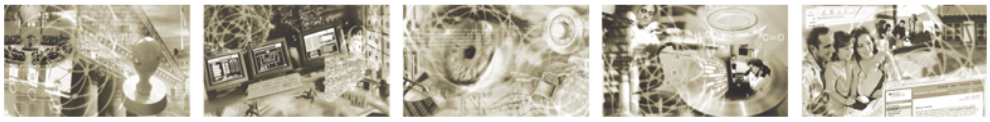
Bundesamt
für Sicherheit in der
Informationstechnik



Band B, Kapitel 10: Überwachung

Bundesamt für Sicherheit in der Informationstechnik
Postfach 20 03 63
53133 Bonn
Tel.: +49 22899 9582-0
E-Mail: hochverfügbarkeit@bsi.bund.de
Internet: https://www.bsi.bund.de

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Zur Erreichung und Optimierung eines geforderten Verfügbarkeitsniveaus kommt dem Monitoring eine besondere Bedeutung zu. Es bildet einerseits eine wesentliche Maßnahme zur Sicherstellung eines dauerhaften und zuverlässigen Betriebs, andererseits bietet es die Grundlage für die Steuerung und Kontrolle spezifizierter Serviceparameter und für ein professionelles IT-Service Management sowie die notwendigen Informationen im Hinblick auf die Optimierung der Verfügbarkeit und zur Unterstützung organisatorischer Prozesse und spezifischer Services im Umfeld der Verfügbarkeit.

Die Abbildung 1 zeigt den Gesamtprozess des Monitoring unter nicht funktionalem und damit auch Verfügbarkeitsfokus. Dabei sind Überwachung, (Früh-) Erkennung und Ereignisreaktion jeweils als Teilprozesse zu verstehen, in denen nicht nur das reine Überwachen der Betriebsparameter, sondern auch deren Auswertung, die Generierung von Alarmen oder die Auslösung von Aktionen geschieht. Der Gesamtprozess wird in der Praxis oft mit dem übergeordneten Begriff „Monitoring“ belegt.

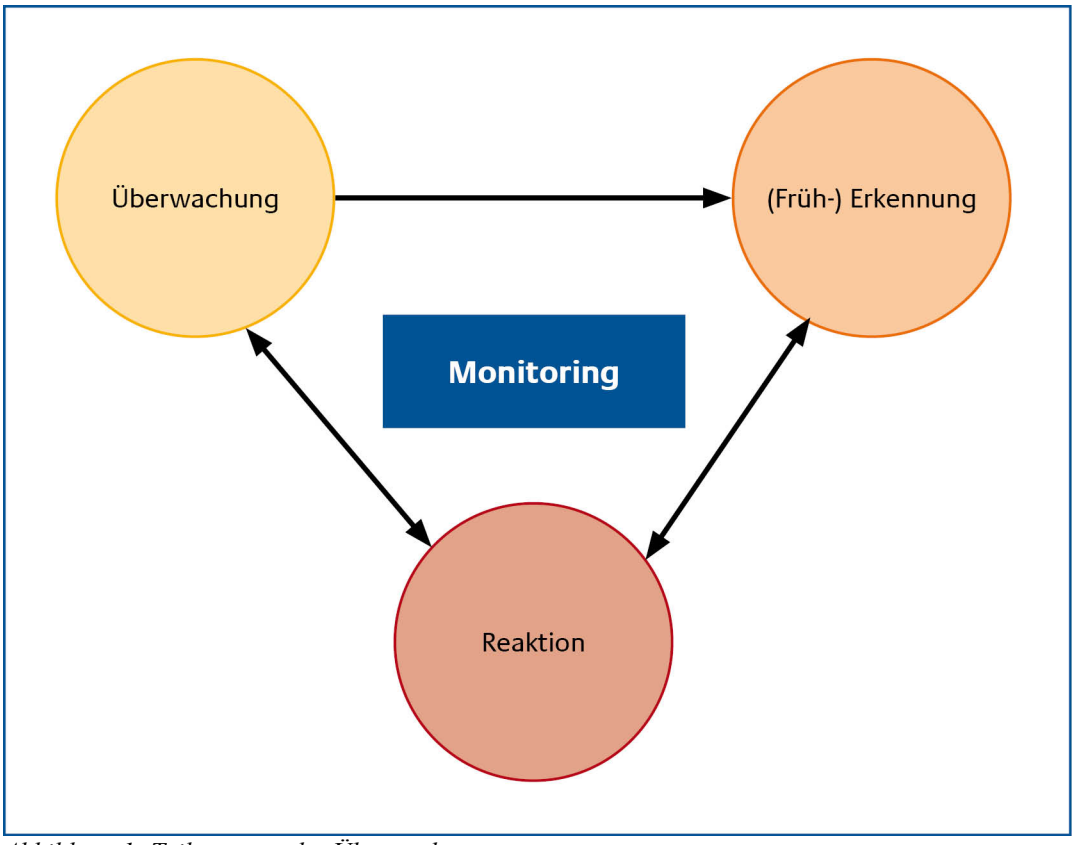


Abbildung 1: Teilprozesse der Überwachung

Ziel des Monitoring unter Verfügbarkeitsfokus ist es, durch geeignete Überwachungsmethoden und -techniken in einem möglichst frühen Stadium potentielle Gefährdungen der Verfügbarkeit zu erkennen und deren Ursachen zu beseitigen. Die Erfassung und Auswertung von Daten aus der Überwachung ermöglichen die Ermittlung von Trends und Prognosen bezüglich des zukünftigen Systemverhaltens. Diesen Prozess bezeichnet man auch als Früherkennung. Auf Grundlage erkannter Ereignisse oder auf der Basis von Prognosen werden Alarme generiert, die ein frühzeitiges Eingreifen zur Verhinderung eines Ausfalls von Systemen gestatten. Das Eingreifen kann manuell oder auch, wie in Fällen geforderter höherer Verfügbarkeit, automatisiert erfolgen.

Das vorliegende Dokument beschreibt die grundlegenden Prinzipien und Technologien des Monitorings sowie die erreichbaren Ziele aber auch die Grenzen der Überwachung, (Früh-) Erkennung und Ereignisreaktion unter Hochverfügbarkeitsaspekten. Ausgehend von der grundsätzlichen Betrachtung möglicher Mess- und Abfragetechniken werden verschiedene Überwachungsbereiche sowie Methoden der (Früh-) Erkennung und Aspekte der Ereignisreaktion dargestellt. Abschließend erfolgt eine weiterführende Darstellung von Monitoringaspekten der Wartung, Inspektion und Instandsetzung.

Zur weiteren Veranschaulichung zeigt die Abbildung 1 ein Monitoring-Szenario, das alle drei Teilprozesse des Monitorings beinhaltet. Ausgehend vom Teilprozess Überwachung (Kapitel 2) sind die weiteren Teilprozesse (Früh-) Erkennung (Kapitel 3) und Reaktion (Kapitel 4) dargestellt. Die Mess- und Kennwerte aus der Überwachung werden zum Teilprozess (Früh-) Erkennung übermittelt und dort ausgewertet. Basierend auf der Auswertung werden im Teilprozess Ereignisreaktion geeignete Maßnahmen zur Sicherstellung der Verfügbarkeit ergriffen.

Durch Monitoring-Clients gesteuert können aus der zentralen Monitoring-Infrastruktur Reports, Auswertungstabellen, ganze Lagebilder oder auch entsprechende visuelle Analysen generiert werden. Reports und Tabellen können wiederum Grundlage für eine weiterführende Analyse und Visualisierung der Systemzustände sein. Die Trennung von Data-Mining-Tool auf dem Client (oder separaten Server) und einem Monitoring-Tool auf dem zentralen Management-Server stellt dabei nur eine mögliche Variante dar. Heute sind Monitoring-Produkte aus dem kommerziellen und Open Source Bereich im Einsatz, die beide Funktionalitäten vereinen.

Darüber hinaus veranschaulicht die Abbildung 1 wie die Aspekte der Instandhaltung, Instandsetzung, Inspektion und Wartung sowie Steuerungsprozesse in den Gesamtprozess des Monitorings integriert werden können.

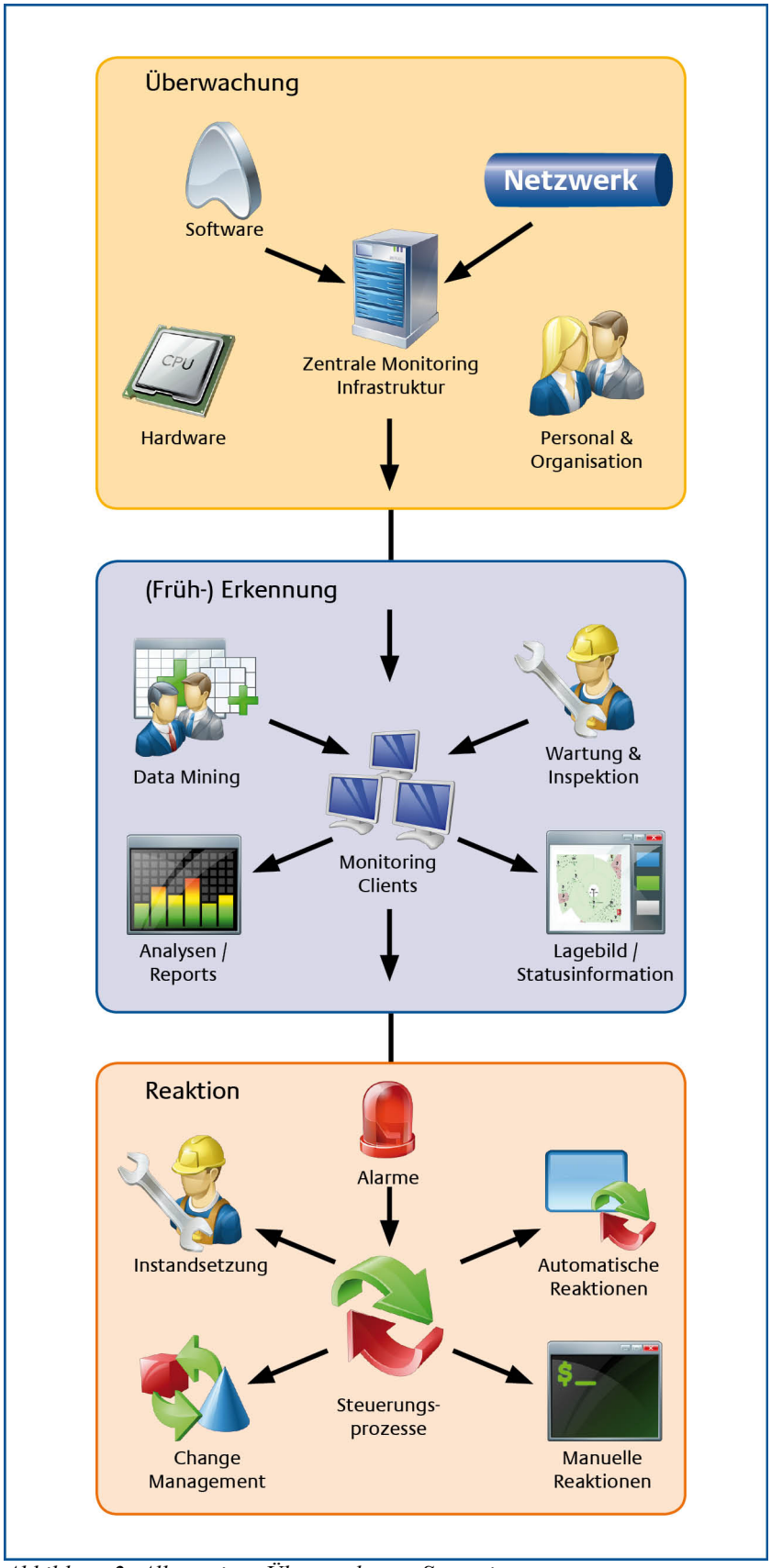


Abbildung 2: Allgemeines Überwachungs-Szenario

2 Überwachung

Die Überwachung kann in die Überwachungsbereiche

- Hardware,
- Software,
- Netzwerk und
- Personal und IT-Organisation

eingeteilt werden. Innerhalb dieser Bereiche ist eine weitere Einteilung hinsichtlich

- der erfassten Messgrößen,
- der Art der Sensorik sowie
- der eingesetzten Übertragungsprotokolle möglich.

Hierbei werden zuerst in jedem Überwachungsbereich Messgrößen vorgestellt, die verfügbarkeitsrelevante Aussagen über die zu überwachenden Komponenten ermöglichen. Darauf folgend wird beschrieben, wie die Messgrößen mittels geeigneter Sensorik erfasst werden können. Wie die einzelnen Messwerte bspw. zu einer zentralen Überwachungsinfrastruktur übermittelt werden, ist anschließend in den Abschnitten zu den Übertragungsprotokollen dargestellt. Diese Einteilung und Vorgehensweise orientiert sich dabei an der in der Studie zur Indikation der Verfügbarkeit von IT-Services (IdV-Services) [IdV 2009] vorgenommenen Strukturierung der Inhalte.

Bei der Planung der Überwachung gilt es zu bestimmen, wie exakt der zeitliche Verlauf einer Messgröße ermittelt werden muss. Bei sich schnell ändernden Größen ist eine hohe Abtastrate (hohe Messrate) erforderlich, um den exakten Zeitverlauf darzustellen (Abtasttheorem). In vielen Fällen ist es nicht erforderlich, Ausreißer oder Peaks zu erkennen, hier reicht es aus, einen Mittelwert über einen bestimmten Zeitraum zu erfassen und als Datenpunkt darzustellen. Dies kann zum einen daran liegen, dass die zeitliche Entwicklung einer Messgröße nur langsam geschieht oder zum anderen soviel „Headroom“ vorhanden ist, dass kurzzeitige Peaks keine Auswirkungen haben und nur Trends von Interesse sind. Bei einer zu geringen Abtastrate besteht die Gefahr, dass kurzzeitige Ereignisse oder aufschlussreiche Transienten nicht detektiert werden. Hohe Abtastraten bedeuten eine hohe Belastung des zu überwachenden Systems als auch des Netzes, das die Überwachungsdaten transportiert. Werden aufgrund von hohen Verfügbarkeitsanforderungen schnelle Reaktionen gefordert, müssen dem entsprechend Ereignisse mindestens ebenso schnell erfasst und erkannt werden. Aus diesem Grund muss sich die Frequenz der Messungen an den geforderten Reaktionszeiten orientieren. In den nachfolgenden Kapiteln werden, zu jedem Überwachungsbereich, Beispiele für die Auswahl von Abtastraten erläutert.

2.1 Hardware

Im Überwachungsbereich Hardware sind IT-Komponenten wie Rechner, Speichersysteme, Netzkoppelelemente sowie Infrastrukturkomponenten wie z. B. Energieversorgung, Klimatisierung und Sicherheitseinrichtungen zusammengefasst. Die Darstellung der Monitoringaspekte dieser IT-

Komponenten erfolgt unter Betrachtung der Messgrößen sowie der verwendeten Sensorik und Überwachungsprotokolle.

2.1.1 Messgrößen

Bei der Überwachung der Hardware werden die grundlegenden Funktionen (Vital-Funktionen) einer Hardware-Komponente überwacht.

Beispiele für Messgrößen von Rechnern sind:

- Anzahl zugewiesener CPU-Kerne,
- Temperatur, Spannung und Stromverbrauch einer CPU,
- Speicherbelegung,
- Link-Status, Auslastung und Übertragungsrate von Netzadaptern.

Beispiele für Messgrößen von Speichersystemen sind:

- Zustand von Festplatten,
- I/O-Raten,
- I/O-Fehler,
- Link-Status, Auslastung und Übertragungsrate von Netzadaptern.

Beispiele für Messgrößen von Infrastrukturkomponenten sind:

- Raumtemperatur und Luftfeuchtigkeit des Serverraums,
- Spannung und Frequenz der Energieversorgung,
- der Ladezustand der USV-Batterien,
- Kühlmittelstand der Klimatisierung.

2.1.2 Sensorik

Die Sensorik stellt Mechanismen und Methoden zur Erfassung der gewünschten Messgrößen zur Verfügung. Bei den Sensoren kann es sich um System-integrierte Software-Sensoren oder eigenständige Hardware-Sensoren handeln. Häufig werden auch Hybrid-Sensoren eingesetzt, die sowohl über eine lokale Erfassungskomponente im System verfügen, als auch über einen zentralen eigenständigen Sensor, der die Daten der dezentralen Einheiten sammelt und konzentriert weiter leitet.

Im Zusammenhang mit der Sensorik stellt die Software-basierte-Überwachung durch sogenannte Agentensysteme eine eigene Form der Überwachung dar. Hierbei werden Software-Sensoren (Agenten) in das zu überwachende Zielsystem integriert. Dies ist heute aufgrund immer leistungsfähigerer Prozessoren möglich. Zum Einsatz kommen dabei Code-Fragmente oder

eigenständige Programme, die zur Messung und Übermittlung vielfältiger Systemparameter eingesetzt werden. Diese Agenten messen als integrierte Sensoren, wie im Kapitel 2.1.1 aufgeführt, verschiedenste Messgrößen. Die Messwerte werden in der Regel an nach gelagerte zentrale Monitoring-Infrastruktur übermittelt. Dieses Prinzip bietet den Vorteil geringer Kosten und Aufwände bei einem hohen Detaillierungsgrad der gewonnenen Informationen. Darüber hinaus lassen sich Software-Sensoren bedarfsgerecht und skalierbar in das Zielsystem integrieren. Hierbei ist zu berücksichtigen, dass die Software basierte Überwachung die Verfügbarkeit des überwachten Systems in der Regel negativ beeinflusst, da die Sensoren die Ressourcen (wie schon im einführenden Absatz zur Sensorik ausgeführt) des Zielsystems belasten.

