Migration gewachsener Umgebungen auf ein zentrales, datenschutzorientiertes Log-Management-System — Erfahrungen am Leibniz-Rechenzentrum

Stefan Metzger, Wolfgang Hommel, Helmut Reiser
metzger@lrz.de, hommel@lrz.de, reiser@lrz.de

Abstract:
Der Umgang mit personenbezogenen Daten, wie sie sich auch in Server- und Systemprotokollen finden, ist in den Datenschutzgesetzen und der aktuellen Rechtsprechung geregelt – allerdings in einer Form, die von juristischen Laien nur mit erheblichem Aufwand interpretiert werden kann. Wir haben uns diese Mühe gemacht und beschreiben in diesem Beitrag die Erstellung einer Log-Richtlinie für ein wissenschaftliches Rechenzentrum, ihre wesentlichen Inhalte, die werkzeuggestützte Umsetzung und Kontrolle sowie unsere im Projektverlauf und Betrieb gewonnenen Erfahrungen.

1 Einleitung

Die Zielsetzung, Protokolle von IT-Diensten und -Komponenten, die personenbezogene Daten beinhalten, maximal sieben Tage zu speichern, ist scheinbar relativ einfach umzusetzen: Serverdienste und Betriebssysteme werden zwar im Allgemeinen nicht entsprechend vorkonfiguriert ausgeliefert, aber mit minimalem Konfigurations- bzw. Implementierungsaufwand lässt sich eine adäquate Gleitlöschung realisieren. Was für ein einzelnes System also mehr oder minder trivial ist und bei seiner Kommissionierung a priori berücksichtigt werden kann, verursacht bei der nachträglichen Umstellung von großen, heterogenen, komplexen und gewachsenen Infrastrukturen einen signifikanten Aufwand, der im Vorfeld oder von außen betrachtet fast nur unterschätzt werden kann.

In diesem Beitrag wird über die Erfahrungen am Leibniz-Rechenzentrum (LRZ) bei der systematischen Umsetzung einer Richtlinie für die Handhabung von Protokolldaten durch alle Server und Dienste berichtet, die das LRZ als IT-Dienstleister der Münchner Hochschulen und als Höchstleistungsrechenzentrum mit europaweitem Einzugsbereich gesammelt hat. Bereits die Ausgangssituation ließ vermuten, dass auf die mehreren Dutzend Administratoren der zahlreichen Dienste, die auf mehreren hundert Servermaschinen, Appliances und aktiven Netzkomponenten betrieben werden, einiges an Arbeit zukommt: Während bei einigen Diensten bereits seit jeher eine kurze Logfile-Speicherdauer umgesetzt wurde, hatten sich bei Weitem nicht alle Administratoren und Dienstverantwortliche explizit mit dem Thema Log-Management auseinander gesetzt, so dass auf manchen Systemen die per Default-Konfiguration erstellten Protokolldateien sogar erst dann durch Automatismen gelöscht wurden, wenn der Plattenplatz auszugehen drohte.

Aktuelle Gerichtsurteile wie beispielsweise [Ber07], nach denen auch IP-Adressen als per-

sonenbezogene Daten aufzufassen sind, und eine verstärkte Ausrichtung des Sicherheitsmanagements am LRZ an ISO/IEC 27001 [ISO05] haben ein Projekt mit dem Inhalt motiviert, die Zielsetzungen im Log-Management in einer organisationsweit verbindlichen Richtlinie festzuhalten und deren werkzeugunterstützte Umsetzung intensiv zu begleiten und zu kontrollieren. In Abschnitt 2 stellen wir unsere Vorgehensweise bei der Ermittlung der Compliance-Anforderungen sowie der Einordnung unserer IT-Dienste vor und beschreiben das Zustandekommen der Richtlinie. Ausgewählte Inhalte dieser Richtlinie und das organisatorische Vorgehen bei ihrer Verabschiedung und Umsetzung sind Gegenstand von Abschnitt 3. In Abschnitt 4 wird auf die technische Umsetzung auf Basis des kommerziellen Werkzeugs *Splunk* eingegangen. Unser Beitrag endet mit einem Resümee des bisher Erreichten und einem Ausblick auf unsere nächsten Schritte in Abschnitt 5.

