erlauben zwar die Ausführung von beliebigen Abfragen, sind aber in der Praxis aufgrund der immens hohen Laufzeit nicht einsetzbar. Es wurde geschätzt, dass eine Google-Abfrage bei Verwendung dieses Verschlüsselungsverfahrens in etwa eine Billion Mal länger dauert als im unverschlüsselten Fall [Cooney 2009].

Mehrbenutzerfähigkeit: Neben der stark eingeschränkten Abfragemächtigkeit bei effizienter clientseitiger Verschlüsselung und serverseitiger Verarbeitung stellt auch das Fehlen von Lösungsansätzen für den verteilten Zugriff auf homomorph verschlüsselte Daten eine Limitation hinsichtlich des Funktionsumfangs dar.

Datenschutz: Die Vertraulichkeit der Daten ist hinsichtlich des zugrunde liegenden Verarbeitungsprinzips hoch, da nur am Client ver- und entschlüsselt wird. Des Weiteren ist, bei Verwendung ausreichend langer Schlüssel, die Sicherheit homomorpher Verschlüsselungsalgorithmen zufriedenstellend. Das eigentliche Problem in Bezug auf die vertrauliche Behandlung ausgelagerter Daten ist, dass der verwendete homomorphe Verschlüsselungsalgorithmus nicht austauschbar ist und somit Daten aus dem Onlinespeicher entfernt werden müssen, falls der Verschlüsselungsalgorithmus geknackt wird.

Marktsituation: Bisher entwickelte homomorphe Verschlüsselungsalgorithmen bieten entweder zu wenig Operationen, um einen genügend großen Funktionsumfang erzielen zu können, oder sind zu rechenaufwendig, um ausreichende Leistungsfähigkeit sicherzustellen. Homomorphe Verschlüsselungsalgorithmen werden deshalb nur im akademischen Umfeld verwendet.

2.3 Clientseitige Verschlüsselung und Verarbeitung

Datenspeicherung: Bei clientseitiger Verschlüsselung und Verarbeitung werden Daten am Client verschlüsselt und gemeinsam mit einem eindeutigen Identifikator an den Server transferiert, um dort gespeichert zu werden. Der dafür notwendige Schlüssel muss vom Benutzer am Client bereitgestellt werden. Obwohl deshalb nicht zwingend erforderlich, wird eine Benutzerauthentifizierung üblicherweise durchgeführt, um den Onlinespeicher vor mutwilliger Erschöpfung der physischen Speichermedien zu schützen.

Abfrageverarbeitung: Um Daten abzufragen, werden die für die Abfragebeantwortung benötigten Daten vom Server in verschlüsselter Form an den Client übertragen, dort entschlüsselt und das Abfrageergebnis berechnet.

Aufgrund dieses Verarbeitungsprinzips muss der Client bei der Abfrageabarbeitung die zuvor bei der Datenspeicherung vergebenen Identifikatoren der benötigten Daten kennen. Um dies zu ermöglichen, werden entsprechende Metadaten am Server verschlüsselt gespeichert und bei der Anmeldung am Onlinespeicher geladen.

Bei verschlüsselten Online-Dateiverwaltungssystemen werden zum Beispiel die Identifikatoren von Ordnern und deren Verschachtelung als Metadaten vermerkt.

Da bei clientseitiger Verschlüsselung und Verarbeitung alle für die Berechnung des Abfrageergebnisses notwendigen Daten vom Server an den Client transferiert werden müssen, ist die Minimierung des Kommunikationsaufwands ein Knackpunkt hinsichtlich der Leistungsfähigkeit solcher Onlinespeicher. Geringe Abfragemächtigkeit wird deshalb häufig zugunsten besserer Leistungsfähigkeit in Kauf genommen.

