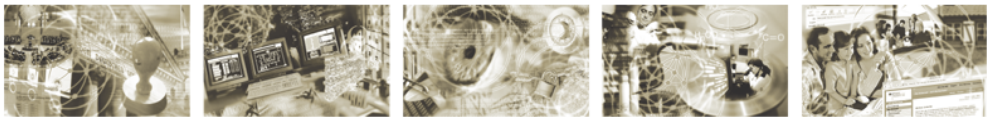
Bundesamt
für Sicherheit in der
Informationstechnik



Band G, Kapitel 2: Definitionen

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
© Bundesamt für Sicherheit in der Informationstechnik 2013

Inhaltsverzeichnis

1 Einleitung....5
2 Metrik und Beispiele zum Mythos der 9-en....21
3 Prinzipielle Möglichkeiten zur Erhöhung der Verfügbarkeit....28
3.1 Erhöhung der MTTF....28
3.2 Verkürzung der MTTR....28
3.3 Der Weg zur Hochverfügbarkeit....29
4 Klassifikation Verfügbarkeitskassen / Systematik / Typifikation....31
Anhang: Verzeichnisse....33
Abkürzungen und Akronyme....33
Glossar....33
Literaturverzeichnis....33

Abbildungsverzeichnis

Abbildung 1: Prozess A als Black-Box....6
Abbildung 2: Analyse einer Betrachtungseinheit als Netz von mehreren Betrachtungseinheiten....7
Abbildung 3: Badewannen-Kurve....12
Abbildung 4: Übergang der augenblicklichen Verfügbarkeit in stationäre Verfügbarkeit....16
Abbildung 5: Abhängigkeit der achieved availability Aa von der Anzahl der Wartungen....18
Abbildung 6: Zustandsblockdiagramm für die Anordnung der Teilkomponenten für Ax....21
Abbildung 7: Zustandsblockdiagramm mit n Elementen in Parallelschaltung....25
Abbildung 8: Redundanz 1 aus 2....26
Abbildung 9: Zustandsblockdiagramm einer Triple Modularen Redundanz (TMR) mit Voter....27
Abbildung 10: Prinzipien und Potentiale....29

Tabellenverzeichnis

Tabelle 1: Ergebnisse der Gesamtverfügbarkeit für serielle Verschaltung von Ax....22
Tabelle 2: Ergebnis der Teileinheiten....24
Tabelle 3: Verfügbarkeitsklassen....31

1 Einleitung

Was ist Verfügbarkeit? Das Wort wird in vielen Sachgebieten mit großer Mehrdeutigkeit benutzt und findet sich in vielen umgangssprachlichen Formulierungen mit Synonymen wie „Bereitschaft“ und „Vorhandensein“ wieder. Im gleichen Zusammenhang werden häufig Begriffe wie Zuverlässigkeit, Verlässlichkeit, Ausfallsicherheit, Funktionsfähigkeit usw. benutzt. Im Folgenden wird Verfügbarkeit für den Kontext dieses Kompendiums definiert. Dabei wird neben mathematischen Festlegungen der Begriff auch verbal abgegrenzt und veranschaulicht.

Wenn nach allgemeinem Verständnis Verfügbarkeit eine Bereitschaft oder das Vorhandensein beschreibt, muss zunächst festgelegt werden, worauf sich die Aussage oder Eigenschaft beziehen soll. In der Vergangenheit und vielfach auch heute noch beziehen sich die Definitionen für Verfügbarkeit oder Zuverlässigkeit fast ausschließlich auf technische Objekte („ein System ist verfügbar ...“, „... eines technischen Produktes.“) und nutzen erst in jüngerer Zeit abstraktere Bezeichnung wie „Einheit“ oder „Fähigkeit einer Einheit“ [ITU-T E.800].

Die Verlagerung der Wertschöpfung von der technischen oder gewerblichen Produktion zur Produktion von Dienstleistung fordert von immateriellen Gütern, wie z. B. Prozessen und Services, neben der funktionellen Beschreibung auch die belastbare Spezifikation nicht funktionaler Eigenschaften wie Verfügbarkeit. Dieses Kompendium beleuchtet Verfügbarkeit aus ganzheitlicher, sowohl unternehmerischer öffentlich ordnender Sicht, sodass hier zunächst die zu betrachtende Einheit des "Was ist, verfügbar“ abgegrenzt wird.

Nachfolgend werden die im Kontext mit diesem Kompendium verwandten Definitionen näher beschrieben:

- Die Bertrachtungseinheit
- Die Umgebung
- Umgebungsbedingungen
- Zuverlässigkeit R(t)
- Mittlere Lebensdauer MTTF
- Fehlerrate
- Verfügbarkeit A(t)
- Mittlere Ausfalldauer MTTR
- Augenblickliche Verfügbarkeit
- Stationäre Verfügbarkeit
- Stationäre mittlere Verfügbarkeit
- Inhärente Verfügbarkeit
- Mittlere Zeit zwischen Wartung und Instandsetzung MTBM
- achieved availability
- Mittlere Instandsetzungszeit MDT
- Operationale Verfügbarkeit

– Mittlere Verfügbarkeit

Die Betrachtungseinheit

Eine Betrachtungseinheit ist ein Prozess oder eine einzelne Komponente eines Prozesses, der bzw. die für sich allein bezüglich ihrer Eigenschaften funktionell und nicht funktionell beschreib- und bewertbar ist.

Der Begriff „Betrachtungseinheit“ wird allgemein als selbsterklärend betrachtet und deshalb in der Literatur sowie in Normen wie DIN 31051:2003-06 verwendet, ohne dass eine genauere Erklärung oder Erläuterung gegeben wird. In der VDI-Norm 4003 ist die Betrachtungseinheit (auch Einheit) der Gegenstand der Zuverlässigkeitsuntersuchung und wird als Produkt verstanden, das ein aus Hardware- oder Software zusammengesetztes Gerät, System, Verfahren, eine Anlage [VDI 3423] oder ein aus mehreren Vorgängen zusammengefügter Prozess oder eine Dienstleistung ist.

Die Definition fordert nicht, dass der Aufbau und die Wirkungsweise der Betrachtungseinheit bekannt ist. Sie ist entsprechend der Systemtheorie eine „Black-Box“, also ein geschlossenes System ohne Betrachtung der inneren Funktionsweise. Mit dieser Herangehensweise wird die Komplexität des Beobachtungsgegenstandes bewusst reduziert.

Von Interesse ist nur das Verhalten der Betrachtungseinheit, das über bestimmte definierte Schnittstellen wirkt. Diese Sicht verdeutlicht Abbildung 1.

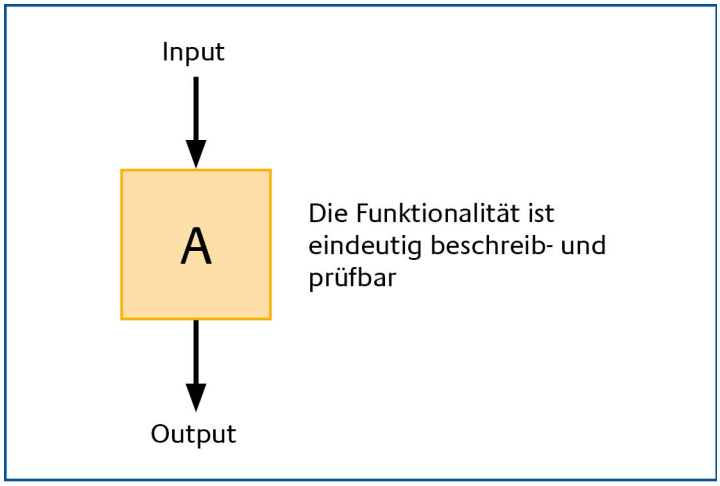


Abbildung 1: Prozess A als Black-Box

Die einzelnen Bestandteile der Black-Box-Betrachtungseinheit in Abbildung 1 interessieren nur, wenn diese in einer tiefer gehenden Analyse wiederum als Betrachtungseinheit ausgewählt werden. Für die Analyse der Betrachtungseinheit sind nur die Schnittstellen und mit welcher Häufigkeit oder Wahrscheinlichkeit die einzelnen Bestandteile oder Teilkomponenten aktiviert werden, wichtig.

Abbildung 2 zeigt die Analyse des Prozesses A aus Zeichnung 1 in der Form einer Geschäftsprozesskette:

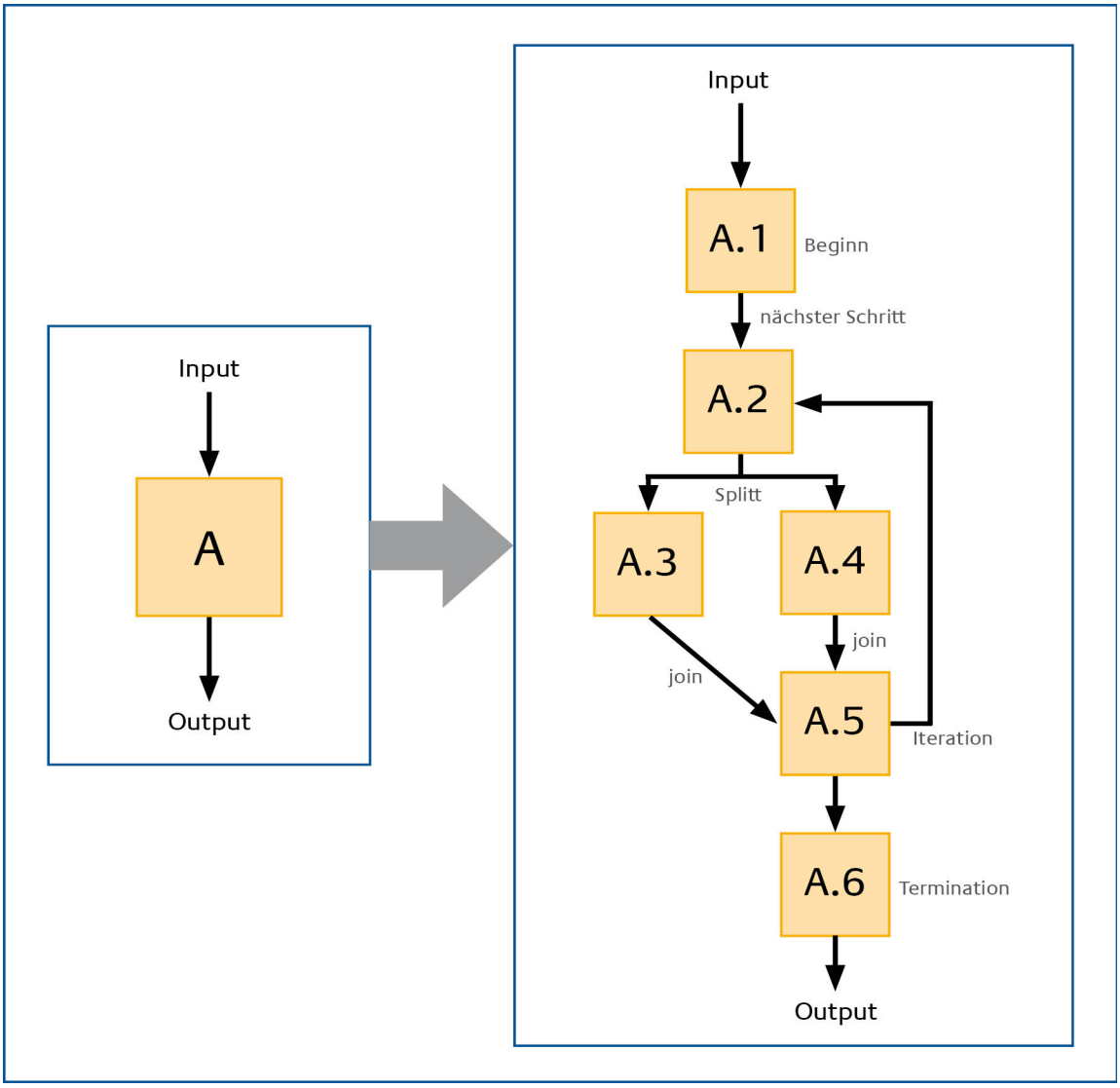


Abbildung 2: Analyse einer Betrachtungseinheit als Netz von mehreren Betrachtungseinheiten

