



PERIMETRIX® SAFESPACE™

TECHNISCHER ÜBERBLICK
UND SYSTEM ARCHITEKTUR

KEEPING SECRETS SAFE





1. ÜBERSICHT

2. DATENKLASSIFIZIERUNG

3. DATA MOVEMENTS

- 3.1. CLIENT
- 3.2. GATEWAY

4. RULE ENGINE

- 4.1. RULE ENGINE IN SAFEEDGE
- 4.2. RULE ENGINE ZUR AUTOMATISCHEN KLASSIFIZIERUNG
- 4.3. DETERMINISTISCHE FAKTEN
- 4.4. PROBABILISTISCHE FAKTEN

5. SYSTEMARCHITEKTUR

- 5.1. ALLGEMEINES
- 5.2. SERVICES



PERIMETRIX

1. ÜBERSICHT

Perimetrix® SafeSpace™ ist eine verteilte Lösung für Unternehmen zum Schutz vertraulicher Informationen. Dies beinhaltet Datenverlust bzw. Datenklau, (versehentliches) Löschen aber auch bewusste Störmaßnahmen.

Hierzu werden deterministische und probabilistische Verfahren zum Erkennen von klassifizierten Daten verwendet. Hinzu kommen Werkzeuge um Kommunikationskanäle, über die klassifizierte Informationen fließen, zu überwachen und eventuelle Verstöße zu erkennen und zu unterbinden:

- Gefährlichen Aktionen auf einem PC, auf dem sensible Daten verarbeitet und/oder gespeichert werden (SafeUse)
- Ausgehende E-Mails mit sensiblem Inhalt (SafeEdge)
- HTTP POST Requests mit sensiblem Inhalt (SafeEdge)
- Instant Messaging Unterhaltungen mit sensiblem Inhalt (SafeEdge)

Das gesamte System ist von Grund auf hoch verfügbar ausgelegt: Sämtliche Dienste unterstützen eingebaute Hochverfügbarkeit und Lastverteilung. Dies ermöglicht eine flexible Anpassung/Erweiterung des Systems in Bezug auf Verarbeitungsressourcen. Außerdem protokolliert das System Benutzeraktionen mit klassifizierten Daten und erlaubt somit eine nachträgliche forensische Analyse von Vorfällen.

2. DATENKLASSIFIZIERUNG

Im Gegensatz zu DLP-Lösungen der ersten und zweiten Generation bietet SafeSpace ein flexibles Model zur Klassifizierung von Daten. Die Erfahrung hat gezeigt, dass ein einfaches Vertraulichkeits-Model nicht ausreicht, um Daten innerhalb eines Unternehmens zu klassifizieren: Ein CTO mag die Berechtigung haben, geheime Entwicklungsdaten zu bearbeiten — geheime Finanzdaten sollten ihm verwehrt bleiben. Werden Daten nur nach Vertraulichkeit klassifiziert (öffentlich/intern/vertraulich/geheim), lässt sich diese Einschränkung nicht realisieren.

