Gottfried Wilhelm
Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering

Konzept und Implementierung eines Analyseverfahrens für DLP-Systeme basierend auf Informationsflüssen

Bachelorarbeit

im Studiengang Informatik

von

Svenja Schulz

Prüfer: Prof. Dr. Kurt Schneider
Zweitprüfer: Prof. Dr. Joel Greenyer
Betreuer: Dipl.-Inform. Stefan Gärtner

Hannover, 28.09.2014

Zusammenfassung

Die Daten eines Unternehmens gewinnen zunehmend an Bedeutung und sind oft entscheidend für den Unternehmenserfolg. Diese Daten sollen daher vertraulich behandelt werden. Immer öfter kommt es jedoch zu Vorfällen wie Datendiebstahl oder der unerwünschten Veröffentlichung interner Daten. Dies kann schwere wirtschaftliche Folgen für die betroffenen Unternehmen haben, da einerseits gegebenenfalls Wettbewerbsvorteile verloren gehen können und andererseits oft auch das Image des Unternehmens unter derartigen Vorfällen leidet.

Um vertrauliche Daten sowohl vor Datendiebstahl als auch vor unerwünschter Offenlegung zu schützen wurden daher *Data-Loss-Prevention*-Systeme entwickelt. Diese Systeme identifizieren vertrauliche Daten anhand zuvor konfigurierter Richtlinien, zeichnen die Zugriffe darauf auf und sind in der Lage, verdächtige Aktionen zu erkennen und gegebenenfalls zu blockieren.

In dieser Arbeit wird eine Methode entwickelt, die die Informationsflüsse im Unternehmen anhand der FLOW-Methode und der Daten des bestehenden *Data-Loss-Prevention*-Systems analysiert. Basierend auf den Erkenntnissen dieser Analyse, werden sowohl organisatorische Maßnahmen, als auch mögliche Anpassungen der Konfiguration des *Data-Loss-Prevention*-Systems und weitere Möglichkeiten zur Verbesserung des Schutzes und der Reduktion der Fehlmeldungen des *Data-Loss-Prevention*-Systems aufgezeigt.

Inhaltsverzeichnis

1. Einleitung

1.1. Motivation

Der Schutz vertraulicher Daten ist auch für Unternehmen ein immer wichtigeres Thema. Insbesondere Unternehmen in einem Umfeld mit hohem Konkurrenz- und Innovationsdruck müssen ihre Daten vor Offenlegung schützen um Schäden zu vermeiden. Dabei ist es nicht ausreichend sich vor externen Angreifern zu schützen, sondern auch unerwünschte Informationsflüsse durch Mitarbeiter sollen unterbunden werden. Ein Problem stellen hierbei nicht nur böswillige Aktionen dar, sondern auch versehentliche Offenlegung der Daten durch *Social Engineering*-Angriffe oder in Folge von mangelndem Bewusstsein und unachtsamen Verhalten der Nutzer im Umgang mit vertraulichen Daten.

Einen Lösungsansatz für diese Probleme stellen Sicherheitstechnologien wie *Data Loss Prevention* (DLP) dar. Ein DLP-System identifiziert schützenswerte Daten und zeichnet sämtliche Zugriffe darauf auf. Ziel ist es dabei, verdächtige Aktivitäten möglichst frühzeitig zu erkennen und gegebenenfalls sogar zu blockieren, um so den Schaden minimieren zu können [BSI12].

Um die Akzeptanz und Effizienz eines solchen Systems zu gewährleisten, müssen die schützenswerten Daten möglichst sicher erkannt werden und die Rate an Fehlmeldungen möglichst gering gehalten werden. Zu häufige Fehlalarme erhöhen zum einen den Administrationsaufwand eines solchen Systems, zum anderen werden auch die Benutzer durch häufige Fehlmeldungen belästigt und in ihrem Arbeitsablauf gestört.

1.2. Ziele der Arbeit

Im Rahmen der vorliegenden Arbeit soll die Ausgangssituation in Hinblick auf *Data Loss Prevention* bei der Continental AG weitergehend analysiert werden. Besonderes Augenmerk liegt dabei auf möglichen Regelmäßigkeiten und Mustern im Verhalten der Mitarbeiter, sowie der Informationsflüsse dieser Mitarbeiter. Darauf aufbauend soll erörtert werden, ob es möglich ist anhand der in der Analyse erfassten Informationen das bestehende DLP-System in Bezug auf die Rate der Fehlmeldungen zu verbessern. Die Reduzierung dieser Rate ist erforderlich, um zum einen die Effektivität des Systems zu gewährleisten und zum anderen die Akzeptanz der Nutzer zu erhöhen.

1.3. Struktur der Arbeit

Zunächst soll auf verwandte Arbeiten und Themengebiete eingegangen werden, um mögliche Ansätze und Verfahren für *Data Loss Prevention* aufzuzeigen. Daraufhin sollen die Grundlagen von Data Loss Prevention und den zum Einsatz kommenden Analyseverfahren dargestellt werden. Im nächsten Schritt sollen die Anforderungen für die erfolgreiche Implementation eines DLP-Systems herausgearbeitet werden und Grenzen eines derartigen Systems aufgezeigt werden. Daraufhin werden die gewählten Analyseverfahren durchgeführt und ein möglicher Verbesserungsansatz skizziert.

Abschließend wird ein Fazit gezogen, sowie ein Ausblick auf weitere Möglichkeiten und denkbare zukünftige Entwicklungen gegeben.

2. Verwandte Arbeiten und Themengebiete

Data Loss Prevention ist ein relativ junges Forschungsgebiet, welches erst in den letzten Jahren an Bedeutung gewonnen hat. Ein Grund hierfür ist die zunehmende Regulierung des Umgangs mit persönlichen Daten durch übergeordnete Institutionen und Behörden, beispielsweise durch die *HIPAA*- [1] oder *PCI DSS* [2]-Richtlinien, die Patienten- beziehungsweise Kreditkartendaten schützen sollen und deren umfassende Umsetzung durch den Einsatz von DLP-Systemen vereinfacht wird. Hinzu kommt auch, dass das Bewusstsein für Datenschutz und Datensicherheit infolge der diversen Vorfälle der letzten Jahre allgemein gewachsen ist.

In diesem Abschnitt soll anhand der bereits bestehenden Arbeiten dargestellt werden, welche anderen Themenbereiche interessante Ansätze für *Data Loss Prevention* liefern.

2.1. Modelle zur Zugriffskontrolle

Vertrauliche Daten sollten schon seit langer Zeit vor unerwünschter Offenlegung geschützt werden. Daher wurden früh Modelle zur Zugriffskontrolle und Rechtevergabe entwickelt, die auch mit klassifizierten Daten arbeiten. Haupteinsatzgebiete waren hier das Militär und andere Regierungsinstitutionen, sowie privatwirtschaftliche Einrichtungen wie Banken oder Versicherungen.

Ein Beispiel für ein derartiges Modell ist das *Chinese-Wall-Modell* [BN89], das auch nach den Entwicklern *Brewer-Nash-Modell* genannt wird. Ein großes Einsatzgebiet dieses Systems sind etwa Unternehmensberatungen, da es Informationsflüsse, welche zu Interessenskonflikten führen, verhindert. So kann ein Berater beispielsweise, sobald er einmal auf die Daten eines Unternehmens zugegriffen hat, nicht mehr auf die Daten seiner Konkurrenzunternehmen zugreifen. Das *Chinese-Wall-Modell* setzt das *Mandatory Access Control*-Konzept um, ermöglicht dem Nutzer aber auch einen gewissen Einfluss, so dass es flexibler und nutzerfreundlicher ist als ein System, das ausschließlich *Mandatory Access Control* verfolgt. Auch der Administrationsaufwand ist geringer. Aufgrund dieser Eigenschaften könnte sich dieses Modell gut mit DLP-Systemen kombinieren lassen. Die Anwendbarkeit dieses Modells im Bereich Data Loss Prevention wird in [JZJ12] näher erörtert.

[1] "Health Insurance Portability and Accountability Act of 1996" vom US-Department of Health & Human Services, Washington, D.C.

[2] Payment Card Industry Data Security Standard

2.2. Intrusion Detection

Viele Verfahren, die in DLP-Systemen verwendet werden, wurden ursrpünglich für Intrusion Detection Systeme oder Antivirenprogramme entwickelt. Dies liegt zum einen an der thematischen Nähe, zum anderen auch an der guten Übertragbarkeit der Verfahren. Aus diesem Grund stammen die meisten etablierten DLP-Systeme von namhaften Herstellern der Antiviren-Branche.

Hier werden oft *Data Mining*-Verfahren eingesetzt, wie etwa die Ausreißererkennung in [XL08], um die Rate an Fehlalarmen in Intrusion Detection Systemen zu reduzieren. *Data Mining*-Verfahren dienen dazu große Datenmengen zu analysieren, wodurch Muster gefunden und neue Erkenntnisse erlangt werden sollen. Einen Überblick über die verschiedenen eingesetzten Verfahren in den Bereichen gibt [WB10].

Eine wichtige Rolle spielt auch das *Association Rule Mining*, insbesondere auch in Kombination mit *Fuzzy Logic*. Analysiert werden diese Verfahren in Zusammenhang mit *Intrusion Detection* in [He14] und [TRM09]. Diese Verfahren wurden auch im Rahmen dieser Arbeit für die Verwendung mit DLP untersucht.

Auch die Erkennung von Anomalien spielt in diesem Bereich eine wichtige Rolle und wird etwa in [MP08] näher erläutert. Auf den ersten Blick ließe sich auch dieses Verfahren auf DLP übertragen, in der Realität ist dies jedoch problematisch. Das grundlegende Problem dabei ist, dass nur Aktionen auf klassifizierten Daten durch das DLP-System aufgezeichnet werden und daher das normale Verhalten nicht bekannt ist. Daher lassen sich Anomalien nicht sicher erkennen.

2.3. IT-Forensik

Die IT-Forensik ist ein Teilgebiet der Forensik, das sich auf die Erhebung, Analyse und Auswertung digitaler Spuren in Daten mit potenziell kriminellem Hintergrund spezialisiert hat. Der hierbei verfolgte Prozess findet immer *post mortem*, also nach dem Vorfall, statt und gliedert sich in vier Schritte:

1. *Identifizierung*
2. *Datensicherung*
3. *Analyse*
4. *Auswertung und Aufbereitung*

Dieser Prozess zeigt bereits Parallelen zu der geplanten und in Kapitel 5 dargelegten Vorgehensweise, was auf eine mögliche Anwendbarkeit forensischer Verfahren im Rahmen dieser Arbeit schliessen lässt. Allerdings wird immer eine *post mortem*-Analyse auf in der Regel sehr großen Datenmengen durchgeführt, was sich für DLP-Systeme weniger eignet, da hier eine Live-Erkennung von Ereignissen gewünscht ist und die vorhandene Datenmenge nicht für alle Verfahren ausreicht.

Sehr häufig sind Log-Daten eine wichtige Grundlage der forensischen Untersuchungen, wie sie auch bei DLP-Systemen auftauchen. Diese werden dann anhand verschiedener Verfahren analysiert und es werden beispielsweise Profile und Assoziationsregeln erstellt. Dabei werden sowohl Daten aus physischen Umgebungen, wie etwa in [FY07], als auch aus digitalen Umgebungen verwendet. Gerade diese Verfahren sind für die vorliegende Arbeit sehr interessant und wurden daher teilweise adaptiert. In [AV02] wird beispielsweise *Association Rule Mining* verwendet um aus Logdaten Profile zu erstellen, die als Grundlage für weitere forensische Maßnahmen dienen. Es wurden auch bereits Verfahren entwickelt, die anhand einer forensischen Analyse Datendiebstahl aufdecken sollen. Diese Verfahren wurden in [HFW11], [LLP12] und [Gri11] entwickelt und vorgestellt.

3. Grundlagen

3.1. Data Loss Prevention

Mit Data Loss Prevention werden Softwaresysteme bezeichnet, die vertrauliche Daten vor Offenlegung schützen sollen. DLP wird dabei nicht nur für vertrauliche Unternehmensdaten eingesetzt, sondern in vielen Fällen auch für Personendaten. Wichtig ist, dass nicht nur vor Angreifern, die bewusst Daten stehlen, gewarnt wird, sondern auch unbeabsichtigte Informationsabflüsse durch Fehlverhalten von Mitarbeitern verhindert werden.

Um diese Aufgabe zu bewerkstelligen muss das DLP-System zunächst die schützenswerten Daten identifizieren. Dies geschieht entweder inhaltsbasiert, kontextbasiert oder mit Verfahren des maschinellen Lernens, welche die Struktur der Dateien mit bereits als vertraulich bekannten Dateien abgleichen [Kas08].