Die analytische Betrachtung ist bei der Bewertung und Modellierung von Betrachtungseinheiten sehr wichtig. Wie Abbildung 2 zeigt, kann der Prozess A in 6 Betrachtungseinheiten aufgegliedert werden, die über unterschiedliche Arten von Beziehungen verbunden sind. Der Prozess A beginnt seine Funktionalität mit der Einheit A.1, von dort aus wird zwingend die Betrachtungseinheit A.2 in der Sequenz aktiviert, von der Betrachtungseinheit A.2 splittet sich der Durchlauf in zwei (denkbar sind natürlich mehrere) Betrachtungseinheiten A.3 und A.4 auf. Die Betrachtungseinheit A.5 kann sowohl von A.3 als auch von A.4 als nächste Einheit in der Kette aktiviert werden (join). Die Betrachtungseinheit A.5 steuert mit einer bestimmten Wahrscheinlichkeit den Fluss zu einer vorhergehenden Einheit A.2 zurück (Iteration) oder zu der Einheit A.6. Mit dieser untergeordneten Betrachtungseinheit terminiert die Betrachtungseinheit A.

Dieses simplifizierte Beispiel verdeutlicht auch, dass der Zuschnitt von Betrachtungseinheiten oder was einer Betrachtungseinheit zuzuordnen ist, nicht ganz trivial ist. So darf eine Betrachtungseinheit nur einen Initiierungspunkt (eine Einheit für den Beginn der Funktionalität) besitzen. Andernfalls ist die Verfügbarkeit nicht „eindeutig beschreib- und bewertbar. Anders ausgedrückt, sonst liegt das Wissen über die statistisch verteilte Aktivierung von Teileinheiten in anderen Betrachtungseinheiten.

Mit einem so spezifizierten Begriff der Betrachtungseinheit lassen sich mit den Verfahren und Methoden der Zuverlässigkeitstheorie wie z. B. in [Walter 2003]) auf organisatorischer Ebene analysieren. Auf technischer Ebene ist damit ebenfalls die Grundlage für die Verfügbarkeitsbetrachtung von Systemen, Geräten, Software und Infrastrukturkomponenten gegeben.

Die Umgebung

Als Umgebung (engl. environment) wird eine Betrachtungseinheit bezeichnet, die technische wie organisatorische Prozesse in Hinblick auf die Anforderungen des jeweiligen Umfeldes betrachtet.

Technische Komponenten können im Allgemeinen nur unter bestimmten physikalischen, chemischen oder biologischen Bedingungen wie z. B. Umgebungstemperatur, Energie, Druck, Feuchte oder Radioaktivität, die erwarteten Ergebnisse erbringen. Leistungen von Menschen setzen darüber hinaus noch besondere Bedingungen an soziale Ruhe, rechtlicher Regelung und öffentliche Ordnung und Sicherheit sowie Versorgung voraus. Zu beachten ist auch, dass jeder technische oder organisatorische Prozess nur eine bestimmte Last bewältigen kann.

Für den Betrieb von technischen Komponenten und der Abwicklung eines Geschäftsprozess müssen alle Vorkehrungen getroffen werden, um die Anforderungen an das Umfeld einzuhalten. Tritt plötzlich ein außergewöhnliches und unvorhersehbares Ereignis ein, das auch durch äußerste Sorgfalt und Vorkehrungen nicht verhindert werden kann, kennt das deutsche Recht den Begriff Höhere Gewalt. Beispiele höherer Gewalt sind z. B. Streiks, Brand, Unwetter, Verkehrsunfälle, Massenkarambolagen, Abstürze von Flugzeugen, Großfeuer, Geiselnahmen, Explosionen, Terroranschläge usw.

Kann durch das Eintreten von höherer Gewalt, eine Betrachtungseinheit wie z. B. ein Service, seine Leistung oder ein Produkt seine Eigenschaften nicht so erbringen, wie berechtigt erwartet wird, muss der Verantwortliche für die Leistungserbringung keinen Schadenersatz leisten. Vereinbarte Pönalen für Minderleistungen sind in solchen Fällen wirkungslos. Ein Blick auf die zuvor aufgeführten Beispiele zeigt, dass solche Umstände in der Lebens- und Wirtschaftsrealität eintreten. Auch wenn dieser rechtliche Schaden die private Wirtschaft vor enormen Folgen schützt, kann der materielle Schaden an Sachen und Vermögen sowie durch Produktionsminderung und der immaterielle Schaden wie Image- und Vertrauensverlust erheblich sein.

Im öffentlich-rechtlichen Sektor gibt es dagegen eine erhebliche Anzahl von Prozessen und Services, die nur und gerade im Falle von der Höheren Gewalt verfügbar sein müssen. Diese besonderen Umstände werden auch als stattlicher Notstand bezeichnet und sind in dem entsprechenden Gesetz der Art nach und bezüglich der Zuständigkeiten geregelt. Die staatlich ordnenden Prozesse müssen unter diesen Bedingungen, Gefahren und Gegebenheiten verfügbar sein, ohne dass im Vorhinein deren Art, Intensität sowie zeitliche und räumliche Ausdehnung genau bekannt ist.

Umso wichtiger ist es zu wissen, für welche normalen und besonderen Bedingungen, wie Last, Gefahren und besonderen Lagen, eine Betrachtungseinheit ausgelegt ist. Vergessen werden dürfen aber auch nicht die administrativen Bedingungen wie Betriebszeiten. Im technischen Bereich wird hierfür auch der Begriff Spezifikation verwandt.

Umgebungsbedingung

Die Umgebungsbedingung, engl. environment, einer Betrachtungseinheit beschreibt überprüfbar die Gesamtheit aller physikalischen, chemischen und biologischen sowie auch aller politischen, sozialen und administrativen Bedingungen, unter denen sie ihre beschriebenen Eigenschaften zusichert.

Die Aufmerksamkeit für eine Betrachtungseinheit liegt im Allgemeinen bei den Eigenschaften, die notwendig zu einer bestimmten Sache dazugehören, weil sie gewünscht und gefordert werden. Es besteht ein inniger Zusammenhang (Inhärenz) zwischen einer Eigenschaft und dem Träger der Eigenschaft. In überwiegender Mehrzahl sind dies sogenannte deterministische Eigenschaften und Merkmale einer Sache, welche direkt messbar sind (u. a. Transaktionsantwort, Abmessungen, Festigkeit, elektrische und Wärmeleitfähigkeit).

Die Begriffsbestimmung als auch das ganze Kompendium beschäftigt sich mit der Tatsache, dass kein technisches Produkt frei ist von der Möglichkeit auszufallen, jedes System früher oder später versagt, eine Sache ihre Eigenschaften früher oder später verliert, kein Prozess immer reibungslos läuft. Es wird dann von einem Fehlverhalten, Ausfall, Versagen, Nichterfüllen oder schlicht von einem Fehler gesprochen.

Abhängig von der Wirkdauer der Fehler gilt folgende Klassifikation:

Fehlermodell:

- Von permanenten Fehlern spricht man, wenn die zugesicherte Eigenschaft oder die geforderte Funktion von der Betrachtungseinheit dauerhaft nicht erfüllt wird. In diesem Falle ist eine Instandsetzung (Reparatur) erforderlich.
- Ein intermittierender Fehler ist das sporadische Fehlverhalten einer Einheit und in vielen Fällen häufiger als permanente Fehler.
- Ein transienter Fehler ist das vorübergehende Nichterfüllen einer Funktion.

Anders als bei den oben beschriebenen deterministischen Eigenschaften eines Produkts, welche direkt messbar sind, unterliegen Fehler einem stochastischen Prozess: Im Allgemeinen kann das Eintreten zeitlich nicht exakt vorhergesagt oder reproduziert werden.

Wegen des stochastischen Auftretens von Fehlern kann zum Beispiel die interessante, weil sehr wichtige und häufig gestellte Frage,:„Überlebt der Landungsrechner eines Flugzeugs die Dauer des Landeanflugs von 10 Minuten oder wann fällt das System frühestens aus?“, deshalb so nicht exakt beantwortet werden und muss umformuliert werden in die Frage: „Wie hoch ist die Ausfallwahrscheinlichkeit während des Landeanflugs?“. Allgemeiner ausgedrückt ist die Frage nach der Wahrscheinlichkeit P, dass das jetzt (t=0) laufende System zu einen Zeitpunkt t in der Zukunft (t>0) ausfällt. Diese statistische Größe wird allgemein Zuverlässigkeit oder mathematisch exakter Überlebenswahrscheinlichkeit genannt.

Zuverlässigkeit R(t) , auch Überlebenswahrscheinlichkeit genannt

- Die Zuverlässigkeit R(t), engl. reliablitiy, einer Betrachtungseinheit ist die Wahrscheinlichkeit, dass die Betrachtungseinheit im Intervall [0,t] alle zugesicherten Eigenschaften bei den beschriebenen Umgebungsbedingungen einhält.
  - Wenn T die Zufallsgröße ist, die die Zeit bis zum ersten Fehler repräsentiert, dann gilt R(t) = P(T>=t),
    - wobei P(E) die kumulative Verteilungsfunktion (vereinfacht Wahrscheinlichkeit) ist, dass das Ereignis E eintritt .
      - Sinnvollerweise ist R(0)=1. d. h. zum Zeitpunkt der Beobachtung funktioniert die Einheit und auch danach mit einer gewissen Wahrscheinlichkeit R(t) > 0
- Umgekehrt ist die Wahrscheinlichkeit, dass bis zu einem vorgegebenen Wert t, also bis zu einem bestimmten Zeitpunkt t ein Ausfall passiert

$F(t) = 1 - R(t)$,

wobei $F(t) \geq 0$, $F(0) = 0$.