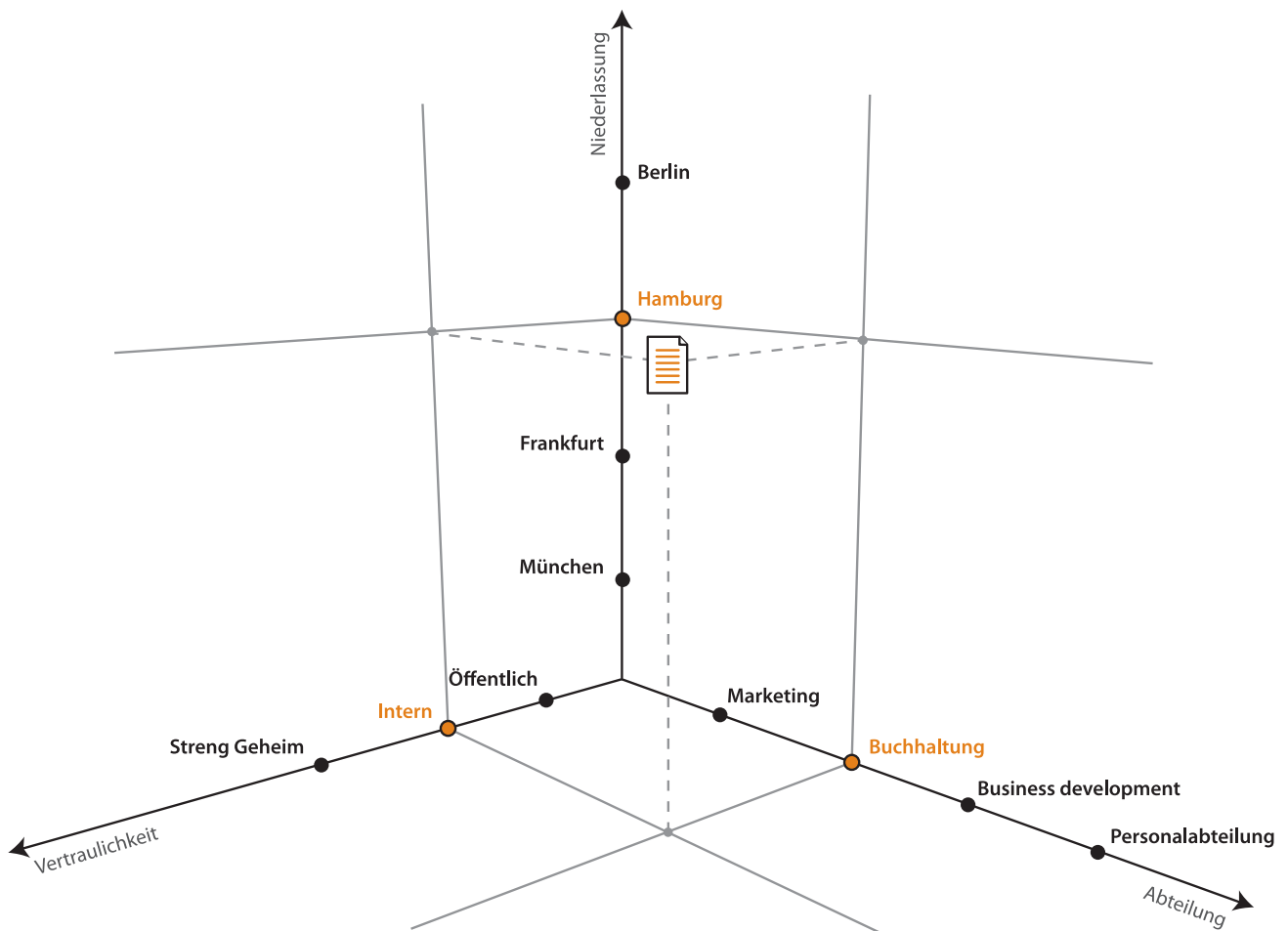


Abbildung 1: Drei verschiedene Dimensionen mit Kategorien

SafeSpace erlaubt es dem Kunden, sein Klassifikationsschema frei zu definieren. Hierzu werden Dimensionen, Kategorien und Level eingesetzt:

Dimensionen dienen zum Gruppieren gleichartiger Kategorien. Als Beispiel sei hier „Vertraulichkeit“ genannt. Andere Beispiele für Dimensionen sind z.B. Fachbereiche oder die geographische Lokation.

Kategorien stellen Punkte innerhalb der jeweiligen Dimension da. In der Dimension „Vertraulichkeit“ kann es z.B. die Kategorien öffentlich, intern, vertraulich und geheim geben. Hierbei können Kategorien auch hierarchisch sein — D.h., die Kategorie „Deutschland“ enthält Unterkategorien „Hessen“, „Hamburg“, „Niedersachsen“ usw.

Levels sind die Kombination von Kategorien in verschiedenen Dimensionen. Sie stellen die elementaren Rechte verschiedener Objekte im SafeSpace System dar.

Als Beispiel gibt es drei Dimensionen: Vertraulichkeit, Abteilung und Land. Mögliche Levels sind z.B. **[Intern, Vertrieb, DE]** oder **[Geheim, Support, CH]**. Objekte im SafeSpace System können mehrere Level besitzen. Dies ermöglicht es, einzelnen Objekten genau in ihren Rechten einzuschränken.

3. DATA MOVEMENTS

Das Grundprinzip von SafeSpace ist die Untersuchung von „Data Movements“. Jede Aktion innerhalb des SafeSpace Systems wird als das verschieben/kopieren von Daten von einem Quell-Container in einen Ziel-Container interpretiert: Als Beispiel sei hier ein User genannt, der eine Datei in Word öffnet. Hierbei werden Daten aus einem Quell-Container (Datei) in einen Ziel-Container (Word) überführt. Dies wird aber nur erlaubt, sofern a) der Benutzer einen Level hat, der es ihm erlaubt, diese Art von Daten (basierend auf Levels) zu bewegen, b) der Benutzer das Recht hat, dem Quell-Container Daten zu entnehmen bzw. diese in den Ziel-Container zu überführen und c) die Container überhaupt Daten mit entsprechendem Level aufnehmen dürfen.

Wichtig ist hierbei, dass Daten ihren Level in diesem Transfer nicht verlieren. Öffnet der Benutzer eine Datei mit dem Level [Intern, Vertrieb, DE] in Word (sofern erlaubt), so „erbt“ Word diesen Level. Versucht der Benutzer diese Datei zu speichern, kann er sie nur an Orten/in Containern speichern, auf die er Rechte hat und die für [Intern, Vertrieb, DE] freigegeben sind.