Im nächsten Schritt werden sämtliche Aktionen, in welche vertrauliche Daten involviert sind, erkannt und mit den Richtlinien des DLP-Systems verglichen. Verletzt ein Zugriff eine oder mehrere dieser Richtlinien, wird der Zugriff aufgezeichnet und eine zuvor festgelegte Maßnahme ergriffen [Bau13]. Dies ermöglicht eine frühzeitige Erkennung und gegebenenfalls sogar eine Vermeidung von unerwünschten Informationsabflüssen, da verdächtige Aktionen blockiert und an einen authorisierten Mitarbeiter zur Freigabe weitergeleitet werden können. Zudem ist es möglich, die Aktion vom Benutzer selbst freigeben zu lassen und ihm eine Warnmeldung anzuzeigen, was das Bewusstsein sowie die Achtsamkeit der Nutzer erhöht und es ihnen zukünftig ermöglicht besser einzuschätzen, welche Daten vertraulich sind und wie mit diesen verfahren werden soll. Um diese positiven Effekte auf das Bewusstsein der Nutzer langfristig zu erhalten und die Effizienz und Akzeptanz des DLP-Systems sicher zu stellen, ist es erforderlich die Rate an Fehlalarmen, also legitimen Verhaltens, das vom DLP-System dennoch gemeldet wird, möglichst gering zu halten.

3.1.1. Klassifizierung von DLP-Systemen

DLP-Systeme lassen sich nach unterschiedlichen Eigenschaften und Kriterien klassifizieren. Einen Überblick über die Möglichkeiten zur Klassifizierung bietet die in Abbildung 3.1 dargestellte Taxonomie.

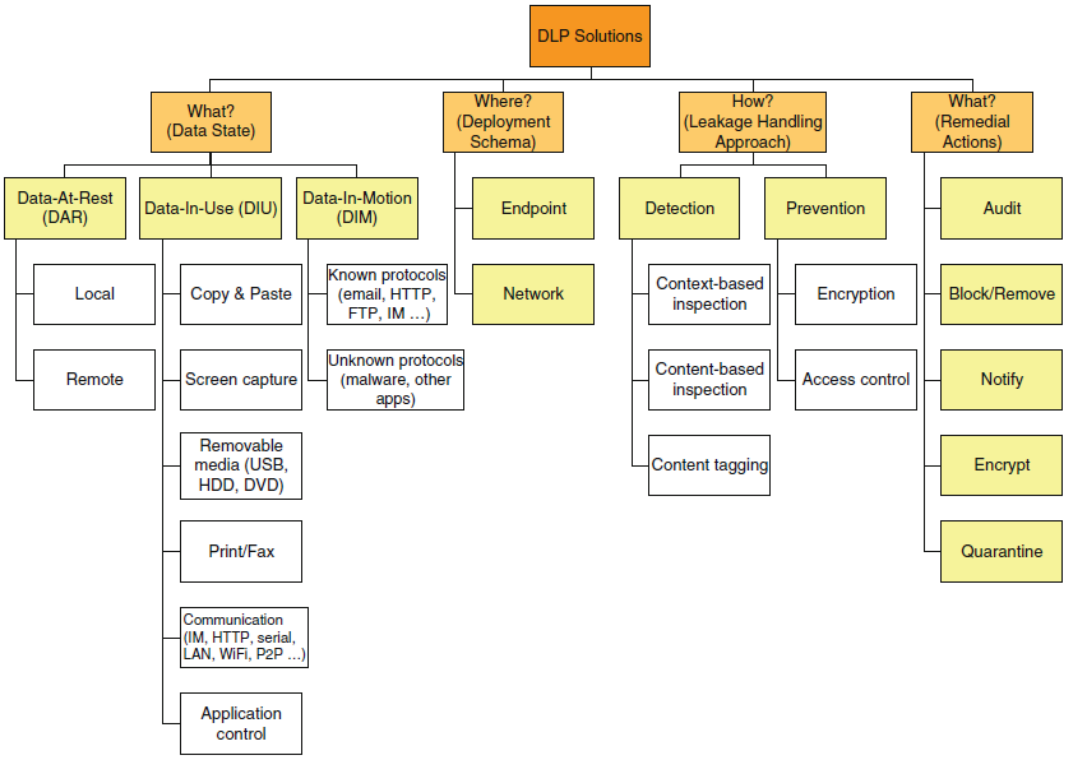

Abbildung 3.1.: Taxonomie von DLP-Systemen [SER12]

3.1.2. Zustand der Daten

Im Rahmen von DLP werden verschiedene Zustände von Daten unterschieden [Bau13]:

- Data at Rest:
  Diese Daten befinden sich an einem festen Speicherort, wie etwa einem lokalen oder Netzlaufwerk und werden zur Zeit nicht verwendet.
- Data in Use:
  Dies sind alle Daten, mit denen der Nutzer gerade interagiert, etwa indem er sie bearbeitet oder druckt.
- Data in Motion:
  Diese Daten durchqueren gerade das Netzwerk, sie werden beispielsweise gerade auf einer Website hochgeladen oder versendet. Dabei können sowohl bekannte, standardisierte Protokolle wie auch unbekannte Protokolle, etwa durch Malware, verwendet werden [SER12].

Je nach Zustand der Daten sind unterschiedliche Maßnahmen nötig, um sie vor Offenlegung zu schützen.

3.1.3. Art des Monitoring

Moderne DLP-Systeme unterscheiden zwischen Endpoint und Network Monitoring.

- Endpoint Monitoring:
  Das Endpoint Monitoring findet direkt auf dem Computer des Benutzers statt. Es muss also auf jedem Client, der genutzt wird, die entsprechende Monitoring Software installiert und für den jeweiligen Nutzer aktiviert werden. Dieser Client bezieht dann die Richtlinien nach denen er arbeitet von einem zentralen Management-Server und speichert sie auch lokal, so dass eine Überwachung auch ohne Netzwerkverbindung sichergestellt ist. Trotz dieses relativ hohen Aufwandes ist ein Endpoint Monitoring unverzichtbar um beispielsweise den Zugriff auf externe Speichermedien wie USB-Sticks oder Speicherkarten zu erkennen. Auch interne Datenträger lassen sich für Datendiebstahl nutzen, was jedoch einen höheren Aufwand erfordert. Interne und externe Speichermedien sind besonders kritisch zu sehen, da problemlos auch sehr große Datenmengen verarbeitet werden können und der weitere Verbleib vollkommen ungeklärt ist. Gerade bei externen Speichermedien ist das unverschlüsselte Speichern von vertraulichen Daten problematisch, da diese sehr unauffällig und klein sind und in Folge dessen oft verloren oder vergessen werden. Auch andere Aktionen, wie etwa das Versenden von E-Mails, können mittels Endpoint Monitoring erfasst werden.
- Network Monitoring:
  Mittels Network Monitoring wird der Netzwerkverkehr überwacht und analysiert. Es eignet sich also insbesondere um den Abfluss von Informationen mittels internetbasierter Wege, beispielsweise Cloud Services, E-Mails oder Dateiuploads, zu vermeiden. Auch der ansonsten unbemerkte Versand von Daten durch Schadprogramme wie Trojaner kann so entdeckt werden [SER12]. Die Arbeitsweise ähnelt dabei stark der von Intrusion-Detection- und -Prevention-Systemen. Hierbei wird der gesamte Netzwerkverkehr auf bestimmte Muster und Auffälligkeiten überprüft, es findet daher hauptsächlich eine inhaltsbezogene Überprüfung statt. Die genauen Richtlinien, nach denen das System arbeitet, werden wie beim Endpoint Monitoring auch üblicherweise über einen zentralen Management Server konfiguriert. Im Gegensatz zum Endpoint Monitoring ist es allerdings nicht nötig die Software auf allen Clients zu installieren. Dennoch gibt es in der Regel mehrere Punkte im Netzwerk, an denen das Monitoring stattfindet [Sym]. Das Network Monitoring kann aufgrund seiner Arbeitsweise nur *Data in Motion* schützen.

Da beide Monitoring-Ansätze über individuelle Stärken und Schwächen verfügen, ist häufig eine Kombination der beiden Ansätze sinnvoll, um die Daten bestmöglich zu schützen.

3.1.4. Handlungsansatz

DLP-Systeme können einen präventiven oder einen auf Erkennung basierenden Ansatz zum Schutz der Daten verfolgen [SER12].

- Erkennung:
  Bei dem erkennenden Ansatz soll unerwünschtes Verhalten erkannt und blockiert oder gemeldet werden. Hierzu gibt es drei unterschiedliche Ansätze:
  - Kontextbasiert:
    Bei diesem Ansatz werden die Kontextinformationen der erfassten Daten analysiert. Dazu zählen unter anderem Dateiname und -format, Speicherort, Sender und Empfänger [Kas08]. Die Verfahren hierfür sind bereits sehr weit entwickelt und erprobt, da sie auch in anderen Schutzsystemen, wie etwa Intrusion-Detection-Systemen und Spamfiltern eingesetzt werden.
  - Inhaltsbasiert:
    Hier wird der Inhalt der Daten mittels verschiedener Techniken analysiert.
    Möglich ist dabei ein Abgleich des Dateiinhalts mit einer zuvor festgelegten Stichwortliste. Wird der konfigurierte Grenzwert an Treffern erreicht, gilt die Datei als vertraulich. Eine weitere Möglichkeit zur Identifikation vertraulichen Dateiinhalts ist die Analyse mittels regulärer Ausdrücke. Mit Hilfe dieser regulären Ausdrücke können beispielsweise Kreditkarteninformationen oder Adressen anhand ihres Aufbaus erkannt werden [SER12]. Für diese beiden Verfahren enthalten die DLP-Systeme in der Regel bereits vorkonfigurierte Stichwortlisten und Ausdrücke für häufig vorkommende Daten, da für den Schutz derartiger Daten bereits vielfältige Regulierungen und Gesetze bestehen. Aus diesem Grund lassen sich diese Verfahren relativ schnell und einfach konfigurieren.
    Eine weitere Technik ist das Speichern von sogenannten Fingerabdrücken. Hierzu wird eine Hash-Funktion auf vertraulichen Daten angewendet und die Werte werden gespeichert. Später erfasste Daten werden dann ebenfalls gehasht und die Werte verglichen [SER12].
    Außerdem kommen auch Verfahren des maschinellen Verfahrens zum Einsatz, die die Daten mit bereits als vertraulich bekannten Daten vergleichen und anhand dieses Vergleiches klassifizieren.
  - Markierung:
    Bei diesem Ansatz werden Daten als vertraulich markiert. Diese Markierung wird gespeichert und kann in der Regel nicht ohne Weiteres geändert oder gelöscht werden. Markiert werden die Daten entweder vom Nutzer selbst oder automatisch. Für die automatische Markierung können einerseits die zuvor genannten kontext- und inhaltsbasierten Verfahren zur Erkennung vertraulicher Daten zum Einsatz kommen, andererseits können auch beispielsweise alle von einer bestimmten Anwendung erstellten Dateien automatisch als vertraulich markiert werden.

Diese Ansätze können in den Richtlinien des DLP-Systems häufig auch kombiniert werden. In der Regel werden mehrere Richtlinien erstellt, so dass etwa unterschiedliche Grenzwerte beim Abgleich mit Stichwortlisten für die unterschiedlichen Nutzergruppen eingestellt werden können. Außerdem ist es möglich, White- und Black-Lists zu erstellen, die zum Beispiel das Senden von Daten an bestimmte E-Mail-Adressen grundsätzlich blockieren.

- Prävention:
  Der präventive Ansatz soll die Offenlegung vertraulicher Daten von vornherein verhindern.
  - Verschlüsselung:
    Es können Richtlinien definiert werden, die festlegen, welche vertraulichen Daten verschlüsselt werden und welche Personen und Anwendungen diese wieder entschlüsseln dürfen.
  - Zugriffskontrolle:
    Die Zugriffsrechte der Nutzer werden anhand von Richtlinien vergeben und gegebenenfalls eingeschränkt.
  - Blockierung von Aktionen:
    Einzelne Aktionen, etwa das Kopieren von sensiblen Daten in die Zwischenablage, können für vertrauliche Daten blockiert werden.
  - Förderung des Bewusstseins: Hierzu zählt die Information und Schulung der Nutzer im Umgang und der Identifizierung von vertraulichen Daten.

3.1.5. Maßnahmen

Sobald ein Verstoß gegen die Regeln des DLP-Systems erkannt wird, bietet dieses in der Regel verschiedene Möglichkeiten zur Reaktion auf das Ereignis. Zunächst werden das Ereignis und die dazu gehörenden Informationen allerdings in jedem Fall in der Log-Datei gespeichert. Daraufhin stehen die folgenden weiteren Maßnahmen zur Verfügung, bei denen die Aktion im Allgemeinen zunächst blockiert wird:

- Prüfung durch den Nutzer:
  Dem Benutzer kann eine Meldung angezeigt werden, die besagt, dass diese Aktion gegen die Richtlinien des DLP-Systems verstößt. Daraufhin kann der Anwender einen Grund angeben, warum er diese Aktion ausführen wollte und die Aktion so zur Durchführung freigeben. Hierbei stehen in der Regel vorkonfigurierte Antworten zur Verfügung, oft hat der Nutzer jedoch auch die Möglichkeit seine Begründung frei zu formulieren. Diese Kommentare können dem DLP-Administrator wertvolle Informationen bezüglich der Konfiguration des Systems, aber auch des Bewusstseins und der Sensibilität der Nutzer geben.
- Benachrichtigen des Vorgesetzten:
  Es ist auch möglich, Verstöße direkt an den Vorgesetzten, den Administrator oder eine andere festgelegte Person zu melden. Dieser kann die Aktion dann blockieren bzw. freigeben und gegebenenfalls Rücksprache mit dem Mitarbeiter

halten und weitere Maßnahmen einleiten.