2 Vorgehen bei der Analyse und Richtlinienerstellung

Die verstärkte Ausrichtung des LRZ-Sicherheitsmanagements nach ISO/IEC 27001 betrifft auch das Thema Log-Management, das in der Norm explizit als Maßnahmenziel A.10.10 gefordert wird. Demnach sind Nutzeraktivitäten zu protokollieren, die erstellten Protokolle vor widerrechtlicher Manipulation zu schützen und regelmäßig auszuwerten. Zusätzlich müssen rechtliche Rahmenbedingungen erfüllt werden, die im Folgenden kurz dargestellt werden sollen, ehe die Vorgehensweise zur Analyse des Log-Managements am LRZ und die daraus resultierende Richtlinienerstellung vorgestellt werden.

Juristische Anforderungen in Deutschland ergeben sich in diesem Bereich aus diversen Gesetzen. Zu nennen sind hier das an der entsprechenden EU-Richtlinie orientierte Bundesdatenschutzgesetz (BDSG) [BDS03], das für das LRZ geltende Bayerische Datenschutzgesetz (BayDSG) [Bay93], das Telekommunikationsgesetz (TKG) [TKG04] sowie das Telemediengesetz (TMG) [TMG07].

Neben einer Definition, welchen Daten als personenbezogene Daten betrachtet werden müssen, regelt das BDSG deren Erhebung, Speicherung, Verarbeitung und Nutzung. Bereits die inhaltliche Abgrenzung von TKG zu TMG und vor allem die Einordnung aller vom LRZ angebotenen Dienste in den jeweils anzuwendenden Gesetzeskontext stellte eine unerwartet schwierige Aufgabe dar. Vereinfacht ausgedrückt bezieht sich der Inhalt des TKG auf die technische Infrastruktur und darin anfallende Verkehrsdaten, während sich das TMG überwiegend mit den übertragenen Inhalten auseinandersetzt. Eine kombinierte Anwendung von Bestimmungen gemäß TKG und TMG ist ebenfalls zulässig. Wireless LAN stellt beispielsweise einen klassischen Telekommunikationsdienst dar. Bei Webseiten, die nur der Veröffentlichung von Inhalten dienen, handelt es sich klar um einen Telemediendienst. Webmail hingegen, bei dem der Versand und Empfang von E-Mails im Vordergrund (TKG) steht, kann unter Berücksichtigung von Aspekten eines Web-Portals (TMG) als Mischdienst betrachtet werden. Die Erhebung von personenbezogenen Daten ist, falls das TKG anwendbar ist, nur zu bestimmten Zwecken wie der Fehlererkennung und -beseitigung sowie der Aufdeckung missbräuchlicher Dienstnutzung oder im Falle einer explizit vorliegenden Genehmigung des Nutzers gestattet. Der Zeitraum, wie lange diese Daten gespeichert werden dürfen, wird in den Gesetzestexten nicht vorgeschrie-

ben. Nach aktueller Rechtsprechung [Dar07] wird hierfür bei Anwendbarkeit des TKG ein Zeitraum von sieben Tagen als angemessen erachtet. Eine Ausweitung dieses Zeitraums ist nur bei einem konkret vorliegenden Vorfall erlaubt. Nur in solchen Fällen darf auch eine entsprechende Protokollierung personenbezogener Daten bei Telemediendiensten erfolgen, da das TMG nur anonymisierte und pseudonymisierte Nutzungsstatistiken, aber keine prophylaktische Protokollierung zur Fehlererkennung erlaubt.

Nachdem die Anforderungen geklärt waren, wurde auf dieser Basis ein kurzer Fragebogen entwickelt und an die Dienstverantwortlichen und Administratoren verteilt. Vereinzelt wurde die Beantwortung des Fragebogens auch in persönlichen Interviews durchgeführt. Ziel war es, den Status Quo des LRZ-Loggings zu ermitteln. Neben der Einhaltung rechtlicher Rahmenbedingungen waren auch technische und organisatorische Aspekte von Interesse. An dieser Stelle exemplarisch aufgezählt ein paar dieser Fragestellungen:

- Zu welchem Zweck werden die Daten erhoben? Ist Dauer für die Speicherung der Daten begrenzt bzw. aus Ihrer Sicht begrenzbar?
- Werten Sie die Daten regelmäßig aus und setzen Sie für die Auswertung der Daten Werkzeuge ein? Wenn ja, welche?
- Werden die Daten lokal und/oder zentral gespeichert? Wer besitzt lesenden bzw. schreibenden Zugriff auf diese Daten?