Die bei der Hardware-basierten Überwachung eingesetzten Sensoren (oder auch Probes) sind durch die Verwendung eigenständiger Betriebsmittel gekennzeichnet. Sie verfügen über eine vom Zielsystem unabhängige Hardware und Software zur Erfassung und Aufbereitung der Messwerte. Bedingt durch die Eigenständigkeit des Hardware-Sensors werden Einflüsse des Monitorings auf das Zielsystem weitestgehend vermieden. Darüber hinaus erlauben diese Systeme die Erfassung und Vorverarbeitung komplexer Messgrößen, da sie einen großen Teil der zu überwachenden HV-Infrastruktur überblicken können (beispielsweise durch das Abhören und Auswerten der Netzwerkkommunikation). Nachteile des Hardware-Sensorik liegen einerseits in den höheren Kosten für Installation und Betrieb der Komponenten, andererseits können Informationen nur mit einem beschränkten Detaillierungsgrad erfasst werden, da diese Systeme naturgemäß außerhalb des zu überwachenden Zielsystems angesiedelt sind. Sie fungieren somit als „Datensammler, welche zumeist an kritischen Punkten des Netzwerkes platziert sind.

Das Hybride Überwachungsverfahren stellt einen Kompromiss aus den beim Hardware-basierten sowie beim Software-basierten Überwachungsverfahren eingesetzten Prinzipien dar. Es erlaubt einen, gegenüber dem Hardware-basierten Überwachungsverfahren verbesserten Detaillierungsgrad der erfassten Messgrößen bei gleichzeitig verringerter Belastung des zu überwachenden Zielsystems. Die in das Zielsystem integrierten Agenten liefern Messgrößen an den außerhalb der überwachten Komponente angesiedelten, autarken Probes. Dieser kann eine Aufbereitung und Übermittlung der Messwerte durchführen, ohne das überwachte System zusätzlich mit Betriebsmittel zu belasten.

Im HV-Umfeld müssen Veränderungen von Messgrößen wie z. B. Spannung und Frequenz der Energieversorgung oder der Link-Status eines Netzadapters schnell erkannt werden. Dieses führt zwangsläufig zu hohen Abtastraten und damit verbunden zu großen Datenvolumina, die zu einer zentralen Monitoring-Infrastruktur übermittelt werden müssen. Die Übermittlung der großen Datenmengen der Überwachung belastet das Netzwerk neben den eigentlichen Nutzdaten erheblich. Sind insbesondere in einer umfangreichen IT-Architektur eine große Anzahl von Komponenten zu überwachen, sollte ein separates Überwachungs-Netz realisiert werden.

Ein Beispiel für geringe Abtastraten stellt die CPU-Temperatur dar. Die Temperaturänderungen der CPU verlaufen relativ langsam, und kurze Transienten sind nicht zu erwarten. In diesem Fall genügt es, die Temperatur im fünf-Minuten-Abstand zu ermitteln. Häufig wird die Temperatur im 10-Sekunden-Takt gemessen, ein Datenpunkt in der Darstellung ist dabei der Durchschnitt aus den Werten der vorangegangenen fünf Minuten.

2.1.3 Übertragungsprotokolle

Die durch unterschiedliche Verfahren, Sensoren und Agenten erfassten Daten müssen in geeigneter Weise an die zentrale Monitoring-Infrastruktur übertragen werden. Darüber hinaus werden Kommunikationskanäle für die Übermittlung von Statusabfragen oder zusätzlichen Informationen benötigt. Hier bieten sich standardisierte Schnittstellen wie beispielsweise das Simple Network Management Protocol *(SNMP)*an (vgl. IT-Grundschutz-Kataloge des BSI: M 2.144 Geeignete Auswahl eines Netzmanagement-Protokolls [BSI GS-Kat2009]). Über einen einheitlichen Zugriff und in *Management Information Bases (MIBs)* genormten Strukturen kann hier ein großes Spektrum an Informationen Netzwerk-intern und System-übergreifend abgefragt werden.

Das *Internet Control Message Protocol (ICMP)* wird in SNMP für Polling-Anfragen genutzt, ist also integraler Bestandteil einer Monitoring-Lösung, die auf SNMP-Basis arbeitet.

Die Strukturen vieler MIBs sind in RFCs genormt. Viele Hersteller liefern darüber hinaus noch weitere Schnittstellen, über die zusätzliche Informationen abgefragt werden können. Diese Strukturen sind in Structures of Management Information *(SMI)* abgelegt, die aktuelle Form davon ist die in RFC 2580 beschriebene SMIv2.

MIBs sind über das SNMP-Protokoll abfragbar. Es gibt drei Versionen des Protokolls, die als SNMPv1 in RFC 1157, SNMPv2c in RFC 1906 und SNMPv3 in RFC 2572 und RFC 2574 beschrieben sind. Das Protokoll SNMPv1 bietet nur schwache Maßnahmen zur Autorisierung und keine Mechanismen zur Wahrung von Vertraulichkeit und Integrität, es ist aber das von Monitoring-Werkzeugen und insbesondere von Komponenten am Markt insgesamt noch am stärksten unterstützte Protokoll. Bei dem Einsatz von SNMP sollte daher die Absicherung der Datenübertragung berücksichtigt werden. Dies soll ein unerlaubtes Ausspähen von Monitoring-Daten und/oder eine Manipulation der Statusinformationen verhindern. Ferner ist zu beachten, dass bei mangelhafter Konfiguration der SNMP-Autorisierung auch unbefugt in die Konfiguration der überwachten Komponente eingegriffen werden kann. SNMP ermöglicht auch die automatische Erkennung von aktiven Netzkomponenten (Router, Switches).

Remote Monitoring *(RMON)* ist eine Erweiterung der MIB und im RFC 2819 und RFC 2021 definiert. Die RMON Spezifikation definiert über die reine Erhebung der Daten hinaus weitere Funktionen zur Aufzeichnung, Filterung und statistischen Aufbereitung der Daten. Mittels Network Probes werden die Daten zu RMON (compliant)-Konsolen übermittelt und können dann zu einer umfassenden Fehler-Diagnose für die Hardware-Komponenten im Netzwerk herangezogen werden. Auch können sog. RMON-Agenten Datenvorverarbeitung im Sinne der Datenvorverarbeitung (siehe Abbildung 1) erledigen.

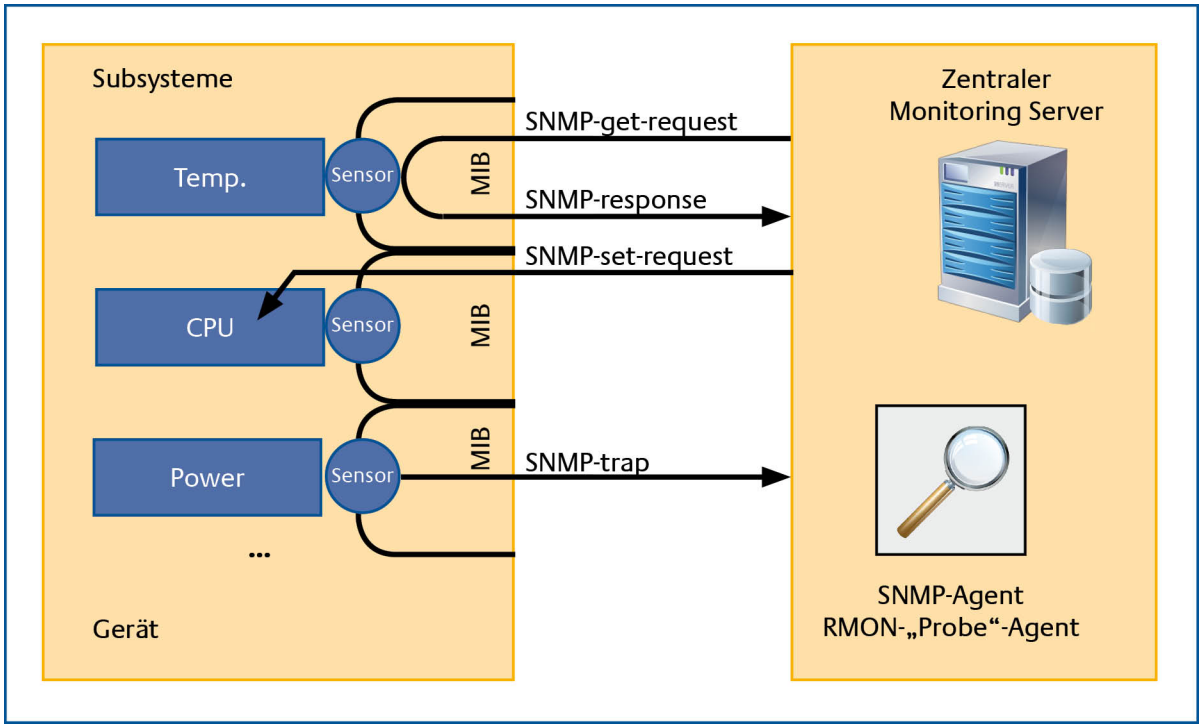


Abbildung 3: Abläufe und Einsatz des Protokolls SNMP

In allen Fällen geschieht die Übermittlung der Messwerte über das bestehende Netzwerk. Zu beachten ist, dass diese Übermittlung zu einer weiteren Belastung des Netzes führt. In HV-Strukturen sollte daher zur Übermittlung der Messwerte ein separates Überwachungs-Netz errichtet werden.

Neben der geringeren Belastung des Wirk-Netzes bietet ein separates Überwachungs-Netz zusätzliche Sicherheit. Überwachungsdaten können einem Angreifer nützliche Informationen über ein Netzwerk und die darin befindlichen IT-Komponenten bereitstellen. Ein separates Netz erschwert den Zugang und Zugriff, insbesondere dann, wenn die Überwachungsdaten nicht verschlüsselt übertragen werde.

Da das Monitoring einen wesentlichen Beitrag zur Aufrechterhaltung der Verfügbarkeit, insbesondere in den hohen Verfügbarkeitsklassen VK3-VK5 (vgl. BSI HV-Kompendium V 2.0 Definitionen [BSI HV-Kom2012]) liefert, sind für dieses Überwachungs-Netze vergleichbare Verfügbarkeitsanforderungen zu stellen wie für das Wirksystem. Mindestens sind zum Erreichen der hohen Verfügbarkeitsklassen die zentralen Monitoring-Komponenten redundant auszulegen.

2.2 Software

Der Überwachungsbereich Software unterteilt sich in die Schichten System-Software (Betriebssysteme, Treiber), systemnahe Software (Services, Middleware) und Anwendungs-Software. Analog zum Kapitel Hardware erfolgt hier eine weitere Gliederung der Betrachtung in die Aspekte Messgrößen, Sensorik und Übertragungsprotokolle. Neben der Messung der Ressourcenauslastung oder des Ressourcenverbrauchs durch die Software kann die Performance sowie die Qualität der Ergebnisse einer Funktion gemessen werden.

2.2.1 Messgrößen

So wie sich der Überwachungsbereich Software in Schichten unterteilt, müssen auch die Messgrößen entsprechend betrachtet werden.

Beispiele für die Messgrößen der System-Software sind:

- Uptime,
- CPU-Nutzung,
- Anzahl der Prozesse und Threads,
- Speichernutzung und Memory-Leaks,
- Schreib- oder Lesefehler.

Beispiele für die Messgrößen der systemnahen Software sind:

- Antwortzeiten,
- Heartbeat,
- Durchsatz,
- Zustände von Warteschlangen.

Beispiele für die Messgrößen der Anwendungs-Software sind:

- Antwortzeiten,
- Qualität der Antworten,
- Anzahl und Art von Fehlern,
- Queue-Größen.

2.2.2 Sensorik

Die Sensorik im Überwachungsbereich Software ist im Wesentlichen analog zur der im Bereich Hardware. Darüber hinaus kann mittels spezieller Testanfragen die Performance und die Qualität überwacht werden.

Das sog. „Endanwender-Monitoring“ (EAM) führt eine Überwachung der für den Endanwender sichtbaren und spürbaren Service-Qualität durch. Die hierbei erfassten Messgrößen gehen über technische Betriebsparameter hinaus und versuchen eine Aussage über die Einhaltung der im Service-Level-Agreement (SLA) vereinbarten Dienste-Qualität zu erhalten. Dies geschieht in der Regel ebenfalls durch Agenten, die die typischen Tätigkeiten eines Endanwenders simulieren und die Antwort- und Reaktionszeiten des Dienstes erfassen. Die zumeist im Netzwerk verteilten dedizierten Agenten, aber auch auf Client-Systemen installierte Agenten können den Verlauf von Transaktionen erfassen und analysieren. Dabei ist zu berücksichtigen, dass bei dieser Methode auch die Performance des Netzwerks mit in die Messung eingeht (vgl. Kapitel Fehler: Referenz nicht gefunden). Die Koordination der im Netzwerk verteilten Agenten und die Sammlung der gewonnenen Messinformationen wird durch spezielle Software so genannten „Distributed Agent Controller“ (DAC) realisiert.

Viele Messgrößen im Bereich der System- und Systemnahen-Software geben unmittelbar Auskunft über den Systemstatus und müssen im HV-Umfeld mit einer hohen Erfassungsrate gemessen werden. Bei Messgrößen wie z. B. Heartbeat, Durchsatz, Antwortzeiten etc. sollte der Zeitraum zwischen zwei Messungen entsprechend kurz sein, insbesondere dann, wenn automatische Reaktionen erfolgen sollen. Die hierfür erforderlichen hohen Abtastraten sorgen für einen erheblichen zusätzlichen Datenstrom im Wirk-Netz. Zum Erreichen der hohen Verfügbarkeitsklassen (VK3-VK5) ist ein separates Überwachungs-Netz notwendig. Die Log-Dateien werden hingegen in der Regel mit deutlich geringerer Frequenz ausgewertet.

2.2.3 Übertragungsprotokolle

Neben denen in Kapitel 2.1.3 dargestellten Protokollen ist das Syslog-Protokoll ein weiterer grundlegender Standard zur Übermittlung von Log-Meldungen in IP-basierten Netzen. Es stellt ein eigenes Standardprotokoll dar, welches keine Möglichkeit der Authentifizierung bietet. Da es von vielen Geräten unterstützt wird, ist es oft Bestandteil von Monitoring-Lösungen.

2.3 Netzwerk

Das Messen in Netzen dient der gezielten Überwachung der Parameter und Leistungsdaten eines Netzwerkes und liefert durch die Kombination verschiedenster Überwachungsverfahren und Mechanismen ein umfassendes Zustandsbild des Netzwerkes. Dies umfasst in der Regel auch die im Netzwerk befindliche Hardware, aktive Komponenten und Netzwerk-Dienste, deren Überwachung bereits in den vorhergehenden Kapiteln betrachtet wurde. Informationen über die Auslastung des Netzwerkes oder vorhandene Latenzzeiten können auftretende Probleme und drohende Verluste der Verfügbarkeit rechtzeitig erkennen lassen, um ggf. Gegenmaßnahmen einzuleiten.

2.3.1 Messgrößen

Im Rahmen der Überwachung von Netzwerken kommen, neben denen im Kapitel 2.1.1 aufgeführten Messgrößen der Hard- und Software, weitere wichtige netzwerkspezifische Messgrößen hinzu, z. B.:

- Paketverzögerung (Latenzzeiten, Einweg- und Umlaufverzögerung),
- Abweichungen der Laufzeit (engl. Jitter),
- Durchsatz (engl. Bulk Transfer Capacity, BTC),
- Paketverluste (Paketverzögerung > Timeout).
- Vertauschung der Ankunftsreihenfolge von Datenpaketen (engl. Paket Reordering)

Ergänzende Messgrößen für ein Monitoring in eigenbetriebenen Weitverkehrsnetzen (WAN) sind u. a:

- WAN Link-Status
- Router Interface Parameter (Auslastung, Frame-Rate, Errors, Anzahl verworfener Frames, Verhältnis Broadcast zu Unicast)

– Router Performance (CPU- und Speicherauslastung, Buffer Hits/Misses)

2.3.2 Sensorik

In IP-basierten Netzwerken sind keine Schnittstellen zum direkten Messen der Messgrößen vorgesehenen. Eine Möglichkeit, Kenntnisse über den Zustand des Netzes zu erlangen, ist die Beobachtung und Analyse mittels Messungen der Datenströme an geeigneten Stellen im Netzwerk. Datenpakete werden an bestimmten Stellen im Netzwerk gesammelt und analysiert (vgl. Kapitel 2.1.2). Diese Vorgehensweise wird als *passives Messen* bezeichnet. Die dazu notwendigen passiven Sensoren beschränken sich auf das „Abhören“ von System- und Leistungsparametern. Im wesentlichen handelt es sich hierbei um sog. TAPs (Traffic Access Point, Mess-Splitter). Ein TAP arbeitet immer auf der OSI-Schicht 1 und ist vergleichbar mit einem sehr hochohmigen Messverstärker. Es sind Kupfer-basierte und sog. Fiberoptic-TAPs zu unterscheiden. Beide Typen sind im Vergleich zu normalen passiven Sensormethoden (Abgriff über einem Hub, Abgriff von einem Switch aus, Mirrorport-Analyse) kostenintensiv, da sowohl für die Anschaffung, als auch für die Einbindung in die Kommunikationsverbindung Aufwendungen anfallen. Für die Verfügbarkeit der überwachten Datenleitung ist von Bedeutung, dass ein Ausfall eines TAP, nicht zu einer Unterbrechung der Datenkommunikation führt, also dieser nicht unbedingt redundant vorgehalten werden muss. Mittels passivem Messen können nur vorhandene Datenströme im Durchfluss untersucht werden und dies erlaubt somit keine Bestimmung von Verzögerungen oder Paketverlusten. Die gewonnenen Erkenntnisse aus der Messung der vorhandenen Datenströme ermöglichen im Rahmen eines Capacity Managements die Parametrisierung und Dimensionierung eines Netzwerkes. In der Regel wird der Datenverkehr dabei nicht gestört. Die Durchführung der Messung erfordert jedoch einen Zugang zu einem entsprechenden Netzknoten und kommt damit nur für Netzbetreiber in Betracht.