So wie R(t) als Zuverlässigkeitsfunktion, ist F(t) die Unzuverlässigkeitsfunktion - wobei F für Fehler, engl. Failure steht – die kumulative Verteilungsfunktion der Ausfallverteilung.

Mit der Dichtefunktion f (t), auch Dichte, abgek. WDF oder PDF von probability density function genannt,

$$f(t) = \frac{dF(t)}{dt} = -\frac{dR(t)}{dt}$$

Formel 1: Dichtefunktion f(t)

wird die „Schärfe“ der Fehlerverteilung beschrieben. Die graphische Darstellung ist die Visualisierung der Fehlerverteilung.

Die Dichtefunktion f(t) :

$$f(t) \geq 0$$

Formel 2: Dichtefunktion f(t)

und

$$\int_0^\infty f(t)\,dt = 1$$

Formel 3: Dichtefunktion f(t)

hat zwei Eigenschaften:

$$F(t) = \int_0^t f(t')\,dt'$$

Formel 4: Erste Eigenschaft der Dichtefunktion f(t)

und

$$R(t) = \int_t^\infty f(t')\,dt'$$

Formel 5: Zweite Eigenschaft der Dichtefunktion f(t)

Anders ausgedrückt, die Zuverlässigkeits (R(t))- und die Unzuverlässigkeitsfunktion (F(t)) bilden die Fläche unter der Kurve f(t), die die Dichte oder die Schärfe der Fehlerverteilung beschreibt.

Weil die gesamte Fläche unter dieser Funktion f(t) gleich 1 ist, gilt für die Zuverlässigkeit und Unzuverlässigkeit

$0 \leq R(t) \leq 1$ und $0 \leq F(t) \leq 1$.

Mit dieser Definition kann man die im technischen Bereich häufig benutzt Größe definiert werden.

Mittlere Lebensdauer, MTTF

Die mittlere Lebensdauer, engl. mean time to failure, ist die durchschnittliche Zeitdauer bis zum ersten Ausfall einer Betrachtungseinheit und mathematisch ausgedrückt der Erwartungswert E(T) der oben definierten Zufallsgröße T, die die Zeit bis zum ersten Fehler repräsentiert.

Bei gegebener Zuverlässigkeit R(t) kann die MTTF folgendermaßen berechnet werden:

$$MTTF = E(T) = \int_0^\infty t\, f(t)\, dt$$

Formel 6: Berechnungsformel für MTTF

Was gleichbedeutend ist mit

$$MTTF = \int_0^\infty R(t)\, dt$$

Formel 7 : Berechnung von MTTF 2. Berechnungsgrundlage

In diesem Zusammenhang ist die Fehlerrate von Interesse.

Fehlerrate λ(t), engl. failure rate, oder hazard rate of failure

Die Fehlerrate λ(t)[1]ist die Anzahl von Ausfällen pro Zeiteinheit und wird mathematisch bestimmt durch die Gleichung:

$$\lambda(t) = \frac{f(t)}{R(t)}$$

Formel 8: Berechnungsformel für die Fehlerrate

Die Fehlerrate λ(t) ist zeitabhängig! Erfahrungsgemäß ändert sich diese Rate über die Lebensdauer, *engl. life cycle* , der Betrachtungseinheit. In vielen Fällen lässt sich die Fehlerrate über die Lebensdauer mit der Badewannenkurve *, engl. bathtube curve* beschreiben, vereinfacht betrachtet ist es auch auf menschliche Produktionsfaktoren anwendbar.

1 Eine bekannte Fehlerrate ist die sog. FIT (Failure In Time) elektronischer Bauteile. Sie gibt die Anzahl der Bauteile an, welche in 10^9 Stunden ausfallen.

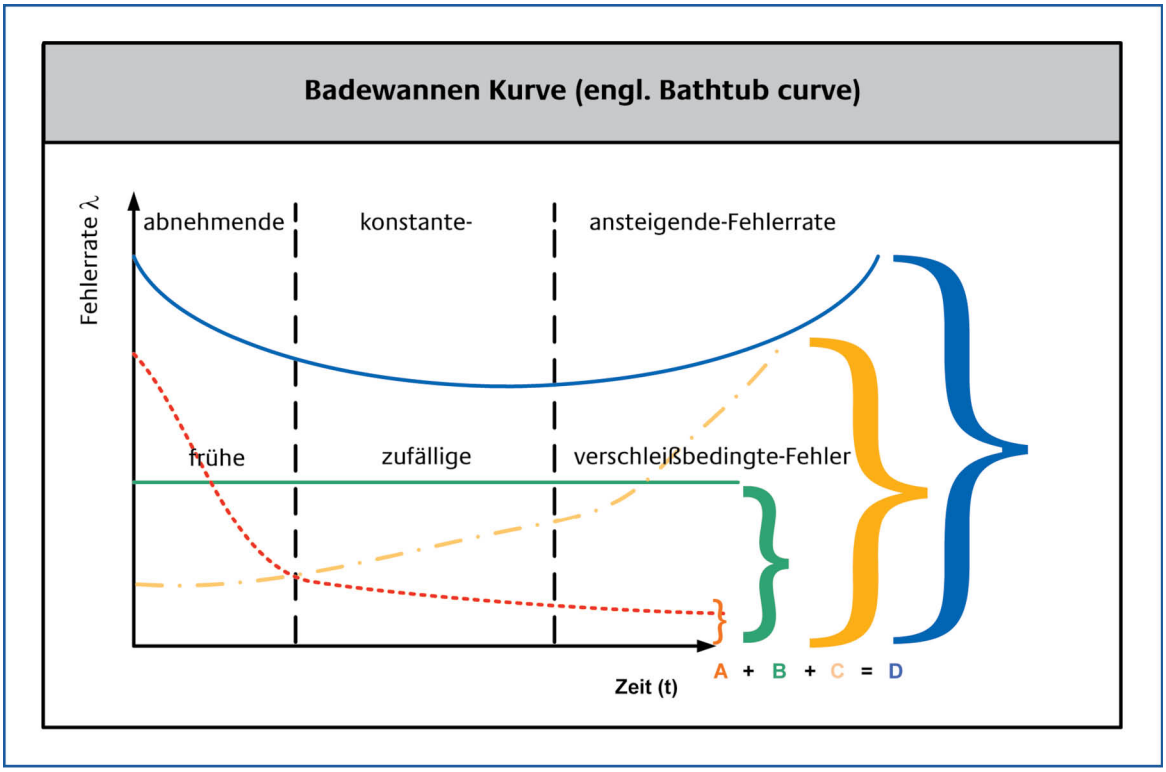


Abbildung 3: Badewannen-Kurve

Die X-Achse gibt die Zeitspanne t von Beginn der Inbetriebnahme der Einheit wieder, während auf der Y-Achse die Ausfallrate λ pro Zeiteinheit aufgetragen wird. Sie beschreibt einen speziellen Verlauf der Fehlerrate aus drei Komponenten:

- Der erste Anteil „frühe Fehler“, engl. early „infant mortality“ failures) ist eine abnehmende Fehlerrate, die sich aus dem allmählichen Einschwingen in die kontinuierliche Produktion ergibt. Design-, Konstruktions-, Produktions- oder Werkstoffmängel fallen häufig gleich zu Beginn auf bzw. führen zu Schäden. Menschen müssen erst geschult und eintrainiert sowie mit der Handhabung erstmalig auftretender Probleme vertraut gemacht werden. Diese Fehlerquellen nehmen im Laufe der Zeit kontinuierlich ab. Nachdem die sogenannten Kinderkrankheiten behoben sind, kann der Betrieb kontinuierlich über die Zeit hinweg funktionieren
- Der zweite Anteil „zufällige Fehler“ , engl. constant (random) failures gibt die statistisch gleichmäßig, konstant über die Zeit verteilten Ausfälle wieder.
- Der dritte Anteil „Abnutzung“ , engl. wear out failures gibt den intuitiv wahrgenommenen kontinuierlichen Anstieg der Ausfälle einer Einheit wieder, die durch Alterung oder abnehmende Aufmerksamkeit erklärt werden.
- Die Addition dieser drei Komponenten ergibt die beobachtete Fehlerrate λ(t)), die dem Verlauf nach über die Lebensdauer mit einem Badewannenprofil vergleichbar ist;
- In der Anfangsphase, im engl. Sprachraum „burn-in“ genannt, ist die Fehlerrate vergleichsweise hoch und nimmt stark ab.
- Im mittleren Bereich, im engl. Sprachraum „useful life“ genannt, ist die Fehlerrate recht konstant und sollte den größten Teil der Lebensdauer umfassen.
- Am Ende, im engl. Sprachraum „burn-out“ genannt, steigt die Fehlerrate wieder kontinuierlich stark an.

Aber wie zuvor erwähnt, folgen nicht alle Betrachtungseinheiten bezüglich der Fehlerrate dieser Lebenskurve. Die berühmte Ausnahme ist Software, die bekanntlich nicht altert, wenn sie nicht gewartet, fortgeschrieben oder sonst verändert wird. Deren Fehlerrate $\lambda(t)$ verhält sich dann bei gleich bleibender Nutzungsart wie der zweite Anteil mit einer konstanten Fehlerrate über die Zeit. Eine gute Einführung in die Thematik der vergleichenden Fehlerraten bei Hard- und Software gibt [Herrman 1999].

Die bisherige Diskussion über die Zuverlässigkeit R(t) beschäftigt sich mit der Wahrscheinlichkeit, dass die Betrachtungseinheit in einem Zeitintervall [0,t] fehlerfrei funktioniert, oder wie die alternative Bezeichnung „Überlebenswahrscheinlichkeit“ ausdrückt, überlebt, „durchhält“ etc.

Für diesen Fokus gibt es viele Berechtigungen. Bei Vorgängen zur Notfallüberbrückung oder Prozessen, die der öffentlich-rechtliche Sektor im Falle von kritischen Lagen aktiviert, steht die Frage im Vordergrund, mit hoher Wahrscheinlichkeit eine bestimmte Zeit zu überbrücken (Betrieb mit Notstrom fortführen, Versorgung aus Reserven oder Notvorräten etc.).

Die umgangssprachlichen Formulierungen für Verfügbarkeit mit Synonymen wie Bereitschaft und Vorhandensein hebt weniger auf ein Zeitintervall bis zum Ausfall ab, sondern im Vordergrund steht die Frage, ob zu einem Zeitpunkt t die Betrachtungseinheit funktioniert, bereit oder vorhanden ist.