Nach Auswertung des Fragebogens ließen sich etwa 200 verschiedene Klassen von Logging-Quellen identifizieren, in denen personenbezogene Daten enthalten waren. Meist wurden diese zur Fehlererkennung und -eingrenzung für einen definierten Zeitraum von wenigen Tagen erhoben; in einigen Fällen mussten jedoch auch die bereits einleitend geschilderten Abweichungen attestiert werden. Die technische Umsetzung, insbesondere im Hinblick auf eine lokale oder zentrale Speicherung, sowie in punkto der hierfür eingesetzten Software und eingestellten Loglevels variierten überraschend stark. Außerdem wurde eine Vielzahl verschiedener Werkzeuge zur – meist nur reaktiv durchgeführten – Auswertung verwendet. Als Konsequenz wurde im Rahmen eines Projektes eine Log-Richtlinie erarbeitet, die neben rechtlichen Aspekten auch technische und organisatorische Punkte umfasst. Exemplarisch werden einige Inhalte im folgenden Abschnitt vorgestellt.

3 Ausgewählte Inhalte der Richtlinie und organisatorische Aspekte der Umsetzung

Das so entstandene, als LRZ-Richtlinie für Server- und Systemprotokolle bezeichnete Dokument wurde mit zwei DIN-A4-Seiten bewusst knapp, verständlich und als „Merkzettel“ geeignet formuliert. Es gibt zunächst einen Überblick über typische Verwendungszwecke von Protokolldateien, der als Checkliste eine individuelle Auseinandersetzung mit den Prinzipien der Datensparsamkeit und Zweckbindung anstoßen soll. Diesem folgt eine Unterscheidung zwischen personenbezogenen und anderen Protokollinhalten, die unter anderem bezüglich des Umgangs mit IP-Adressen, Benutzernamen und ähnlichen Informatio-

nen sensibilisieren soll. Da sich die Gerichtsurteile zu IP-Adressen als personenbezogenen Daten noch nicht flächendeckend herumgesprochen hatten, wurden diese nämlich bei den Interviews von einigen Gesprächspartnern nicht intuitiv als personenbezogen bezeichnet.

Die Richtlinie verweist anschließend auf das separate Dokument mit den gesammelten juristischen Hintergrundinformationen, mit dessen Hilfe Interessierte optional ihr Wissen vertiefen können. Als Quintessenz daraus findet sich die explizite Vorgabe in der Richtlinie, dass Protokolldateien mit personenbezogenen Daten unter dem Grundsatz der Datensparsamkeit und der Zweckdefinition und -bindung bei Anwendbarkeit des TKG maximal sieben Tage aufbewahrt werden dürfen, sofern mit den Betroffenen (z. B. Kunden und Anwendern sowie Mitarbeitern) keine explizite andere Vereinbarung getroffen wurde und somit deren Einverständnis vorliegt.

Die Dokumentation der Richtlinie schließt mit zwei Verweisen auf weitere Dokumente: Ein spezifisch für die LRZ-Umgebung erstellter technischer Leitfaden erläutert anhand konkreter Beispiele die relevanten Konfigurationsoptionen u. a. für *rsyslog* unter Linux und die Ereignisprotokollierung auf Windows-Servern, wobei zwischen befristeter lokaler Protokollierung und der Nutzung des in Abschnitt 4 beschriebenen Werkzeugs *Splunk* unterschieden wird. Diese Anleitungen sind einerseits für die Umkonfiguration bestehender Systeme relevant und werden andererseits in den Templates für neue (virtuelle und physische) Servermaschinen umgesetzt. Das andere Dokument, auf das verwiesen wird, konkretisiert mit den Betroffenen abzuschließende Vereinbarungen und regelt beispielsweise auch die Erstellung von Online-Datenschutzerklärungen für Webanwendungen, deren Benutzer nicht explizit authentifiziert werden.

Die Richtlinie wurde nach ihrem Entwurf zunächst im Rahmen des abteilungsübergreifenden, LRZ-internen Security-Arbeitskreises diskutiert und den Abteilungsleitern als Vorabinformation und zur Abschätzung des Umsetzungsaufwands vorgelegt. Nach der Einarbeitung von Rückmeldungen zur weiteren Verbesserung der Verständlichkeit wurde die Richtlinie von der LRZ-Leitung ratifiziert und in Kraft gesetzt. Da zu diesem Zeitpunkt noch nicht alle Dienste und Maschinen policykonform konfiguriert waren, wird seitdem eine Liste bekannter Verstöße geführt und der Status der Umsetzungsarbeiten kontinuierlich überwacht – die Anzahl nicht konformer Systeme und Dienste ist von ursprünglich über 40 auf unter zehn zurückgegangen, bei denen sich die erforderlichen Maßnahmen als sehr aufwendig erweisen und noch nicht abgeschlossen werden konnten. Über Systemmanagement-Skripte, die beispielsweise auch einmal pro Nacht die Konformität lokal installierter Firewalls darauf überprüfen, dass z. B. SSH-Logins nur von Managementstationen aus möglich sind, wird seither auch ausgewertet, ob auf Servern in den dafür üblichen Verzeichnissen Logfiles abgelegt sind, die älter als eine Woche und nicht in der Ausnahmeliste für Maschinen mit entsprechenden Benutzervereinbarungen aufgeführt sind.