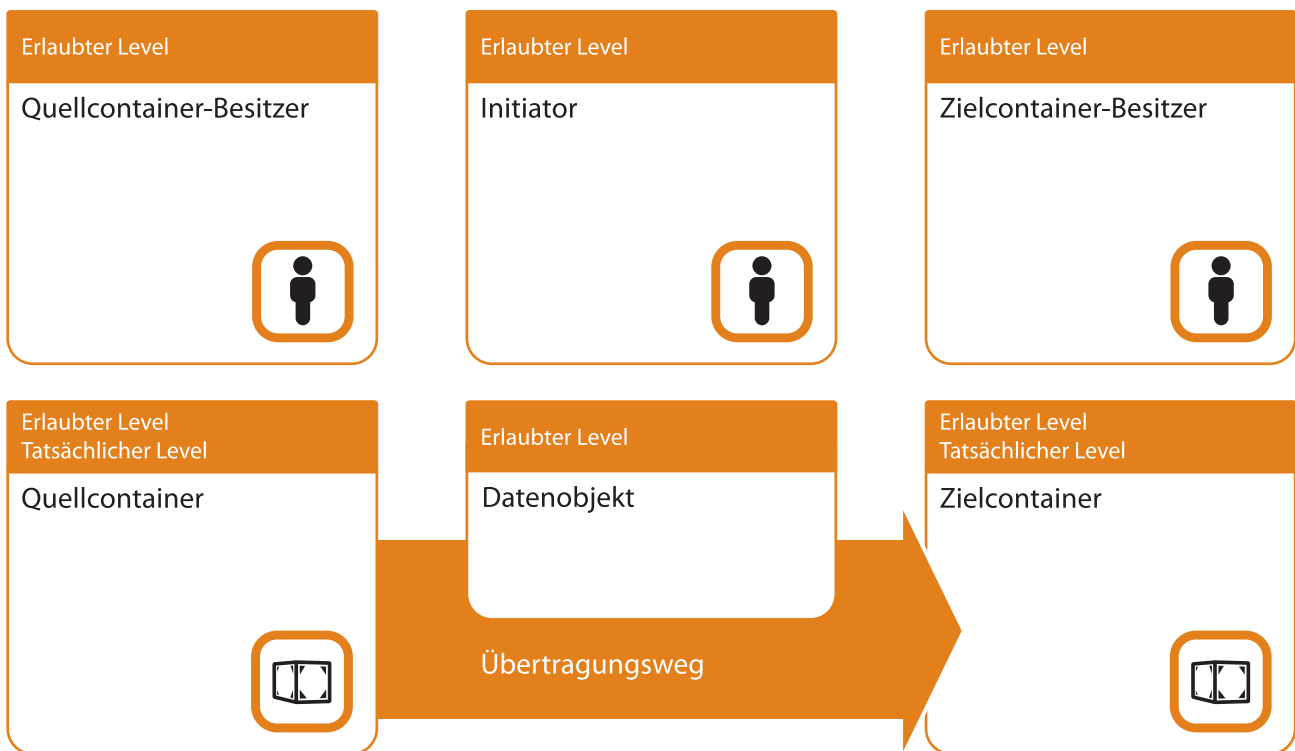


Abbildung 2: Data Movements im SafeSpace System

Dieses Prinzip kommt sowohl am Client (SafeUse) als auch am Gateway (SafeEdge) zum Tragen.

Natürlich stellt sich die Frage: „Was ist ein Container?“

Hier bietet das SafeSpace System eine umfangreiche Liste von Containern, die es dem Unternehmen erlaubt, genau zu definieren, welche Arten von Daten wie bewegt und bearbeitet werden können. Grundsätzlich ist hier zwischen Client und Gateway zu unterscheiden — die folgenden Aufzählungen sind Beispiele und nicht vollständig.

3.1. CLIENT

Network Node (Socket)

Jegliche Kommunikation von einem Socket kann als Container klassifiziert werden. D.h. zum Beispiel keine „geheimen“ Daten an Port 80 — oder auch „alles was von Port 4658 (Homegrown Application) kommt wird automatisch als [Vertrieb] klassifiziert“.

Process (Applikationen)

Applikationen können als Container mit verschiedenen Levels definiert werden. Dies ermöglicht z.B., dass Konstruktionsdaten mit einem CAD Programm ganz normal bearbeitet werden können — mit einem HEX Editor jedoch nicht. Dabei können Applikationen über den Prozessnamen und/oder über Signaturen definiert werden. Dies bedeutet auch, dass entsprechend abhängige Child Prozesse bzw. Dateien, Ordner oder Pipes die der Prozess anlegt, den Level erben.