Verfügbarkeit A(t) , engl. availability

Die Verfügbarkeit A(t), engl. availability, einer Betrachtungseinheit ist die Wahrscheinlichkeit, dass die Betrachtungseinheit alle zugesicherten Eigenschaften bei den beschriebenen Umgebungsbedingungen zum beliebigen Zeitpunkt t einhält oder fehlerfrei funktioniert.

Für nicht wieder herstellbare Betrachtungseinheiten gilt $A(t) = R(t)$.

Hochverfügbare Prozesse und Systeme werden nach einem Ausfall schnellstmöglich wieder hergestellt. Bei instandsetzungsfähigen Betrachtungseinheiten gilt deshalb $A(t) >= R(t)$, da die ausgefallene Betrachtungseinheit zum Zeitpunkt t schon wieder in den funktionsfähigen Zustand zurückgekehrt sein kann. Dazu wird analog zur MTTF der Begriff der mittleren Ausfalldauer MTTR definiert.

Mittlere Ausfalldauer MTTR , engl. mean time to repair

Die mittlere Ausfalldauer einer Betrachtungseinheit gibt die durchschnittliche Länge des Zeitintervalls an, das zwischen dem Ausbleiben einer Eigenschaft und dem wieder Vorhandensein aller Eigenschaften der Betrachtungseinheit liegt.

Die Bezeichnung Ausfalldauer ist exakter als der im technischen Umfeld häufig benutzte Ausdruck Reparaturzeit, da die Instandsetzung neben der reinen Reparatur auch noch die Zeit für die Erkennung und die Lokation des Fehlers umfasst. In manchen Fällen besteht die Wiederherstellung auch in dem kompletten Austausch oder Übergang auf eine andere funktionsfähige Betrachtungseinheit.

Die mittlere Ausfalldauer ist aber nicht nur von technischem oder rein mathematisch-statistischem Interesse, sondern spielt bei der Modulation der Verfügbarkeit von Geschäfts- sowie bei Vorsorge- oder Sicherungsprozessen eine wichtige Rolle. Dies vor allem unter dem Gesichtspunkt, was an Ausfallzeit vertretbar oder überbrückbar ist.

Für viele Verantwortungsträger steht nicht der abstrakte Wahrscheinlichkeitsausdruck in der Verfügbarkeitsdefinition im Vordergrund, sondern die Frage, welche maximalen Ausfallzeiten vertretbar sind. Dies ist vor dem Hintergrund zu sehen, dass es für das Geschäfts- und Prozessziel zu verantworten ist, wenn ein negatives Ereignis maximal x-mal in der Zeit Δt eintritt. Diese pragmatische und rein operationale Sicht ignoriert unbewusst den statistischen Charakter von

Ausfällen und fließt trotzdem regelmäßig als Vereinbarung über Verfügbarkeit (z. B. als eine Dienstgütevereinbarung (DGV), *engl. Service Level Agreement (SLA)*, in Verträge ein. Damit verbunden sind dann erhebliche Strafzahlungen (Pönale) bei Minder- oder Bonuszahlungen bei Besserleistungen. Obwohl die vertraglich vereinbarte Beobachtungsdauer für eine so definierte Verfügbarkeit von im Allgemeinen nur einem Jahr kaum statistische Signifikanz sichert, ist ein solches Vorgehen im Wirtschaftsleben in der Praxis häufig anzutreffen.

Bei Prozessen, die der öffentlich-rechtliche Sektor im Falle von kritischen Lagen aktiviert, spielt die Frage der maximal überbrückbaren Ausfallzeit eine entscheidende Rolle. Weil die Dauer der besonderen Situation häufig nicht vorhersehbar ist, muss hierfür eine Abschätzung oder pure Annahme getroffen werden. Diese Zeitspannen lassen keine seriöse Aussage über die Anzahl zu, sondern unterstellen lediglich, dass die Ausfälle sich nicht häufen.

Unterschiedliche Interpretationen der Ausfalldauer MTTR, so z. B. ungeplante und geplante Ausfälle, führen zu unterschiedlichen Varianten in der Definition der Verfügbarkeit, wie die folgende Darstellung auch aufzeigt:

Die Verfügbarkeit beschreibt den Zustand der Betrachtungseinheit über nur zwei diskrete Werte, namentlich „funktioniert“ oder „funktioniert nicht“, und kann deshalb durch folgende Zustandsfunktion X(t) mit den binären Werte 0 und 1 beschrieben werden:

$$X(t) = \begin{cases} 1, & \textit{die Betrachtungseinheit zeigt alle zugesicherten Eigenschaften} \\ 0, & \textit{sonst} \end{cases}$$

Formel 9: Zustandsfunktion X (t)

Mit dieser Funktion X(t), die den Status einer instandsetzungsfähigen Betrachtungseinheit zum Zeitpunkt t beschreibt, kann die Verfügbarkeit wie die Zuverlässigkeit als eine Wahrscheinlichkeit beschrieben werden. Nach [Barl75][Barlow & Proschan (1975) und [Barl65](1965) gibt es drei grundlegende Varianten der Verfügbarkeitsinterpretation und damit auch Verfahren der Bestimmung:

- augenblickliche Verfügbarkeit , engl. availability funktion
- stationäre Verfügbarkeit , engl. steady state availability oder auch limiting availability
- mittlere Verfügbarkeit, engl. average availability auch mean availability oder average up-time availability

Augenblickliche Verfügbarkeit , engl. instant availability , oder point availability,

Die augenblickliche Verfügbarkeit einer Betrachtungseinheit zum Zeitpunkt t ist definiert als: A(t) = P (X (t) = 1).

Das ist die Wahrscheinlichkeit, dass die **Betrachtungseinheit** zum Zeitpunkt t funktioniert. Weil es fast immer unmöglich ist, einen expliziten Ausdruck für A(t) zu erhalten, insbesondere kurz nach der Installation oder gegen Ende der Lebensdauer einer Betrachtungseinheit (bathtub curve), gibt es andere Vorschläge für die Bestimmung.

Lange nach der Installation einer Betrachtungseinheit, und nachdem sie dann für eine längere Zeit gearbeitet hat, erreicht die Einheit einen stationären Zustand. Die Systemtheorie spricht auch von einem eingeschwungenen oder stabilen Zustand (engl. steady state). Ohne externe Eingriffe ist das Verhalten der Einheit über die Zeit konstant, zeitunabhängig. Anders ausgedrückt strebt die

Wahrscheinlichkeit A(t) nach längerer Zeit gegen einen festen zeitunabhängigen Wert A (oder mathematisch ausgedrückt: für große t gegen einen Grenzwert A, den lim A(t) für t gegen ∞). Diese Erfahrungen führen zu der im praktischen Gebrauch üblichen Definition von Verfügbarkeit, der stationären Verfügbarkeit.

Stationäre Verfügbarkeit As

Die stationäre Verfügbarkeit As, wird in der Literatur eigentlich nur unter dem englischen Ausdruck steady state availability bekannt , oder limiting availability

In instandsetzungsfähigen Betrachtungseinheiten wechseln die Phasen Funktion und Ausfall ständig gegenseitig ab. Dabei strebt die Wahrscheinlichkeit A(t) für große t gegen den Grenzwert, der stationäre Verfügbarkeit genannt wird:

$$A_s = \lim_{t \to \infty} A(t)$$

Formel 10: Die stationäre Verfügbarkeit

Mit dieser Definition ergibt sich die:

Stationäre mittlere Verfügbarkeit A∞ , engl. limiting average availability

Die stationäre mittlere Verfügbarkeit A∞ ist der Grenzwert der mittleren Verfügbarkeit

$$A_\infty = \lim_{t \to \infty} \overline{A(t)}$$

Formel 11: Die Stationäre mittlere Verfügbarkeit ist der Grenzwert der mittleren Verfügbarkeit

Oft ist das Rechnen einfacher mit der:

Unverfügbarkeit UA , engl.unavailability

Die Unverfügbarkeit UA ist definiert als

UA := 1 – A,

womit auch gilt:

$$UA = \frac{MTTR}{MTTF + MTTR}$$

Formel 12: Unverfügbarkeit

womit umgekehrt ebenso gilt:

$$A = \frac{MTTF}{MTTF + MTTR}$$

Formel 13: Berechnung von A

Diese Größen kennzeichnen die Wahrscheinlichkeiten, dass die Betrachtungseinheit verfügbar ist, nachdem sie für eine lange Zeit unter vorgegebenen völlig konstanten Randbedingungen arbeitete,

insbesondere lange nach dem „Einschwingverhalten“, und ist sehr signifikant für die Kennzeichnung der Verfügbarkeit von instandsetzungsfähigen Betrachtungseinheiten. Die folgende Zeichnung stellt zwei idealisierte Beispiele dar:

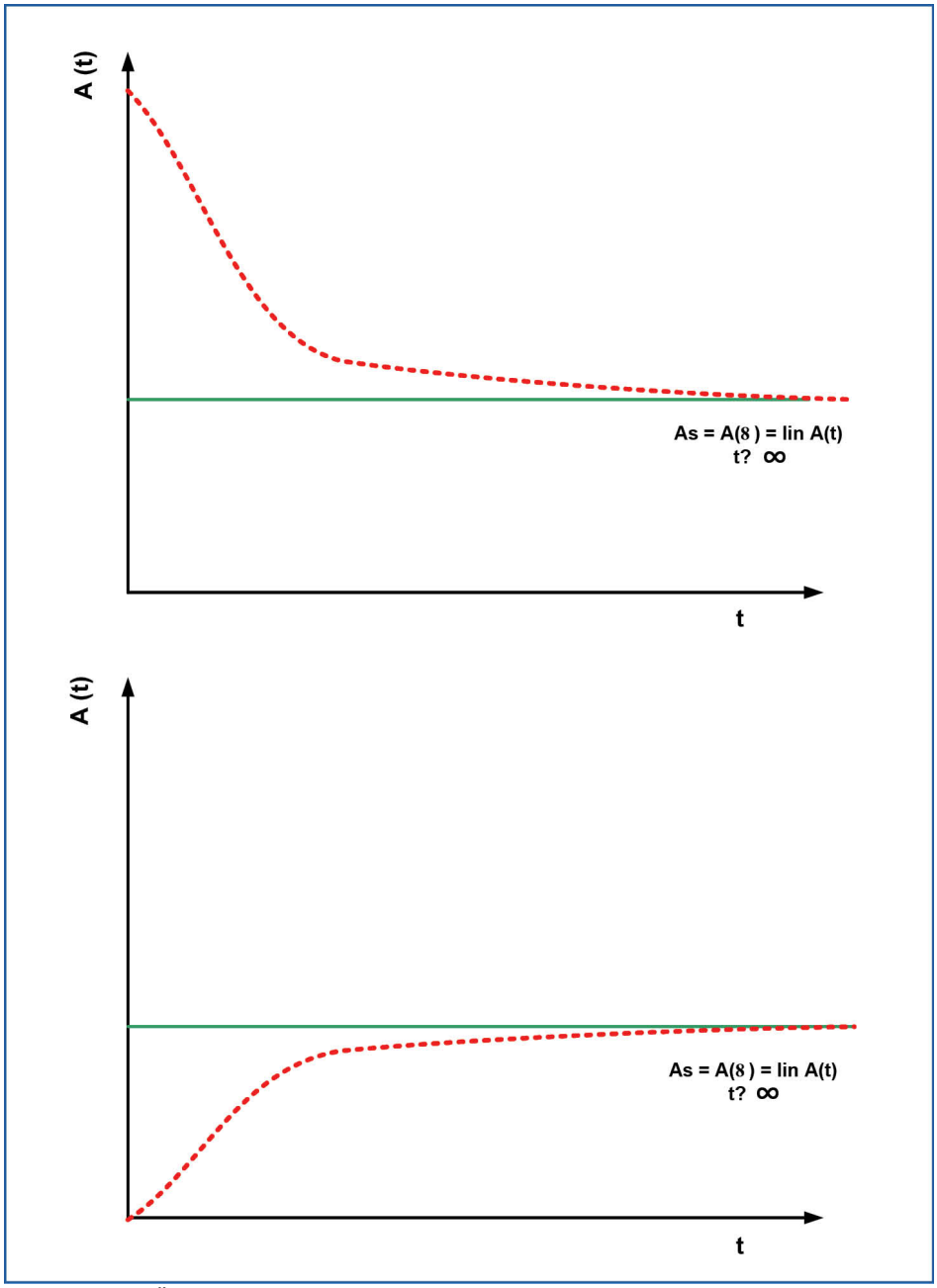


Abbildung 4: Übergang der augenblicklichen Verfügbarkeit in stationäre Verfügbarkeit

Wie die Erläuterungen zur Fehlerrate anhand der *bathtube curve* zeigte, ist die Konstanz der Verfügbarkeit oder Unabhängigkeit der Verfügbarkeit von der Lebenszeit häufig nicht immer gegeben, auch wenn der mathematische Ausdruck dieses sehr leicht vermuten lässt.

Es gibt einige Variationen der stationären Verfügbarkeit, abhängig von den geplanten und ungeplanten Ausfallzeiten, wovon die wichtigsten im Folgenden dargelegt werden.

Inhärenter Verfügbarkeit Ai, engl. inherent availability,

Von Inhärenter Verfügbarkeit ist die Rede,wenn nur die Verfügbarkeit betrachtet wird, die „der Sache anhaftet/ innewohnt“, also ihr inhärent ist.

- Die inhärente Verfügbarkeit betrachtet nur die Zeit, in der die Betrachtungseinheit fehlerfrei funktioniert und die Zeit für die Instandsetzung aus dem Fehlerzustand. Es wird von einer idealen Betrachtungseinheit in einer idealen Umgebung ausgegangen, für die keine geplanten Ausfälle für Wartung, Verwaltung oder Logistik erforderlich sind.
- Die inhärente Verfügbarkeit kann entsprechend ausgedrückt werden als:

$$A_i = \lim_{t \to \infty} A(t) = \frac{MTBF}{MTBF + MTTR}$$

Formel 14: Berechnung der inhärenten Verfügbarkeit

Die inhärente Verfügbarkeit basiert alleine auf der Fehlerverteilung und der Wiederherstellungsverteilung. Sie ist deshalb in erster Linie ein Verfügbarkeitswert, der die Wartungsfähigkeit/-freundlichkeit der Einheit ausdrückt. Beachtet wird hier aber nicht das gesamte Wartungsmanagement zur Verfügbarkeitssicherung. Um auch letztgenannten Aspekt zu berücksichtigen wird, kann der folgende Begriff definiert werden:

Mittlere Zeit zwischen Wartung und Instandsetzung MTBM, engl. mean time between maintenance,

- Die mittlere Zeit zwischen Wartung und Instandsetzung ist die durchschnittliche Zeitdauer bis zum ersten geplanten oder ungeplanten Ausfall einer Betrachtungseinheit.

Mittlere Wartungs- und Instandsetzungszeit , engl. mean active maintenance time,

$$\overline{M}$$

Formel 15: Symbol zur Darstellung der mittleren Wartungs- und Instandsetzungszeit

- Die Mittlere Wartungs- und Instandsetzungszeit ist die durchschnittliche Zeit für ungeplante Ausfälle und geplante Wartung, ohne Berücksichtigung von Zeiten für Logistik und Administration. Es wird vorausgesetzt, dass alle Mittel für die Instandsetzung und Wartung immer sofort (100 %) verfügbar sind.

Damit erhält man einen Ausdruck für die Verfügbarkeit, der für die Betrachtungseinheit in der optimal aufgestellten Umgebung erreichbar (achievable) ist. Sie berücksichtigt nur Eigenschaften der Betriebseinheit und der –umgebung.

achieved availability Aa,

- Die achieved availability Aa ist definiert als:

$$A_a = \frac{MTBM}{MTBM + \overline{M}}$$

Formel 16: Berechnungsformel für achieved availability

Falls Wartungsarbeiten schlecht organisiert oder zu häufig anfallen, können diese negativen Einfluss auf die achieved availability *Aa* haben, obwohl dadurch möglicherweise die MTBF erhöht wird. Die folgende Zeichnung stellt für einen speziellen Fall die Änderung der achieved availability *Aa* in Abhängigkeit von den Wartungsintervallen dar.

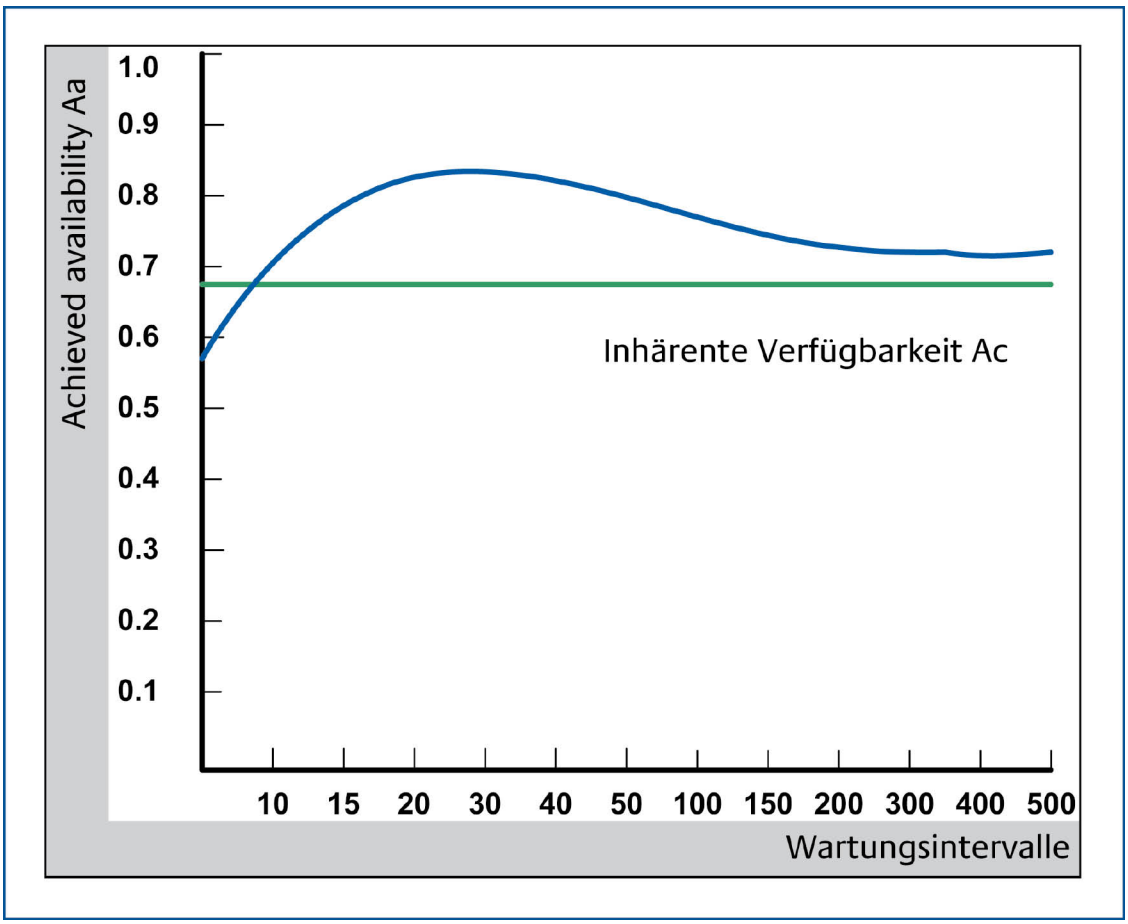


Abbildung 5: Abhängigkeit der achieved availability Aa von der Anzahl der Wartungen

Keine oder sehr seltene Wartung führen zu häufigen ungeplanten Ausfällen, sodass die Verfügbarkeit *Aa* niedriger ist als die inhärente Verfügbarkeit *Ai* .Wenn die präventive Wartungen zunehmen, erreicht die Verfügbarkeit *Aa* ein Maximum und nähert sich dann im Allgemeinen der inhärenten Verfügbarkeit *Ai* an. Ein Ziel von Verfügbarkeit getriebener Wartung ist es den „peak“ der Kurve herauszufinden und auf dieser Wartungshäufigkeit zu arbeiten. Es ist wichtig, den Verlauf dieser Kurve Aa zu kennen. Sonst ist es nicht möglich zu bestimmen, was angemessen und möglich ist, wie der Name dieses Verfügbarkeitsbegriffes betont.

Eine umfassende Sicht auf die Verfügbarkeit muss aber alle Quellen und Ursachen für Ausfälle, sowohl geplante, wie ungeplante, solche im technischen, wie auch im organisatorischen Umfeld enthalten.

Mittlere Instandhaltungszeit MDT engl. mean maintenance down time,

- Die mittlere Instandhaltungszeit MDT ist die mittlere Ausfallzeit für alle Maßnahmen, die sicherstellen, dass die Betrachtungseinheit funktionsfähig erhalten bleibt oder bei einem ungeplanten Ausfall wieder hergestellt wird.
- Sie umfasst die mittlere Wartungs- und Instandsetzungszeit sowie logistische und administrative Ausfallzeiten.

Die DIN-Norm DIN 31051 strukturiert die Instandhaltung in die vier Grundmaßnahmen Wartung, Inspektion, Instandsetzung und Verbesserung. Die mittlere Instandhaltungszeit MDT umfasst den gesamten zeitlichen Aufwand der Instandhaltung nach der DIN-Norm.

Operationale Verfügbarkeit Ao , auch Service Verfügbarkeit genannt, engl. operational availability,

Dieser Verfügbarkeitsbegriff ist insbesondere für die wirtschaftliche und administrative Bewertung von Prozessen relevant.