Diese Maßnahmen können in Response Rules konfiguriert werden und dann bei einer oder mehreren Richtlinien eingesetzt werden. Eine mögliche Response Rule mit Meldung an den Nutzer und zeitlicher Beschränkung zur Freigabe zeigt Abbildung 3.2 anhand eines Screenshots aus dem Symantec-System.

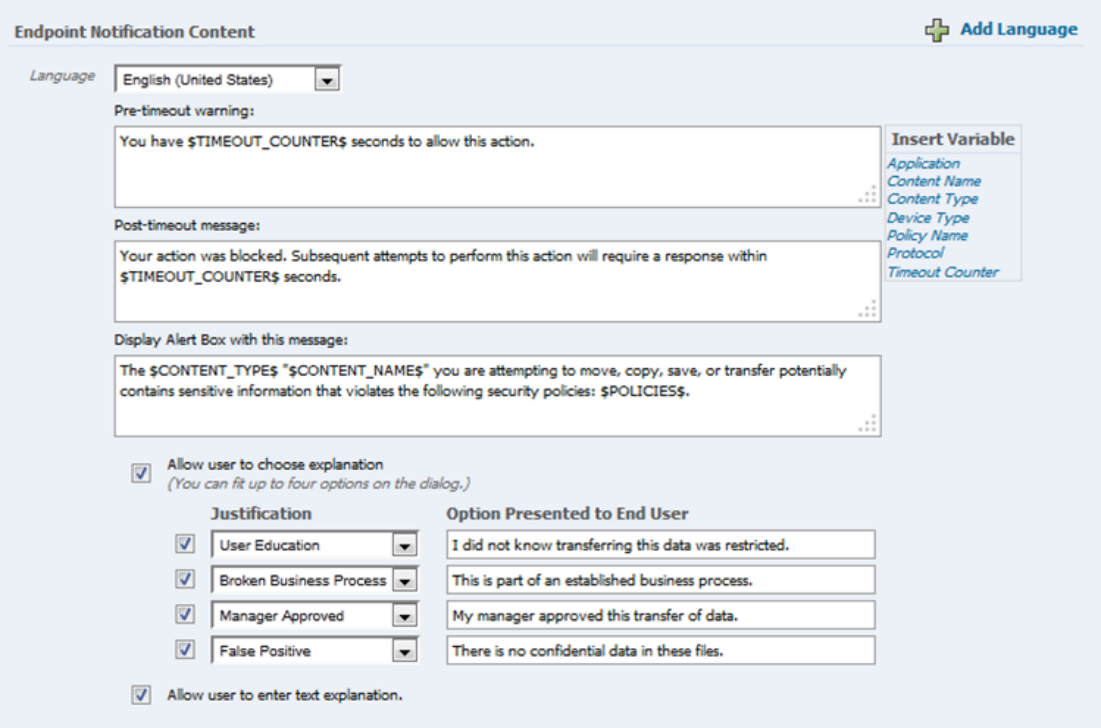

Abbildung 3.2.: Konfigurationsmöglichkeiten der Response Rules bei Symantec

3.2. FLOW

Der FLOW-Ansatz dient der systematischen Erfassung von Informationsflüssen in der Softwareentwicklung. Ziel ist es, durch die Modellierung und Analyse der Informationsflüsse und ihrer Eigenschaften neue Erkenntnisse zu gewinnen und bestehende Prozesse zu optimieren. Dieses Ziel soll durch einen Prozess mit den drei Phasen Erheben, Analysieren und Verbessern erreicht werden. Wird dieser Prozess mit dem Ziel der kontinuierlichen Verbesserung betrieben, muss er wiederholt durchgeführt werden und bildet so einen Kreislauf [SS12]. Dieser Kreislauf wird in Abbildung 3.3 dargestellt.

Da das FLOW-Projekt sehr praxisorientiert ist und Informationsflüsse auch abseits der Softwareentwicklung eine wichtige Rolle spielen, bietet es sich an, die Analyse- und Modellierungsmethoden des FLOW-Ansatzes auch auf andere Themengebiete anzupassen und zu übertragen. Insbesondere ist eine Analyse der Informationsflüsse zur

Unterstützung der Implementation eines DLP-Systems empfehlenswert, da so bestehende Prozesse auch auf Schwachstellen bezüglich der Informationssicherheit geprüft werden können. Doch auch eine Verbesserung der Informationsflüsse kann die Arbeit des DLP-Systems unterstützen und die unerwünschte Offenlegung vertraulicher Daten vermeiden.

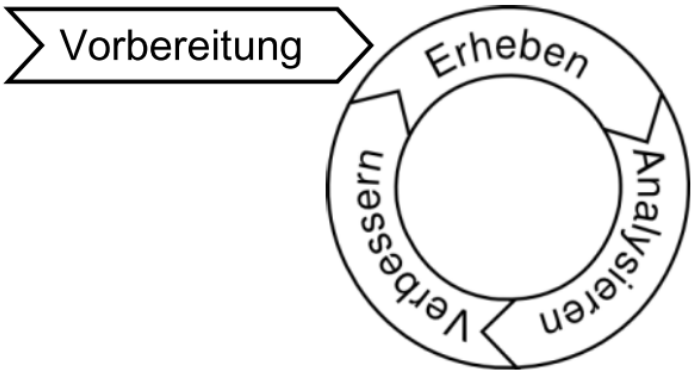

Abbildung 3.3.: FLOW-Prozess mit Vorbereitungsphase [SS12]

3.2.1. Definitionen

Die Definition des Begriffes Information variiert je nach der Domäne, in der er verwendet wird. Im Umfeld von FLOW wird Information als *potentielles personenunabhängiges Wissen* verstanden [St12].

Der FLOW-Ansatz unterscheidet Informationen anhand der Metapher der Aggregatzustände “fest” und “flüssig” [Sch06], wobei eine Information als fest gilt, wenn sie jederzeit von allen abgerufen und verstanden werden kann. Dies trifft in der Regel hauptsächlich auf analoge oder digitale Dokumente zu.

Ein großer Vorteil fester Informationen ist die langfristige Zugänglichkeit, insbesondere auch für Dritte, die so auch problemlos auf die Information zugreifen können. Zudem ist es möglich wiederholt die fixierte Information abzurufen. Allerdings ist es auch sehr aufwändig die Informationen zu erfassen, zu verarbeiten und zu speichern [St12]. Auch die Internalisierung fester Information ist langwieriger und schwieriger als bei flüssigen Informationen.

Als flüssige Informationen gelten alle Informationen, die nicht fest sind [Sch06]. Dies betrifft insbesondere das Wissen und Gedächtnis von Personen, aber auch z.B. Chat-Protokolle und Notizzettel. Diese Informationen sind im Gegensatz zu festen Informationen zwar leicht abruf- und speicherbar, gehen hingegen aber auch leicht verloren, verändern sich oder werden missverstanden. Auch ist ein Zugriff von Dritten auf diese Informationen nicht ohne Weiteres möglich.

Sowohl flüssige als auch feste Informationen sind in der Lage zu “fließen”. Als fließen

versteht man dabei den Übergang der Information von einem Informationsspeicher zu einem anderen. Hierbei ist es jedoch in der Regel nicht so, dass der Informations- oder Wissensgehalt beim sendenden Speicher abnimmt, wie man es durch das Bild der Metapher annehmen könnte, sondern die Information wird in der Regel vervielfältigt.

Die Informationsflüsse können je nach Art der involvierten Informationsspeicher in vier Formen unterschieden werden: Sozialisation, Externalisierung, Internalisierung und Kombination. Unter Sozialisation versteht man einen Informationsfluss zwischen zwei Personen, wie etwa in einem direkten Gespräch. Als Externalisierung gilt ein Informationsfluss, in dem eine Person ihr Wissen "verfestigt", also z.B. in einem Dokument aufschreibt. Internalisierung hingegen bezeichnet die Aufnahme und das Verstehen von fester Information durch eine Person [St12]. Als Kombination gelten Informationsflüsse, in denen nur feste Informationen involviert sind, etwa das Kopieren oder Versenden von Dateien. Für DLP-Systeme sind hauptsächlich Informationsflüsse der Formen Externalisierung und Kombination, aber auch Internalisierung interessant, da nur diese überhaupt durch das System erfasst werden können.

3.2.2. Notation

Um FLOW-Modelle abzubilden, wurde eine Notation [SS12] entwickelt, die intuitiv und daher leicht verständlich ist. Die einzelnen Symbole finden sich in Abbildung 3.4. Da das Ziel von FLOW-Modellen ein besseres Verständnis der Informationsflüsse und nicht die vollständige Abbildung dieser ist, sollte man zunächst nur grob modellieren und gegebenenfalls später verfeinern [St12]. Die Modelle sollten, wie auch die Notation, möglichst einfach und übersichtlich gehalten werden.

Feste Informationsspeicher werden als einzelnes oder als Stapel von mehreren Dokumente dargestellt. Dies folgt der Definition, nach der feste Informationen in der Regel schriftlich fixiert sind. Ein Stapel von Dokumenten steht dabei immer für einen Dokumenttyp anstatt für ein konkretes Dokument. Flüssige Informationsspeicher hängen laut Definition meist direkt mit Personen zusammen und werden daher durch ein oder mehrere Smileys dargestellt. Eine Unterscheidung zwischen konkreten Personen und Rollen ist hier nur durch die Bezeichner möglich.

Als Symbol für Informationsflüsse dienen Pfeile, die optional auch farblich unterschiedlich gestaltet sein können um Erfahrungen von reinen Informationen abzugrenzen. Flüssige Informationsflüsse entspringen immer einem flüssigen Informationsspeicher und werden durch einen Pfeil mit gestrichelter Linie dargestellt. Von festen Speichern können entsprechend auch nur feste Informationsflüsse abgehen, die durch einen Pfeil mit durchgezogener Linie dargestellt werden. Für Informationsspeicher und -quellen, deren Art unbekannt ist, wurden zusätzliche Symbole entwickelt.

Zusätzlich gibt es ein Aktivitätssymbol, das zwei Aufgaben erfüllt. Mit ihm lassen sich einerseits mehrere Informationsspeicher und -flüsse zusammenfassen und verdecken, so dass eine hierarchische Struktur gebildet werden kann. Es kann aber auch das FLOW-Modell mit anderen, bereits bestehenden Prozessdarstellungen verknüpfen. Das Aktivitätssymbol verfügt über vier Verbindungspunkte, davon drei eingehende und

ein ausgehender. Von links erreichen die Aktivität die eingehenden Informationsflüsse, deren Informationen in der Aktivität verarbeitet werden. Die Ausgabe erfolgt dann auf der rechten Seite. Von unten erreichen unterstützende Informationen die Aktivität, von oben steuernde.

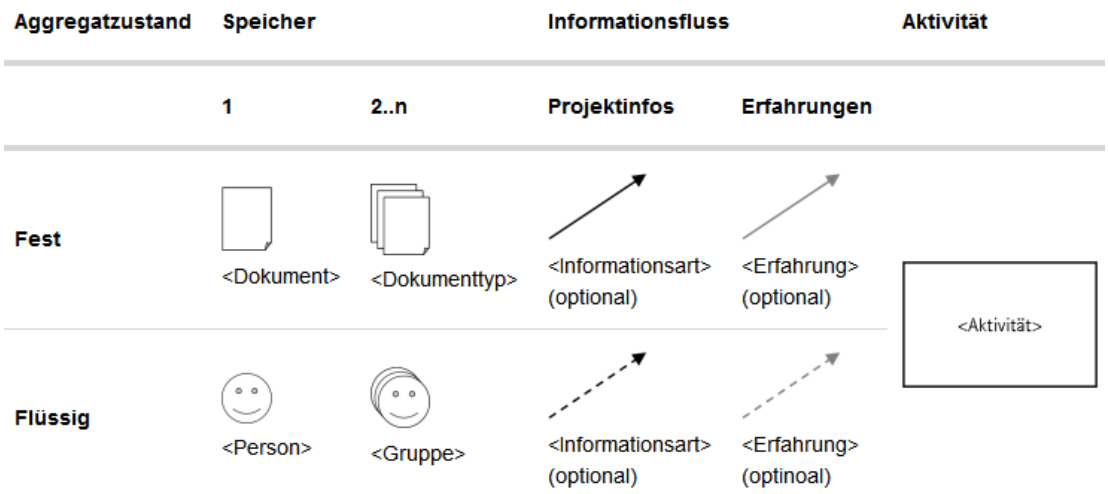

Abbildung 3.4.: Übersicht FLOW-Notation [SS12]

3.2.3. Vorbereitung

Die Vorbereitungsphase bildet die Grundlage für alle weiteren Aktivitäten im Rahmen des FLOW-Prozesses. Hier ist es zunächst wichtig, zu klären welches Ziel mit FLOW erreicht werden soll und Rahmenbedingungen wie den Zeitraum und den zu betrachteten Bereich festzulegen. Daraufhin werden weitere Projektparameter, wie etwa die beteiligten Personen und Rollen oder weitere Randbedingungen festgehalten und die passenden FLOW-Techniken ausgewählt. Sobald diese Arbeiten abgeschlossen sind, kann mit dem eigentlichen FLOW-Prozess begonnen werden.