Eine weitere Möglichkeit ist das gezielte Injizieren von sog. Test- oder Probepaketen in das Netzwerk. Diese Methode wird als *aktives Messen* bezeichnet und erlaubt die Bestimmung von Verzögerungen und Paketverlusten, insbesondere zur Überwachung von Weitverkehrsstrecken (WAN), wenn diese von Dritten betrieben werden. Bei diesem Szenario werden zwischen zwei Endpunkten Datenpakete im Netzwerk ausgetauscht. Im Rahmen der Einweg-Messung werden die Messsysteme, in Form von aktiven Sensoren (sog. Probes und Agenten) an zwei Punkten im Netz angeschlossen. Die Laufzeiten und Verluste der Pakete werden durch Messen der Zeit zwischen dem Aussenden und Empfangen der Pakete ermittelt. Bei der Umlauf-Messung wird bspw. mit dem Programm *ping* ein Paket von einem Punkt zu einem zweiten Punkt gesendet. Der zweite Punkt beantwortet dieses Paket mit einem neuen Paket. Auch hier wird die Zeit zwischen dem Aussenden und Empfangen der Pakete gemessen. Hier muss jedoch berücksichtigt werden, dass der zweite Punkt mit einer Zeitverzögerung antwortet. Nutzer des Netzes und Diensteanbieter, die in der RRegel keinen Zugang zu einem Netzknoten haben, erlaubt das aktive Messen, das Verhalten der Datenströme im Netz zu analysieren.

Zur Bestimmung der Momentanwerte der Netzwerk-Performance werden Probepakete mit einer hohen Rate einige Millisekunden bis einige Sekunden in das Netz injiziert. Dieser Messvorgang wird in regelmäßigen Abständen wiederholt. Soll im Rahmen des aktiven Messens ein Verbindungsausfall in einem Netzwerk detektiert werden, dürfen die Abstände zwischen zwei Messvorgängen nicht größer als die Zeit sein, in der ein Ausfall erkannt werden soll.

Im Bereich des passiven Messens sollen u. a. Erkenntnisse für das Capacity Management gewonnen werden. Häufig kommt es aufgrund von kurzzeitigen Spitzenbelastungen zu Kapazitätsengpässen

und damit verbundenen Performance-Einschränkungen. Zur Detektion solcher kurzzeitigen Ereignisse ist eine hohe Messrate erforderlich.

In allen Fällen geschieht die Übermittlung der Test- oder Probepakete über das bestehende Netzwerk. Zu beachten ist, dass diese Übermittlung zu einer weiteren Belastung des Netzes führt. Diese Messungen können nicht über ein Überwachungs-Netz durchgeführt werden.

2.3.3 Übertragungsprotokolle

Im Kapitel 2.1.3 sind bereits die Übertragungsprotokolle zur Übermittlung der Messwerte der Hardware-Komponenten in einem Netzwerk beschrieben worden. Das Switched Monitoring *(SMON)* ist neben RMON ein weiteres Verfahren, um den Datenverkehr von aktiven Netzelementen (Routern, Switches) zu erfassen. Es ist im RFC 2613 beschrieben. Das Protokoll stellt eine Erweiterung des RMONs dar (u. a. durch Statistikfunktionen für VLANs). SMON-Parameter wie Paketverzögerungen, Jitter, und Paketverluste können wiederum über SNMP einer zentralen Institution zur Verfügung gestellt werden.

2.4 Personal und IT-Organisation

Die Verfügbarkeit von IT-Services wird nicht ausschließlich durch die Verfügbarkeit von Hard- und Software bestimmt. Eine weitere wesentliche Rolle spielen das eingesetzte IT-Personal sowie die zu Grunde liegende IT-Organisation mit ihren Prozessen. Die Definition von IT-Prozessen kann bspw. auf der Basis generischer Prozessmodelle wie CobiT und ITIL erfolgen und ist in [BSI HV-Kom2012] Kapitel Organisation und Personal beschrieben.

Oft muss bspw. aufgrund der Komplexität von IT-Lösungen Fachkunde und Expertenwissen herangezogen werden, um i) Situationen richtig zu beurteilen, ii) notwendige Entscheidungen zu treffen, iii) manuell einzugreifen und iv) dieses in Hochverfügbarkeitsumgebungen in kürzester Zeit. Die notwendige Fachkunde und die entsprechenden Personen müssen „verfügbar“ sein.In der Studie zur Indikation der Verfügbarkeit von IT-Services (IdV-Services) [IdV 2009] sind Messgrößen und Sensorik aufgeführt, die herangezogen werden können, den Personaleinsatz den Anforderungen entsprechend zu überwachen und zu gestalten.

2.4.1 Messgrößen

Die Messgrößen für das Personal müssen hinsichtlich der Größen, die das Personal direkt betreffen und denen, die die organisatorischen Aspekte des Personaleinsatzes betrachten, unterschieden werden.

Direkte Messgrößen beziehen sich beispielsweise auf die Anzahl und die Qualifikation der Mitarbeiter. Die IT-Organisation betreffende Messgrößen betrachten Aspekte wie die Rollen- und Rechteplanung, Betriebs- und Verfahrensanweisungen, Einsatzplanung, Urlaubsregelungen und Vertretungsregelungen.

Beispiele für direkte Messgrößen für Personal sind:

- Anzahl der Mitarbeiter,
- Fluktuationsrate,

- Qualifikationslevel und Zertifizierungen,
- Abhängigkeiten von Schlüsselpersonen.
- Beispiele für die Messgrößen von organisatorischen Aspekten sind:
- Reifegrade,
- Key Performance Indicators (KPIs),
- Anzahlen der Incidents,
- Verletzung von Service Level Agreements (SLAs),
- Konsistenz der Schichtpläne,
- Anreisezeit,
- Güte der Arbeitsanweisungen.

2.4.2 Sensorik und Übertragungsprotokolle

Die Sensorik der Überwachungsbereiche Personal und IT-Organisation ist grundsätzlich andersartig zu betrachten als die der bisher aufgeführten Überwachungsbereiche und stellt besondere Anforderungen an die Monitoringarchitektur. In diesen Bereichen kann keine Messtechnik im Sinne von technischen Komponenten, die physikalische Größen aufnehmen, eingesetzt werden. Die Sensorik muss sich vielmehr auf die menschliche Erfassung von geeigneten Prozess-spezifischen Indikatoren als Messgrößen ausrichten. Die erfassten Messgrößen entstammen hierbei in weiten Teilen linguistischen Wertebereichen und müssen nach der Erfassung in der Regel in konkrete Wertebereiche transformiert werden um einen Überwachungsprozess mit exakten Angaben zu Messgrößen, Messverfahren, Messzyklen und Auswertungen zu etablieren.

In diesem Sinne kann bei der Betrachtung der Sensorik grundsätzlich zwischen internen Sensoren (internes Monitoring) und externen Sensoren (externes Monitoring, Revision, Audit) unterschieden werden. Das *interne Monitoring* beruht im Wesentlichen auf einer Selbsteinschätzung durch die beteiligten Personen oder Organisationsbereiche. Dabei werden die Messgrößen durch interne Mitarbeiter erfasst. Die Organisation erfasst und überwacht hierbei selbst die Elemente und Aktivitäten, die direkt von ihr gesteuert werden.

Beim *externen Monitoring* erfolgt die Erfassung und Aufbereitung durch externe Personen oder Organisationen, wodurch in der Regel eine höhere Güte und Konsistenz der Überwachung realisiert werden kann.

Technische Sensoren in Form von funktionalen Softwarekomponenten können im Bereich Personal und IT-Organisation verwendet werden, sofern bereits Software-basierte Lösungen eingesetzt werden. In diesem Fall kann durch eine geeignete Erfassung und Auswertung vorhandener Datenbestände, wie beispielsweise einer Configuration-Management-Database (CMDB) eine nahtlose Integration in die zentrale Monitoring-Infrastruktur erreicht werden. Die Sensorik beruht in diesem Fall auf einer selektiven Aufbereitung vorhandener Datenbasen.

Von Übertragungsprotokollen im Sinne von technischen Protokollen kann in diesem Überwachungsbereich nur mittelbar gesprochen werden. Sofern technische Sensoren im Sinne des vorhergehenden Absatzes eingesetzt werden können, lassen sich die gleichen technischen

Übertragungsprotokolle wie in den anderen Überwachungsbereichen einsetzen. In weiten Teilen werden in diesem Bereich aber auch Papier-basierte oder mündliche „Übertragungsprotokolle“ eingesetzt, da sowohl die erfassten Messgrößen als auch insbesondere die daraus abgeleiteten (Früh-) Erkennungs- und Reaktionsprozesse eine hohe Komplexität aufweisen und einen direkten Bezug zu der, bzw. direkte Auswirkungen auf die, die hochverfügbare IT-Architektur betreibende Organisation haben.

2.5 Nutzer im Monitoring

Auch der Nutzer kann im Monitoring als (Quasi-) Sensor betrachtet werden. Im HV-Umfeld ist es selbstverständlich zu spät, wenn der Nutzer den Ausfall eines IT-Services bemerkt. Der Nutzer kann jedoch im Rahmen der Früherkennung einen wesentlichen Beitrag leisten.

2.5.1 Subjektives Empfinden

Ein wesentlicher Vorteil des Nutzers als Sensor liegt darin, dass der Nutzer mit realen Daten und mit natürlichem Verhalten den IT-Service benutzt. Er kann somit Anomalien in seiner alltäglichen Arbeit erkennen, die Software-Agenten mit kontinuierlichen, aber statischen Testanfragen nicht in der Lage sind zu erkennen. Der Nutzer kann damit „transzendente“ Ereignisse, also Ereignisse, die die empirische Erfahrung überschreiten und somit auch nicht durch die technischen Sensoren detektiert werden, aufgrund seines subjektiven Empfindens „erfühlen“.

Beispiele für subjektive Empfinden sind:

- Erleben eines Services,
- Veränderung der „gefühlten Verfügbarkeit“,
- Performance-Schwankungen,
- nicht reproduzierbare Phänomene,
- Veränderungen der Zufriedenheit.

Jedoch entsteht die Frage, wie das subjektive Empfinden eines Nutzers erfasst (gemessen) und aufbereitet werden kann, so dass im Monitoring-Prozess Früherkennung handlungsrelevante Erkenntnisse gewonnen werden können. Die mögliche Indikatoren liegen in der Regel nur in linguistischer Form und ohne jegliche Skala oder Normierung vor (z. B. „heute ruckelt es aber heftig“, „es dauert länger als üblich“).

Zur Erreichung einer Bewertbarkeit müssen Indikatoren gefunden werden, die das subjektive Empfinden kategorisieren und es dem Nutzer erlauben, auf einer vorgegebenen Skala sein Empfinden zu bewerten. Mögliche Wertekategorien können mittels Beobachtung, Vergleich oder Abschätzung gewonnen werden. Darüber hinaus sollte der Nutzer auch die Möglichkeit haben, wirklichen Freitext anzugeben, um damit Phänomene zu beschreiben, die bisher für die Erhebung noch keine Rolle gespielt haben.

Die Indikatoren sollten in Form von Fragebögen, bspw. als Online-Fragebögen im Intranet, den Nutzern angeboten werden. Ferner sollte dazu ein Prozess etabliert werden, der den Nutzer kontinuierlich auffordert, bestimmte Indikatoren mit seinem subjektive Empfinden zu beschreiben.

Beispiele:

- „Gefühlte“ Verfügbarkeit, Skala: schlecht-mäßig-mittel-gut-sehr gut,
- Zufriedenheit, Skala: nicht-teilweise-vollständig,
- Performance, Skala: 0-1-2-3-4-5-6.

Die Auswertung kann mit den im Kapitel Fehler: Referenz nicht gefunden Früherkennung beschriebenen Methoden erfolgen.

2.5.2 Messwerte und Kenngrößen

In diesem Abschnitt werden Messwerte und Kenngrößen aufgeführt, die aus dem Nutzerumfeld erfasst werden und als Indikatoren zur Früherkennung dienen können.

Mit der Hilfe von geeigneten Sensoren und Analysemethoden (vgl. Kapitel 3.2.3.1) können bspw. die folgenden Größen erfasst und zur Früherkennung herangezogen werden:

- Beobachtung und Auswertung zur Optimierung der Benutzer-/Bedienprozesse („REFA“),
- Durchschnittliche Bearbeitungszeiten und/oder Nutzeraktivitäten,
- Kontinuität von Bearbeitungsprozessen,
- Anzahl der Tickets am UHD/SPOC,
- Anzahl und Verhältnis der Anfragen bei First-, Second-, Third-Level Support.

Hier ist zu bemerken, dass einige Kenngrößen eine direkte Beurteilung der Leistung des Nutzers und damit seines Arbeitsverhaltens erlauben. Die Kenngrößen müssen auf jeden Fall anonymisiert werden und die Auswertung sollte zuvor mit der Arbeitnehmervertretung und dem Datenschutzbeauftragten geklärt werden. Es soll nicht die Leistung des Nutzers, sondern die der IT bewertet werden.

Zur Gewährleistung einer kontinuierlichen Datenerfassung sollten mit den o. g. Instanzen oder Prozessen Schnittstellen definiert werden. Die Schnittstellen sollten hinsichtlich möglicher Übertragungsprotokolle (z. B. elektronisch oder Papier), der Formate und geeigneter Zeitintervalle beschrieben sein. In Analogie zum technischen Bereich führt auch hier eine hohe „Abtastrate“ zu einer hohen zusätzlichen Belastung des Nutzers sowie der Übertragungs- und Auswerteprozesse.

2.6 Zusammenfassung

In der Tabelle 1 sind die verwendeten Typen der Überwachung noch einmal zusammengefasst. Die Tabelle zeigt die verwendeten Typen der Überwachung mit deren Charakteristika. Für die verschiedenen Überwachungsbereiche wird die Art der Überwachung, die Natur der Messwerte, die Arten der Skalen, die Möglichkeiten der Sensorik, die Form des Messens und die dazugehörigen Übermittlungsmöglichkeiten zu einer zentralen Monitoring-Infrastruktur dargestellt.

Überwach-ungsbereich	***Art***	***Messwert***	***Skala***	***Sensorik***	***Form***	***Übermitt-lung***
Hardware Software Netzwerk	Vitalfunk-tionen Performance	physikalisch	ordinal kardinal	Probes Agenten Hybrid Taps Mirrorports	passiv aktiv	SNMP RMON SMON Syslog
Endan-wender	Qualität	Kennwert	nominal ordinal kardinal	Agenten Mensch	aktiv	SNMP RMON SMON
Management Personal	Performance	Kennwert	nominal ordinal kardinal	Mensch	intern extern passiv	Bericht-erstattung

Tabelle 1: Überwachungstypen

3 (Früh-) Erkennung

Im Gesamtprozess Monitoring ist die (Früh-) Erkennung ein weiterer Prozess, der die Aufgabe hat, neben dem Erkennen von akuten Störungen und damit verbundenen Verfügbarkeitsbeeinträchtigungen mögliche zukünftige Störungen frühest möglich zu prognostizieren. Die (Früh-) Erkennung bildet die Basis für Reaktionen, um damit geeignete Maßnahmen kurzfristig oder sogar im Vorfeld der Beeinträchtigung zu ergreifen.

Die zuvor erläuterte Erhebung von Daten und deren Weiterleitung an ein Überwachungssystem, bilden die Grundlage für die in diesem Abschnitt behandelte (Früh-) Erkennung. Generell ist der Prozess (Früh-) Erkennung in die drei Bereiche Anzeigen, Erkennen und Früherkennen einzuteilen.

3.1 Anzeigen

Die Aufgabe dieses Bereichs ist die Darstellung von Mess- und Kennwerten, die im Bereich der Überwachung erfasst wurden. Diese Werte werden ohne weitere Verarbeitung und Interpretation angezeigt und es liegt einzig in der Verantwortung des Überwachungspersonals, diese zu interpretieren und geeignete Schlüsse zu ziehen.

Beispiele für solche Anzeigen sind Uhreninstrumente (analog/digital), Balkenanzeigen, Signale etc. Diese Anzeigegeräte können sich unmittelbar in der Nähe des zu überwachenden Systems befinden, können aber zur Überwachung mehrerer Systeme in einem entfernten Armaturenbrett (Uhreninstrumente vgl.Abbildung 4) installiert sein.

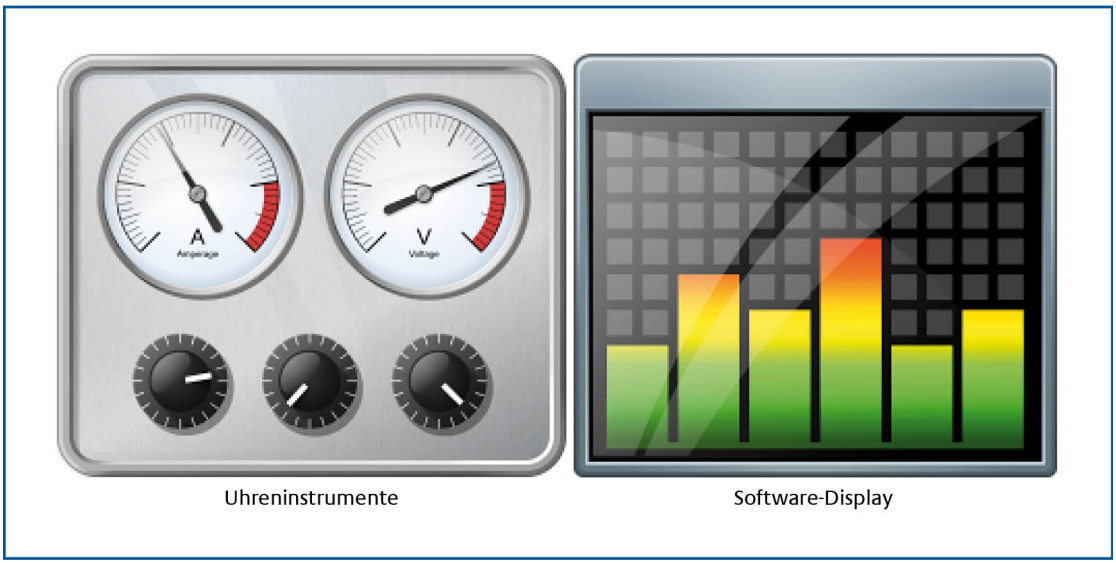


Abbildung 4: Anzeigevarianten

Eine moderne Variante der Anzeigeinstrumente besteht darin, die einzelnen Instrumente in Form von Graphiken auf einem Multifunktions-Display oder einem Software-Display zu emulieren (vgl. Abbildung 4 Software-Display).