On-Access Blocked Devices

Dies umfasst Geräte, die zwar ein Filesystem besitzen, auf die aber auch unter Umgehung der Filesystem Funktionen zugegriffen wird. Dies umfasst z.B. das direkte Auslesen/schreiben von CDs bzw. der Zugriff auf Fest-/Wechselplatten beim Imagebackup oder bei der Defragmentierung:

- CD/DVD Laufwerke bzw. Brenner
- Floppy
- Festplatte
- Wechseldatenträger



PERIMETRIX

Preventively Blocked Device

Geräte, die weder ein (sichtbares) Filesystem bilden und auf die unter Umgehung der Volume APIs zugegriffen wird, werden als „Preventively Blocked Device“ definiert. Diese werden von Windows als „Black-Box“ behandelt. Daher wird der Zugriff auf diese Devices während der Arbeit mit klassifizierten Daten unterbunden. Preventively Blocked Devices werden mittels der Device Class oder der Device Instance ID definiert und umfassen u.a.:

- IEEE1394 (FireWire)
- Bluetooth
- Infrarot (IrDA)
- Modems, Seriel (inkl. Multiportkarten)
- Netzwerk
- PCMCIA
- Ports
- Drucker (normal/PnP)
- Biometrische Geräte
- USB
- HID (Human Interface Devices)

Clipboard

Das Clipboard als Zwischenspeicher kann für bestimmte Daten freigegeben werden. Dabei bleibt die Markierung der Daten erhalten. Öffnet man klassifizierte Daten in Word, kopiert Teile davon ins Clipboard und den Clipboard Inhalt in eine neue Wordpad Instanz, so bekommt auch Wordpad diesen Level. D.h. Wordpad kann nur in Containern speichern, die Daten der Klassifizierung der ursprünglichen Daten aufnehmen können.

Datei/Ordner

Dateien und Ordner sind offensichtliche Container. Hierbei können diese explizit Level(s) besitzen bzw. implizit aufgrund eines Pfades. Hierbei erlaubt die Nutzung von Environmentvariablen auch die flexible Anpassung an heterogene Systemumgebungen

Filesystem

Filesysteme können als Container dienen. Sie können generisch (z.B. NTFS oder FAT) aber auch genau definiert werden (Volume Serial Number). Hiermit kann z.B. verhindert werden, dass Daten auf einem Filesystem ohne „echte ACLs“ (FAT) gespeichert werden.

Trash

Dieser Pseudo-Container dient zum Abfangen von Löschaktionen. Sofern z.B. die Bewegung aus einem Ordner- in den Trashcontainer nicht erlaubt ist, kann man eine Datei nicht löschen. Dies ist unabhängig vom Windows Explorer — hier werden direkt die entsprechenden SystemCalls gefiltert.

Manual Categorisation

Dieser Pseudo-Container dient zum Begrenzen von manuellen Klassifizierungen. Bei einem Inventory Counting Run (siehe Rule Engine zur automatischen Klassifizierung) macht die Rule Engine dem Security Verantwortlichen Vorschläge zur Klassifizierung. Entscheidet dieser Element manuell zu klassifizieren, so kann er nur Dimensionen/Kategorien nutzen, die im „Manual Categorisation“ Container enthalten sind.

Cryptex

Cryptexe stellen eine Art Verschlüsselungscontainer dar. Diese werden z.B. genutzt, um verschlüsselte Ablage von Daten zu erzwingen oder klassifizierte Daten über nicht sichere Medien (USB-Stick) zu übertragen. Auch bei E-Mail Anhängen oder Attachments innerhalb von Outlook PST/OST Dateien können Cryptexe eingesetzt werden. Sie ermöglichen es, dass Attachments verschiedener Level innerhalb einer OST/PST Datei gespeichert werden können — D.h., die OST/PST Datei ansich ist „levellos“, die Attachments an den jeweiligen Mails sind Cryptexe mit jeweils unterschiedlichen Levels.

3.2. GATEWAY**URLs/HTTP Post**

Hier wird das Herausschicken sensibler Daten über HTTP Post (z.B. Webmail) unterbunden. Auch hier wird natürlich wieder aufgrund einer Vielzahl von Informationen (User, Quelle und Ziel entschieden) aber auch aufgrund des Inhaltes entschieden.

E-Mail

Ähnlich wie bei URLs kommen hier die Informationen über User, Quelle, Ziel und Inhalt zum Tragen.

Instant Messaging

Eine vielfach übersehene Gefahr sind Instant-Messaging Plattformen. Nahezu jede dieser Plattformen bietet heute die Möglichkeit Dateien über proprietäre Kanäle zu verschicken. Instant Messaging Container erlauben es, Datentransfers ganz zu unterbinden bzw. auf Empfänger-/gruppen abhängig vom Level einzuschränken.