3.2.4. Erhebung der benötigten Informationen

Bei der Erhebung der Daten lässt sich der *bottom-up*- oder der *top-down*-Ansatz verfolgen. Der *top-down*-Ansatz bietet sich vor allem für Soll-Modelle an, während sich der *bottom-up*-Ansatz vorrangig für Ist-Modelle eignet.

Es gibt verschiedene Verfahren um die benötigten Informationen zu erfassen, die sich anhand der Informationsquelle unterscheiden lassen. Die Verfahren aus dem Bereich Elicitation verwenden das Wissen der beteiligten Personen um die Informationsflüsse, wohingegen die ableitenden Verfahren mit festen Informationsquellen, etwa bereits bestehenden Ablaufbeschreibungen, arbeiten und aus diesen die Informationsflüsse ableiten. Als Verfahren kommen hierfür in Frage [St12]:

- Elicitation-Verfahren
  - *“Selber machen”:*
    Die erhebende Person führt die Aktion, die das Modell erfassen soll, selbst aus und erfasst dabei die benötigten Informationen.
  - *Beobachten:*
    Der FLOW-Experte beobachtet die zu erfassende Situation.
  - *Interview:*
    Die an der Situation beteiligten Personen werden von dem FLOW-Experten zu der Situation befragt.
  - *Fragebogen:*
    Die erhebende Person erstellt einen Fragebogen mit allen relevanten Fragen und verteilt diesen an die beteiligten Personen.
- Ableitende Verfahren
  - *Aus bestehenden Modellen:*
    Die für Informationsflüsse relevanten Teile bereits bestehender Modelle werden erfasst und zu einem neuen FLOW-Modell verbunden.
  - *Aus Kommunikationsereignissen:*
    Aus Kommunikationsereignissen, deren Aufzeichnung und Log-Daten werden Informationsflüsse abgeleitet.

Jedes der genannten Verfahren hat Vor- und Nachteile bezüglich der Kosten, des Vorbereitungsaufwandes, der Durchführbarkeit und der Exaktheit des Ergebnisses. Daher muss das Verfahren sorgfältig ausgewählt werden. Auch eine Kombination mehrerer Verfahren ist möglich.

3.2.5. Analyse

Auch in der Analysephase stehen unterschiedliche Methoden zur Auswahl, die auch miteinander kombiniert werden können [SS12].

- *Visualisierung:*
  Das FLOW-Modell wird anhand der FLOW-Notation visualisiert und vom Experten manuell untersucht.
- *Simulation:*
  Bei der Simulation werden zusätzliche Elemente in das Modell eingefügt und der Ablauf simuliert, um Probleme, wie etwa das Vergessen wichtiger Informationen, aufzudecken.
- *Mustersuche:*
  Es werden manuell, z.B. anhand des Musterkataloges [Ge08] oder automatisch Muster im FLOW-Modell gesucht.

3.2.6. Verbesserung

In der Verbesserungsphase werden die Erkenntnisse aus den anderen Phasen genutzt, um die Informationsflüsse zu optimieren. Hierzu können die Aggregatzustände der Informationsquellen und -flüsse verändert werden, aber auch die Informationsflüsse an sich können verändert [St12], beispielsweise zusammengefasst oder abgekürzt, werden. Auch ganze Aktivitäten und Muster können komplett oder teilweise angepasst werden.

3.2.7. Anwendung für DLP

Es bieten sich verschiedene Möglichkeiten an, die FLOW-Methode auch im Rahmen von DLP einzusetzen.

Um ein DLP-System gut zu konfigurieren, ist es nötig, die im Unternehmen stattfindenden Abläufe und Informationsflüsse sehr gut zu kennen. Diese Kenntnisse ermöglichen erst eine maßgeschneiderte Konfiguration der Richtlinien und Maßnahmen des DLP-Systems. Um diese Informationsflüsse und damit auch die Betriebsabläufe umfassend und systematisch zu erfassen, bieten sich insbesondere die verschiedenen Elicitation- und Analyseverfahren von FLOW an. FLOW kann auch dazu dienen, die bereits bestehenden Richtlinien und Ereignisse eines DLP-Systems zu verfeinern und zu filtern.

Werden alle drei Phasen des FLOW-Prozesses durchlaufen, können im Rahmen der Verbesserungsphase auch die bestehenden Informationsflüsse verändert werden um mögliche Sicherheitslücken zu vermeiden. Dies stellt eine präventive Herangehensweise dar.

3.3. Assoziationsanalyse

Die Assoziationsanalyse ist ein Teilgebiet des Data Mining und der künstlichen Intelligenz, das in der praktischen Anwendung eine große Bedeutung hat. Ein wichtiges Anwendungsgebiet ist die Analyse und Vorhersage von Kundenverhalten, beispielsweise als Grundlage für Crossmarketing. Ziel der Suche nach Assoziationsregeln ist es, in den vorhandenen Daten Regeln zu finden, die Korrelationen, Beziehungen und Gemeinsamkeiten in den Daten beschreiben. Eine Assoziationsregel besteht dabei immer aus einem Bedingungs- und einem Konsequenzteil und kann so als Wenn-Dann-Regel aufgefasst werden [AIS93]. Ein Beispiel für eine solche Regel ist “Wenn ein Kunde Mehl kauft, dann kauft er mit einer Wahrscheinlichkeit von 65% auch Eier”.

Die wichtigsten Eigenschaften einer Assoziationsregel sind Support, Confidence und Lift. Anhand dieser Kriterien lässt sich die Relevanz und Aussagekraft der Regel einschätzen.

Der Support stellt den prozentualen Anteil der Datensätze, in denen die Regel anwend-

bar ist, an den gesamten Daten dar. Somit beschreibt der Support die Wahrscheinlichkeit der Regel in Bezug auf die gesamte Datenbasis [Boll96]. D steht dabei in der Formel für die gesamte Datenmenge, bestehend aus unterschiedlichen Datensätzen, bezeichnet mit t.

$$support(\{Mehl\} \rightarrow \{Eier\}) = \frac{|\{t \epsilon D | (Mehl \cup Eier) \subseteq t\})|}{|D|}$$

Die Confidence wird berechnet, indem man die Anzahl der regelerfüllenden Datensätze mit der Anzahl der Datensätze, in denen nur der Bedingungsteil, nicht aber die Konsequenz, erfüllt wird in Beziehung setzt [AIS93].

$$confidence(\{Mehl\} \rightarrow \{Eier\}) = \frac{support(\{Mehl, Eier\})}{support(\{Mehl\})}$$

Der Lift zeigt an, wie sehr die Confidence den Erwartungswert übertrifft [Boll96]. Dazu vergleicht man den Support der Regel mit den miteinander mutliplizierten Supports der einzelnen Items.

$$lift(\{Mehl\} \rightarrow \{Eier\}) = \frac{support(\{Mehl \cup Eier\})}{support(\{Mehl\}) \times support(\{Eier\})}$$

Eine Assoziationsregel soll immer einen vom Nutzer festgelegten minimalen Support und minimale Confidence erfüllen [Boll96].

4. Anforderungen und Umsetzung von DLP-Systemen

4.1. Organisatorische und technische Rahmenbedingungen

DLP-Systeme haben ihren Ursprung in sehr stark strukturierten Umgebungen. Die ersten Systeme wurden im Bereich des Militärs eingesetzt, wo eine starke Hierarchie, festgelegte und strukturierte Arbeitsabläufe und klar definierte Zugriffsstrukturen den Einsatz erleichtern. Auch die Klassifikation von Dokumenten war in diesem Bereich bereits lange vor Einführung von DLP-Lösungen üblich.

Um ein DLP-System erfolgreich einzusetzen, ist es von besonderer Bedeutung zu analysieren, welche Daten im Unternehmen vertraulich sind und woran sich diese erkennen lassen. Diese Erkenntnisse müssen dann in die Konfiguration des DLP-Systems einfließen. Auch eine gezielte Kommunikation dieser Kriterien an die Benutzer ist sehr hilfreich, da sie so zum einen selbst erkennen können, welche Daten sensibel sind. Zum anderen kann dies das Verständnis für Meldungen des DLP-Systems positiv beeinflussen.

Ein DLP-System kann eine strukturierte Rechtevergabe nach dem sogenannten *Principle of Least Privilege* nicht ersetzen. Daher sollten bereits vor der Installation eines DLP-Systems Rechte nach festgelegten Richtlinien und Abläufen vergeben werden, um Datenlecks und unberechtigten Zugriff auf sensible Daten zu vermeiden. Zudem ist es von Vorteil, die Redundanz bei der Ablage der Daten möglichst gering zu halten [SER12].

Auch strukturierte Arbeitsabläufe sind für den Einsatz eines DLP-Systems förderlich. Diese bilden die Grundlage für eine spätere Bewertung der Vorfälle. Zudem können festgelegte Arbeitsabläufe und andere Vorgaben auch bereits Sicherheitslücken vermeiden.

Vorschriften und Richtlinien bezüglich des Umgangs mit vertraulichen Daten geben den Nutzern Sicherheit und können Schäden durch unbedarftes Verhalten abwenden. Eine mögliche Richtlinie könnte beispielsweise besagen, dass E-Mails mit Dateianhängen oder Daten auf USB-Sticks immer verschlüsselt werden müssen. Auch regelmäßige Schulungen in Bezug auf Datenschutz und Datensicherheit sind förderlich um die Awareness zu erhöhen.

Oft ist auch ein Datenaustausch mit externen Personen oder Einrichtungen erforderlich. Hierbei ist es wichtig festzuhalten, welche Daten auf welchem Wege ausgetauscht werden dürfen. Hierbei sind möglichst homogene Richtlinien wünschenswert, um möglicherweise verdächtige Abweichungen festzustellen.

Um umfassend vor unerwünschten Datenabflüssen zu schützen, ist auch eine fachgerechte Entsorgung alter Datenträger unbedingt erforderlich. Hierbei sollte man sich an den Standards des US Department of Defense (DoD 5220.22-M (E), DoD 5220.22-M (ECE)) oder des BSI ([**?**] BSI-TL 03420) orientieren.

4.2. Rechtliche Grundlagen

Der Betrieb eines DLP-Systems ist aus rechtlicher Sicht problematisch, da umfassende personenbezogene Daten aufgezeichnet werden und die Mitarbeiter und ihre Arbeitsleistung, je nach Konfiguration der Richtlinien des Systems, überwacht werden können [Ca11]. Eine Einwilligung der betroffenen Mitarbeiter in die Aufzeichnung der Daten ist daher unabdingbar. Diese Einwilligung muss freiwillig geschehen ([BDSG] §4a Abs. 1), das heißt es dürfen dem Mitarbeiter keine Nachteile entstehen, sollte er diese Einwilligung nicht erteilen. Auch der Betriebsrat hat aufgrund der umfangreichen Datenerhebung ein Mitspracherecht ([BetrVG] §87 Abs.1 Nr.6).

Da von dem DLP-System nur Aktionen aufgezeichnet werden, die gegen die Richtlinien verstoßen, folgt es dem Grundsatz der Datensparsamkeit ([BDSG] §3). Allerdings ist hierbei die genaue Konfiguration der Richtlinien zu beachten, schließlich ist es möglich, diese so allgemein zu konfigurieren, dass nahezu jede Aktion gegen eine Richtlinie verstößt.

4.3. Technische Grenzen

Grundsätzlich lässt sich jedes Sicherheitssystem umgehen, so dass es nie eine hundertprozentige Sicherheit geben kann. Man sollte sich daher nach der Implementierung eines DLP-Systems nicht in falscher Sicherheit wiegen.

Insbesondere lässt sich ein DLP-System in der Regel umgehen, indem man die geschützten Dateien mittels eines starken Verschlüsselungsalgorithmus verschlüsselt. Eine andere Alternative um DLP-Systeme zu umgehen, stellen teilweise auch steganographische Verfahren dar, mit denen die vertraulichen Daten etwa in harmlos aussehenden Bilddateien versteckt werden können.

Auch der Einsatz von SharePoint-, Cloud-, Citrix- oder Virtualisierungsystemen erschwert den erfolgreichen Einsatz eines DLP-Systems. Besonders problematisch sind diese Systeme bei einem reinen Endpoint-Monitoring, da die Zugriffe nicht sicher erfasst werden. Auch bei einem Network-Monitoring kann es jedoch zu Problemen kommen.

Je kritischer und wertvoller die Daten für das Unternehmen sind, desto eher lohnen sich auch aufwändigere Verfahren, um Daten zu stehlen. Insbesondere das physische Umfeld kann nicht mittels des DLP-Systems überwacht werden. Daher kann auch ein sehr strikt konfiguriertes DLP-System beispielsweise nicht erkennen und somit auch nicht verhindern, dass sensible Daten fotografiert oder abgeschrieben werden.

Diese Grenzen machen deutlich, dass ein DLP-System immer von organisatorischen Maßnahmen, wie Schulungen und sinnvollen Richtlinien im Umgang mit vertraulichen Daten, begleitet werden sollte. Die Analyse der Informationsflüsse wie in 5.2.4 dargestellt, kann die Entwicklung dieser Maßnahmen unterstützen und auch Fehlverhalten der Mitarbeiter im physischen Umfeld erkennen.