3.2 Erkennung

Das Erkennen von Ereignissen geht deutlich über das reine Anzeigen von Mess- und Kennwerten hinaus. Die Werte müssen analysiert und interpretiert werden. Sobald Werte kritische Schwellwerte erreichen, sollen geeignete Reaktionen erfolgen. Dazu werden in Abhängigkeit der Dringlichkeit Meldungen und Alarme erzeugt, die durch den Teilprozess Reaktion „aufgefangen“ werden und zu Maßnahmen der Sicherung oder Wiederherstellung der Verfügbarkeit führen. Ein Ziel der Erkennung ist es, Fehler unmittelbar nach ihrem Auftreten zu erkennen, die Ursachen eines Fehlers zu ermitteln und ggf. die Instandsetzung zu alarmieren – dieses mit dem Ziel, die Mean Time to Repair (MTTR) zu minimieren. Dazu werden geeignete Methoden, Visualisierungstechniken sowie Techniken der Berichterstattung vorgestellt.

3.2.1 Manuelles Erkennen

Werden Mess- und Kennwerte nur angezeigt, bedarf es einer Beobachtung dieser Werte durch geschultes und fachkundiges Personal (vgl. Abbildung 9). Das Personal muss z. B. wissen, wann Messwerte kritisch zu bewerten sind und muss entsprechend reagieren. Häufig besteht die Gefahr, dass das Erreichen von kritischen Werten zu spät oder gar nicht erkannt wird. Ferner ist eine ständige Beobachtung der Werte durch das Personal erforderlich.

3.2.2 Automatisiertes Erkennen

Dieser Abschnitt beschreibt u. a. das automatisierte Erkennen von kritischen Werten und damit verbundene Auslösen von Alarmen. Dazu werden in diesem Kapitel verschiedene Analysemethoden dargestellt, die es erlauben, Fehler zu erkennen, Zusammenhänge von Fehlermeldungen zu ermitteln und die Ursachen für Fehler aufzudecken. Im darauf folgenden Abschnitt werden Visualisierungsmethoden vorgestellt, die die Ergebnisse der Analysen in einer für das Überwachungspersonal geeigneten Form anzeigen.

3.2.2.1 Analysemethoden

Die Ansprüche an die Analyse und Auswertung zielen auf die Aufdeckung von Zusammenhängen, Abhängigkeiten und zeitliche Entwicklungen hin. Häufig werden die Methoden der deskriptiven Statistik, um empirische Daten durch Tabellen und Grafiken übersichtlich darzustellen und zu ordnen, sowie durch geeignete grundlegende Kenngrößen zahlenmäßig zu beschreiben, angewendet.

Ein wichtiger Gesichtspunkt ist in dieser Hinsicht, dass erzeugte Analysen nicht nur zur Beurteilung der aktuellen Situation herangezogen werden, sondern vielmehr im zunehmenden Maße auch in eine mittel- bis langfristige Planung und Steuerung der IT-Architektur einfließen. Über diesen Aspekt hinaus lassen sich (auf den ersten Blick) unerklärliche Phänomene in der Regel durch eine sorgfältige Analyse der Messwerte erklären und beseitigen.

3.2.2.1.1 Schwellwertanalyse

Feste (auch fixe) Schwellwerte werden für einzelne, meist kritische, Komponenten (Messobjekte) der IT-Services anhand von Daumenregeln, Herstellervorgaben oder vereinbarten Service Level Agreements festgelegt. Die Erreichung der Schwellwerte wird im Rahmen der Überwachung geprüft. In der Regel werden zwei Schwellwerte pro Messobjekt definiert. Der erste Wert löst bei

Erreichung eine Warnmeldung aus, der zweite einen Alarm. An den definierten Schwellwerten ist damit auch der Status des Messobjekts (der Komponente) geknüpft.

Einige Monitoringwerkzeuge ermöglichen die dynamische Anpassung von Schwellwerten anhand von statistischen Verfahren (z. B. Standardabweichungen). Auch bei diesem Verfahren werden in der Regel zwei Schwellwerte bestimmt. Der erste Wert löst bei Erreichung eine Warnmeldung aus, der zweite einen Alarm. An den definierten Schwellwerten ist damit auch der Status des Messobjekts (der Komponente) geknüpft. Kombiniert wird dieses Verfahren mit festen Schwellwerten, die die Grenzen für die dynamische Anpassung nach oben bzw. nach unten hin, je nach Definition und Messparameter, darstellen.

Die Self-Monitoring, Analysis and Reporting Technology (SMART) ist bspw. ein in Festplatten integriertes Verfahren zur Selbstüberwachung, Analyse und Statusmeldung von Festplatten. Es ermöglicht das permanente Überwachen von Festplatten-spezifischen Parametern und somit das frühzeitige Erkennen drohender Defekte. Die dazu gehörige Software orientiert sich an vom Festplattenhersteller vorgegebenen Schwellwerten für die einzelnen Parameter.

3.2.2.1.2 Auswertung von Event-Logs

Die Auswertung von protokollierten Ereignissen erfolgt ggf. unter Anwendung von Tools (z. B. SIEM-Tools - Security Information and Event Management oder Web-basierte Syslog Management Software). Diese Tools ermöglichen Log-Dateien von mehreren Komponenten in einer Datenbank zusammenzufassen und an einer zentralen Stelle zu analysieren (bspw. lässt sich analysieren, wie viele Anmeldungen, Reboots, Fehler-Events etc. in einem bestimmten Zeitraum auf mehreren Rechnern aufgetreten sind). Die Ergebnisse der Auswertung ermöglichen eine schnellere Beseitigung eines Fehlers und reduzieren damit die MTTR.

3.2.2.1.3 Ereigniskorrelation (Event Correlation)

Ein Fehler kann eine Flut von Ereignissen auslösen. Diese Flut von Ereignissen muss effizient verwaltet werden, um eine optimale Steuerung und letztendlich eine optimale Bearbeitung aller Ereignisse sicherstellen zu können. Dabei ist der Informationsinhalt von Ereignissen durch die Unterdrückung nicht gewollter, redundanter Events und das Hinzufügen neuer, mehr informativer Events mittels der Nutzung von Event-Reduzierung-Strategien (wie z. B. Deduplizierung) zu verbessern. Die Event-Correlation-Funktion setzt Ereignisse untereinander in Beziehung.

3.2.2.1.4 Ursachenanalyse (Root Cause Analysis)

Die Ursachenanalyse für Fehler kann automatisch-deterministisch erfolgen oder probabilistisch (wahrscheinlichkeitstheoretisch). Die Grundlage für die Ursachenanalyse bildet der Erkennungsprozess (auch Discovery-Prozess genannt). Ziel der Ursachenanalyse ist die Reduzierung der MTTR und die Erhöhung der Betriebseffizienz. Dabei werden voneinander abhängige Störungsströme untersucht.

So werden bei Störungen z. B. die eingehenden Meldungen mittels einer Eventkorrelation ausgewertet. Es wird geprüft, ob die Meldungen voneinander abhängen und evtl. eine gemeinsame Ursache haben. Eine Aufgabe der Konsolidierung besteht darin, die Vielzahl der Meldungen zu High-Level-Alarme zu verdichten. Ist z. B. die Klimatisierung in einem Server-Raum gestört, melden, die sich darin befindenden IT-Systeme, eine kritische Temperaturerhöhung. Die Eventkorrelation soll in diesem Fall erkennen und melden, dass die Klimatisierung gestört ist und dass die weiteren Meldungen davon abhängen.

3.2.2.2 Visualisierung

Das Überwachungspersonal muss mit Hilfe der visuellen Techniken den aktuellen Gesamtzustand des IT-Systems erkennen können. Im Normalbetrieb ist beispielsweise der Detaillierungsgrad der Darstellung sehr gering. So werden z. B. nur die Komponenten des IT-Systems dargestellt und deren Zustand durch Hinterlegung mit grüner Farbe signalisiert. Kritische Zustände oder Ereignisse werden hingegen mit gelber oder roter Farbe hinterlegt. Zusätzlich besteht die Möglichkeit für das Überwachungspersonal durch Wechsel auf weitere Darstellungsebenen, den Detaillierungsgrad hinsichtlich dieser Ereignisse zu erhöhen. Unter Konsolidierung versteht man das Aufbereiten der aktuellen Messwerte, so dass das Überwachungspersonal mittels weniger Indikatoren einen vollständigen Gesamtüberblick über die aktuelle Situation erhält. Zum Beispiel wird nur der Min-, Max- und Mittelwert eines Indikators zyklisch aktualisiert oder Systemzustände in Form von interpretierten Messgrößen als Aussagen wie „Rot, Gelb, Grün“ oder „OK, Warnung, Kritisch“ dargestellt. In vielen Fällen werden zusätzliche akustische Signale erzeugt, die dazu dienen, die Aufmerksamkeit des Überwachungspersonals zu lenken.

Hier kommen vor allem einfache und intuitive Methoden der graphischen Darstellung (2D-Standarddiagramme, Zustandssymbole- und farbliche Tabellen) zur Darstellung von Systemzustände zum Einsatz.

In den folgenden Abschnitten werden die beiden wesentliche Methoden der Darstellung von Systemzuständen erläutert.

3.2.2.2.1 Statustabellen

Die Abbildung 5 zeigt eine Tabelle mit Service Details, wie sie beispielsweise von einer typischen Monitoring-Lösung erzeugt werden. Die Tabelle liefert eine detaillierte Übersicht der Statusinformationen von Services mehrerer überwachter Systeme. Eine farbliche Hinterlegung ermöglicht somit einen schnellen Überblick über die relevanten Servicezustände.

Host	Service	Status	Last Check	Duration	Attempt	Service Information
bogus-router	PING	CRITICAL	07-15-2001 13:59:39	4d 3h 43m 17s	1/3	CRITICAL - Plugin timed out after 10 seconds
bogus1	Something...	CRITICAL	07-15-2001 14:00:38	4d 3h 58m 49s	1/3	(Service Check Timed Out)
	PING	CRITICAL	07-15-2001 14:02:36	4d 3h 58m 49s	1/3	CRITICAL - Plugin timed out after 10 seconds
bogus2	PING	CRITICAL	07-15-2001 13:59:09	4d 3h 44m 27s	1/3	CRITICAL - Plugin timed out after 10 seconds
	Something...	CRITICAL	07-15-2001 13:59:39	4d 3h 42m 26s	1/3	(Service Check Timed Out)
bogus3	PING	CRITICAL	07-15-2001 14:00:38	4d 3h 42m 7s	1/3	CRITICAL - Plugin timed out after 10 seconds
	Something...	CRITICAL	07-15-2001 13:57:36	4d 3h 30m 35s	1/3	(Service Check Timed Out)
bogus4	PING	CRITICAL	07-15-2001 13:59:09	4d 3h 43m 35s	1/3	CRITICAL - Plugin timed out after 10 seconds
	Something...	CRITICAL	07-15-2001 13:59:39	4d 3h 42m 26s	1/3	(Service Check Timed Out)
bogus5	PING	CRITICAL	07-15-2001 14:00:43	4d 3h 41m 7s	1/3	CRITICAL - Plugin timed out after 10 seconds
	Something...	CRITICAL	07-15-2001 13:57:36	4d 3h 30m 25s	1/3	(Service Check Timed Out)
netware3	Total Cache Buffers	WARNING	07-15-2001 13:59:48	4d 3h 28m 24s	3/3	Total cache buffers = 21193
netware4	Total Cache Buffers	WARNING	07-15-2001 14:01:01	4d 3h 27m 14s	3/3	Total cache buffers = 22691
nt3	Physical Memory Use	CRITICAL	07-15-2001 14:02:28	3d 1h 21m 44s	3/3	Physical memory problem - 506.4 MB (99%) of 511.4 MB used
printer1	PING	CRITICAL	07-15-2001 14:02:46	1d 1h 35m 15s	1/3	CRITICAL - Plugin timed out after 10 seconds
	Printer Status	CRITICAL	07-15-2001 14:01:20	1d 1h 35m 54s	1/3	Timeout: No response from 134.84.92.77

Abbildung 5: Statustabellen

3.2.2.2.2 Netzwerk-Darstellung

Die Abbildung 6 skizziert beispielhaft die Darstellung eines Netzwerkes, wie sie in einer Netzwerk-Management-Software erzeugt wird. Es soll nicht nur die aktuelle Situation widerspiegeln, sondern darüber hinaus zur Fehlerdiagnostik beitragen und aufgrund verschiedener Analysen auch Prognosen für das zukünftige Verhalten des IT-Systems visualisieren. Diese Anforderungen an die Netzwerk-Darstellung gehen dabei schon in Bereiche des visuellen Data Minings (vgl. Kapitel 3.2.3).

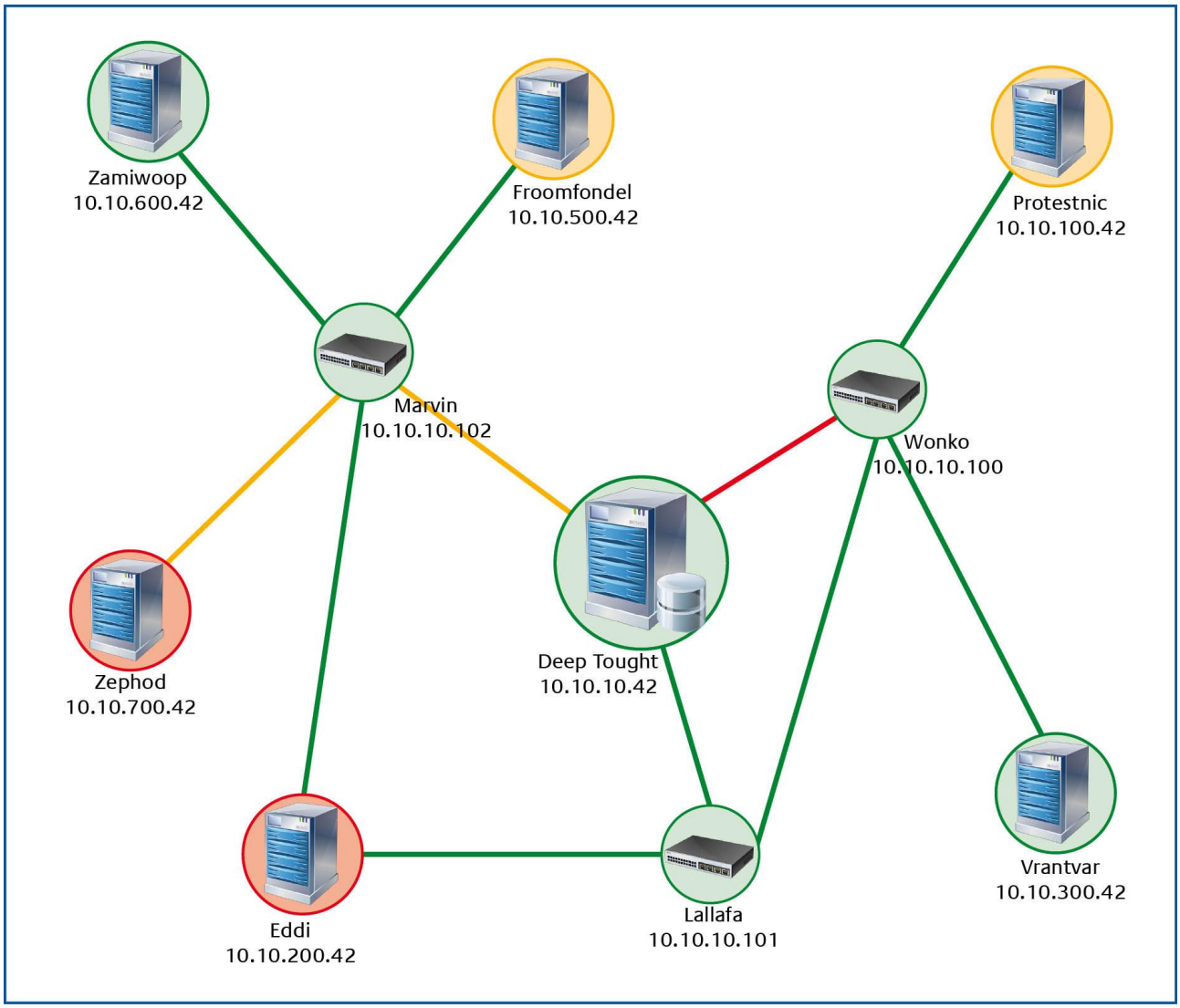


Abbildung 6: Netzwerkdarstellung

3.2.2.3 Alarme

Werden Ereignisse erkannt, müssen in der Regel zur Wahrung oder Wiederherstellung der Verfügbarkeit Reaktionen erfolgen. In manchen Fällen kann das Überwachungspersonal über die Nutz-Netzwerkverbindung direkt eingreifen und beispielsweise IT-Systeme rekonfigurieren oder

rebooten. Im Rahmen eines so genannten Out-of-Band-Managements greift das Überwachungspersonal über eine zusätzliche Schnittstelle („Hintertür“) außerhalb des regulären Datenpfades (Out-of-Band-Infrastructure, z. B. serielle Schnittstelle oder separates Management-Netz) ein. In den Fällen, in denen das „remote“ Eingreifen durch das Überwachungspersonal nicht mehr möglich ist, muss die Instandsetzung (vgl. Kapitel 4.2) aktiviert werden. Je nach Art des Ereignisses werden z. B. Meldungen bei geringer Dringlichkeit oder Alarme bei erhöhter ausgelöst. Das Überwachungspersonal entscheidet über die Dringlichkeit, mit der das Ereignis behandelt werden muss. Ferner ist es die Aufgabe des Überwachungspersonal zu entscheiden, welche Einheit für die Instandsetzung aktiviert wird.

Im Rahmen der Erkennung werden auch Ereignisse behandelt, die nicht aufgrund eines Fehlers aufgetreten sind, sondern auch Zeitereignisse wie z. B. das Ende eines Wartungs- oder Inspektionsintervalls. In diesen Fällen löst das Überwachungspersonal eine Meldung an die Inspektion oder Wartung (vgl. Kapitel3.2.4 und 4.3) aus.