Mehrbenutzerfähigkeit: Um mehreren Benutzern Zugriff auf gemeinsame Daten zu ermöglichen, muss entweder der verwendete Schlüssel unter diesen Benutzern verteilt werden oder es muss für jeden Benutzer eine eigene Kopie der Daten angelegt werden. Falls die Schlüssel verteilt werden, entsteht durch die Schlüsselverwaltung ein beträchtlicher Mehraufwand. Des Weiteren muss beim Entziehen von Zugriffsrechten eine erneute Verschlüsselung der von der Änderung betroffenen Daten durchgeführt werden, um das Lesen der Daten für den nicht länger zugriffsberechtigten Benutzer unmöglich zu machen. Falls gemeinsame Daten dupliziert werden, entsteht erheblicher Mehraufwand dadurch, dass die Daten zuerst am Client des autorisierenden Benutzers entschlüsselt, dann an den Client des autorisierten Benutzers transferiert und dort abschließend wiederum verschlüsselt werden müssen. Des Weiteren ist das Entziehen einmal gewährter Zugriffsrechte nur in Bezug auf zukünftige Datenänderungen möglich, da der autorisierte Benutzer bereits eine Kopie der aktuellen Version der Daten besitzt. Aufgrund des hohen Aufwands wird bei clientseitiger Verschlüsselung und Verarbeitung häufig keine Mehrbenutzerkontrolle angeboten bzw. nur in eingeschränkter Form, ohne die Möglichkeit des Rechteentzugs.

Datenschutz: Die Vertraulichkeit ausgelagerter Daten ist bei clientseitiger Verschlüsselung und Verarbeitung umfassend gewährleistet, da Daten nie im Klartext den vertrauenswürdigen Client verlassen. Im Unterschied zu dem im vorigen Abschnitt beschriebenen Verarbeitungsprinzip besteht des Weiteren die Möglichkeit, den verwendeten Verschlüsselungsalgorithmus auszutauschen, falls dieser zum Beispiel aufgrund eines Quantensprungs in der Rechnerleistung unsicher werden sollte. Mittels einmaliger, erneuter Verschlüsselung unter Verwendung eines neuen Verschlüsselungsalgorithmus kann der Onlinespeicher ohne Einbußen hinsichtlich der Vertraulichkeit der Daten weiter verwendet werden.

Marktsituation: Verschlüsselte Onlinespeicher mit clientseitiger Verschlüsselung und Verarbeitung werden bereits in Form von Online-Dateiverwaltungssystemen, wie zum Beispiel Swiss Disk [SwissDisk 2011] oder True Crypt [TrueCrypt 2011], angeboten. Die zufriedenstellende Leistungsfähigkeit und der hohe Grad an Vertraulichkeit der ausgelagerten Daten gehen bei diesen Systemen insofern zulasten der angebotenen Funktionalität, als sie nur als Backup-Systeme, nicht aber zur Suche in und nach Dateien anhand von Dateieigenschaften oder des Dateiinhalts verwendbar sind.

2.4 Gegenüberstellung

Die Güte der besprochenen Rahmenarchitekturen hinsichtlich des Schutzes der Vertraulichkeit, der angebotenen Funktionalität und der Leistungsfähigkeit sind zusammenfassend in Tabelle 1 gegenübergestellt.

	Rahmenarchitektur		
	(I)	(II)	(III)
Vertraulichkeit	-	+	+
Funktionalität	+	-	~
Leistungsfähigkeit	+	~	~

Tab. 1: Gegenüberstellung der Rahmenarchitekturen

3 Encrypted Private Online Disc (EPOD)

Das Institut für Data & Knowledge Engineering der Johannes Kepler Universität Linz hat, gefördert von der Internet Foundation Austria, ein verschlüsseltes Online-Dateiverwaltungssystem mit clientseitiger Verschlüsselung und Verarbeitung entwickelt, das im Unterschied zu bisherigen Systemen die effiziente Suche nach Dateien und Ordnern anhand ihrer Eigenschaften und Inhalte ermöglicht und somit den gewohnten Komfort lokaler Datei-Explorer bietet [EPOD 2011]. Um diese Funktionalität ohne drastische Einbußen in der Leistungsfähigkeit zu ermöglichen, ist es notwendig, Dateien und Ordner durchsuchen zu können, ohne diese zuvor an den Client transferieren zu müssen. EPOD erreicht dies durch den Einsatz individuell auf zu erwartende Abfragen zugeschnittener Indexstrukturen, die gemeinsam mit den ausgelagerten Dateien und Ordnern verschlüsselt am Server gespeichert werden.

Der EPOD-Client (vgl. Abb. 2) vergibt systemweit eindeutige Nummern (IDs) für Dateien und Ordner, die als Identifikatoren dienen. Die Datei- und Ordner-IDs werden für die Repräsentation der Ordnerstruktur verwendet. Die Strukturinformation wird verschlüsselt am Server gespeichert und beim Start des EPOD-Clients vom Server geladen. Aufgrund einer eigens entwickelten, komprimierten Repräsentation der Strukturinformation dauert es nur wenige Sekundenbruchteile, um diese zu laden.