4.4. Neue Sicherheitsprobleme

Der Betrieb eines DLP-Systems kann auch zu neuen Sicherheitsproblemen führen, da in dem System viele personenbezogene Daten gespeichert sind [Tor12]. Zudem identifiziert ein derartiges System, wenn es gut konfiguriert ist, auch sämtliche vertraulichen Daten des Unternehmens. Dies sind für Angreifer potenziell sehr interessante Informationen, so dass es für sie sehr vielversprechend sein kann, das DLP-System anzugreifen. Sobald man damit erfolgreich ist, kann man daraufhin versuchen gezielt die vertraulichen Daten zu stehlen [Tor12]. Das Unternehmen hat dadurch wesentlich weniger Zeit um auf den Angriff zu reagieren, da das gezielte Stehlen weniger Dateien zum einen wesentlich schneller erfolgt als das Stehlen großer Datenmengen in der Hoffnung, dass auch interessante Daten enthalten sind. Zum anderen fällt auch der Transfer kleiner Datenmengen weniger auf, so dass der Angriff schwieriger zu bemerken ist. Zusätzlich verfügen die Angreifer durch die Daten des DLP-Systems bereits über detaillierte Informationen über das Unternehmen, seine Datenstruktur und Mitarbeiter. Diese Informationen sind eventuell selbst schon sehr interessant für den Angreifer und können anderenfalls auch eine gute Grundlage für *Social Engineering*-Angriffe bilden.

4.5. Ausgangssituation

Die IT-Abteilung der Continental AG beschäftigt sich seit Oktober 2013 mit dem Thema Data Loss Prevention. Zur Zeit wird die *Data Loss Prevention*-Software der Firma Symantec in einem Probebetrieb verwendet. Es besteht aus verschiedenen Modulen für unterschiedliche Aufgabenbereiche, die zentral über die Weboberfläche des Management Servers verwaltet werden [Sym]. Dieses System bietet sowohl Module für Endpoint- als auch für Network-Monitoring, wobei zur Zeit nur das Endpoint-Monitoring eingesetzt wird.

Vertrauliche Daten erkennt das System anhand einer Stichwortliste mit etwa 120 Wörtern. Dabei wird nur zwischen “vertraulich” und “nicht vertraulich” unterschieden, eine weitergehende Unterscheidung oder Klassifikation findet momentan nicht statt.

Da der Betrieb von DLP-Systemen eine weitgehende Überwachung der Mitarbeiter ermöglichen kann, muss dem Einsatz derartiger Software vom Betriebsrat zugestimmt werden. Aufgrund der entfallenden Zustimmungspflicht wird die Software momentan hauptsächlich bei leitenden Angestellten eingesetzt. Zurzeit werden die Aktivitäten

von 15 Personen aus unterschiedlichen Unternehmensbereichen durch das System erfasst.

Die Mitarbeiter werden aufgrund ihres Arbeitsbereiches in die zwei Gruppen *“Material Development Users”* und *“Non Material Development Users”* unterteilt, für die jeweils unterschiedliche Richtlinien gelten. Zur Zeit sind drei Richtlinien aktiv:

- *Material Development Users: Material Related Terms:*
  Diese Richtlinie löst bei Mitarbeitern, die der Gruppe “Material Development Users” zugeordnet wurden, bei Aktionen auf Dateien mit mindestens 7 enthaltenen Stichworten ein Ereignis aus.
- *Non Material Development Users: Material Related Terms:*
  Diese Richtlinie gilt für alle Mitarbeiter, die nicht in der Materialentwicklung tätig sind und löst bereits bei drei Stichworten ein Ereignis aus.
- *Continental Keywords All Matches:*
  Diese Richtlinie gilt zusätzlich zu den anderen beiden für alle Mitarbeiter und löst aus, sobald ein Stichwort gefunden wird.

Bei der Anzahl an Stichworten macht es keinen Unterschied, ob ein Wort mehrfach aufgetaucht ist oder mehrere unterschiedliche Stichwörter in der Datei enthalten sind. Auch der Schweregrad der einzelnen Ereignisse wird nicht weiter unterschieden.

In den Log-Daten des DLP-Systems werden alle derartigen Ereignisse, deren Art, der auslösende Computer und Benutzer, das Datum, die gefundenen Stichwörter sowie je nach Art des Ereignisses weitere Informationen gespeichert. Mögliche Ereignisarten sind dabei E-Mail, Schreiben auf Wechselmedien, Schreiben auf Netzlaufwerke, Drucken und die Übertragung via HTTPS/SSL. Eine weitere Reaktion auf diese Ereignisse, etwa eine Meldung an den Benutzer, erfolgt momentan noch nicht, was auch durch die hohe Anzahl an Meldungen begründet wird. Aufgrund der hohen Vertrauenswürdigkeit der Nutzer im Testbetrieb wird davon ausgegangen, dass diese Meldungen zu einem großen Teil Fehlmeldungen sind.

5. Ansatz und Methode

5.1. Ansatz

Der in dieser Arbeit entwickelte Ansatz hat als Ziel, das normale Verhalten der Nutzer zu identifizieren und Abweichungen davon zu erkennen. Hierzu sind 4 Schritte nötig:

1. *Identifizierung:*
   Zunächst werden die unterschiedlichen Benutzergruppen im Unternehmen identifiziert und die entsprechenden Personas erstellt. Anhand der Personas werden dann die Benutzer für die weitergehende Analyse ausgewählt.
2. *Beobachtung und Analyse:*
   Das Verhalten wird vom DLP-System erfasst und in Log-Dateien gespeichert. Diese Logdaten werden auf Assoziationsregeln untersucht.
3. *Interview:*
   Die Informationsflüsse im Unternehmen werden erfasst, indem man die in der Identifizierungsphase ausgewählten Benutzer interviewt. Hierzu wird die FLOW-Methode verwendet.
4. *Extraktion der Regeln:*
   Die Ergebnisse der bisherigen Analysen werden kombiniert und weiter analysiert, um Regeln zu finden. Daraus lassen sich neue Regeln für das DLP-System erstellen oder bereits bestehende Regeln verfeinern.

Diese Schritte lassen sich auch wiederholt ausführen, sodass ein Kreislauf gebildet wird. So lässt sich das Regelwerk des DLP-Systems nach und nach immer weiter verbessern.

Der gesamte Ablauf des Lösungsansatzes wird in Abbildung 5.1 visualisiert. Der Nutzer greift im Rahmen seiner Tätigkeit lesend oder schreibend auf ein Dokument zu. Diese Zugriffe werden durch das DLP-System überwacht und im Falle eines Verstoßes gegen die konfigurierten Richtlinien in der Log-Datei festgehalten. Daraufhin würde ein weiteres Programm auf diese Logdateien zugreifen und die enthaltenen Daten anhand der herausgefundenen Verhaltensregeln und -muster filtern. Datensätze, die diesen Mustern entsprechen, könnten gelöscht werden und abweichende Daten je nach Ausmaß der Abweichung mit einem höheren oder niedrigeren Schweregrad versehen werden. Dem Projektleiter oder DLP-Administrator wird daraufhin die gefilterte Log-Datei angezeigt, so dass weniger Ereignisse untersucht werden müssen, die zudem auch bereits stärker differenziert sind. Diese Person kann daraufhin mit dem Nutzer Rücksprache halten, so dass fehlende Verhaltensregeln und -muster entdeckt und dem Regelwerk hinzugefügt werden können. Aufgrund der leicht veränderbaren und

verständlichen Repräsentation des Regelwerkes bietet sich das Drools-Framework für die praktische Umsetzung an.

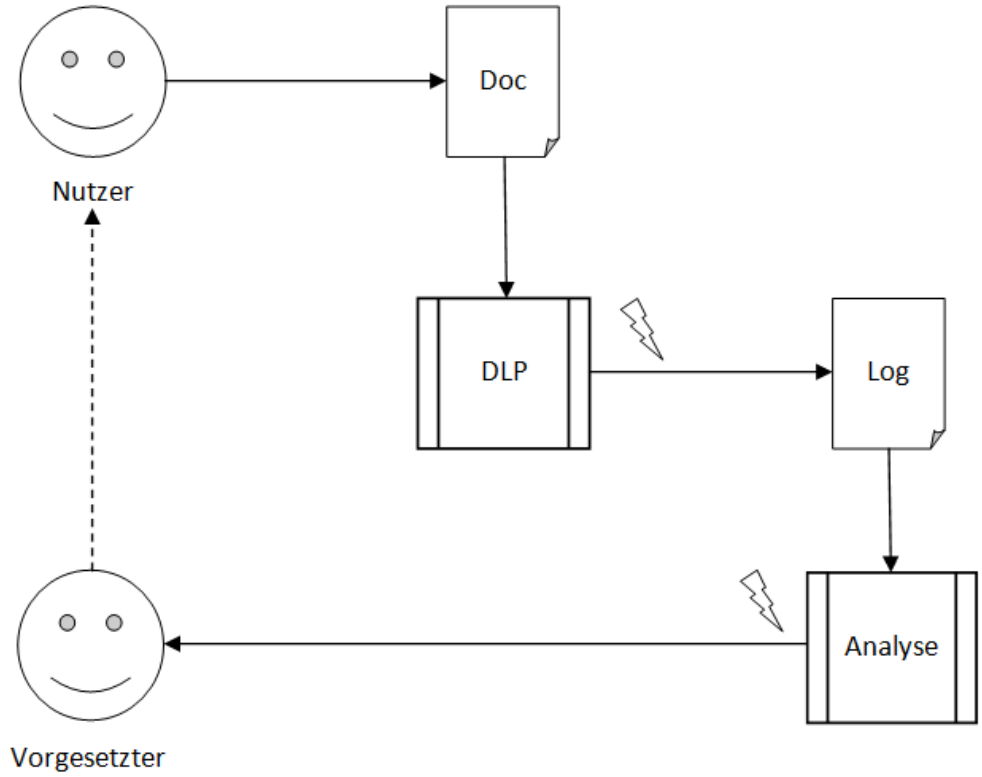

Abbildung 5.1.: Ablauf des möglichen Lösungsansatzes

Ein DLP-System, welches diesen Ablauf verfolgt, hat einen Aufbau wie er in Abbildung 5.2 dargestellt wird. Die Umgebung, in der das System eingesetzt wird, wird überwacht und die Einhaltung der im System konfigurierten Richtlinien überprüft. Vorgänge, die gegen diese Richtlinien verstoßen, werden im Datenbestand gespeichert. Daraufhin werden diese Daten auf Muster hin untersucht. Wird ein Vorgang dabei als verdächtig erkannt, wird er geblockt und an den Vorgesetzten gemeldet. Dieser kann dann die Aktion sperren oder freigeben. Um das System fortlaufend zu verbessern und auch sich ändernde Arbeitsabläufe zu berücksichtigen, kann anhand der Ergebnisse aus der Analyse die Überwachung verfeinert werden, etwa indem der Grenzwert an Treffern aus der Stichwortliste in den Richtlinien angepasst wird.

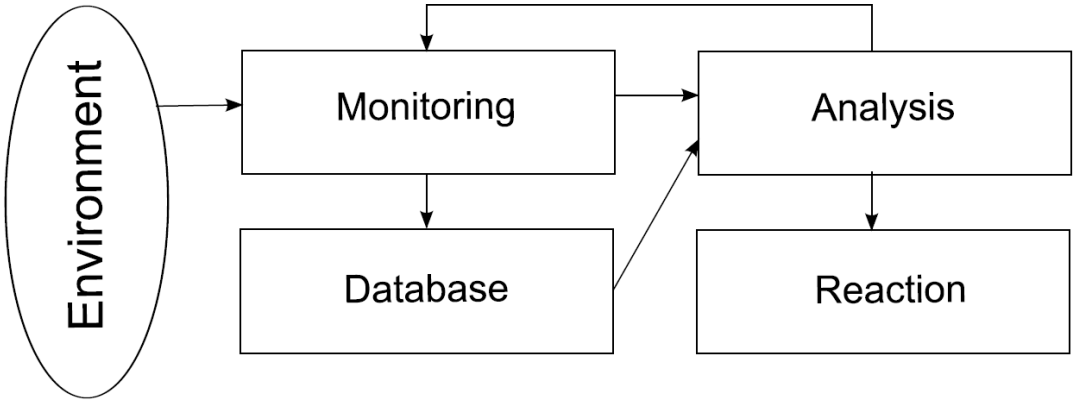

Abbildung 5.2.: Aufbau eines DLP-Systems mit der vorgeschlagenen Arbeitsweise

5.2. Durchführung des Ansatzes bei der Continental AG

Im Fokus dieses Abschnitts steht die exemplarische Durchführung des Ansatzes bei der Continental AG, insbesondere die Erfassung der Informationsflüsse und die Analyse der Daten des vorhandenen DLP-Systems. Hierzu wurden zunächst das bestehende DLP-System und die vorhandenen Daten betrachtet. Daraufhin wurden verschiedene Verfahren auf ihre Eignung zur Analyse dieser Daten geprüft.

Ein einschränkender Faktor der Analyse war dabei die vergleichsweise kleine Datenmenge infolge des stark eingegrenzten Probebetriebs. Ein weiteres Problem ist, dass keine gelabelten Trainingsdaten zur Verfügung stehen. Es ist weder bekannt, welche der Datensätze Fehlmeldungen darstellen, noch ob alle relevanten Ereignisse vom DLP-System erfasst wurden. Aufgrund dieser Rahmenbedingungen sind beispielsweise Verfahren des maschinellen Lernens nicht einsetzbar.