Ereignisse, die vom „Single Point of Contact“ (SPoC) aufgrund einer Nutzermeldung generiert wurden, können ebenfalls zu einem Alarm führen, sofern sie nicht im Rahmen eines Incident-Management-Prozesses zufriedenstellend behandelt werden können.

Die Alarmierung der Instandsetzungseinheiten darf nicht über die zu überwachenden IT-Systemen erfolgen. Im Fall eines Fehlers ist die kurzfristige Alarmierung über die ausgefallenen Komponenten nicht möglich. Vielmehr müssen alternative Kommunikationsmöglichkeiten wie z. B. Telefon (wenn nicht VoIP), Mobiltelefon, Pager oder Bote verwendet werden.

Des Weiteren obliegt es dem Überwachungspersonal, auch über die Durchführung der notwendigen Instandsetzungsarbeiten zu wachen. Dazu muss es von der Instandsetzung informiert werden, inwiefern sich die Störung zeitnah beseitigen lässt oder ob ggf. eskaliert werden muss oder sogar der IT-Notfall proklamiert werden muss. Auch der Erfolg einer Instandsetzung muss dem Überwachungspersonal gemeldet werden. Einige kommerzielle Monitoring-Systeme beinhalten Komponenten zur Nachverfolgung der Störungsbeseitigung und zur Erstellung von Management-Berichten.

3.2.3 Früherkennung

Die Aufgabe des Bereichs Früherkennung besteht darin, aus den erhobenen Daten Ereignisse in einem so frühen Stadium zu erkennen, so dass es mittels geeigneter Gegenmaßnahmen zu keinen oder nur geringfügigen Beeinträchtigungen der Verfügbarkeit kommt. Neben der Darstellung von Prognosemethoden wird hier auch auf Auswertungsmöglichkeiten des Data Mining eingegangen.

3.2.3.1 Prognosemethoden

Im Fokus stehen in diesem Abschnitt Methoden, die es ermöglichen, Aussagen über das zukünftige Verhalten des IT-Systems oder der Komponenten zu geben. Im Wesentlichen basieren die Prognosen auf Methoden der mathematischen Statistik, auf Simulation und Berechnungsmodellen.

3.2.3.1.1 Richtlinien und Daumenregeln

Hierbei handelt es sich um allgemeine Richtlinien oder Regeln zur Vorhersage des Performance-Verhaltens. Dabei werden oft nur die zentralen Komponenten separat betrachtet, um Verfügbar-

keitsverluste zu vermeiden. Daumenregeln helfen, Grenzen abzustecken. Sie basieren u. a. auf Erfahrungen, Angaben von Herstellern und Benchmarkwerten. Dieses Verfahren liefert ungenaue Ergebnisse.

3.2.3.1.2 Zeitreihenanalyse

Die Zeitreihenanalyse ist eine Methode zum Erkennen von Trends und zur Erstellung von Prognosen aus aufgezeichneten Daten. Die Zeitreihenwerte sind z. B. chronologisch geordnete Werte eines konsolidierten Reports (z. B. IP-Verkehr gemessen über die Zeit).

3.2.3.1.3 Historische Daten

Dieses Verfahren beinhaltet die Sammlung (z. B. durch Polling) und Protokollierung von historischen Daten als Grundlage für Trendanalysen und Berichte. Dabei werden u.a. Ressourcen- und Performance-Messdaten von Anwendungen, Datenbanken, Netzwerken und Systemen eingesammelt, protokolliert und zusammengefasst. Diese historischen Daten können anschließend statistischen Analyse- und Berichtsprogrammen zur Verfügung gestellt werden.

3.2.3.1.4 Trendanalyse

Die Trendanalyse dient der Identifizierung von Trends und zur dynamischen Erkennung von Ressourcen- und Performance-Engpässen, die zu Verfügbarkeitsverlusten führen können und damit zur Bestimmung der optimalen Konfiguration, der Ersetzung und /oder der Erweiterung der entsprechenden Komponenten. Dieses Verfahren trifft Vorhersagen basierend auf den gesammelten Historiedaten. Dazu werden Langzeitbeobachtungen der entsprechenden Komponenten mit Hilfe von Monitoring-Werkzeugen vorgenommen und die gewonnenen Daten entsprechend verdichtet. Häufig werden lineare Extrapolierungsansätze verwendet. Dieser Ansatz, den Zusammenhang zwischen Last und Performance-Maßen durch eine lineare Gleichung zu beschreiben, berücksichtigt nicht den rasanten Anstieg von Leistungsmaßen in der Nähe der Sättigungspunkte. Darüber hinaus werden bei der Trendanalyse in der Regel nicht das Zusammenwirken unterschiedlicher Arbeitslasten berücksichtigt.

3.2.3.1.5 Lastsimulation

Lastsimulation beinhaltet die Nachbildung der realen IT-Services, der mit ihnen verbundenen Datenbestände sowie des Benutzerverhaltens. Ziel ist das Erkennen von Engpässen zur Vermeidung von Verfügbarkeitsverlusten. Diese Lastsimulation kann in Form von Tests in Referenzumgebungen durchgeführt werden. Im Betrieb werden dafür die Lastdaten erhoben. Die Synthetisierung erfolgt mittels Lastgenerator. Messwerkzeuge analysieren das IT-Serviceverhalten als Gesamtsystem hinsichtlich der Verfügbarkeit und ermöglichen so dessen Bewertung, Dimensionierung und optimale Konfiguration.

3.2.3.1.6 Analytische Modellierungsverfahren

Die analytische Modellierung (Verifikation, stochastisch, deterministisch) beschreibt die Performance des IT-Service durch eine Reihe von Gleichungen. Das Performance-Verhalten wird durch mathematische Größen und Beziehungen zwischen ihnen beschrieben. Das analytische stochastische Modell berücksichtigt zufällig auftretende Ereignisse. Die Erstellung einer solchen Modellierung kann sehr aufwändig sein, da in der Regel eine Vielzahl von Abstraktionen, Vereinfachungen und Annahmen nötig sind, um das Performanceverhalten auf einige wenige Parameter zu reduzieren. Die Genauigkeit der Modellierung hängt sehr stark davon ab, wie genau sich die Realität durch die Modellierungselemente beschreiben lässt.

3.2.3.1.7 Simulative Modellierungsverfahren

Die simulative Modellierung (Black-Box-Test, deterministisch) beschreibt das Performanceverhalten mittels mathematischer Größen und Beziehungen zwischen ihnen. Diese Modelle beinhalten aber auch Größen, die sich in Abhängigkeit von der Zeit dynamisch ändern (Ereignis-diskret oder kontinuierlich). Sie ermöglichen genauere Performance-Ergebnisse als analytische Modelle, da sie mehr Details des realen IT-Service nachbilden können. Das Verfahren ist in der Regel aufwändiger als die analytische Modellierung. Die simulative Modellierung sollte hauptsächlich zur Analyse und Bewertung von wichtigen Details eingesetzt werden.

3.2.3.2 Data Mining

Klassisch wird unter Data-Mining verstanden, neues, gültiges und handlungsrelevantes Wissen ohne konkrete Fragestellung zu entdecken. Die Praxis erfordert jedoch eine präzise Beschreibung der Problemstellung.

Allgemein können komplexe IT-Systeme, wie z. B. ein Netzwerk, als kybernetisches System verstanden werden. Daraus folgt, dass die Rückkopplungen aller beteiligter Komponenten betrachtet und analysiert werden müssen. Untersuchungsgegenstand ist somit das Aufspüren von Regeln und Mustern bzw. statistischen Auffälligkeiten der eingehenden Messgrößen, Meldungen, Warnungen und Fehlermeldung. Heute kennt man unterschiedliche Verfahren, dieser Datenflut Herr zu werden. Dieses „Methodenbesteck“ ist unter dem Begriff Data Mining zusammengefasst. Im Umfeld der Früherkennung kommen nun größtenteils automatisierte Auswertungen von Datenbeständen mit Hilfe des Data-Mining-Ansatzes zum tragen. Grundlagen des Data-Minings sind Methoden der deskriptiven und induktiven Statistik. Darauf setzen die eigentlichen Data-Mining-Verfahren, wie künstliche neuronale Netze, Clustering-Verfahren, Diskriminanzanalysen, Multifaktorielle Regressionsanalyse, Einsatz genetischer Algorithmen usw. auf. Ziel dabei ist das Aufspüren von Regeln und Mustern bzw. statistischen Auffälligkeiten.

Beispielhaft für andere Data-Mining-Methoden soll hier aufgeführt werden, welche skalierbaren Möglichkeiten die Analyse von Zeitreihen bietet:

- Die einfachste Beschreibung einer historischen Zeitreihe (z. B. durch mathematische Funktion),
- Vorhersage von künftigen Zeitreihenwerten (Prognose) auf der Basis der Kenntnis ihrer bisherigen Werte (z. B. Vorhersage des weiteren Verlaufes der CPU-Auslastung),
- Erkennung globaler Trends (z. B. langfristige Erhöhung des IP-Verkehrs),
- Erkennung von Abhängigkeiten von äußeren Parametern (z. B. Pausenzeiten der Nutzer, Wetter, Sonnenaktivität etc.),
- Eliminierung von seriellen Abhängigkeiten oder Trends in Zeitreihen, um einfache Parameter wie Mittelwerte verlässlich zu schätzen.

Diese lassen erste Schlüsse über das Vorliegen von Trends, Periodizitäten, Saisonalitäten, Ausreißern, Instationaritäten (z. B. der Varianz) sowie sonstiger Auffälligkeiten zu. Zukünftige Probleme können so bereits im Vorfeld aufgezeigt werden um sie im Rahmen z. B. des Capacity-Managements proaktiv zu vermeiden.

Eine weitere Form des Data Minings ist die Visuelle Datenexploration. Die Grundidee der visuellen Datenexploration ist die geeignete Darstellung der Daten in visueller Form, um einen Einblick in

die Struktur der Daten zu bekommen. Dies ermöglicht darüber hinaus Schlussfolgerungen aus den visualisierten Daten zu ziehen, sowie direkt mit den Daten zu interagieren und Hypothesen über die Daten aufzustellen. Dabei kann mit stark inhomogenen und verrauschten Daten gearbeitet werden und die Datenexploration kann auch durch Nicht-Spezialisten durchgeführt werden [Keim 2002].

Die visuelle Datenexploration lässt sich nach [Keim 2002] in 3 Schritte untergliedern: „Overview", „Zoom and Filter" und „Details-on-demand". Zunächst braucht das Überwachungspersonal einen Überblick über die Daten (overview). Das Personal kann interessante Muster in den Daten erkennen und sie anschließend mit Zoom- und Selektionstechniken (zoom and filter) genauer untersuchen. Für die genaue Analyse der Muster kann auf Details der Daten zugegriffen werden (detail-on-demand).

In der Abbildung 2 sind diesbezüglich exemplarisch einige methodischen Ansätzen dargestellt. Baumvisualisierungstechniken und hyperbolische Visualisierungstechniken kommen dabei häufig bei Netzwerk-Verkehrsanalysen zum Einsatz. Hier werden detaillierte Übersichten und Auswertung der Netzwerkkommunikation gefordert. Neben dem Netzwerkverkehr einzelner Komponenten und Verbindungen lassen sich auch Informationen über die Belastung einzelner Netzsegmente oder Muster der Kommunikationsbeziehungen ableiten. Für die Darstellung von Zuständen innerhalb von Verzeichnisstrukturen (Dateityp, Alter der Datei etc.) können die Sunburst oder auch die Beamtree-Technik verwendet werden. Sowohl Daten der Verkehrsanalyse, als auch Informationen zu Verzeichnisstrukturen können als Grundlage einer mittelfristigen Kapazitätsplanung eingesetzt werden und so einer Bedrohung der Verfügbarkeit durch Überlast entgegen wirken.

In der Leitlinie „**Integration verfügbarkeitsrelevanter Datenbasen"** wurde ein Konzept erarbeitet, das es ermöglicht, die verfügbarkeitsrelevanten Daten in einer Behörde, einem Unternehmen oder einer anderen Institution zu identifizieren, bezüglich ihrer Verfügbarkeits-relevanz zu analysieren und in einem zentralen Datenpool zu integrieren. Mit Hilfe dieser konsolidierten Datenbasis können unter Nutzung von o. g. Data-Mining- und Analyseverfahren die derzeitige Verfügbarkeitssituation und Prognosen für Ausfälle in der Zukunft aussagekräftig dargestellt werden. Diese Aussagen bilden dann eine wesentliche Voraussetzung für die zu treffenden Entscheidungen bzgl. der zu ergreifenden Maßnahmen.

3.2.3.3 Alarme und Meldungen

Auch im Bereich der Früherkennung müssen aufgrund von Erkenntnissen Alarme und Meldungen ausgelöst werden. Die Instandsetzung muss z. B. alarmiert werden, wenn mit kurzfristig eintretenden Fehlern zu rechnen ist. Hingegen sind Meldungen z. B. an Prozesse des IT-Managements zu erzeugen, wenn es sich um Erkenntnisse handelt, die längerfristiges planerisches Handeln erfordern (z. B. drohende Kapazitätsengpässe, saisonale Abhängigkeiten oder langfristige Trends).

3.2.4 Inspektion

Die [DIN 31051:2003-06] (vgl. Kapitel 4.2) versteht unter *Inspektion* Maßnahmen zur Feststellung und Beurteilung des IST-Zustandes einer Betrachtungseinheit einschließlich der Bestimmung der Ursachen der Abnutzung und dem Ableiten der notwendigen Konsequenzen für eine künftige Nutzung.

Im HV-Umfeld und insbesondere im Rahmen der Überwachung soll die Inspektion Phänomene (früh-) erkennen, die das Überwachungs-System nicht in der Lage ist zu erkennen. Weil bspw. für bestimmte Überwachungsbereiche keine Sensorik vorhanden ist oder es zu kostspielig oder sogar unmöglich ist, diese automatisch zu überwachen.

Die Inspektion ist im Bereich der Früherkennung angesiedelt. Aufgrund der relativ großen zeitlichen Abstände (geringe Abtastrate) zwischen den Inspektionen ist sie nicht dazu geeignet, Ereignisse kurzfristig nach deren Eintreten zu erkennen. Im Rahmen der Früherkennung trägt die Inspektion jedoch dazu bei, Störungen zu vermeiden.

Die Inspektion geschieht durch regelmäßige Untersuchungen, z. B. durch in Augenscheinnahme oder manuelle Messungen durch fachkundiges Personal. Veranlasst wird die Inspektion durch Ablauf eines Zeitraums oder durch Erreichen bestimmter Werte von Betriebsparameter wie Betriebsstunden, Energieverbrauch, Ladezyklen oder Laufleistungen.

Die o. g. Betriebsparameter können überwacht werden und bei Bedarf (Ereignis) kann durch den Überwachungs-Prozess „(Früh-) Erkennung“ die manuelle Inspektion veranlasst werden. Die Inspektion muss dokumentiert werden und das Ergebnis der Inspektion muss Überwachungs-Prozess „(Früh-) Erkennung“ zurückgemeldet werden. Die Überwachung entscheidet aufgrund der Ergebnisse der Inspektion, inwiefern Maßnahmen notwendig sind und veranlasst diese, z. B. durch Alamierung der Instandsetzung.

Im Gegensatz zur Definition der Inspektion in der [DIN 31051:2003-06] sollte im HV-Umfeld die „Ableitung der notwendigen Konsequenzen“ nicht durch die Inspektion erfolgen. Im HV-Umfeld können die Ursachen für Ereignisse so vielfältig sein und viele weitere IT-Management-Prozesse an deren Beseitigung beteiligt sein, dass eine zentrale Steuerung der Maßnahmen notwendig ist.

Werden durch die Inspektion Mängel festgestellt, die z. B. durch Maßnahmenvorgaben des Herstellers beseitigt werden können so können diese auch im Rahmen der Inspektion beseitigt werden.

Beispiele für Inspektionen sind:

- Säurefüllstand der USV-Akkumulatoren,
- Doppelboden des Rechenzentrums,
- Trafostation,
- Brandschottung und Kabeldurchführungen.

4 Reaktion

In diesem Kapitel werden die verschiedenartigen Möglichkeiten von Ereignisreaktionen im Überwachungs-Umfeld beleuchtet. Neben der Darstellung und Analyse verschiedener Reaktionsarten werden insbesondere auch Reaktionen betrachtet, die im Rahmen einer Instandsetzung, Wartung oder Steuerung Rückkopplungen auf die überwachte HV-Architektur realisieren. In jedem Fall sind im Vorfeld Prozesse aufzusetzen, die sich an etablierten generischen Prozessmodellen wie ITIL oder CobiT orientieren. Damit wird gewährleistet, dass Reaktionen im Rahmen eines Steuerkreises geordnet und durch das Incident Management getriggert ablaufen (vgl. HV-Kompendium Organisation [BSI HV-Kom2009]).

4.1 Reaktionsarten

Im HV-Umfeld verlangen Alarme des Überwachungs-Systems eine sehr kurzfristige Reaktion. Diese Reaktion kann autonom, automatisiert oder manuell erfolgen. Redundanz und Fail-Over-Mechanismen auf Komponentenebene ermöglichen quasi-unterbrechungsfreie Services. Im Folgenden werden auch die damit verbundenen auf Systemebene realisierten Methoden der Ereignisreaktion und deren Steuerung im Rahmen eines IT-Service-Managements betrachtet.

4.1.1 Autonome Reaktionen

Autonomie im Umfeld der Hochverfügbarkeit kann als Eigenständigkeit bei der Erfüllung der Verfügbarkeitsanforderungen verstanden werden (HV-Autonomie). Während die Zielvorgaben für das Design und die Ausgestaltung der Architektur durch die geforderte Verfügbarkeitsklasse von außen vorgegeben werden, erfüllen die Teilkomponenten diese Aufgabe autonom und transparent für den Anwender (vgl. Prinzipien der HV-Autonomie [BSI HV-Kom2009]).

Eine Eigenschaft autonomer Systeme ist die so genannten Selbststabilisierung. Das System führt eine eigenständige, automatisierte Behebung von Fehlerzuständen durch. In diesem Fall sind keine weiteren Reaktionen durch das Überwachungspersonal oder der Instandsetzung notwendig.