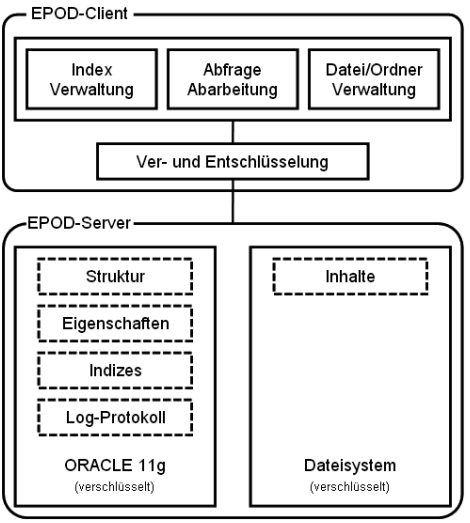

Abb. 2: Architektur des EPOD-Systems

Neben der Strukturinformation werden auch die gängigen Eigenschaften von Dateien und Ordnern, wie Name, Typ oder Änderungsdatum, und die Inhalte von Dateien am Server gespeichert. Zur Identifikation der Eigenschaften und des Inhalts einer Datei bzw. eines Ordners wird die Datei- bzw. Ordner-ID verwendet. Navigiert der Benutzer durch die ausgelagerten Ordner und Dateien oder öffnet er eine Datei, so werden die für die Darstellung in der Benutzeroberfläche benötigten Datei- und Ordnereigenschaften bzw. der Dateiinhalt dynamisch vom Server mittels der aus der Strukturinformation ersichtlichen IDs geladen. Dadurch wird der Kommunikationsaufwand auf ein Minimum reduziert und eine kurze Antwortzeit erreicht.

Grundsätzlich werden Daten, wie die Strukturinformation und die Datei- und Ordnereigenschaften, am EPOD-Server in eigenen Tabellen einer relationalen Datenbank abgelegt. Einzige Ausnahme stellt der Inhalt von Dateien dar. Dieser wird aus Effizienzgründen am Dateisystem des EPOD-Servers gespeichert und über FTP zugänglich gemacht. Die Verwaltung der Strukturinformation, der Datei- und Ordnereigenschaften und des Dateiinhalts wird von der Komponente *Datei/Ordner Verwaltung* realisiert.

Um die effiziente Suche zu ermöglichen, werden beim Speichern von Dateien und Ordnern Indizes von der Komponente *Index Verwaltung* erstellt und verschlüsselt am Server abgelegt. Die unterstützten Indexstrukturen (Hash-Indizes, B*-Bäume und invertierte Listen) wurden so gewählt, dass jede in der Benutzeroberfläche spezifizierbare Abfrageart optimal unterstützt wird.

Zum Beispiel wird für die Beantwortung von Abfragen der Art »*Finde alle Dateien, die an einem bestimmten Datum erstellt wurden*« ein Hash-Index mit den Erstellungsdaten von Dateien als Hash-Keys und den IDs der jeweiligen Dateien als Hash-Values verwendet.

Um eine solche Abfrage zu beantworten, verschlüsselt die Komponente *Abfrage Abarbeitung* den jeweiligen Datumswert und schickt diesen an den Server. Der Server gibt die zu dem verschlüsselten Datumswert im Hash-Index vermerkte Liste von verschlüsselten Datei- und Ordner-IDs zurück, die abschließend am Client entschlüsselt wird.

Punktabfragen über Datei- und Ordnereigenschaften können somit mit nur einer Kommunikation zwischen Client und Server beantwortet werden. Neben Punktabfragen unterstützt das EPOD-System auch Bereichsabfragen über Datei- und Ordnereigenschaften durch B*-Bäume und die Volltextsuche im Sinne einer Wortstammsuche im Inhalt von Dateien durch invertierte Listen.

Die Verwendung von Indexstrukturen stellt naturgemäß einen erheblichen Aufwand dar. Um die Antwortzeit beim Speichern von Dateien und Ordnern so kurz wie möglich zu gestalten, wird nur die Strukturinformation neuer Dateien und Ordner unmittelbar am Server gespeichert, die Datei- und Ordnereigenschaften, der Dateiinhalt und die Indexstrukturen aber im Hintergrund aktualisiert. Insbesondere die Unterstützung der Volltextsuche ist, wie auch bei unverschlüsselten Online-Dateiverwaltungssystemen, sehr zeitaufwendig, da der Inhalt von Dateien vor der Indizierung von irrelevanten Teilen wie Steuerzeichen und Stopwords bereinigt werden muss.