Aus diesen Gründen wurde ein anderer Ansatz gewählt, mit dem das normale Verhalten und die üblichen betrieblichen Prozesse abgebildet werden sollen. Abweichungen von diesem Verhalten stellen dann einen Verstoß dar. Dieser Ansatz setzt ein homogenes Arbeitsumfeld vorraus, so dass für unterschiedliche Unternehmensbereiche und Benutzergruppen individuelle Richtlinien entwickelt werden müssen. Um das Normalverhalten sowie die Prozesse zu erfassen, fiel die Wahl auf den FLOW-Ansatz, der durch Association Rule Mining ergänzt wird. Zur Vorbereitung und Anpassung dieser Verfahren ist auch eine vorhergehende Analyse der Daten nötig.

5.2.1. Voranalyse

Die Voranalyse diente einer ersten Einschätzung und bildete die Grundlage für die Anpassung der weiteren Maßnahmen. Dies ist eine vorbereitende Tätigkeit, die nur

einmalig durchgeführt wird und daher nicht Bestandteil des in Abbildung 5.2 dargestellten Aufbaus ist.

Um die Daten, die vom DLP-System erfasst wurden, zu analysieren, wurden diese zunächst in der Weboberfläche des DLP-Systems selbst betrachtet und gefiltert und später zur weitergehenden statistischen Analyse exportiert. Durchschnittlich traten in den Richtlinien “Non Material Development Users: material related terms” und “Material Development Users: material related terms” im Durchschnitt 6,7 Ereignisse pro Tag auf. Da die Datenmenge aufgrund der kleinen Nutzergruppe und des vergleichsweise kurzen Untersuchungszeitraumes relativ klein und nicht repräsentativ ist, müssen die Schlussfolgerungen daraus mit Vorsicht betrachtet werden. Dennoch bieten sie erste Anhaltspunkte für eine Verfeinerung der Richtlinien.

Die Log-Daten zeigten ein sehr heterogenes Verhalten der Nutzer. Dies wurde später auch in den Interviews (5.2.4) bestätigt.
So ist die Anzahl an Ereignissen je nach Nutzer sehr unterschiedlich. Der User mit den meisten Ereignissen hat etwa einen Anteil von fast 20%, während der Nutzer mit dem geringsten Anteil bei deutlich unter 1% liegt. Allgemein teilen sich die Nutzer in eine Gruppe mit sehr wenigen Ereignissen und eine mit mehr Ereignissen, sowie die Mitarbeiter mit leitenden Tätigkeiten in den Entwicklungsabteilungen, die besonders viele Ereignisse verursachen. Bei ihnen ist auch der Anteil an lesenden Zugriffen geringfügig höher.
Des weiteren gibt es in der Regel nur sehr wenige Wiederholungen von Dateien in den einzelnen Ereignissen. Bevor auch lesende Zugriffe erkannt und geloggt wurden, trat jede Datei nur ein bis zwei mal in den Log-Daten auf. Es waren jedoch bereits aus der Benennung der Dateien einzelne Dokumentklassen ersichtlich, etwa Präsentationen oder Stellenbeschreibungen. Diese Information floss auch in die Erstellung des Fragebogens (A.2) mit ein. Bei den lesenden Zugriffen auf Dateien kommt es hingegen häufiger zu Wiederholungen. Hier lässt sich erkennen, dass meist pdf-Dateien betroffen sind, häufig etwa Berichte. Auch Dateien mit organisatorischem Inhalt, wie etwa eine Excel-Datei zur Raumreservierung, tauchen gehäuft auf. Dies entspricht durchaus den Erwartungen. Auch hier spielt die geringe Datenmenge eine Rolle, da Wiederholungen und Muster über einen längeren Beobachtungszeitraum und bei mehr involvierten Personen sehr viel wahrscheinlicher sind.

Die Anzahl an Treffern aus der Stichwortliste lag bei den einzelnen Ereignissen in diesen beiden Richtlinien zwischen 3 und 333, mit einem Median von 14 bei einer Standardabweichung von 58.

Lesende Zugriffe wurden ursprünglich von den konfigurierten Richtlinien nicht erfasst. Da diese aber mutmaßlich einen großen Anteil an den gesamten Aktivitäten haben und so auch ein wichtiger Faktor für Informationsflüsse sind, wurde die Richtlinie im Rahmen der Arbeit angepasst. Die daraufhin gesammelten Daten bestätigten diese Annahme, die lesenden Zugriffe machen über die Hälfte aller geloggten Ereignisse aus. Es folgen Schreibzugriffe auf Netzlaufwerke und E-Mails. Die weiteren Ereignisarten Drucker/Fax, Wechselmedien und HTTPS/SSL machen einen deutlich kleineren Anteil aus. Die genauen Prozentwerte finden sich in Abbildung 5.3. Diese Daten beziehen sich auf die gesamte Datenmenge, allerdings gab es auch bei einzelnen

Nutzern keine auffälligen Abweichungen von dieser Verteilung. Sollten also bei einem Nutzer anteilsmäßig auffällig viele Ereignisse aus den Bereichen Drucker/Fax, Wechselmedien oder HTTPS/SSL auftauchen, könnte dies als Hinweis auf unerwünschtes Verhalten dienen.

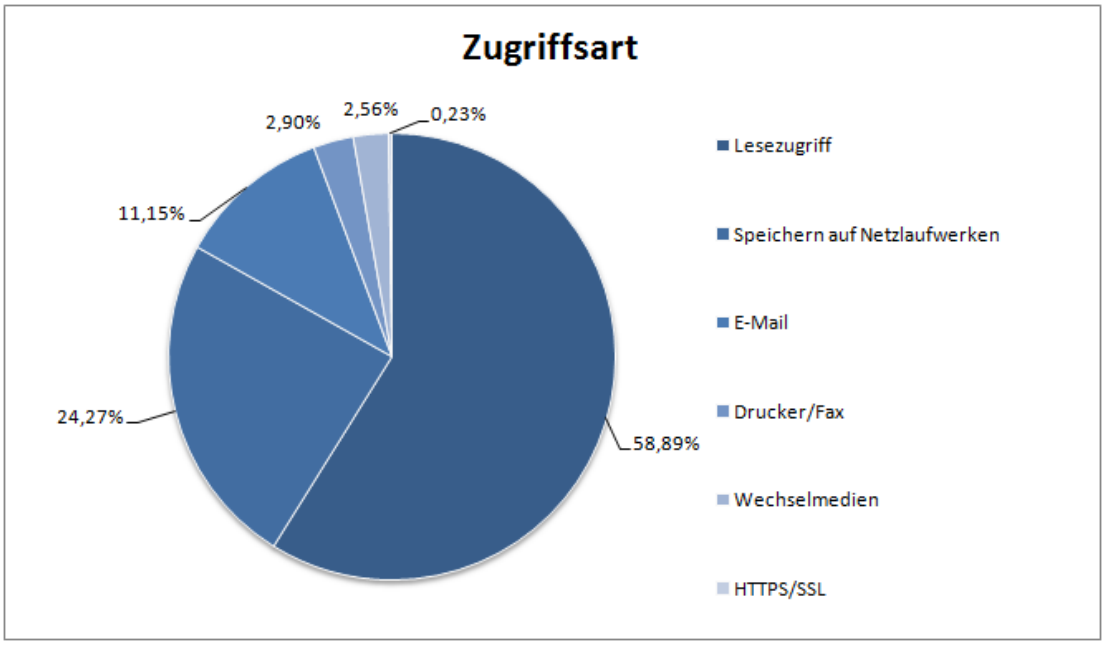

Abbildung 5.3.: Anteile der unterschiedlichen Zugriffsarten

Ein Großteil der geloggten Ereignisse bezog sich auf Aktionen, die mit pdf-Dokumenten gearbeitet haben. Auch Excel- und Powerpoint-Formate traten häufig auf. Word-Dokumente traten wider Erwarten vergleichsweise selten auf, was neben unerwarteten Arbeitsabläufen auch durch die Verwendung eines SharePoint-Systems erklärt werden kann. Andere Dokumenttypen traten nur sehr selten auf und wurden daher in Abbildung 5.4 als “Sonstige” zusammengefasst. Zu bemerken ist hierbei, dass selbst Zip-Archive mit Daten, die Stichwörter enthalten, vom System erkannt werden, auch wenn hier die Erkennungsquote niedriger ist als bei anderen Dateien.

5.2.2. Identifizierung

Im Vorfeld der Befragung wurden die Daten des DLP-Systems, sowie die Berufsbezeichnungen der bisherigen Nutzer des Systems analysiert. Aus diesen Daten wurden dann unterschiedliche Personas erstellt.

- *Leitende Angestellte:*
  Leitende Angestellte haben Zugriff auf Daten aus sehr vielen unterschiedlichen Projekten. Besonders häufig finden lesende Zugriffe statt. Ein Großteil des Arbeitspensums besteht aus Verwaltungsaufgaben und Meetings. Außerdem fin-

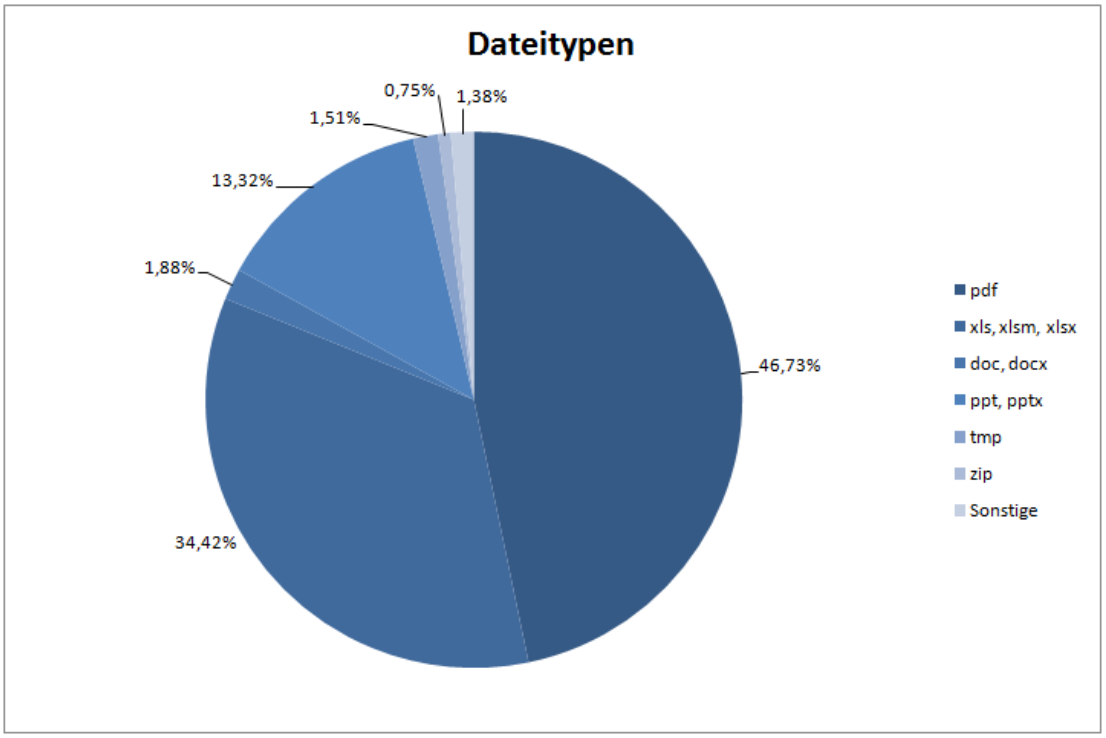

Abbildung 5.4.: Anteile der unterschiedlichen Dateiarten

den häufig Dienstreisen statt und es wird von unterwegs oder anderen Unternehmensstandorten aus gearbeitet.

- *Unterstützende Angestellte:*
  Unterstützende Angestellte stellen häufig eine Schnittstelle zwischen Externen, Fachangestellten und leitenden Angestellten dar. Sie haben daher Zugriff auf viele unterschiedliche Daten, arbeiten jedoch nicht sehr häufig mit diesen Daten. Es gibt wenige Zusammenhänge zwischen den Daten.
- *Fachangestellte:*
  Fachangestellte arbeiten mit vertraulichen Daten aus wenigen verschiedenen Projekten parallel. Hierbei werden oft auch neue Dateien erstellt und es erfolgen sowohl schreibende als auch lesende Zugriffe. Die Daten werden in der Regel nur mit Mitgliedern des jeweiligen Projektes geteilt, worunter jedoch auch Unternehmensexterne sein können.

Diese Personas dienten der Kategorisierung der Nutzer und bildeten die Grundlage für die Auswahl der zu befragenden Personen. Ausgewählt wurden daraufhin 4 Personen aus den unterschiedlichen Personas, die unterschiedliche Arbeitsabläufe und Unternehmensbereiche abdecken. Zur Vorbereitung der Interviews wurden dann ein weiteres Mal die aufgezeichneten Daten des DLP-Systems betrachtet, um so im Interview ggf. auf Auffälligkeiten einzugehen.