- Die operationale Verfügbarkeit Ao einer Betrachtungseinheit ist die Wahrscheinlichkeit, dass die Betrachtungseinheit alle zugesicherten Eigenschaften bei den beschriebenen Umgebungsbedingungen einhält oder fehlerfrei funktioniert.

$$A_o = \frac{MTBM}{MTBM + MDT}$$

Formel 17: operationale Verfügbarkeit

Wenn bei Abstimmungen zwischen unterschiedlichen organisatorischen Bereichen die Verfügbarkeit von z. B. Prozessen diskutiert wird, führt die unterbliebene Klarstellung, welche Art der (stationären) Verfügbarkeit Gegenstand der Absprache ist, häufig zu falschen Erwartungshaltungen.

Mittlere Verfügbarkeit , engl. average availability auch mean availability oder average up-time availability

Der Vollständigkeit halber wird hier noch ein Verfügbarkeitsbegriff erläutert, der im praktischen Umfeld nur in besonderen Fällen angewandt wird:

$$\overline{A(t)}$$

Formel 18: mittlere Verfügbarkeit

- Die mittlere Verfügbarkeit ist der Anteil der Zeit, während der die Betrachtungseinheit funktionierte an der Gesamtzeit, oder der Mittelwert der augenblickliche Verfügbarkeit A(t) über die Zeitperiode [0,t] und ergibt sich als

$$\overline{A(t)} = \frac{1}{t}\int_o^t A(t')dt' \text{ für } t>0.$$

Formel 19: Berechnung der augenblicklichen Verfügbarkeit über die Zeitperiode [0,t]

Damit erhält man ein Maß, wie sich die Gesamtgüte der Verfügbarkeit abhängig von der Lebenszeit verhält. Dieser Wert ist nützlich bei der Planung der Instandhaltung von Bedeutung. So z. B bei der Beantwortung der Frage, wann die Betrachtungseinheit ersetzt werden könnte.

2 Metrik und Beispiele zum Mythos der 9-en

Die zuvor definierten Begriffe der Verfügbarkeit dienten der eindeutigen Festlegung des Sachverhaltes. Ihre Beschreibung als Wahrscheinlichkeit liefert zugleich eine Metrisierung dieser Eigenschaften auf einer Skala zwischen 0 für nicht und 1 für immer verfügbar. Im nicht-mathematischen Umfeld wird der Wert allerdings üblicherweise als Prozentzahl zwischen 0 % und 100 % angegeben. Der Wert kann empirisch durch die Ermittlung der Ausfallhäufigkeit über längere Zeit oder von vielen Betrachtungseinheiten in Abhängigkeit von der Lebenszeit bestimmt werden (Black-Box-Test). Der analytische Ansatz aus der Ableitung der Verfügbarkeitswerte der einzelnen Teile der Betrachtungseinheit ist die Alternative. Dazu wird die Blackbox-Betrachtungseinheit aus Abbildung 1 rekursiv als Grey-Box analysiert, indem die Zuverlässigkeit der Gesamteinheit aus den Zuverlässigkeitswerten der einzelnen Teileinheiten unter Berücksichtigung ihrer funktionellen Abhängigkeit und statistisch verteilten Aktivierung ermittelt wird. Dieses rekursive Vorgehen ist so lange durchzuführen, bis für alle einzelnen Teileinheiten empirisch gesicherte oder von anderer Stelle dokumentierte Verfügbarkeitswerte vorliegen.

Zur Darstellung der Abhängigkeiten der einzelnen Teilkomponenten benutzt man sogenannte Zustandsblockdiagramme, wie in Abbildung 6. Die folgenden Beispiele mit Verfügbarkeitswerten und Blockdiagrammen sollen die Anschauung der zuvor definierten abstakten Begrifflichkeiten unterstützen. Dabei wird vorausgesetzt, dass die einzelnen Teile statistisch unabhängig sind. Das bedeutet praktisch, dass sie nicht gleichzeitig von einer Gefahr betroffen werden und sich im Fehlerfalle nicht gegenseitig beeinflussen.

Wenn mehrere Teilkomponenten i bis n nacheinander, also in Reihe oder Serie, geschaltet sind, müssen alle funktionieren, damit die gesamte Einheit X funktioniert:

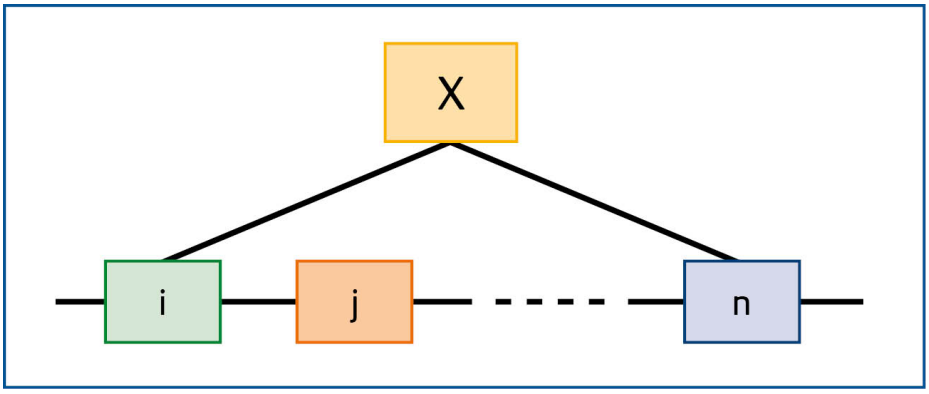


Abbildung 6: Zustandsblockdiagramm für die Anordnung der Teilkomponenten für Ax

Somit ist:

$$A_x = \prod_{a=i}^{n} A_a = A_i * A_j * \ldots * A_n$$

Formel 20: Berechnung der Teilkomponenten für Ax

Die folgende Tabelle zeigt einige Ergebnisse für die Gesamtverfügbarkeit Ax bei vorgegebenen Verfügbarkeiten Ai der Teilkomponenten i bis n, die in Reihe geschaltet sind:

Verfügbarkeit der Gesamteinheit	***Durchschnittliche Gesamtausfallzeit bei 7x24 Betrieb pro Jahr***	***Teilkomponente***					
		i	***j***	***k***	***l***	***m***	***n***
0,90000	36,5 Tage	0,9					
0,81000	69,4 Tage	0,9	0,9				
0,72900	99 Tage	0,9	0,9	0,9			
0,65610	125,6 Tage	0,9	0,9	0,9	0,9		
0,59049	149,6 Tage	0,9	0,9	0,9	0,9	0,9	
0,53144	171,1 Tage	0,9	0,9	0,9	0,9	0,9	0,9
0,99000	3,7 Tage	0,99					
0,98010	7,3 Tage	0,99	0,99				
0,97030	10,8 Tage	0,99	0,99	0,99			
0,96060	14,4 Tage	0,99	0,99	0,99	0,99		
0,95099	17,9 Tage	0,99	0,99	0,99	0,99	0,99	
0,94148	21,4 Tage	0,99	0,99	0,99	0,99	0,99	0,99
0,99900	8,8Std	0,999					
0,99800	17,5Std	0,999	0,999				
0,99700	1,1 Tage	0,999	0,999	0,999			
0,99601	1,5 Tage	0,999	0,999	0,999	0,999		
0,99501	1,8 Tage	0,999	0,999	0,999	0,999	0,999	
0,99401	2,2 Tage	0,999	0,999	0,999	0,999	0,999	0,999
0,99990	52,6Min	0,9999					
0,99980	1,8Std	0,9999	0,9999				
0,99970	2,6Std	0,9999	0,9999	0,9999			
0,99960	3,5Std	0,9999	0,9999	0,9999	0,9999		
0,99950	4,4Std	0,9999	0,9999	0,9999	0,9999	0,9999	
0,99940	5,3Std	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
0,99999	5,3Min	0,99999					
0,99998	10,5Min	0,99999	0,99999				
0,99997	15,8Min	0,99999	0,99999	0,99999			
0,99996	21Min	0,99999	0,99999	0,99999	0,99999		
0,99995	26,3Min	0,99999	0,99999	0,99999	0,99999	0,99999	
0,99994	31,6Min	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
0,99000	3,7 Tage	0,99					
0,99990	52,6Min	0,9999					
0,98990	3,7 Tage	0,99	0,9999				

Tabelle 1: Ergebnisse der Gesamtverfügbarkeit für serielle Verschaltung von A_x

Die Tabelle zeigt die Verfügbarkeit / Gesamtausfallzeit der Gesamteinheit bei gegebener Verfügbarkeit der einzelnen Komponenten i, j, …, n

Anmerkung:

Ein SI-Jahr entspricht 365,25 Tagen und somit 31557600 Sekunden.

Damit wird also gezeigt, dass sich die ergebende Verfügbarkeit, wenn mehrere Teilkomponenten i, j, ..., n mit vorgegebener Verfügbarkeit nacheinander, also in Reihe, funktionieren müssen. Bemerkenswert ist, wie stark die Verfügbarkeit abnimmt. Am Schluss der Tabelle wird demonstriert, wie gering die Gesamtverfügbarkeit der Einheit mit 0,9899 ist, wenn zwar hochverfügbares Teil (vier Neunen) mit einem normal Verfügbaren Teil (zwei Neunen) in Reihe geschaltet wird. Es bestätigt sich nicht nur das die Kette so stark wie ihr schwächstes Glied ist, genauer ist die Kette schwächer als ihr schwächstes Glied. Als vergleichbaren Trivialfall betrachten wir eine Einheit, die aus n gleichen Teileinheiten besteht. Die Gesamteinheit arbeitet nur dann einwandfrei, wenn alle Teilkomponenten fehlerfrei arbeiten. Eine konkrete Anwendung ist ein Chip mit mehreren Millionen Transistoren, bei dem noch keine automatische Fehlerlokation und –Behebung eingebaut ist. Die Frage ist, welche Verfügbarkeit müssen die alle identischen Transistoren besitzen, damit die Gesamtheit Chip die gewünschte Verfügbarkeit erreicht.