Um sicherzustellen, dass zum Zeitpunkt der Beendigung des EPOD-Clients noch nicht fertig abgearbeitete Aufgaben beim nächsten Start wieder fortgesetzt werden können, wird ein Log-Protokoll geführt, das am Server in verschlüsselter Form abgelegt wird. In diesem Log-Protokoll ist für jede Datei und jeden Ordner vermerkt, ob die Eigenschaften und der Dateiinhalt bereits gespeichert und die Indexstrukturen bereits aktualisiert wurden. Um Inkonsistenzen in den persistierten Daten aufgrund eines unerwarteten Abbruchs während der Abarbeitung einer einzelnen Aufgabe, zum Beispiel des Schreibens der Eigenschaften einer Datei, zu verhindern, werden alle Schreiboperationen, die für die Erledigung der Aufgabe notwendig sind, innerhalb einer Datenbanktransaktion ausgeführt, die im Falle des unerwarteten Abbruchs zum Rückführen der Datenbank in den letzten konsistenten Zustand führt.

4 In Zukunft umfassender Schutz

Aufgrund der Verwendung individuell auf zu erwartende Abfragearten zugeschnittener Indexstrukturen und des dynamischen Ladens von Datei- und Ordnereigenschaften sowie von Dateiinhalten zeichnet sich das EPOD-System durch gute Leistungsfähigkeit bei gleichzeitig hoher Abfragemächtigkeit aus. In der nächsten Ausbaustufe des EPOD-Systems ist eine Erweiterung für den verteilten Zugriff auf ausgelagerte Dateien geplant, um hinsichtlich der angebotenen Funktionalität mit derzeit bestehenden, unverschlüsselten Online-Dateiverwaltungssystemen gleichzuziehen und darüber hinaus umfassenden Schutz der Vertraulichkeit der ausgelagerten Daten zu gewährleisten.

5 Literatur

[Agrawal et al. 2004] *Agrawal, R.; Kiernan, J.; Srikant, R.; Xu, Y.:* Order Preserving Encryption for Numeric Data. SIGMOD, Paris, 2004, pp. 13-18.

[Chen et al. 2010] *Chen, Y.; Paxson, V.; Katz, R. H.:* What's New About Cloud Computing Security? Electrical Engineering and Computer Sciences, Berkeley, CA, 2010.

[Cooney 2009] *Cooney, M.:* IBM Touts Encryption Innovation, 2009, *www.computerworld.com*; Zugriff am 31.05.2011.

[EPOD 2011] *EPOD:* Encrypted Private Online Disc, *http://epod.dke.uni-linz.ac.at*; Zugriff am 31.05.2011.

[Gentry 2009] *Gentry, C.:* Fully Homomorphic Encryption Using Ideal Lattices. Proc. of the 41st ACM Symposium on Theory of Computing, New York, NY, 2009, pp. 169-178.

[Oracle 2011] *Oracle:* Transparent Data Encryption, *www.oracle.com/technetwork/database/options/advanced-security/index.html*; Zugriff am 31.05.2011.

[Richardson 2008] *Richardson, R.:* CSI Computer Crime and Security Survey 2008. Computer Crime Institute (2008), *http://gocsi.com/sites/default/files/uploads/CSIsurvey2008.pdf*; Zugriff am 31.05.2011.

[Schneier 1996] *Schneier, B.:* Applied Cryptography: Protocols, Algorithms and Source Code in C. John Wiley & Sons, Inc., Kanada, 1996.

[SwissDisk 2011] *Swiss Disk:* Private Secure Online Sync and Storage, *www.swissdisk.com*; Zugriff am 31.05.2011.

[Takabi et al. 2010] *Takabi, H.; Joshi, J. B. D.; Ahn, G. J.:* Security and Privacy Challenges in Cloud Computing Environments. IEEE Computer and Reliability Societies, 2010, pp. 24-31.

[TrueCrypt 2011] *TrueCrypt:* Free Open-Source On-The-Fly Encryption, *www.truecrypt.org*; Zugriff am 31.05.2011.

Dr. Michael Karlinger
Mag. Klaus Ettmayer
o. Univ.-Prof. Dr. Michael Schrefl
Johannes Kepler Universität Linz
Institut für Wirtschaftsinformatik
Data & Knowledge Engineering
Altenberger Str. 69
A-4040 Linz
{karlinger, ettmayer, schrefl}@dke.uni-linz.ac.at
www.dke.jku.at