5.2.3. Beobachtung und Analyse

Im Rahmen der Beobachtung und Analyse wurden die Ereignislogs des DLP-Systems mit Hilfe der Software RapidMiner analysiert. RapidMiner ging aus der seit 2001 an der Technischen Universität Dortmund entwickelten Software YALE hervor und bietet umfassende Lösungen zur automatischen Analyse von Daten mittels Verfahren des Data Minings und des maschinellen Lernens [RM]. Große Vorteile dieser Software sind die weite Verbreitung, der große Umfang an Verfahren und die Verfügbarkeit als Open-Source-Software.

5.2.3.1. Vorbereitung der Daten

Um die Daten des DLP-Systems mit Hilfe von RapidMiner weitergehend zu analysieren, mussten diese zunächst angepasst werden. Als Grundlage diente hierbei die vom DLP-System bereitgestellte XML-Datei, da diese im Gegensatz zur CSV-Datei sämtliche benötigten Informationen enthält. Diese XML-Datei musste zunächst in eine CSV-Datei umgewandelt werden, um die Verarbeitung durch RapidMiner zu ermöglichen. Zusätzlich mussten die enthaltenen Daten zusammengefasst und die Darstellung der Daten geändert werden, da für das Association Rule Mining binominale Daten benötigt werden. Besonderes Augenmerk lag in der Analyse auf den vorgekommenen Stichwörtern, dem Dateiformat und der Art des Ereignisses. Um die vom DLP-System bereitgestellten Exportdaten für den Analyseprozess passend zu transformieren, wurde ein Java-Programm geschrieben.

5.2.3.2. Analyseprozess

Im Analyseprozess wird zunächst die vom Java-Programm erstellte CSV-Datei importiert und die enthaltenen Werte in binomiale Daten umgewandelt. Dies ist erforderlich, da die enthaltenen 0 und 1 beim Import als numerische Werte interpretiert werden. Aus diesen Daten werden daraufin *Frequent Item Sets* [1] generiert, die häufig vorkommende Kombinationen in den Daten aufzeigen. Diese *Frequent Item Sets* bilden dann die Grundlage für die Erstellung der Assoziationsregeln. Den gesamten Prozess zeigt Abbildung 5.5 anhand eines Screenshots.

5.2.3.3. Ergebnisse

Die aus dem Analyseprozess resultierenden Ergebnisse lieferten entgegen den Erwartungen kaum neue Erkenntnisse und daher auch nur wenige für eine Verbesserung anwendbare Assoziationsregeln. Die Assoziationsregeln bildeten hauptsächlich bereits bekannte und schlüssige Abhängigkeiten in den Daten ab. Dies hängt vermutlich auch mit der geringen Datenmenge zusammen, so dass sowohl Support als auch Confidence relativ gering gewählt werden mussten.

[1] Häufig in den Daten vorkommende Wertekombinationen

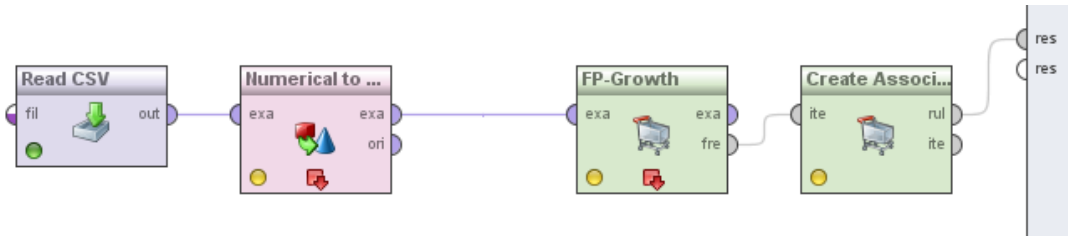

Abbildung 5.5.: Analyseprozess aus RapidMiner

5.2.4. FLOW-Analyse

Um die Informationsflüsse im Unternehmen systematisch zu erfassen, wurde die FLOW-Methode eingesetzt. Das Hauptaugenmerk lag in diesem Fall auf festen und verfestigenden Informationsflüssen, da nur diese für das DLP-System zu erfassen sind. Zur Informationserhebung wurde der *bottom-up*-Ansatz gewählt, da dieser sich sehr gut für die Erhebung von Ist-Zuständen eignet und zu Beginn keinen Gesamtüberblick erfordert. Als Verfahren kamen sowohl Interviews als auch die Ableitung aus Kommunikatonsereignissen zum Einsatz. Für die Ableitung aus Kommunikationsereignissen dienten die Log-Daten des DLP-Systems als Grundlage. Diese wurden sowohl mittels deskriptiver Statistik, als auch mittels Associaton Rule Mining genauer analysiert, wie in Abschnitt 5.2.1 und 5.2.3 beschrieben.

5.2.4.1. Fragebögen

Die Erfassung der benötigten Informationen für die FLOW-Informationsflüsse erfolgt anhand eines Fragebogen, sowie des dazu gehörenden Erhebungsbogen, in dem die Informationen bereits systematisch und übersichtlich erfasst werden können. Grundlage bildeten die von Kai Stapel und Prof. Kurt Schneider in [SS12] entwickelten FLOW-Frage- und Erhebungsbögen, die jedoch im Rahmen dieser Arbeit leicht angepasst wurden, um besser in den Kontext von Data Loss Prevention zu passen. Hierzu wurden sie um weitere Fragen und Faktoren ergänzt. Insbesondere wurden im Erhebungsbogen zusätzlich die Frequenz und Vertraulichkeit der einzelnen Aktionen bzw. Dokumente abgefragt, um so weitere wichtige Erkenntnisse für das DLP-System zu erlangen. Die Frequenz einzelner Aktionen diente vor allem zum Vergleich der in Folge der Interviews erwarteten Häufigkeit der jeweiligen Dateien und Aktionen mit den tatsächlich vom DLP-System erfassten Daten.

Der Fragebogen wurde um neue Fragen ergänzt, die auf Grundlage der vorhergegangenen Erkenntnisse aus der Analyse der Log-Daten gestaltet wurden.

Die angepassten Erhebungs- und Fragebögen finden sich in Anhang A.1 beziehungsweise A.2

5.2.4.2. Ergebnisse

Zusammenfassend lässt sich sagen, dass die Befragungen insgesamt ein sehr heterogenes Verhalten zeigten. Dennoch wurden auch einzelne Parallelen im Arbeitsablauf erkannt.

Beispielsweise wurden E-Mails von allen Befragten als Hauptbestandteil ihrer täglichen Arbeit eingeschätzt. Viele Abläufe in der Organisation und Kommunikation weisen jedoch von Mal zu Mal große Unterschiede auf. Hier muss hinterfragt werden, ob einheitliche Regelungen für häufig wiederkehrende Abläufe getroffen werden sollten. Dies könnte auch positive Effekte auf das DLP-System haben.

Alle Befragten waren sich bewusst, dass sie regelmäßig mit vertraulichen Daten umgehen und diese vor unbefugtem Zugriff geschützt werden sollten. Auch die Einschätzung, welche Daten vertraulich sind, stimmten bei den einzelnen Befragten stark überein. Die tatsächliche Verhaltensweise im Umgang mit diesen Daten zeigte jedoch große Unterschiede auf. Hierbei sollte nicht unerwähnt bleiben, das einige dieser Verhaltensweisen, etwa das unverschlüsselte Senden von Daten via E-Mail, zu Sicherheitslücken führen. Die Einführung und deutliche Kommunikation einheitlicher Richtlinien kann hier auch bereits ohne den Einsatz des DLP-Systems zu einem Sicherheitsgewinn führen. Die Einhaltung der Richtlinien ließe sich jedoch auch durch das DLP-System überprüfen und wenn gewünscht teilweise sogar erzwingen.

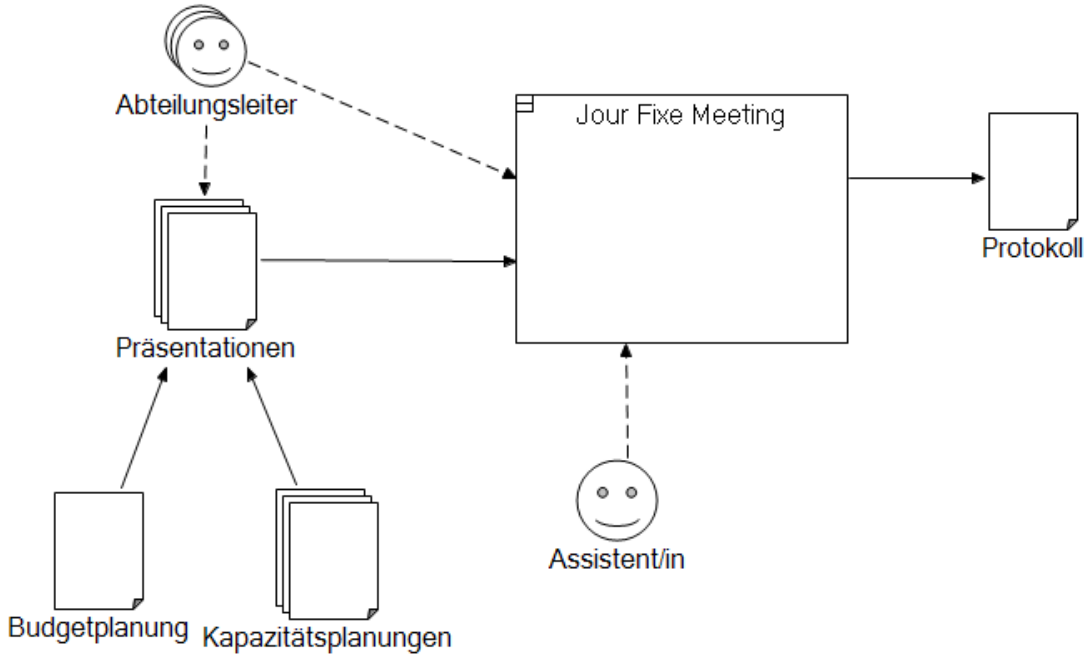

Abbildung 5.6.: Flow-Diagramm Jour Fixe Meeting

Häufig sind die befragten Personen auch an unterschiedlichen Meetings beteiligt, eines davon wird als FLOW-Diagramm in Abbildung 5.6 dargestellt. Auch hier gibt es deutliche Parallelen und Wiederholungen im Ablauf, wobei jedoch die verwendeten Informationswege von Mal zu Mal stark variieren. So wird das Protokoll solcher Meetings beispielsweise entweder direkt per Mail verschickt, auf einem SharePoint-System oder ContiView hochgeladen, auf einem Netzlaufwerk gespeichert oder als Download-Link verschickt.

6. Fazit und Ausblick

Zur Zeit bringt der Betrieb eines DLP-Systems einige Probleme mit sich. Neben der rechtlichen Problematik, ist auch der administrative Aufwand für die Implementation und den erfolgreichen Betrieb eines derartigen Systems sehr hoch. Dennoch gewinnt dieses Thema an Bedeutung, da der Innovationsdruck gerade bei Industrieunternehmen stetig zunimmt und das geistige Eigentum so an Bedeutung gewinnt. Daher ist weiterer Forschungsaufwand notwendig um die bestehenden Systeme zu verbessern und einen guten Trade-Off zwischen Nutzerfreundlichkeit für die Mitarbeiter und der Sicherheit der Daten zu erlangen.

Des Weiteren stellen neue Techniken und Entwicklungen neue Herausforderungen an die bestehenden DLP-Systeme. Der Einsatz starker Verschlüsselungsverfahren etwa nimmt zu. Zudem wird vermehrt von mobilen Geräten gearbeitet und auch *Bring Your Own Device* (BYOD) und *Software as a Service* (SaaS) gewinnen an Bedeutung, was neue Probleme für den Schutz vertraulicher Daten mit sich bringt. Die bestehenden Techniken reichen dafür nicht aus und müssen an die aktuellen Entwicklungen angepasst werden.

Um diesen Problemen und Herausforderungen künftig Rechnung tragen zu können, wurde im Rahmen dieser Arbeit eine Methodik entwickelt, mit der sich die im Unternehmen bestehende Situation bezüglich der Informationsflüsse und Verhaltensmuster umfassend erfassen und analysieren lässt. Auf Grundlage dieser Erkenntnisse können dann gegebenenfalls organisatorische Maßnahmen ergriffen werden. Zudem ist es möglich diese in die Konfiguration des DLP-Systems einfliessen zu lassen.

Für die Verbesserung der bestehenden Systeme können jedoch auch Ansätze aus anderen Bereichen verfolgt werden, wie auch in dem Abschnitt “Verwandte Arbeiten und Themengebiete” (Kapitel 2) bereits festgestellt wurde. Auch die eingehende Prüfung und Modellierung der Informationsflüsse kann den Betrieb des DLP-Systems unterstützen und vor Datenverlust schützen. So können nicht nur gegebenenfalls die Richtlinien des DLP-Systems angepasst werden, sondern auch in Hinblick auf die Datensicherheit problematische Informationsflüsse erkannt und verändert werden.

Da jedoch ein vollständiger Schutz nicht erreicht werden kann, ist es unabdingbar, die Mitarbeiter für den korrekten Umgang mit vertraulichen Daten zu sensibilisieren und zu schulen. Richtlinien anhand derer sensible Daten erkannt werden können und die die korrekten Verhaltensweisen aufzeigen, sollten erstellt und umfassend kommuniziert werden. So kann bereits vor Datenverlust geschützt werden und auch der Betrieb des DLP-Systems wird erleichtert, da bei korrektem Umgang mit den Daten weniger Ereignisse gemeldet werden. Zudem ist es auch nur so möglich, Nutzer bei groben Verstößen gegebenenfalls abzumahnen oder andere Maßnahmen zu ergreifen.