Zurzeit sind autonome Systeme im obigen Sinne nur als Gegenstand der Forschung oder als Prototypen zu finden. Häufig dennoch als autonome Systeme bezeichnete Systeme beinhalten jedoch nur automatische Reaktionen. Die Selbststabilisierung ist dann nicht vollumfänglich implementiert.

Im weitesten Sinne kann in IT-Diensten in dienstorientierten Architekturen (SOA) (vgl. IT-Dienste und Dienstleistungen [BSI HV-Kom2009] sowie [BSI SOA2010]) von autonomen Reaktionen gesprochen werden. Die Reaktionen sind aus der Sicht des Dienstnehmers autonom. Solange die SLAs erfüllt werden, ist es für ihn wenig von Interesse, wie der Dienstleister die Autonomie realisiert, auch dann, wenn die Reaktionen des Dienstleisters automatisch oder manuell erfolgen.

4.1.2 Automatische Reaktionen

Automatische Reaktionen können sowohl auf der Komponenten-Ebene als auch auf der Überwachungs-Ebene stattfinden.

Auf der Komponentenebene überwachen sich die Komponenten häufig mittels eigener Methoden selbst und stellen bspw. fest, dass eine Komponente ausgefallen ist und übernehmen deren Aufgaben. Hierzu zählen Fail-Over-Mechanismen (vgl. Cluster-Architekturen, Prinzipien der HV - Ferhlertoleranz [BSI HV-Kom2009]) oder Re-Routing-Mechanismen im Bereich des Routings in Netzwerken (vgl. HV-Kompendium Netzwerk [BSI HV-Kom2009]).

Klassische Beispiele für automatische Reaktionen auf der Überwachungs-Ebene sind das automatisierte Ausführen von Skripten, das automatisierte Rebooten von Systemen oder der automatisierte „Umzug“ von virtualisierten Systemen auf andere Host-Systeme. Diese Reaktionen bedürfen keinem manuellem Eingreifen, sie werden vielmehr durch Erkenntnisse des Überwachungs-Prozesses „(Früh-) Erkennung“ automatisch ausgelöst.

Automatisierte Fehlerreaktionen sind insbesondere im HV-Umfeld sorgfältig zu planen und umzusetzen, da diese innerhalb des komplexen HV-Umfeldes zu weiteren Schäden oder Beeinträchtigungen führen können. Das Potenzial automatisierter Fehlerbehebung wird in der Regel bereits durch das Design der HV-Komponenten und deren Mechanismen zur Fehlerbehebung (z. B. Fail-Over, Clustering, Load-Balancing) ausgeschöpft. Die Wirksamkeit der Überwachung und Fehlerreaktion ist daher heutzutage noch in einem großen Maße von der menschlichen Kompetenz des Administrations- und Überwachungspersonals abhängig. Wie hier bei der Erläuterung der autonomen und vollautomatischen Fehlerreaktion deutlich wurde geht jedoch der Trend eindeutig zu sich selbst steuernden IT-Infrastrukturen. Dies ist heutzutage nur in Teilbereichen implementiert. Bis die jeweiligen Konzepte und Techniken in ca. 5 bis 10 Jahren die Praxis vollständig durchdringen werden, muss das IT-Personal weiterhin einen entscheidenden Beitrag leisten, komplexe Fehlerursachen zu überblicken. Auf Ausfälle, die durch entsprechend implementierte Automatismen abgefangen werden, muss reagiert werden, um weiterhin die Verfügbarkeit zu gewährleisten. Darüber hinaus wird es auch auf unabsehbare Zeit nötig sein, den Faktor Mensch als übergeordnete Kontrollinstanz zu erhalten.

4.1.3 Manuelle Reaktionen

Manuelles Eingreifen ist immer dann erforderlich, wenn Ereignisse aufgetreten sind, die sich durch automatische Reaktionen aufgrund der Fehlerart nicht korrigieren lassen. Hierzu zählen u. a. Totalausfälle von Systemen aufgrund von Hardware- oder Software-Problemen. Zu beachten ist, dass selbst bei automatischen Fail-Over-Mechanismen zur Wiederherstellung der Verfügbarkeit manuelle Reaktionen notwendig sind. Hat z. B. der Ausfall einer Komponente ein Fail-Over ausgelöst, so muss das Überwachungs-System einen entsprechenden Alarm generieren. Aufgrund des Fail-Over-Mechanismus ist zwar die Serviceerbringung in der Regel nicht, oder nur geringfügig beeinträchtigt, allerdings muss zur Wahrung der „Hochverfügbarkeit“ die ausgefallene Komponente kurzfristig wieder Instand gesetzt werden (manuelle Reaktion). Der so ausgelöste Alarm muss durch einen Prozess der Instandsetzung aufgefangen werden und adäquate Reaktionen müssen erfolgen.

Werden bestimmte Systemzustände erkannt, wie z. B. der Ausfall einer Komponente oder das Über- oder Unterschreiten von Grenzwerten, werden durch den Überwachungs-Prozess „(Früh-) Erkennung“ Alarme generiert und ein direktes manuelles Eingreifen initiiert. Gewöhnlich können die Ereignisse durch das Überwachungs-System kategorisiert und hinsichtlich ihrer Dringlichkeit bewertet werden. Das Überwachungspersonal fängt die Alarme auf und eskaliert sie aufgrund entsprechender Alarmierungspläne. Das Überwachungs-System kann darüber hinaus

entsprechendes Administrations- oder Instandsetzungspersonal direkt alarmieren und mittels Telefon oder Pager Nachrichten hinsichtlich des Ereignisses übermitteln.

Insbesondere im HV-Umfeld muss die manuelle Reaktion zeitnah erfolgen. Dies bedeutet, dass das Instandsetzungspersonal (vgl. Kapitel 4.2) mit entsprechender Fachkunde und in ausreichender Anzahl zur Verfügung stehen muss. Darüber hinaus sind entsprechende Melde- und Eskalationswege zu definieren (vgl. HV-Kompendium Organisation [BSI HV-Kom2009]). Die Abwicklung der manuellen Reaktion sollte auf Basis eines gemanagten und optimierten Prozesses erfolgen. Als Grundlage für die Prozessdefinition kann ein geeignetes Prozessmodell wie z. B. CobiT und ITIL herangezogen werden.

Manuelle Reaktionen erfolgen in der Regel auch als Reaktion auf ein im Überwachungsbereich Personal und Organisation festgestelltes Ereignis, da diese die Steuerung der Organisation selber betreffen.

4.2 Instandsetzung

Die Instandsetzung ist eine manuelle Reaktion und wird im Kontext der Instandhaltung gesehen. Die Definition der Instandhaltung gemäß der [DIN 31051:2003-06] lässt sich auch im IT-Umfeld und insbesondere im HV-Umfeld anwenden. In diesem Kapitel werden die Bezüge der einzelnen Aspekte der Instandhaltung zur Überwachung und darüber hinaus auch zur HV-Organisation hergestellt.

Die Instandhaltung umfasst gemäß der [DIN 31051:2003-06] alle *„Maßnahmen zur Bewahrung und Wiederherstellung des Soll-Zustandes sowie zur Feststellung und Beurteilung des Ist-Zustandes von technischen Mitteln eines Systems“*.

Diese Maßnahmen werden seitens der [DIN 31051:2003-06] in vier Grundmaßnahmen untergliedert:

- Instandsetzung
- Inspektion
- Wartung
- Schwachstellenbeseitigung oder auch Verbesserung

Im HV-Umfeld wird nicht nur ein vermehrtes Augenmerk auf die kurzfristige Wiederherstellung (Instandsetzung) gelegt, sondern vielmehr spielen die weiteren Grundmaßnahmen der Instandhaltung, die Inspektion, die Wartung und die Schwachstellenbeseitigung eine wesentliche Rolle. Die Grundmaßnahmen Inspektion, Wartung und Schwachstellenbeseitigung dienen im HV-Umfeld zur Vermeidung von Ausfällen, damit zur Erhöhung der Verfügbarkeit. Die Inspektion ist aufgrund ihrer rein erkennenden Funktion im Kapitel (Früh-) Erkennung (vgl. Kapitel Fehler: Referenz nicht gefunden) näher betrachtet, die Wartung und Schwachstellenbeseitigung in diesem Kapitel.

Unter *Instandsetzung* werden nach der [DIN 31051:2003-06] Maßnahmen zur Rückführung einer Betrachtungseinheit in den funktionsfähigen Zustand, mit Ausnahme von Verbesserungen, verstanden. Gemeint ist die Wiederherstellung des SOLL-Zustands z. B. mittels Reparatur.

In der Regel wird die Instandsetzung durch den Überwachungs-Prozess „(Früh-) Erkennung“ ausgelöst (vgl. Kapitel 3.2.2.3). Gerade im HV-Umfeld wird von der Instandsetzung erwartet, den SOLL-Zustand eines IT-Systems schnellstmöglich wieder herzustellen. Um dieses zu realisieren, muss die Instandsetzung hinsichtlich des Personals und der Prozesse organisiert werden.

4.2.1 Personalaspekte

Zur schnellen Reaktion muss zeitnah qualifiziertes Personal in ausreichender Anzahl zur Verfügung stehen. Im Rahmen eines Personalkonzeptes müssen Aspekte wie Personalbeschaffung und Aus- und Weiterbildung geklärt werden (vgl. HV-Kompendium Organisation [BSI HV-Kom2009]). In diesem Konzept sind als Parameter die Anforderungen der HV zu berücksichtigen. Für einen „7*24“ Betrieb bspw. muss genügend Personal zur Verfügung stehen, dass ein entsprechender Dienstplan erstellt werden kann.

In vielen Fällen kann die Instandsetzung durch eigenes Personal wie z. B. durch Administratoren erfolgen. In einigen Fällen muss auf externes Personal wie z. B. Service Personal eines Herstellers oder eines Dienstleiters zurückgegriffen werden. Um den kurzfristigen Zugriff auf externes Personal zu erhalten, sollten im Vorfeld mit den Unternehmen entsprechenden Vereinbarungen in Form von SLAs getroffen werden.

4.2.2 Organisationsaspekte

Um den HV-Anforderungen gerecht zu werden müssen im Rahmen der Instandsetzung organisatorische Prozesse und Zuständigkeiten definiert werden, die im Fall einer Alarmierung eine geeignete Reaktion, idealerweise im Sinne eines ITIL „IT-Service Continuity Managements“ oder CobiT „ServiceDesk & IncidentManagement“, ermöglichen. Hierbei sind die folgenden Fragestellungen zu berücksichtigen:

- Wer nimmt die Alarmierung entgegen?
- Wer führt die notwendigen Arbeiten aus (Hardware, Software, Infrastruktur etc)?
- Wie erfolgt die Anreise?
- Wer entscheidet über Reparatur oder Ersatzbeschaffung?
- Wer veranlasst die Ersatzbeschaffung?
- Wer erstellt wann einen Lagebericht (Erfolg oder Misserfolg der Instandsetzung)?
- An wen wird der Lagebericht gemeldet?
- Wer protokolliert und dokumentiert den Vorgang?

Darüber hinaus müssen bei der Erstellung von Dienst- und Einsatzplänen, die Aspekte wie ausreichende Anzahl und Qualifikation des Personals, Vertretungsregelungen, Anreise- und Rüstzeiten sowie Melde- und Eskalationswege berücksichtigt werden.

In vielen Fällen erfolgt die Reparatur durch den Austausch von Komponenten. Dazu muss eine Ersatzbeschaffung organisiert werden. Für kritische Komponenten (z. B. SPoF-Komponenten) müssen Bevorratungen getroffen werden, so dass sie in kurzer Zeit ausgetauscht werden können. Weniger kritische Komponenten müssen mittels zuvor mit Lieferanten geschlossenen SLAs kurzfristig beschafft werden können.

4.3 Wartung

Unter *Wartung* versteht die [DIN 31051:2003-06] (vgl. Kapitel 4.2) Maßnahmen zur Verzögerung des Abbaus des vorhandenen Abnutzungsvorrates der Betrachtungseinheit.

Die Wartung im HV-Umfeld dient in erster Linie zur Erhöhung der Zuverlässigkeit von Komponenten und damit zur Erhöhung der Gesamtverfügbarkeit eines IT-Systems. Darüber hinaus kann durch die Wartung eine Reduzierung der Instandhaltungskosten erzielt werden.

Die Wartung umfasst z. B. Nachstellen, Schmieren, funktionserhaltendes Reinigen, Konservieren, Nachfüllen oder Ersetzen von Betriebsstoffen oder Verbrauchsmitteln (z. B. Kraftstoff, Schmierstoff oder Wasser) und planmäßiges Austauschen von Verschleißteilen (z. B. Filter oder Dichtungen) oder ganzer Komponenten (z. B. Festplatten oder USV-Batterien) im Rahmen der manuellen Reaktion.

Auch die eingesetzte Software muss gewartet werden. Dazu zählt das Einspielen von Up-Dates, Up-Grades oder Patches (vgl. ITIL Change Management Prozess [ITIL-ServTrans2007]).

Die Wartungs-Intervalle orientieren sich meistens an den Herstellerangaben oder -empfehlungen. Darüber hinaus kann die Überwachung aber auch aufgrund von eigenen Erfahrungen mit bestimmten Komponenten weitergehende Wartungsarbeiten veranlassen oder die Intervalle verkürzen oder verlängern.

Der Prozess „(Früh-) Erkennung“ der Überwachung überwacht die Wartungs-Intervalle und stößt die erforderlichen Wartungsarbeiten an. Bei den Ereignissen handelt es sich in der Regel nicht um Fehler oder Störungen. Vielmehr werden anhand eines Wartungsplans die Wartungsarbeiten unabhängig von einem Defekt oder einem Ausfall durchgeführt. In diesem Fall werden Zeiträume überwacht.

Der Ersatz von defekten Teilen gehört zur Instandsetzung und nicht zur Wartung. Kleinere Defekte können jedoch häufig im Zuge von regelmäßigen Wartungsarbeiten behoben werden (sog. kleine Instandsetzung).

Beispiele für Wartung sind:

- Klimatisierung,
- Netzersatzanlage,
- Löschanlage,
- Umspanneinrichtungen,
- Rechner-Reinigung (insb. Reinigung der Lüfter),
- Sicherheits-Patches für Web-Server,
- Schwachstellenbeseitigung.

Laut [DIN 31051:2003-06] unterstützen die Maßnahmen der *Schwachstellenbeseitigung* die technische Betrachtungseinheit in der Weise, dass das Erreichen einer festgelegten Abnutzungsgrenze nur noch mit einer Wahrscheinlichkeit zu erwarten ist, die im Rahmen der geforderten Verfügbarkeit liegt.

Die Schwachstellenbeseitigung ist im HV-Umfeld eine wesentliche Maßnahme zur Erhöhung der Verfügbarkeit. Die Schwachstellen können von unterschiedlicher Natur sein. Zum einen kann die eingesetzte Komponente aufgrund ihrer Unzuverlässigkeit eine Schwachstelle darstellen. Zum anderen kann die HV-Konzeption mit ihrer IT-Architektur eine Schwachstelle sein. Die hier genannte Schwachstellenbeseitigung im Rahmen der Instandhaltung hat eine „technische Betrachtungseinheit“, also in der Regel eine Komponente im Fokus. Gab es z. B. bei einem Rechnertyp in der Vergangenheit häufig Ausfälle des Netzteils, so wird man mittels der Schwachstellenbeseitigung diesen Rechnertyp austauschen. Schwachstellen auf konzeptioneller Ebene können durch die Schwachstellenbeseitigung im Rahmen der Instandhaltung meistens nicht erkannt und beseitigt werden.

Die Schwachstellenbeseitigung im HV-Umfeld sollte in IT-Management-Prozessen (z. B. ITIL Problem- und Change-Management, vgl. Kapitel 5.6) integriert sein. Durch die Integration in Management-Prozesse ist ein koordiniertes Handeln möglich, das die gesamte Verfügbarkeit des Systems erhöht und auch Schwachstellen im Design und in der Architektur erkennt und beseitigt.

4.4 Steuerung

Die Steuerung als eine Art der Reaktion führt eine Rückkopplung auf die überwachten IT-Komponenten durch. Die Rückkopplung kann automatisiert oder manuell innerhalb existierender Betriebstoleranzen erfolgen. Um diese Reaktion zu realisieren, müssen bereits während des Betriebes ausreichende Redundanzen vorhanden sein.

Die in den vorhergehenden Abschnitten beschriebenen Überwachungs-Prozesse zur Überwachung, (Früh-) Erkennung und Reaktion und der konsolidierten Beobachtung der System- und Service-Parameter einer HV-Infrastruktur stellen einen unverzichtbaren Bestandteil der Planung, Realisierung und des Betriebes hoch verfügbarer IT-Strukturen dar. Nur eine leistungsfähige Überwachung gewährleistet die dauerhafte Aufrechterhaltung der definierten und benötigten Service Qualität. Für die Aufrechterhaltung und die fortlaufende Optimierung bzw. Anpassung der Überwachungs-Prozesse an veränderte Anforderungen (z. B. IT-Governance, Compliance, Performance) bedarf es geeigneter IT-Management-Prozesse zur Überwachung der IT-Performance. Basis dafür bilden bspw. die CobiT-Prozesse ME1-4 [CobiT4.0] in den höheren Reifegradstufen.

Dem Nutzen der Überwachung sind aber auch technische und organisatorische Grenzen gesetzt, die teilweise durch ergänzende Technologien oder nachgelagerte Prozesse ausgeglichen werden müssen. Darüber hinaus existieren prinzipbedingte Einschränkungen, die im Rahmen einer Restrisikobetrachtung in die HV-Konzeption einfließen müssen und Inhalte des Business Continuity Management darstellen. Diese Aspekte sollen in den nachfolgenden Abschnitten zusammenfassend betrachtet werden.

5 Weiterführende Aspekte

Das vorliegende Kapitel betrachtet spezifische Überwachungsbereiche sowie weiterführende Aspekte der Überwachung im technischen und organisatorischen Bereich.