4. RULE ENGINE

Bisher sind wir von einem einfachen Entscheidungsmodell ausgegangen: Sofern der Benutzer, Quelle und Ziel die ausreichenden Level besitzen, ist eine Datenbewegung erlaubt. Dies ist auf dem Gateway nur zum Teil richtig.

SafeEdge bietet die Möglichkeit den Inhalt einer Transaktion (z.B. E-Mails) mittels verschiedener inhaltsbasierter Verfahren zu untersuchen. Würde man dies bei jeder Transaktion durchführen wäre der Ressourceneinsatz immens. Hinzu kommen Anforderungen des Kunden, Entscheidungsprozesse anzupassen.

Beide Anforderungen werden mittels einer Rule Engine erfüllt:

4.1. RULE ENGINE IN SAFEEDGE

Eine Rule Engine (auch Expertensystem oder Workflow Engine) entscheidet aufgrund von Regeln und Fakten: Regeln „erzeugen“ aufgrund von Fakten neue Schlussfolgerungen (Fakten) oder stoßen Aktionen an. Bei Fakten gibt es Initialfakten, die „immer“ gelten bzw. Fakten, die sich bei jedem Durchlauf ändern. Wichtig ist hierbei, dass eine Rule Engine nicht automatisch alle Fakten abfragt — dies geschieht nur, wenn dies zur weiteren Entscheidungsfindung notwendig ist. Hier erlaubt die Rule-Engine den flexiblen Umgang mit diesen „unscharfen“ Fakten.

Kann z.B. bei einer E-Mail schon aufgrund von Sender/Empfänger entschieden werden, ist die ressourcenhungrige Textanalyse nicht notwendig und wird auch nicht angefordert.

Da Regeln nebenläufige Aktionen anstoßen können, hat der Kunde die Freiheit beliebige Aktionen innerhalb der Entscheidungsfindung anzustoßen: Z.B. darf der CTO geheime Informationen per E-Mail versenden — tut er dies mitten in der Nacht, wird zumindest ein weiterer Log Eintrag kreiert.

Jedoch können nicht nur Aktionen angestoßen werden, sondern auch externe Daten mit eingebunden werden. Möchte der Kunde z.B. die Mondphase in die Entscheidung einfließen lassen, ist dieses wenig sinnvoll aber machbar. Grundsätzlich erlaubt die verwendete Rule Engine (Jess, <http://herzberg.ca.sandia.gov/>) den Aufruf beliebiger Java Funktionen bzw. die Abfrage beliebiger Funktionen über Java.



Dies ermöglicht es, Anforderungen an den Workflow flexibel umzusetzen. Ein Beispiel ist die „User Intention“. Auch wenn einem Benutzer bestimmte Aktionen gestattet sind, kann z.B. auf 1000 dieser Aktionen pro Stunde reagiert werden — durch Ablehnung, Logeinträge, Bandbreitendrosselung...