Damit entsprechende technische Maßnahmen nicht nur als Schikane oder Einmischung in den Dienstbetrieb missverstanden werden, wurde auch die Möglichkeit zur Nutzung eines zentralen Log-Management-Systems geschaffen, auf das im nächsten Abschnitt eingegangen wird. Zur Einführung der technischen Möglichkeiten wurden Workshops für die LRZ-Administratoren veranstaltet. Die Inhalte der Richtlinie werden zudem in den jährlich stattfindenden IT-Sicherheitseinweisungen für alle Mitarbeiter behandelt.

4 Werkzeugunterstützte Umsetzung und Kontrolle der Leitlinie

Aus Sicht des LRZ besteht heutiges Log-Management neben der Einhaltung juristischer Anforderungen auch aus einem zentralen, hierarchisch aufgebauten Logging-Service, dazu proaktiver und wenn möglich automatischer Echtzeit-Auswertung sowie einer System- und damit Logfile-übergreifenden Korrelation von protokollierten Ereignissen. Der Aufbau und Betrieb eines zentralen Log-Servers wirft bereits im Hinblick auf die Ereignisweiterleitung einige Fragestellungen auf. Zu klären ist beispielsweise, ob die Log-Daten weiterhin lokal gespeichert oder ausschließlich zentral abgelegt werden und ob alle Daten an den zentralen Server weitergeleitet werden oder bereits eine lokale Vorfilterung durchgeführt wird. Nach zentraler Ablage der Daten ist eine strikte Mandantentrennung sicherzustellen, damit der Zugriff auf die Daten nur einem hierzu berechtigten Personenkreis möglich ist.

Bereits aus den hier genannten Punkten lässt sich die Notwendigkeit, ein geeignetes Werkzeug zu verwenden, direkt ableiten. Das LRZ hat sich für den Einsatz der Software *Splunk* entschieden. Ausschlaggebend für die Wahl waren positive Erfahrungen mit der Software im Rahmen einer durch eine LRZ-Abteilung durchgeführten, praxisnahen Produktevaluierung. Splunk bietet sehr flexible Möglichkeiten an, Daten zu integrieren. Log-Events können über eine einfach zu konfigurierende *Syslog*-Weiterleitung oder durch Überwachung entfernt liegender Ereignis-Speicher eingebunden werden. Dennoch wird am LRZ eine agentenbasierte Integration der Ereignisse, realisiert über einen dedizierten Forwarder, präferiert. Der Forwarder bietet neben lokalem Caching von Events, das bei temporärer Unerreichbarkeit des zentralen Servers wichtig ist, auch eine Ereignisfilterung.

Ein integriertes Rollen- und Rechte-Konzept sowie die Möglichkeit, zentrale Identity-Management-Lösungen anzubinden, erlauben die strikte Trennung von Log-Daten, die Einschränkung der Sichtbarkeit dieser Daten und insbesondere auch Beschränkung jeder Nutzerinteraktion mit dem System. Das am LRZ erarbeitete Konzept sieht einen mehrstufigen Ansatz vor. Zunächst wird auf Abteilungs- bzw. Gruppen-Ebene separiert. Damit ist die Sichtbarkeit von Log-Daten für einen überwiegenden Teil der LRZ-Mitarbeiter bereits auf Organisationsebene beschränkt. Innerhalb der Abteilungen und Gruppen kann der Zugriff bei Bedarf weiter separiert werden.