5.1 Intrusion Detection Systeme in der Überwachung

Intrusion Detection Systeme (IDS) weisen in weiten Teilen Parallelen zu Überwachungs-Architekturen auf. Auch Intrusion Detection Systeme erfassen unter Verwendung von Sensoren bestimmte Messgrößen und überwachen Systemparameter. Sie führen darüber hinaus ebenfalls eine Auswertung im Sinne einer (Früh-) Erkennung durch und leiten automatisierte oder manuelle Reaktionen ein. Der Fokus der IDS liegt jedoch nur indirekt auf der Sicherstellung der Verfügbarkeit von IT-Systemen und IT-Komponenten bzw. Netzwerken, vielmehr besteht die Aufgabe eines IDS darin, Angriffe auf ein Netzwerk oder IT-System zu erkennen. Dazu verwenden sie spezialisierte Mechanismen, Messgrößen und Auswerteverfahren zur Erkennung von Anomalien oder Angriffsversuchen durch Auswertungen, Mustervergleiche oder die Anwendung spezifischer Algorithmen. Intrusion Detection Systeme unterstützen so die Überwachung und Einhaltung von Service-Parametern hinsichtlich der Vertraulichkeit und Integrität (vgl. [BSI IDS2002]). Darüber hinaus können Sie bei Bedarf Angriffe erkennen, durch Alarme melden oder durch die teilweise automatisierte Einleitung von geeigneten Gegenmaßnahmen abwehren oder sogar verhindern.

Eindringversuche in Netzwerke oder Systeme sowie Angriffe von innen oder außen können ebenfalls die Verfügbarkeit erheblich beeinträchtigen oder sogar zu einem Verfügbarkeitsverlust führen, dem durch die Mechanismen der in diesem Dokument beschriebenen Überwachungs-Prinzipien nicht oder nur unzureichend begegnet werden kann. Intrusion Detection Systeme können eine wertvolle Ergänzung und Unterstützung klassischer Überwachungs-Systeme darstellen, da über die vom Überwachungs-System erfassten technisch bedingten Ereignisse hinaus auch durch Angriffe herbeigeführte Beeinträchtigungen erkannt oder verhindert werden. IDS können klassische Überwachungs-Systeme daher nicht ersetzten, sie aber insbesondere im HV-Umfeld sinnvoll ergänzen.

5.2 Überwachung in Storage Area Networks

Ein Storage Area Network (SAN) stellt eine zentrale Instanz zur Datenspeicherung zur Verfügung und ist strukturell analog zu einem Lokal Area Network (LAN) aufgebaut. Es beinhaltet ebenso wie ein LAN Komponenten, die die Funktion von Hubs, Switchen oder Routern erbringen und verknüpft Festplattensubsysteme mit Server-Systemen mittels breitbandigen Netzwerk-verbindungen. Darüber hinaus erlaubt ein SAN, Speichersysteme physikalisch getrennt von Server-Systemen und über große Distanzen zu betreiben (vgl. HV-Kompendium Kapitel Speichertechnologien [BSI HV-Kom2009]).

Grundsätzlich können aufgrund der gleichartigen Struktur in einem SAN die gleichen Messgrößen wie für die bisher betrachteten IT-Komponenten wie auch für das Netzwerk erfasst, übermittelt und ausgewertet werden. Aufgrund der spezifischen Besonderheiten der SAN-Komponenten sowie der SAN-Architektur ist aber eine spezifische Betrachtung des SAN-Überwachung erforderlich. So sollte auf den im Rahmen der Verkehrsanalyse in LANs üblichen Zugriff auf Daten mittels

Mirrorports an den Switches innerhalb eines SAN verzichtet werden, da der gespiegelte Port bei Volllast des Swichtes nicht in jedem Fall die vollständigen Informationen des Fiber Channel Protokolls enthält. Statt dessen sollten im SAN aufgrund der dort üblichen Fiber Channel Technik sog. Fiberoptic-TAPs (vgl. Kapitel 2.3.2) zur Verkehrsmessung eingesetzt werden. Diese speziellen TAPs teilen das eingehende Licht mittels Prismen in zwei Strahl-Systeme auf. Der eine Strahl wird hierbei ungehindert weiter geleitet, während der zweite Strahl einer Analyseeinheit zugeführt werden kann. Durch den Einsatz der Fiberoptic-TAPs wird verhindert, dass der Lichtstrahl des Fiber Channel unterbrochen und bei einem Ausfall eines TAP's oder einer Analyseeinheit der Datenstrom beeinträchtigt oder gar unterbrochen wird.

Die in den Analyseeinheiten gewonnene Informationen können über ein LAN oder ein separates Überwachungsnetz mittels SNMP zu einer zentralen Überwachungs-Infrastruktur übertragen werden. Das SAN-Fabric ist in der Regel nicht darauf ausgerichtet, Überwachungsdaten zu transportieren, die Daten der Überwachung der SAN-Komponenten werden stattdessen in der in Kapitel 2.1.3 beschrieben Form übermittelt.

Zu beachten ist, dass in SAN-Architekturen aus Performance-Gründen häufig eine Class 3-Kommunikation eingesetzt wird. Bei dieser Kommunikationsart findet keine Bestätigung der Übermittlung statt. Komponenten wie Speichersysteme, HBAs oder Switches können im SAN bei dieser Kommunikationsart Datenverluste weder erkennen noch verhindern oder beheben. Die Erkennung findet auf Betriebssystem-Ebene in den Hosts statt und führt in der Regel dazu, dass Daten erneut gesendet oder gelesen werden müssen. Aus diesem Grund sollte die SAN-Überwachung auf die SCSI-Ebene der Host-Systeme erweitert werden.

Häufig beinhalten SAN-Management-Systeme bereits spezielle Überwachungs-Komponenten. Diese Software gestattet es, SAN-spezifische Messgrößen (z. B. SAN-Bandbreitennutzung pro Endgerätekommunikation, Reaktionszeiten oder Lastverteilung) zu erfassen und zu überwachen, die im LAN oder WAN nicht vorkommen und mit den dort üblichen Überwachungs-Lösungen nicht erfasst werden.

Aufgrund der in der Regel zentralen Bedeutung eines SANs für die HV-Architektur und der oben aufgeführten Besonderheiten kommt der Planung der Überwachung im SAN, insbesondere vor dem Hintergrund der weiteren Aspekte der Informationssicherheit besondere Bedeutung zu (vgl. M 2.353 Erstellung einer Sicherheitsrichtlinie für SAN-Systeme [BSI GS-Kat2009]).

5.3 Überwachung in virtueller Umgebung

Die Überwachung in virtuellen Umgebungen muss sowohl aus physikalischer als auch aus virtueller Sicht erfolgen. Zum einen müssen die physikalischen Host-Systeme wie im Kapitel 2.1 beschrieben überwacht werden, zum anderen muss eine Überwachung der virtuellen Systeme im Rahmen der Software-Überwachung (vgl. Kapitel 2.2) realisiert werden. Darüber hinaus ergeben sich aus Überwachungs-Sicht noch weitere Messgrößen, die sich aus der speziellen Kombination von „Software“-Maschinen auf „Hardware“-Maschinen ergeben. Die Überwachungsdaten aus der physikalischen und der virtuellen Sicht müssen bei der Überwachung virtueller Systeme korreliert werden. Nur dies erlaubt Aussagen über die Auslastung oder mögliche Engpässe zu treffen.

In einer virtuellen Umgebung ist z. B. die Angabe der CPU-Auslastung des Host-Systems für die Beurteilung der virtuellen Maschine wenig aussagekräftig. Um CPU-Engpässe zu erkennen, sind

Zusatzinformationen notwendig. Entscheidend ist vielmehr, wie viele physikalische CPU-Ressourcen vom „Scheduler“ zu einem Zeitpunkt der virtuellen Maschine zugeteilt waren. Darüber hinaus sind Informationen über Wartezustände ein Anzeichen für CPU-Engpässe. Die „CPU-Ready time“, also die Zeit, die eine virtuelle Maschine im „Ready-to-run“ Status warten muss, bevor sie CPU-Ressourcen zugeteilt bekommt, ist ein wichtiges Indiz dafür. Die CPU-Ready time hat einen direkten Einfluss auf die Antwortzeit eines Services. Viele Hersteller von Virtualisierungs-Software stellen Mechanismen zur Verfügung, die es erlauben, so genannten „Scheduler trace“ Daten zu analysieren. Diese Daten geben Aufschluss darüber, wie der „Scheduler“ im Zusammenhang mit den Statusinformationen der Prozesse oder der CPU Zuteilungsentscheidungen getroffen hat.

Häufig kommt es in virtuellen Umgebungen durch Mehrfach- oder sogar Überbelegung des Arbeitsspeichers zu Engpässen. In diesen Fällen werden die virtuellen Maschinen aufgefordert, unbenutzten oder nicht mehr benötigten Speicher freizugeben. Reicht die damit erzielte Wirkung nicht aus, werden die virtuellen Maschinen aufgefordert, Prozesse mit hoher Speichernutzung zu beenden. Dieser Mechanismus wird „Ballooning“ genannt. Im HV-Umfeld muss verhindert werden, dass aufgrund von Speicherengpässen Prozesse beendet werden. Ballooning kann nur im Nachhinein mittels Auswertung von durch Balloning-Technik freigegebenen Hauptspeicher festgestellt werden und zur Früherkennung von weiteren Engpässen genutzt werden. Ballooning muss auf jeden Fall als ein Hinweis auf Speicherengpässe gedeutet werden. Darüber hinaus ist es sinnvoll zu kontrollieren, wie viel von dem für die virtuellen Maschinen allozierten Hauptspeicher tatsächlich genutzt wird. Nur so lässt sich im Rahmen der Früherkennung sagen, wann und wo ein Ressourcenengpass entstehen wird, wann der Speicher auf physischer oder virtueller Ebene ausgeschöpft sein wird und auf welche Services sich der Engpass auswirken wird.

Der Einsatz virtueller Systeme bietet im Überwachungs-Umfeld Techniken, die eine automatisierte Reaktion beispielsweise durch „rapid virtual machine cloning or migration“ ermöglichen.

5.4 WAN-Überwachung

Die Überwachung über ein WAN unterliegt einigen Besonderheiten, die in diesem Kapitel betrachtet werden sollen.

Häufig werden WANs mittels VPN-Lösungen in Paket-vermittelten Netzen realisiert. In diesem Fall stehen die IT-Komponenten des Transportnetzes nicht für eine direkte Überwachung zur Verfügung. Die Überwachung der Komponenten ist nur an den Tunnelendpunkten durch Überwachung der VPN-Gateways möglich. Ebenso sollte die VPN-Performance sinnvollerweise auch an den Endpunkten überwacht werden.

Häufig stellen Netzbetreiber von so genannten „leased lines“ ihre Überwachungsdaten in konsolidierter Form den Nutzern zur Verfügung. Reaktionen sind nur dahin gehend möglich, den Netzbetreiber zu benachrichtigen oder wenn möglich, eine alternative Verbindung zu nutzen. Maßnahmen wie die der Instandsetzung oder der Rekonfiguration sind innerhalb dieses Netzes nicht möglich. Bei einem Verlust der Verfügbarkeit dieses Netzes kann nur auf Ausweichverbindungen (z. B. Border Gateway Protocol vgl. HV-Kompendium Band 2, Kapitel 3 Netzwerke [BSI HV-Kom2012]) ausgewichen werden. Im Bereich der Hochverfügbarkeit müssen mit dem Anbieter von Weitverkehrsnetzen vertragliche Vereinbarungen in Form von *Service Level Agreements* (SLA) getroffen werden, die nicht nur eine geforderte Performance garantieren, sondern auch eine schnelle

und geeignete Reaktion bei Ereignissen. Die WAN-Überwachung erlaubt in diesem Fall die Überwachung der Einhaltung der vereinbarten Service-Parameter.

Häufig ist aufgrund der großen Distanzen in Weitverkehrsnetzen ein separates Überwachungs-Netz aus Kostengründen nicht möglich. Die Überwachungsdaten werden somit zusätzlich zu den Nutzdaten über das WAN übermittelt. Dieses führt zu einer weiteren Belastung des WANs. Ferner stehen bei einem Leitungs- oder Komponentenausfall im WAN keine Überwachungsdaten der restlichen Leitungen und Komponenten mehr zu Verfügung. Auch die steuernde Reaktion ist in diesem Fall nur noch bis zum letzten erreichbaren Netzknoten möglich.

Kapazitätsengpässe werden aufgrund der daraus folgenden verzögerten Übermittlung unter Umständen spät erkannt. Dieses erschwert zusätzlich die Fehlereingrenzung oder -suche. Im Rahmen der Instandsetzung muss, insbesondere dann, wenn bei den Netzknoten kein Vorort-Personal vorhanden ist, mit erheblichen Anfahrt- und Rüstzeiten gerechnet werden.

5.5 Zentrales Überwachungs-System

Die Auswertung der durch Überwachungs-Sensoren bereitgestellten Daten ist umso leistungsfähiger, je mehr Informationen aus der HV-Umgebung konzentriert zur Verfügung stehen.

Erst durch die Konzentration der Überwachungsdaten an einem (logischen) Punkt ist es möglich, eine leistungsfähige Früherkennung oder sogar Data-Mining zu realisieren sowie die vielfältigen Reaktionen koordiniert zu steuern.

Es bietet sich deshalb ein zentraler *Log- oder Monitorserver* an, auf dem alle anfallenden Daten gespeichert und konsolidiert werden (vgl. Abbildung 2). Werden auf diesem System auch noch zusätzlich Aufgaben des Systemmanagements mit übernommen, so wird von einem *Network and Systems Management Server* gesprochen.

Ein Network and Systems Management Server soll eine Reihe von Aufgaben erfüllen:

- Entgegennahme von Überwachungsdaten,
- sichere Archivierung der Überwachungsdaten,
- übersichtliche Darstellung der aktuellen Situation und ggf. des Systemstatus,
- Erkennung und Visualisierung von Trends,
- Auswertung von Überwachungsdaten z. B. anhand von Mustern und Regeln,
- automatische Alarmierung bei besonderen Ereignissen.

Dabei ist deutlich, dass ein zentralisierter Architekturansatz verschiedene Nachteile birgt:

- Eine zentrale Managementplattform stellt einen SPoF dar, der aber mittels redundanter Auslegung vermieden werden kann.
- Durch die zumeist hohe Anzahl von Daten zuleitenden Agenten und der z. T. fehlenden Datenverdichtung kann es ggf. zu sehr hohen Netzlasten kommen, welches möglicherweise eine Erhöhung der Bandbreiten erforderlich macht.

5.6 Überwachung und ITIL

Überwachung und ITIL stehen in einem engen Zusammenhang. Beispielsweise ist ein Service Operation Prozess [ITIL-ServOp2007] definiert, der Events erkennt, diese einordnet und geeignete Maßnahmen festlegt. Events sind in diesem Fall Benachrichtigungen, die von einem Überwachungs-Tool erzeugt werden. Dieser Prozess trägt in ITIL die Bezeichnung *Event Management*. Ein Event ist in ITIL definiert als jedes erkennbare Auftreten, das für das Management der IT-Infrastruktur oder die Bereitstellung des IT Service sowie für die Bewertung der Auswirkung einer Serviceabweichung von Bedeutung ist. Das Event Management stellt in erster Linie Mechanismen zur frühzeitigen Erkennung von Incidents bereit. Ein Event in ITIL ist das Analogon zu einem Ereignis im Rahmen der Überwachung, das zur Früherkennung herangezogen wird.

Incidents sind in ITIL nicht geplante Unterbrechungen eines IT Service oder eine Qualitätsminderung eines IT Service. Der Service Operation Prozess *Incident Management* [ITIL-ServOp2007] ist in ITIL für alle Incidents verantwortlich. Wichtigstes Ziel dieses Prozesses ist die möglichst schnelle Wiederherstellung des normalen Servicebetriebs. Incident Management ist in diesem Fall ein übergeordneter Prozess, der die Überwachung als Tool zur Erkennung von Incidents nutzt, selbst aber auch für die Behandlung von Incidents im Rahmen einer Reaktion sorgt. Ein Incident ist das Analogon zu einem Ereignis in Form eines Fehlers im Überwachungs-Umfeld. Der Prozess Incident Management umfasst die folgenden Prozessaktivitäten:

- Identifizierung
- Erfassung
- Kategorisierung
- Priorisierung
- Eskalation
- Untersuchung und Diagnose
- Lösung und Wiederherstellung
- Abschluss

Außerhalb des Regelbetriebs der HV-Systeme auftretende, sich zu einer Katastrophe entwickelnde Incidents, werden in ITIL im Rahmen des IT Service Continuity Managements behandelt und nicht in der Überwachung. Überwachung kann ITIL als Tool zur Erkennung von Major Incidents und Katastrophen dienen.

Zur Fehlerdiagnostik und Problemlösung können weitere Informationen aus einer zentralen Konfigurations-Datenbank bei Bedarf angezeigt werden. Z. B. bildet im *ITIL-Framework* das *Configuration Management* in einer Meta-Datenbank *Configuration Management Database (CMDB)* ein logisches Modell der IT-Infrastruktur ab (vgl. HV-Kompendium Organisation [BSI HV-Kom2009] und [ITIL-ServTrans2007]). In Form von in Relation zueinander stehenden Configuration Items werden alle relevanten Informationen zur Hardware, Software, Netzwerk, Versionsstände, Dokumentation, Verantwortliche, SLAs, Verfahrensanweisungen etc. erfasst.

5.7 Nebenwirkungen und Grenzen der Überwachung

Eine allumfassende Überwachung aller Systemparameter ist in der Regel nicht realisierbar, da mit steigendem Umfang und Tiefe der Erfassung die Beeinträchtigung der zu überwachenden Systeme zunimmt („probe-effect“). Einerseits findet durch das Einbringen von Sensoren in die Zielsysteme ein Eingriff in diese Systeme statt, der das Systemverhalten und auch die Verfügbarkeiten beeinflusst. Die Sensoren benötigen Ressourcen, die dem Zielsystem im Bedarfsfall nicht mehr zur Verfügung stehen. Andererseits stören durch das Überwachungs-System durchgeführte aktive Messungen (z. B. Live-Requests, Pings, Tests von Antwortzeiten) das Verhalten und die Performance der überwachten Infrastruktur. Bei der Planung und Realisierung von Überwachungs-Konzepten ist daher ein sinnvoller Kompromiss zu finden, der einerseits eine ausreichende Überwachung der HV-Struktur erlaubt, andererseits aber nicht zu einer störenden Beeinträchtigung der Systemverfügbarkeiten führt.