Weil:

$$A_{Gesamt} = \left(A_{Teil} \right)^{n},$$

Formel 21: Berechnung von A Gesamt

ist

$$A_{Teil} = \left(A_{Gesamt} \right)^{\frac{1}{n}}.$$

Formel 22: Berechnung von A Teil

Für mehrere Millionen (10, 20, ..50) Teileinheiten ist das Ergebnis in folgender Tabelle zusammengestellt:

Verfügbarkeit Gesamteinheit	***Duchschnittliche Gesamtausfallzeit in Tagen bei 7x24 Betrieb pro Jahr***	***Anzahl Teileinheiten in Millionen***				
		10	***20***	***30***	***40***	***40***
0,90	36,53	0,9999999895	0,9999999947	0,9999999965	0,9999999974	0,9999999979
0,91	32,87	0,9999999906	0,9999999953	0,9999999969	0,9999999976	0,9999999981
0,92	29,22	0,9999999917	0,9999999958	0,9999999972	0,9999999979	0,9999999983
0,93	25,57	0,9999999927	0,9999999964	0,9999999976	0,9999999982	0,9999999985
0,94	21,92	0,9999999938	0,9999999969	0,9999999979	0,9999999985	0,9999999988
0,95	18,26	0,9999999949	0,9999999974	0,9999999983	0,9999999987	0,9999999990
0,96	14,61	0,9999999959	0,9999999980	0,9999999986	0,9999999990	0,9999999992
0,97	10,96	0,9999999970	0,9999999985	0,9999999990	0,9999999992	0,9999999994
0,98	7,3	0,9999999980	0,9999999990	0,9999999993	0,9999999995	0,9999999996
0,99	3,65	0,9999999990	0,9999999995	0,9999999997	0,9999999997	0,9999999998

Tabelle 2: Ergebnis der Teileinheiten

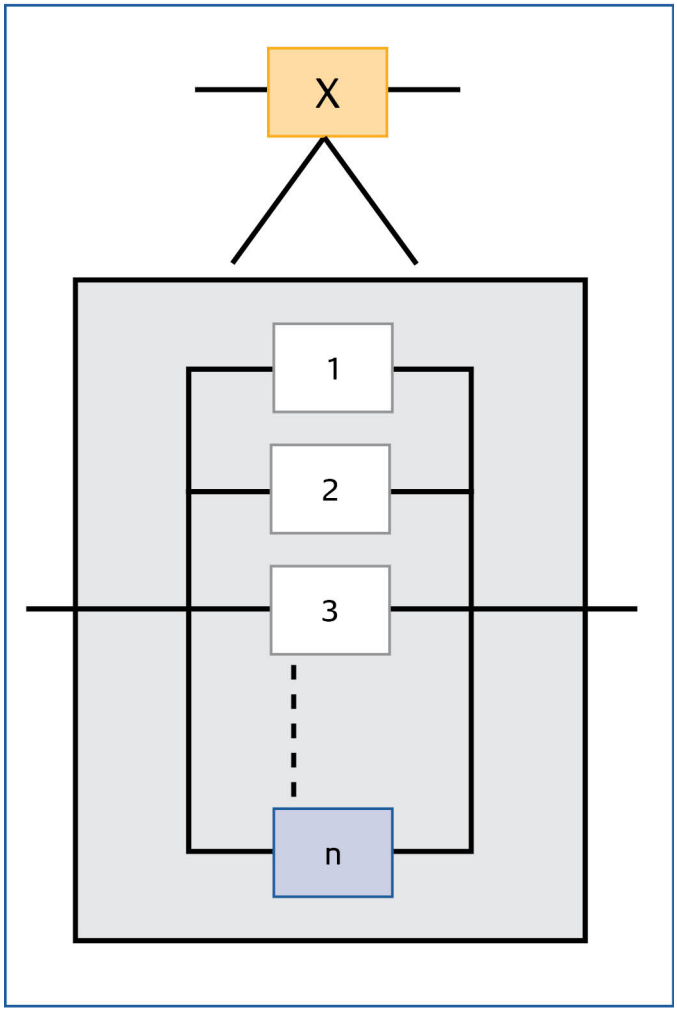


Abbildung 7: Zustandsblockdiagramm mit n Elementen in Parallelschaltung

Tabelle 2 zeigt die Verfügbarkeitsanforderung an jede Teilkomponente, abhängig von der Anzahl der Teileinheiten.

Anmerkung: Ein SI-Jahr entspricht 365,25 Tagen und 31557600 Sekunden.

Die Beispiele zeigen, wie extrem verfügbar die einzelnen Teileinheiten ausgelegt werden müssen, wenn diese nacheinander oder alle gleichzeitig funktionieren müssen, damit für die gesamte Einheit noch eine akzeptable Verfügbarkeit erreicht werden kann. Der Fertigungsaufwand für die Teileinheiten ist so hoch, dass ein fehlertolerantes Design mit Redundanzen wirtschaftlicher ist. Das zeigt das folgende Zahlenbeispiel.Sind zur Erfüllung der Funktionen einer Betrachtungseinheit nur k aus n Teilelementen notwendig, bilden n-k Elemente die Reserve. Bei einer solchen Struktur spricht man von Redundanz. Das Zustandsblockdiagramm wird in diesem Fall zu einer Parallelschaltung.

Die Wahrscheinlichkeit, dass genau k ausgewählte Komponenten intakt (die Komponenten 1,..,k), die anderen Komponenten defekt sind (die Komponenten k+1,...,n) ist:

$$P_{k\ aus\ n} = P_1 * P_2 * .....P_k * (1 - P_{k+1}) * (1 - P_{k+2}) * ... * (1 - P_n)$$

Formel 23: Wahrscheinlichkeit das k Komponenten intakt sind und die anderen Komponenten defekt sind.

Es gibt (n über i) Möglichkeiten, i Komponenten aus n Komponenten auszuwählen, also

$$P_{k\ aus\ n} = \sum_{i=k}^{n} \binom{n}{i} P^{i} * (1 - P)^{n-i}$$

Formel 24: Möglichkeit i Komponenten aus n Komponenten auszuwählen

Gehen wir von einer Redundanz 1 aus 2 aus: Die geforderte Funktion ist erfüllt, wenn mindestens eines der Elemente E1 oder E2 ausfallfrei arbeitet:

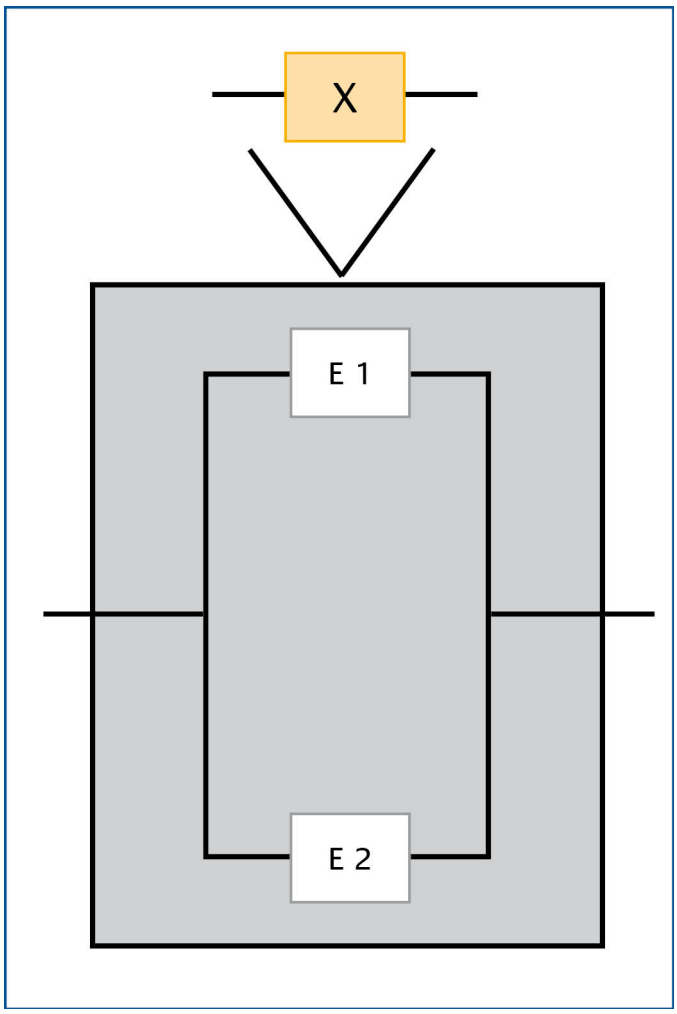


Abbildung 8: Redundanz 1 aus 2

Da E1 und E2 unabhängig voneinander arbeiten und ausfallen, erhalten wir:

$A = A_{E1} + A_{E2} - A_{E1} \times A_{E2}$

Der Spezialfall gleicher Elemente mit konstanter Ausfallrate führt zu:

MTBF = 3/2 MTBF.

Besteht eine Einheit aus zwei oder *n* Teilkomponenten, von denen nur eine alleine notwendig für die Gesamtaufgabe ist, so müssen alle Teilsysteme ausfallen, damit das Gesamtsystem ausfällt:

$1 - A = (1 - A_1) \times (1 - A_2) \times \ldots \times (1 - A_n)$

Unter der Annahme, dass die Teilsysteme alle gleich oder zumindest ähnlich hinsichtlich ihrer Einzelverfügbarkeit sind ($V_1 = V_2 = \ldots = V_n$), ergibt sich daraus für ein Cluster mit *n* Knoten:

Ages = 1 - (1 – Ai)n

Mit dem bisher Beschrieben sind hochverfügbare (weil fehlertolerante) Systeme analysierbar. Hardware kann z. B. durch Hinzufügen von Redundanz fehlertolerant gemacht werden. Laufen z. B. zwei Implementierungen einer Schaltung parallel (dual modular redundancy, DMR), so kann eine Entscheidungseinheit einen Fehler durch Vergleichen der Ausgänge der beiden Komponenten feststellen, jedoch nicht korrigieren. Fügt man eine weitere Instanz der Komponenten hinzu (triple modular redundancy, TMR), so kann eine Entscheidungseinheit einen Fehler korrigieren. Wird die fehlerhafte Einheit als defekt markiert, ist ein Fehler weiter erkennbar (wie bei DMR).

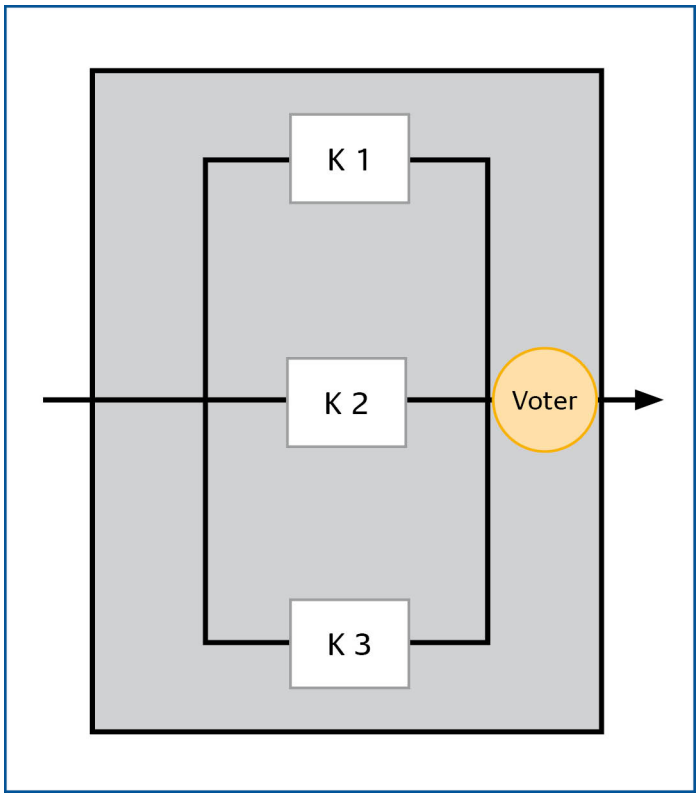


Abbildung 9: Zustandsblockdiagramm einer Triple Modularen Redundanz (TMR) mit Voter

Im TMR-Beispiel seien die Verfügbarkeiten der einzelnen Module K1, K2 und K3 gleich A. So ergibt sich die Gesamt-Verfügbarkeit als

$$A_{TMR} = (\quad A^3 + 3 A^2 \quad * (1 - A)) * A_{Voter}$$

Formel 25: Gesamt-Verfügbarkeit

mit A = 0,90 und AVoter = 0,99

erhält man A_{TMR} = 0,96.