A. Anhang

A.1. Erhebungsbogen

Siehe die folgende Seite. Als Grundlage diente der Erhebungsbogen in Flow-Methode - Methodenbeschreibung zur Anwendung von FLOW [SS12]

Erhebungsbogen

ID	
Kürzel	
Datum	

Steuerung

Funktion/Person	Inhalt	Fest/flüssig	Form/Medium	Vertraulich	Frequenz

Von

Funktion/Person	Inhalt	Fest/flüssig	Form/Medium	Vertraulich	Frequenz

Aufgabenbezeichnung:

Person(en):

Frequenz:

Besonderes:

An

Funktion/Person	Inhalt	Fest/flüssig	Form/Medium	Vertraulich	Frequenz

Unterstützung

Funktion/Person	Inhalt	Fest/flüssig	Form/Medium	Vertraulich	Frequenz

A.2. Fragebogen

Siehe die folgenden 4 Seiten. Als Grundlage diente der Fragebogen in Flow-Methode - Methodenbeschreibung zur Anwendung von FLOW [SS12]

Fragebogen

Einleitung

Vielen Dank, dass Sie sich die Zeit nehmen und mit mir diesen Fragebogen bearbeiten.

Mein Name ist Svenja Schulz und ich untersuche im Rahmen meiner Bachelorarbeit (Studiengang Informatik, Leibniz Universität Hannover) die Informationsflüsse bei Continental. Ziel ist es, übliche Prozesse im Rahmen ihrer Tätigkeit zu erfassen, um diese aus den Meldungen des Data Loss Prevention Systems zu entfernen und dieses System so zu verbessern.

Die bei dieser Befragung erhobenen Daten werden ausschließlich anonymisiert zur Auswertung im Rahmen der Bachelorarbeit verwendet. Die Teilnahme ist selbstverständlich freiwillig.

Sollten Sie jetzt oder später noch Fragen haben stehe ich Ihnen gerne zur Verfügung. Sie erreichen mich per E-Mail unter svenja.schulz@conti.de.

Fragen für den FLOW-Bogen

Was sind ihre Hauptaufgaben?

Für jede Hauptaufgabe:

Wie oft führen Sie diese Aufgabe durch?

Mit wem arbeiten sie dabei zusammen?

Woher erhalten Sie die dafür benötigten Informationen?

In welcher Form erhalten Sie die Informationen, beispielsweise in schriftlicher oder mündlicher Form?

Welche Software/Programme/Werkzeuge nutzen Sie um diese Aufgabe durchzuführen?

(Welche Vorgaben oder Richtlinien gibt es, an die sie sich bei dieser Aufgabe halten?)

Wer unterstützt Sie bei diesen Aufgaben??

An wen können/sollen Sie diese Aufgabe oder Teile davon delegieren?

Was für Daten/Dokumente entstehen dabei?

Würden Sie diese als vertraulich einschätzen?

Wer greift auf die entstehenden Daten oder sonstige Ergebnisse zu? (Mitarbeiter aus der eigenen Abteilung, aus anderen Abteilungen, von kooperierenden UN, externe Personen, Behörden)

Fragen zu Vertraulichkeit von Daten allgemein

Gibt es bestimmte Zeiträume oder Situationen in denen Sie besonders häufig mit vertraulichen Daten umgehen, Bspw. zum Jahres- oder Quartalsende, vor besonderen Ereignissen?

Wie häufig arbeiten Sie mit den nachfolgend genannten Dokumenttypen? Würden Sie diese als vertraulich einschätzen?

Präsentationen

Geschätzte Häufigkeit:
- ☐ ___ mal pro Tag / Woche / Monat
- ☐ Selten
- ☐ Nie
- ☐ Weiß nicht

Vertraulich?
- ☐ Sehr vertraulich
- ☐ Weniger vertraulich
- ☐ Nicht vertraulich
- ☐ Weiß nicht

Planungs- oder Zielvereinbarungen (xls)

Geschätzte Häufigkeit:
- ☐ ___ mal pro Tag / Woche / Monat
- ☐ Selten
- ☐ Nie
- ☐ Weiß nicht

Vertraulich?
- ☐ Sehr vertraulich
- ☐ Weniger vertraulich
- ☐ Nicht vertraulich
- ☐ Weiß nicht

Pflichtenheft (xls)

Geschätzte Häufigkeit:
- ☐ ___ mal pro Tag / Woche / Monat
- ☐ Selten
- ☐ Nie
- ☐ Weiß nicht

Vertraulich?
- ☐ Sehr vertraulich
- ☐ Weniger vertraulich
- ☐ Nicht vertraulich
- ☐ Weiß nicht

Flyer (pdf)

Geschätzte Häufigkeit:
- ☐ ___ mal pro Tag / Woche / Monat
- ☐ Selten
- ☐ Nie
- ☐ Weiß nicht

Vertraulich?	☐ Sehr vertraulich ☐ Weniger vertraulich ☐ Nicht vertraulich ☐ Weiß nicht
Whitepaper (pdf)	
Geschätzte Häufigkeit:	☐ ___ mal pro Tag / Woche / Monat ☐ Selten ☐ Nie ☐ Weiß nicht
Vertraulich?	☐ Sehr vertraulich ☐ Weniger vertraulich ☐ Nicht vertraulich ☐ Weiß nicht
Vorträge (ppt)	
Geschätzte Häufigkeit:	☐ ___ mal pro Tag / Woche / Monat ☐ Selten ☐ Nie ☐ Weiß nicht
Vertraulich?	☐ Sehr vertraulich ☐ Weniger vertraulich ☐ Nicht vertraulich ☐ Weiß nicht
Jahresberichte (doc)	
Geschätzte Häufigkeit:	☐ ___ mal pro Tag / Woche / Monat ☐ Selten ☐ Nie ☐ Weiß nicht
Vertraulich?	☐ Sehr vertraulich ☐ Weniger vertraulich ☐ Nicht vertraulich ☐ Weiß nicht
Stellenbeschreibungen (doc)	
Geschätzte Häufigkeit:	☐ ___ mal pro Tag / Woche / Monat ☐ Selten ☐ Nie ☐ Weiß nicht
Vertraulich?	☐ Sehr vertraulich ☐ Weniger vertraulich ☐ Nicht vertraulich ☐ Weiß nicht

Messdaten (xls)

Geschätzte Häufigkeit:
- ☐ ___ mal pro Tag / Woche / Monat
- ☐ Selten
- ☐ Nie
- ☐ Weiß nicht

Vertraulich?
- ☐ Sehr vertraulich
- ☐ Weniger vertraulich
- ☐ Nicht vertraulich
- ☐ Weiß nicht

Inwiefern können Sie einschätzen, welche Daten vertraulich sind?

Wodurch?

Fragen zur Person

Was ist ihre Tätigkeitsbezeichnung?

Wie lange sind Sie für Ihren gegenwärtigen Aufgabenbereich zuständig?

In welcher Abteilung arbeiten Sie?

Literaturverzeichnis

[AIS93] Rakesh Agrawal, Tomasz Imielinski, Arun Swami,
Mining association rules between sets of items in large databases,
ACM SIGMOD Record: pp. 207-216, 1993

[Al11] Sabah Al-Fedaghi,
A Conceptual Foundation for Data Loss Prevention
International Journal of Digital Content Technology and its Applications (5): pp. 293-303

[AV02] Rakesh Agrawal, Tomasz Imielinski, Arun Swami,
Investigative profiling with computer forensic log data and association rules,
2002 IEEE International Conference on Data Mining, Proceedings: pp. 11-18, 2002

[Bau13] Simon Bauer,
Sicherheitsfragen in Zusammenhang mit Data Loss Prevention und Vorschläge zur Erweiterung des Österreichischen Informationssicherheitshandbuchs,
Johannes Kepler Universität Linz, Institut für Informationsverarbeitung und Mikroprozessortechnik, 2013

[BDSG] *Bundesdatenschutzgesetz in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 1 des Gesetzes vom 14. August 2009 (BGBl. I S. 2814) geändert worden ist*

[BetrVG] *Betriebsverfassungsgesetz in der Fassung der Bekanntmachung vom 25. September 2001 (BGBl. I S. 2518), das zuletzt durch Artikel 3 Absatz 4 des Gesetzes vom 20. April 2013 (BGBl. I S. 868) geändert worden ist*

[BN89] David F.C. Brewer, Michael J. Nash,
The Chinese Wall Security Policy,
Security and Privacy: pp. 206-214, 1989

[Boll96] Toni Bollinger,
Assoziationsregeln – Analyse eines Data Mining Verfahrens,
Springer, Informatik-Spektrum 19: pp. 257-261, 1996

[BSI12] Bundesamt für Sicherheit in der Informationstechnik,
IT-Grundschutz-Katalog,
2012

[Ca11] Tracey Caldwell,
Data Loss Prevention - Not Yet a Cure,

Computer Fraud and Security: pp. 5-9, 2011

[FY07] Simon Fong, Zhuang Yan,
A Security Model for Detecting Suspicious Patterns in Physical Environment,
Third International Symposium on Information Assurance and Security: pp. 221-226, 2007

[Ge08] Xiaoxuan Ge,
FLOW Patterns: Beschreibung und Diskussion von Informationsflussmustern in der Softwareentwicklung,
Leibniz Universität Hannover, Fachgebiet Software Engineering, 2008

[Gri11] Jonathan Grier,
Detecting data theft using stochastic forensics,
Digital Investigation 8: pp. 71-77, 2011

[He14] Liang He,
An Improved Intrusion Detection based on Neural Network and Fuzzy Algorithm,
Journal of Networks Vol. 9: pp. 1274-1280, 2014

[HFW11] Yi Hu, Charles Frank, James Walden, Emily Crawford, Dhanuja Kasturiratna,
Profiling file repository access patterns for identifying data exfiltration activities,
IEEE Symposium on Computational Intelligence in Cyber Security: pp. 122-128, 2011

[JZJ12] Ma Jun, Wang Zhiying, Ren Jiangchun, Wu Jiangjiang, Cheng Yong and Mei Songzhu,
The Application of Chinese Wall Policy in Data Leakage Prevention,
International Conference on Communication Systems and Network Technologies, 2012

[Kas08] Hannes Kasparick,
Data Leakage Prevention,
Fachhochschule Oberöstereich, Fakultät für Informatik, Kommunikation und Medien, 2008

[LK10] Simon Liu, Rick Kuhn,
Data Loss Prevention,
IT professional, 12(2): pp. 10-13, 2010

[LLP12] Seokhee Lee, Keungi Lee, Jong Hyuk Park, Sangjin Lee,
An on-site digital investigation methodology for data leak case,
Security and Communication Networks 2012

[MP08] Wes Masri, Andy Podgurski,
Application-based anomaly intrusion detection with dynamic information flow analysis,

Computers & Security (27): pp. 176-187, 2008

[RM] RapidMiner,
RapidMiner Press Kit, http://rapidminer.com/about-us/press-kit/
Zuletzt aufgerufen am 26.09.2014

[Sch06] Kurt Schneider,
Aggregatzustände von Anforderungen erkennen und nutzen,
GI Softwaretechnik-Trends, Band 26, Heft 1: pp. 22-23, 2006

[SER12] Asaf Shabtai, Yuval Elovici, Lior Rokach,
A survey of data leakage detection and prevention solutions,
Springer, 2012

[SK13] George J. Silowash, Christopher King,
Insider threat control: Understanding data loss prevention (DLP) and detection by correlating events from multiple sources,
Technical report, Software Engineering Institute (SEI), 2013

[SS12] Kai Stapel, Kurt Schneider,
FLOW-Methode - Methodenbeschreibung zur Anwendung von FLOW,
Leibniz Universität Hannover, Fachgebiet Software Engineering, 2012

[St12] Kai Stapel,
Informationsflusstheorie der Softwareentwicklung,
Dissertation, Leibniz Universität Hannover, Fachgebiet Software Engineering, Verlag Dr. Hut, 2012

[StS12] Kai Stapel, Kurt Schneider,
Managing Knowledge on Communication and Information Flow in Global Software Projects,
Expert Systems - Special Issue on Knowledge Engineering in Global Software Development, 2012

[Sym] Symantec,
Data Loss Prevention,
http://www.symantec.com/products-solutions/families/?fid=data-loss-prevention/
Zuletzt aufgerufen am 26.09.2014

[Tor12] Tore Torsteinbø *Data Loss Prevention Systems and Their Weaknesses*,
University of Agder, Faculty of Engineering and Science, Department of Information Technology, 2012

[TRM09] Arman Tajbakhsh, Mohammad Rahmati, Abdolreza Mirzaei
Intrusion detection using fuzzy association rules,
Applied Soft Computing (9): pp. 462-269, 2009

[WB10] Shelly Xiaonan Wu, Wolfgang Banzhaf,
The use of computational intelligence in intrusion detection systems: A review,
Applied Soft Computing (10): pp. 1-35, 2010

[XL08] Fu Xiao, Xie Li,
Using Outlier Detection to Reduce False Positives in Intrusion Detection,
IFIP International Conference on Network and Parallel Computing: pp. 26-33, 2008

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig und ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 28.09.2014

_____

Svenja Schulz