Im Umfeld hoch verfügbarer Infrastrukturen sollte daher geprüft werden, ob die Überwachung über ein weitestgehend separates Überwachungs-Netzwerk durchgeführt werden kann. Ein solches Überwachungs-Netzwerk bildet auch die Voraussetzung für die Realisierung höherer HV-Klassen, da nur so eine ausreichend dichte und umfassende Erfassung der Messgrößen ohne übermäßige negative Auswirkungen auf die HV-Architektur selber erreicht werden kann.

Mit zunehmender Komplexität der HV-Architektur und steigender HV-Klassen wächst auch die Abhängigkeit der Erreichung der gewünschten Hochverfügbarkeit von der Überwachungs-Infrastruktur, dies gilt insbesondere sofern neben der Überwachung auch automatische oder gar autonome Reaktionen realisiert werden. In diesen Fällen stellt daie Überwachung einen festen Bestandteil der HV-Architektur dar. Ein Ausfall der Überwachung kommt einem Kontroll- und Steuerungsverlust gleich und zieht eine gravierende Beeinträchtigung der Hochverfügbarkeit der HV-Architektur selber nach sich. Für die Überwachungs-Infrastruktur sind daher nahezu gleichwertige Verfügbarkeitsanforderungen wie für die primären HV-Komponenten zu stellen.

Auch bei beliebigem Ausbau der Überwachung sowie des Überwachungs-Netzwerkes realisiert die Überwachung aufgrund physikalisch-technischer Grenzen immer nur eine sowohl temporal als auch topographisch begrenzte Sicht der Systemzustände der HV-Architektur. Grundsätzlich kann die Überwachung zu jedem Zeitpunkt immer nur eine Momentaufnahme sowie isolierte Sicht auf einzelne Messgrößen liefern. Eine wesentliche Aufgabe der (Früh-) Erkennung besteht daher auch darin, die fragmentierte Sicht der Überwachung zu einem umfassenden und vollständigen Abbild des Systemzustandes der HV-Architektur zu ergänzen.

Eine Überwachungs-Infrastruktur, die einem Optimierungsprozess unterworfen ist, strebt neben der zuvor formulierten Vollständigkeit der Überwachung auch hin zu einer weitestgehenden Automatisierung oder sogar Autonomie der Reaktion um sich einer sich selbst stabilisierenden HV-Architektur anzunähern. Dieser Prozess wird einerseits von der Anforderung nach immer kürzeren und manuell kaum noch zu erbringenden Reaktionszeiten getrieben. Anderseits kann mit fortschreitender Autonomie eine Reduzierung der zur Sicherstellung der geforderten Verfügbarkeit erforderlichen (strukturellen) Redundanz und somit eine Reduzierung der Kosten erreicht werden. Sowohl automatische, auf Redundanzen basierende automatische, als auch autonome Reaktionen bergen aber auch die Gefahr einer Destabilisierung der HV-Architektur, da sie selbst fehlerhaft ablaufen können und sich zunächst der menschlichen Überwachung und Steuerung entziehen.

6 Übersicht

Die nachfolgende Tabelle veranschaulicht den Zusammenhang zwischen Verfügbarkeitsklasse und den dafür notwendigen Überwachungs-Prozessen und Maßnahmen.

Klasse	***Überwachung***									***Erkennung***					***Reaktion***				
	Komponenten[1]	Netzwerk	Software	EAM	Infrastruktur	separates Netz	Abtastrate	Nutzer als Sensor	Zentrales Mgmt	Anzeigen	Erkennen	Früherkennen	Data Mining	Inspektion	Instandsetzung	Wartung	autonome Reakt.	automatische Reakt.	manuelle Reakt.
VK1	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	-	+
VK2	+	+	+	+	+	-	+	-	-	++	+	-	-	-	++	+	-	+	++
VK3	++	++	++	+	+	++	++	+	+	++	++	+	-	+	++	+	-	++	++
VK4 VK5	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++

Tabelle 2: Übersicht der Maßnahmen für die verschiedenen Verfügbarkeitsklassen

6.1 Beispiel-Szenarien

Die nachfolgenden Beispiele veranschaulichen die in Tabelle 2 aufgeführten Maßnahmen für die Realisierung der Überwachung in verschiedenen Verfügbarkeitsklassen und orientieren sich an den Empfehlungen des CobiT Sub-Prozesses DS13.3 [CobiT4.0].

1 - : sporadisch, empfohlen, gering; + : teilweise, erforderlich, hoch; ++ : vollständig, zwingend, sehr hoch

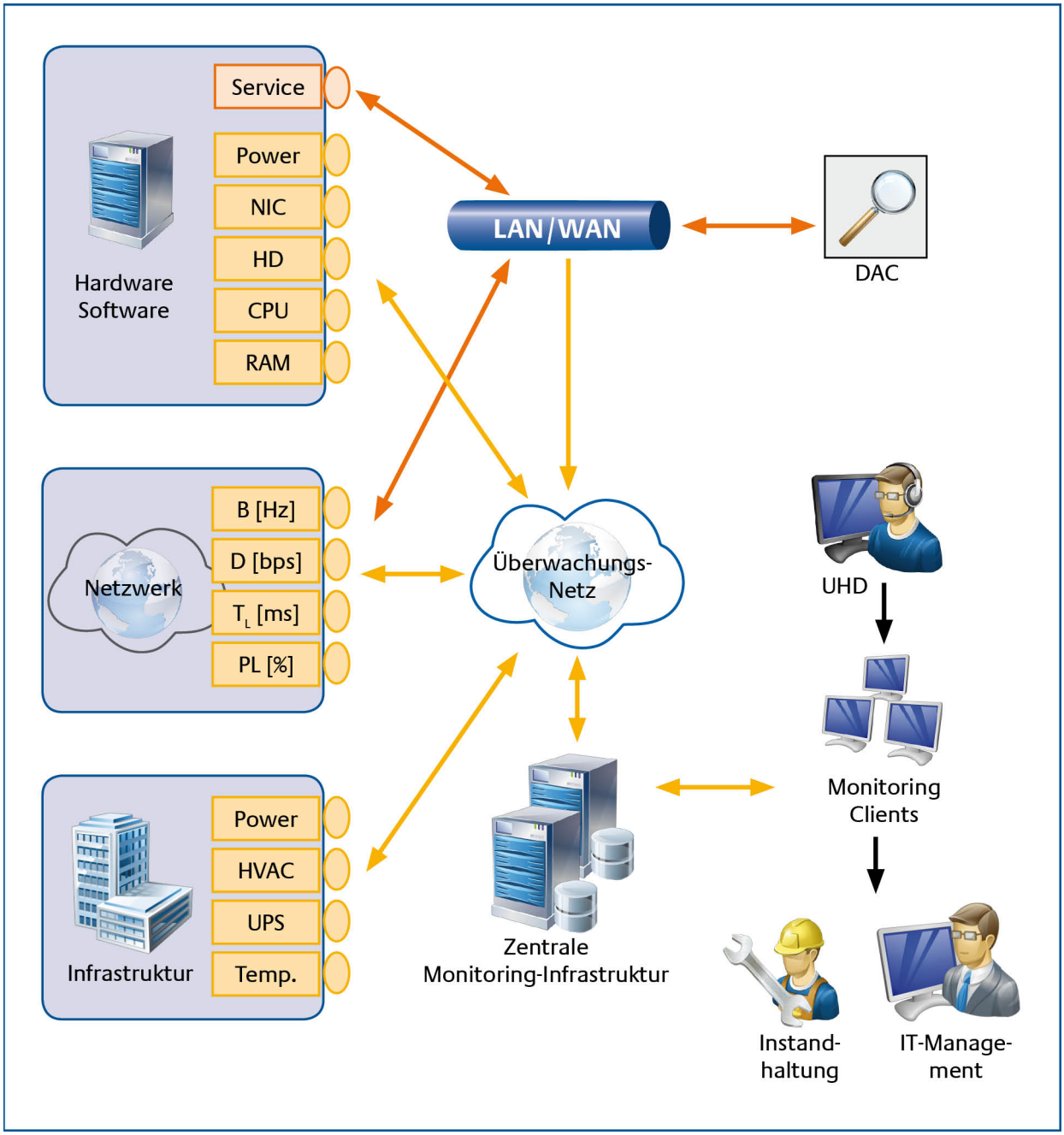


Abbildung 7: Beispielszenario für die Verfügbarkeitsklasse 4-5

In Abbildung 7 ist ein beispielhaftes Überwachungs-Szenario für die Verfügbarkeitsklassen VK 4 und VK5 dargestellt. Dieses Szenario beinhaltet eine vollständige Überwachung aller IT-Komponenten, der Infrastruktur, des Netzwerks sowie eine Performance-Überwachung der IT-Services. Zur Überwachung der IT-Komponenten ist zusätzlich ein separates Überwachungs-Netz errichtet worden, welches das Wirk-Netz vom erheblichen Datenverkehr der Überwachung befreit. Die Überwachung der IT-Services erfolgt über das Wirk-Netz durch Distributed Agent Controler (DAC), was den Vorteil hat, dass somit gleichzeitig die Netzwerk-Performance eine Netzes überwacht werden kann, zu dessen Knoten kein Zugang besteht (z. B. VPN via Internet). Sämtliche Messwerte laufen in einer redundant ausgelegten zentralen Überwachungs-Infrastruktur zusammen. Die Messwerte werden mittels Überwachungs-Clients in einer Überwachungszentrale angezeigt, ausgewertet und bei Bedarf werden entsprechende Alarme (vgl. Kapitel 3.2.2.3) generiert.

Die Doppelpfeile des separaten Überwachungs-Netzes symbolisieren, dass über die Überwachung hinaus eine automatisierte Reaktion durch die Zentrale Überwachungs-Infrastruktur möglich ist.

Darüber hinaus ist dargestellt, dass ebenso die Meldungen des UHD in der zentralen Überwachungs-Infrastruktur erfasst und ausgewertet werden. Ferner ist eine Schnittstelle des Monitorings zu den IT-Management-Prozessen dargestellt.

Die Abbildung 8 zeigt ein beispielhaftes Überwachungs-Szenario für die Verfügbarkeitsklasse VK

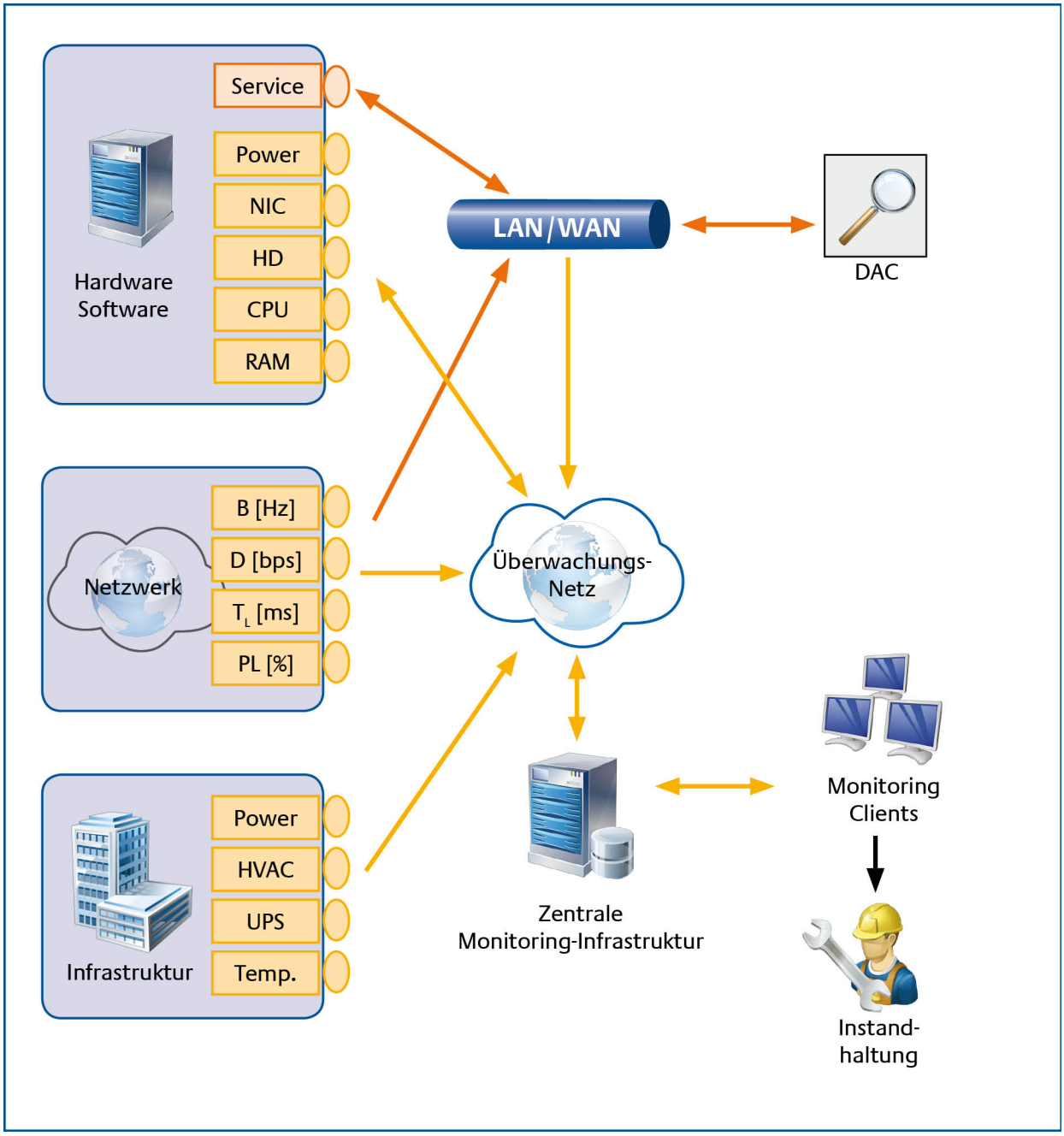


Abbildung 8: Beispielszenario für die Verfügbarkeitsklasse 3

3. Im Vergleich zum Beispiel-Szenario der VK 4/5 werden hier nicht sämtliche IT-Komponenten sondern nur die Kritischen überwacht. Aufgrund der in dieser Verfügbarkeitsklasse notwendigen

hohen Abtastraten und dem damit verbundenen hohen Datenaufkommen ist ein separates Überwachungs-Netz notwendig.

Eine automatisierte Reaktion ist im Bedarfsfall auch nur für die kritischen Komponenten möglich. Es fehlen in diesem Szenario die organisatorischen Schnittstellen zum UHD und den IT-Steuerungsprozessen.

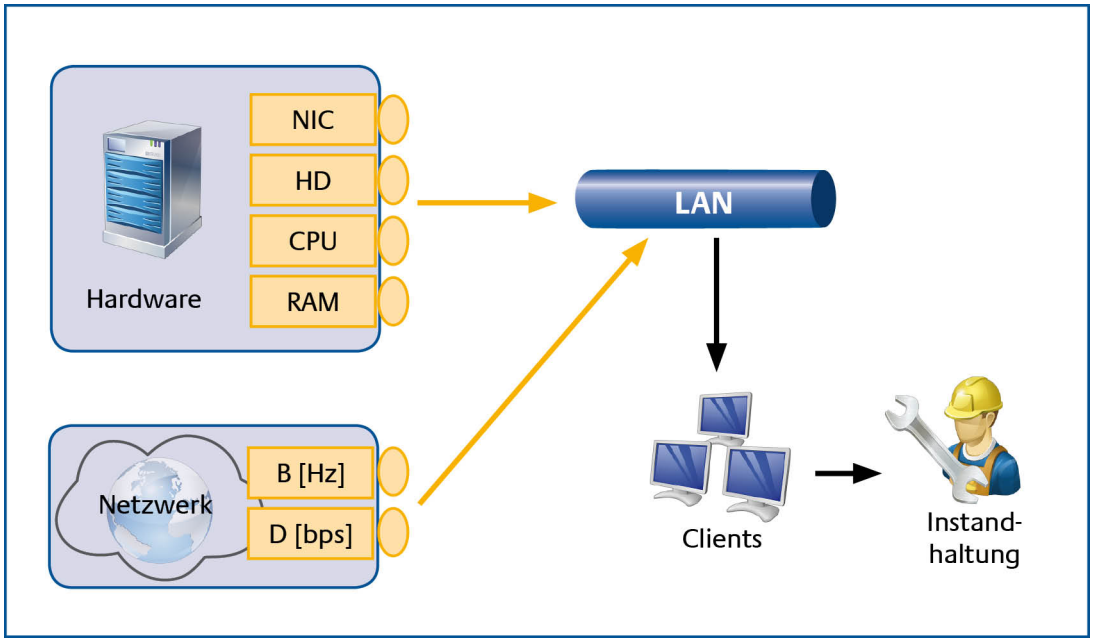


Abbildung 9: Beispielszenario für die Verfügbarkeitsklasse 2

Für die Verfügbarkeitsklasse VK2 ist in der Abbildung 9 ein Szenario dargestellt, das aufgrund des geringen Überwachungs-Datenaufkommens kein separates Überwachungs-Netz hat. Die Anzahl der zu überwachenden Komponenten und die Abtastraten sind in diesem Fall gering. Die Überwachungsdaten werden durch das Wirk-Netz zu verteilten Überwachungs-Clients übermittelt. Von dort wird bei Bedarf die Instandsetzung als einzige Reaktionsmöglichkeit aktiviert. Falls im Rahmen der Überwachung weitergehende Kenntnisse erzielt werden, werden diese durch keinen geregelten Prozess dem IT-Management zugeleitet. Eine Rückkopplung oder Steuerung erfolgt eher zufällig.

Anhang: Verzeichnisse

Abkürzungsverzeichnis

Ein komplettes Verzeichnis hierzu findet sich in Band AH, Kapitel 5

Glossar

Ein komplettes Verzeichnis hierzu findet sich in Band AH, Kapitel 6

Literaturverzeichnis

Ein komplettes Verzeichnis hierzu findet sich in Band AH, Kapitel 7