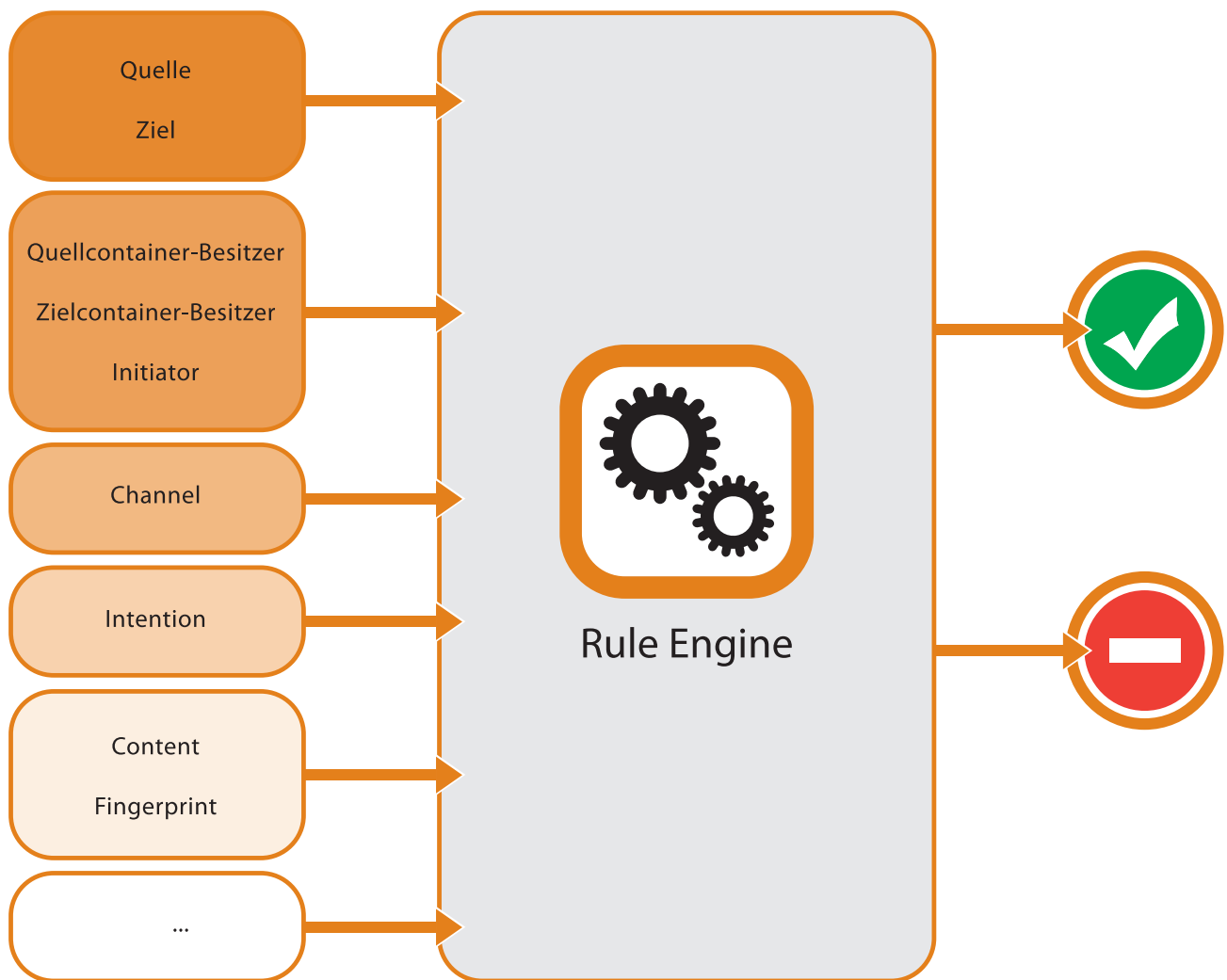


Abbildung 3: Rule Engine in SafeEdge

4.2. RULE ENGINE ZUR AUTOMATISCHEN KLASSIFIZIERUNG

Die Rule Engine wird ebenfalls zum Klassifizieren vorhandener Daten genutzt. SafeUse benötigt klassifizierte Daten, um Entscheidungen treffen zu können. Nicht klassifizierte (oder neue) Dateien werden mittels eines „Inventory Counting Runs“ gefunden und mittels der Rule Engine klassifiziert. Dabei macht die Rule Engine dem Security Verantwortlichen Vorschläge, wie die Daten zu klassifizieren sein. Die letztendliche Entscheidung über die Annahme, Ablehnung oder manuelle Änderung der Klassifizierung trifft der Security Verantwortliche und damit die beste Rule Engine überhaupt — der Mensch.