Basis für die effiziente Auswertung von Daten bildet wie bei den bisher verwendeten Werkzeugen eine Suche, die mittels einer integrierten Befehlssyntax sehr einfach parametrisiert werden kann. Vom Nutzer initiierte Suchen lassen sich, falls sie häufiger benötigt werden, in frei definierbaren Zeitintervallen oder in Echtzeit automatisch wiederholen und die Suchergebnisse direkt im System oder in Form von Dashboards oder Reports übersichtlich darstellen. Selbst das automatische Auslösen von definierten Aktionen bei Eintreten bestimmter Events, beispielsweise der Versand einer Alarm-E-Mail an den zuständigen Administrator bis hin zum Start von selbsterstellten Skripten, lassen sich konfigurieren. Für eine Vielzahl von Anwendungsfällen (Windows, Linux, Netzkomponenten) existieren vorgefertigte, frei verfügbare Applikationen, die sich per Mausklick in die eigene Splunk-Umgebung integrieren lassen und vordefinierte Suchabfragen bieten. Logfile-übergreifende Suchen sind nur ausgewählten Teams (z. B. Security-Team) erlaubt.

Betrachtet man abschließend den operativen Aufwand, so lässt sich festhalten, dass sowohl die Installation als auch das Update der Software sowie die technische Integration von Log-Daten in das System aus Sicht des LRZ keine große Herausforderung darstellen.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Projekt am LRZ vorgestellt, das innerhalb eines halben Jahres die Datenschutzkonzepte für Log-Daten grundlegend überarbeitet und eine Datenschutzrichtlinie erstellt sowie technisch umgesetzt hat. Die Vorgehensweise im Projekt war, mit Hilfe von Interviews den Ist-Stand zu ermitteln. Dabei wurden über 200 Klassen von Logging-Quellen identifiziert. Die Ergebnisse aus der Analyse wurden datenschutzrechtlich bewertet und aufbauend darauf wurden die Richtlinie erstellt sowie Konzepte zur technischen Unterstützung entwickelt. Die Policy konnte dann von der Leitung des LRZ in Kraft gesetzt werden. Nach Abschluss des Projektes und somit kurz nach Umsetzung der Richtlinie liegt die Anzahl der (noch) nicht konformen Quellen, bzw. der Dienste, bei denen eine explizite Vereinbarung mit den Betroffen geschlossen werden muss, inzwischen bei unter zehn. Auch die anfängliche Befürchtung, dass die Beschränkung der Speicherdauer negative Auswirkungen auf den Dienstbetrieb haben könnte, haben sich in keinster Weise bestätigt. Die vom LRZ gewählte Vorgehensweise lässt sich sehr einfach auch auf andere Rechenzentren oder Institutionen übertragen und in der in dieser Arbeit beschriebenen Form anwenden.

Danksagung

Die Autoren danken den Mitgliedern des Münchner Netzwerk-Management Teams (MNM-Team) für hilfreiche Diskussionen und wertvolle Kommentare zu früheren Versionen dieses Artikels. Das MNM-Team ist eine Forschungsgruppe der Münchener Universitäten und des Leibniz-Rechenzentrums der Bayerischen Akademie der Wissenschaften unter der Leitung von Prof. Dr. Dieter Kranzlmüller und Prof. Dr. Heinz-Gerd Hegering.

Literatur

[Bay93] Bayerisches Datenschutzgesetz (BayDSG). In *Bayerisches Gesetz- und Verordnungsblatt (GVBL)*, Seite 498, 23. Juli 1993. Zuletzt modifiziert am 27.7.2009 (GVBL S. 400).

[BDS03] Bundesdatenschutzgesetz (BDSG). In *Bundesgesetzblatt (BGBl. I)*, Seite 66, 14. Januar 2003. Zuletzt modifiziert am 14. August 2009 (BGBl.I S. 2814).

[Ber07] Landgericht Berlin. *Urteil vom 6. September 2007 - Az. 23 S 3/07*, September 2007.

[Dar07] Landgericht Darmstadt. *Urteil vom 6. Juni 2007 - Az.: 10 O 562/03*, Juni 2007.

[ISO05] DIN ISO/IEC 27001: Informationstechnik – IT-Sicherheitsverfahren – Informationssicherheits-Managementsysteme – Anforderungen (ISO/IEC 27001:2005), 2005.

[TKG04] Telekommunikationsgesetz (TKG). In *Bundesgesetzblatt (BGBl. I)*, Seite 95, 26. Juni 2004. Zuletzt modifiziert am 17. Februar 2010 (BGBl.I S. 78, 79f.).

[TMG07] Telemediengesetz (TMG). In *Bundesgesetzblatt (BGBl. I)*, Seite 10, 26. Februar 2007. Zuletzt modifiziert am 05. Juni 2010 (BGBl.I S. 692).