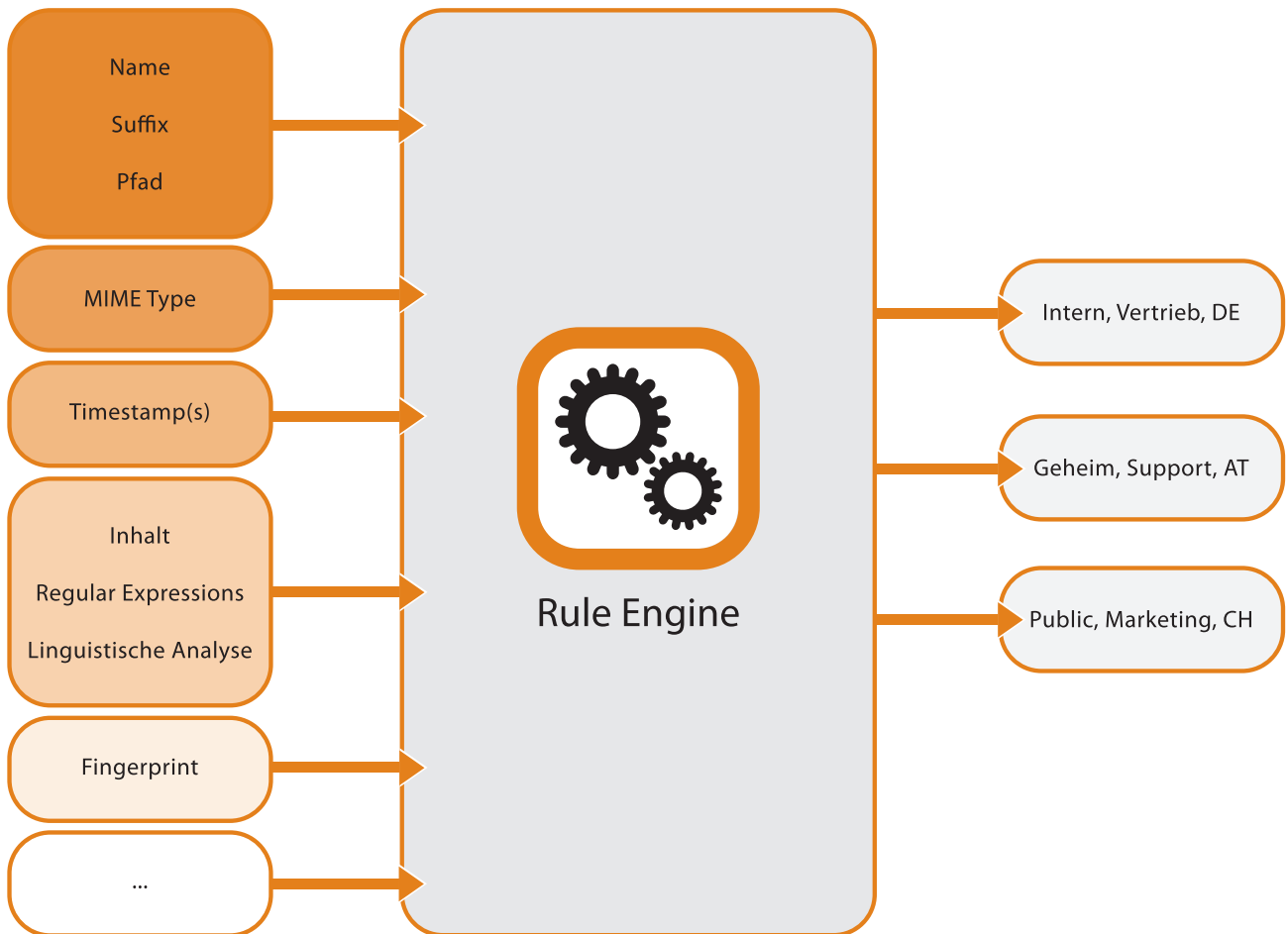


Abbildung 4: Rule Engine zur automatischen Klassifizierung

4.3. DETERMINISTISCHE FAKTEN

Einige Fakten, auf Basis derer die Rule Engine entscheidet, sind streng deterministisch. Dies gilt zum Beispiel für Datum, Uhrzeit, Dateibesitzer oder E-Mail Empfänger. Das Ermitteln dieser Fakten stellt keinen erhöhten Ressourcenaufwand dar. Daher werden diese Fakten als „initiale Fakten“ der Rule Engine zur Verfügung gestellt (forward-chaining).

4.4. PROBABILISTISCHE FAKTEN

Fakten, deren Bereitstellung nicht in allen Fällen notwendig ist oder deren Erzeugung ressourcenintensiv ist, werden erst bei Bedarf der Rule Engine bereitgestellt (backward-chaining). Dies umfasst u.a.

- **Textextraktion:** Extrahieren von Textinformationen aus Binärdaten (z.B. Microsoft Office oder OpenOffice). Hierbei werden auch eventuelle Container (Archive) „ausgepackt“.
- **Linguistische Analyse:** Basierend auf den Textdaten wird die Sprache des Textes erkannt und der Inhalt auf Grundformen zurückgeführt. Dies umfasst die Vereinfachung bzgl. Zeit, Fall, Geschlecht, Mehrzahl ...
- **Pattern/Regular Expressions:** Auf Basis der linguistischen Analyse können einfache Pattern oder Regular Expressions zur Erkennung von Mustern (z.B. Aktennummern) eingesetzt werden. Hierbei werden auch Schwellwerte bzw. der Abstand der gefundenen Abschnitte in die Entscheidungsfindung einbezogen.
- **Fingerprint:** Basierend auf den Referenzfingerprints vorhandener Daten können hier auch nicht-klassifizierte Daten bzw. Teilstücke zugeordnet werden.

5. SYSTEMARCHITEKTUR

5.1. ALLGEMEINES

Perimetrix® SafeSpace™ besteht aus Netzwerkdiensten, die verschiedene Funktionen wahrnehmen und Enduser Agenten, die auf den PCs eine sichere Umgebung zum Umgang mit klassifizierten Daten bereitstellen. Die Netzwerkdienste sind plattformunabhängig und können frei im Netz verteilt werden.

Die Enduser Agenten (SafeUse) sind z.Z. für Windows XP und Windows Vista verfügbar. Die Netzwerkdienste (SafeEdge und Backend) sind bis auf wenige Ausnahmen auf jeder Java-fähigen Plattform lauffähig. SafeSpace nutzt eine zentrale relationale Datenbank zur Speicherung der Konfigurationen und der dynamischen Daten (Ereignisse, Benachrichtigungen, Vorfälle ...). Dabei ist SafeSpace unabhängig von der Datenbank und unterstützt mehr als 20 verschiedene Produkte wie Oracle, Microsoft SQL Server, DB/2 aber auch MySQL oder Postgress. Die zentrale Managementkonsole ist webbasierend und verzichtet somit auf einen dediziert zu installierenden Management Client auf dem Administrator PC.

Grundsätzlich ist festzuhalten, dass die meisten Dienste nicht direkt sondern über Loadbalancer-Dienste angesprochen werden. Dies ermöglicht es dem Gesamtsystem dynamisch Services (und damit Ressourcen) hinzuzufügen oder zu verändern, ohne es zu beeinträchtigen.

5.2. SERVICES

Name Lookup Service

Dies ist der zentrale Namesdienst. Alle anderen Dienste nutzen ihn, um die physikalische Lokation anderer Dienste zu lokalisieren. Beim Start anderer Dienste registrieren sich diese beim Name Lookup Service und sind damit im Gesamtsystem sicht- und erreichbar.

SMTP Sensor

Dieser Dienst dient der Integration in SMTP Messaging Umgebungen. Er empfängt eingehende E-Mails und leitet diese an die Decision Engine zur Analyse weiter.

SMTP Mail Delivery Service

Hier werden E-Mails an den next-hop/Relay Mailserver ausgeliefert, sofern die Zustellung erlaubt wurde. Auch E-Mails, die vom System geblockt und vom CSO nachträglich freigegeben wurden, werden über diesen Dienst ausgeliefert.



Network Packet Capturer

Die Untersuchung von HTTP und IM Verbindungen basiert auf diesem Dienst. Hierbei werden „rohe“ Netzwerkpakete untersucht. Die hierzu notwendige Library (libpcap.so) ist plattformabhängig.

Decision Engine

Die Decision Engine (Rule Engine) ist das „Gehirn“ des Systems und analysiert einkommende Daten von Sensoren und entscheidet über ihr Schicksal. Abhängig von der Regelbasis werden verschiedene andere Dienste kontaktiert, die tiefergehende Untersuchungen des Inhaltes vornehmen. Intern nutzt die Decision Engine Jess als hochperformante Rule Engine basierend auf dem Rete Netzwerk Algorithmus.

Morphological Categoriser

Dieser Dienst wird von der Decision Engine kontaktiert, um den Inhalt einer sprachlichen Analyse zu unterziehen. Dabei untersucht dieser Dienst sprachabhängig den Inhalt und klassifiziert die Daten entsprechend. Hierbei werden auch linguistische Methoden angewandt, um eventuelle sprachliche Veränderungen (z.B. andere Zeiten bzw. Verb- oder Verlaufsformen) zu erkennen. Ein Teil dieser Funktionalität ist aus Geschwindigkeitsgründen als native Bibliothek ausgelegt und bei Bedarf von Perimetrix für entsprechende Plattformen verfügbar.

Digital Fingerprint Categoriser

Dieser Dienst kategorisiert für die Decision Engine einkommende Dokumente basierend auf vorher berechneten Fingerprints klassifizierter Dokumente.

Digital Fingerprint Caculator

Zusammen mit dem Digital Fingerprint Categoriser klassifiziert dieser Dienst bekannte klassifizierte Dateien.

Keystore Service

Zentrale Schlüsseldatenbank für Cryptex© Container.

Cryptostore Service

Verarbeitet Cryptex© Ver- und Entschlüsselungs-Anfragen.

Text Extraction Service

Wird von der Decision Engine, Morphological Categoriser und Digital Fingerprint Categoriser kontaktiert um den MIME Type zu erkennen, Text zu extrahieren und eventuelle (rekursive) Dekomprimierung durchzuführen.



Workstation Events Collector Service

Dieser Dienst verwaltet Benutzer PCs und empfängt Ereignisse von diesen.

Scheduled Task Execution Service

Ähnlich wie „cron“ verwaltet dieser Dienst zeitabhängige Aktionen. Er wird zum Beispiel für zeitaufwendige Prozesse wie die Inventarisierung von Daten aber auch zum Erstellen von Reports genutzt.

Workstation File Classification Service

Dieser Dienst ist vergleichbar mit der Decision Engine auf dem Gateway. Allerdings wird dieser zur automatisierten Inventarisierung von Client Daten genutzt.



Perimetrix Systems GmbH

Birkenstrasse 4a
85107 Baar-Ebenhausen
Germany

Phone: +49 (8543) 3399 700
Fax: +49 (8543) 3399 704

info@perimetrix-systems.com
www.perimetrix-systems.com

KEEPING SECRETS SAFE