Die vorstehend beschriebene Metrik ist rein statistischer Natur und eignet sich besonders für die Messung und Bewertung der Service-Qualität (Service-Delivery) von technischen IT-Ressourcen.

3 Prinzipielle Möglichkeiten zur Erhöhung der Verfügbarkeit

Die oben dargestellte Unterscheidung der stationären Verfügbarkeit unterbleibt in vielen Fällen und es wird, wie schon hergeleitet, einfach gleichgesetzt.

$$A_s = \lim_{t \to \infty} A(t) = \frac{MTTF}{MTTF + MTTR}$$

Formel 26: Gleichsetzen der Formeln

Aus dieser Vereinfachung wird deutlich, dass es zwei prinzipielle Möglichkeiten gibt, die Verfügbarkeit zu erhöhen:

Erhöhung der MTTF der Betrachtungseinheit

Verkürzung der MTTR der Betrachtungseinheit

Die zwei Ansätze sollen im Folgenden kurz beschrieben werden, wobei die Verfahren nicht immer eindeutig zugeordnet und auch nicht isoliert betrachtet werden können. Nur durch die Kombination mehrerer Möglichkeiten wird Hochverfügbarkeit erreicht.

3.1 Erhöhung der MTTF

Die mittleren Zeit bis zum Ausfall kann durch mehrere Ansätze erhöht werden, die in einem gesonderten Dokument / Kapitel ausführlicher als Prinzipien zur Erreichung von Hochverfügbarkeit erläutert sind: Mit den Eigenschaften der Fehlertoleranz und Robustheit wird Fehlern und Mängel von innen und sowie schwankenden Bedingungen der Umwelt und störenden Einflüssen von außen entgegengewirkt. Dabei erhöht sich die Lebenszeit bis zum Ausfall. Trotz aller Sorgfalt zur Absicherung gegen Fehler können solche nicht ausgeschlossen werden. Um dagegen vorzubeugen nutzt man redundante Implementierungen, weil die Wahrscheinlichkeit gering ist, dass negative Ereignisse gleichzeitig in den redundanten Komponenten auftreten. Bei großflächigen Schadensereignissen muss die Redundanz geographisch verteilt werden, bei globaler Ausdehnung kann durch heterogene Ausprägungen bzw. Diversität vorgebeugt werden. Die Zahlenbeispiele in den Tabellen zeigten, wie stark die Verfügbarkeit durch parallele Nutzung der redundanten Einheiten erhöht werden kann. Für weitere prinzipielle Ansätze zur Erhöhung der MTTF verweisen wir auf das Kapitel 7: „HV-Prinzipien“ im Band G des HV-Kompendium.

3.2 Verkürzung der MTTR

Ein Teil der zuvor beschrieben Ansätze sind auch wirksam zur Verkürzung der MTTR. So verringern redundante Implementierungen die Zeit für die Instandsetzung. Auch andere Prinzipien können den Ablauf der Wiederinbetriebnahme optimieren, wie Skalierbarkeit, Virtualisierung oder vereinfachen und absichern, wie Transparenz, Separation. Vorbeugende Wartung dient der Robustheit und Fehlertoleranz. In extremen Situationen sind nicht immer komplett vorher planbar und erfordern beim Eintreten besondere Entscheidungen, wobei eine vorbereitete Priorisierung die Entscheidungsfindung und damit die MTTR verkürzt. Auch diese Aufzählung ist an dieser Stelle nur exemplarisch.

3.3 Der Weg zur Hochverfügbarkeit

Muss eine so hohe Verfügbarkeit erreicht werden, dass der Prozess nahezu unterbrechungsfrei zur Verfügung steht, müssen alle relevanten Prinzipien sorgfältig aufeinander abgestimmt werden. Welche Prinzipien für Verfügbarkeit Relevanz besitzen wird ausführlich im HV-Kompendium V1.6 Band G, Kapitel G7: Prinzipien der Verfügbarkeit“ dargestellt. Die nachstehende Darstellung beschreibt den Zusammenhang zwischen Prinzipien der Verfügbarkeit und Potentialen zur Erhöhung der Verlässlichkeit.

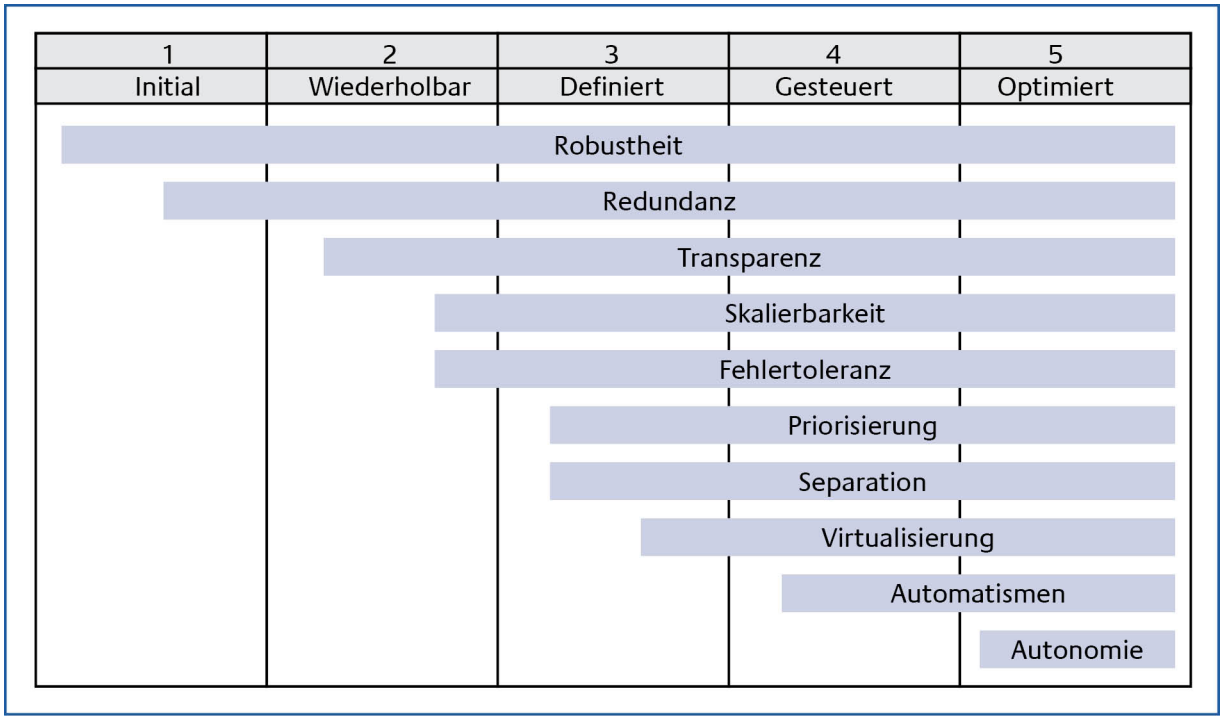


Abbildung 10: Prinzipien und Potentiale

Die Darstellung beschreibt, welche Prinzipien auf welcher Potentialstufe greifen, das bedeutet, dass Maßnahmen zur Automatisierung erst dann sinnvol wirken, wenn robuste und redundante Komponenten vorhanden sind. Es ist die Kreativität im Design-Prozess, die alle Maßnahmen sinnvol kombiniert, um zur Erhöhung der MTTF und zur Verringerung der MTTR beizutragen. Um das Ziel der Transparenz zu erreichen, muss die MTTR nahezu gegen 0 streben. Das bedeutet aber, dass manuelle Entscheidungen und Handlungen zur Instandsetzung nicht erforderlich sind. Die Operationen zur Wiederherstellung – ohne die möglichen Schadensereignisse zu kennen – sind alle geplant und laufen automatisch ab, einschließlich des Prozesses zur Fehlererkennung. Damit dieses für die Hochverfügbarkeit charakteristische Prinzip der Automatismen allerdings wirken kann, muss insbesondere das Prinzip der Redundanz abhängig von den zulässigen Umgebungsbedingungen in den verschiedenen Ausprägungen umgesetzt werden.

Eine noch höhere Steigerung der Verfügbarkeit strebt das Prinzip der Autonomie an. So sollen Prozesse einen besonderen Grad zur selbstständigen Erkennung von drohenden Ereignissen zur Verfügbarkeitsminimierung erlangen und über die Kompetenz verfügen, vorbeugend oder mindestens reaktiv Entscheidungen zur Abwendung eigenständig zu treffen und umzusetzen, sich also für Ihren Bestimmungszweck selbst zu verwalten.

Die Ansätze zum Design hochverfügbarer Prozesse mit ihren ganz spezifischen Methoden und höheren Aufwendungen sind mit Anforderungen der Sicherheit oder Schadensauswirkung erst ab

maximal zulässigen Ausfallzeiten im Minutenbereich wirtschaftlich zu rechtfertigen, also ab 99,99% oder anders ausgedrückt: ab vier Neunen.

4 Klassifikation Verfügbarkeitskassen / Systematik / Typifikation

Wie zuvor dargelegt hat die Verfügbarkeit als Attribut einer Betrachtungseinheit einen Wert zwischen 0 und 1. In der Praxis hat sich die Einteilung in 6 Verfügbarkeitskassen bewährt, wobei die Anzahl der Neunen zur Unterteilung dient (siehe Tabelle 3). Die Übergänge sind fließend, aber dieses Unterscheidungsmerkmal grenzt die Anforderungen und Erwartung an die Verfügbarkeit und Methoden zur Qualitätserreichung im Allgemeinen sehr gut gegeneinander ab. Eine Abbildung auf den IT-Grundschutz ist in Band 1, Kapitel 1: Einführung des HV-Kompendiums tabellarisch dargestellt.

Verfügbarkeitsklasse (VK)	***Bezeichnung Betrachtungseinheit, Prozess, System, Einheit, Komponente***	***Mindestverfügbarkeit***	***Ausfallzeit pro Monat[#□]***	***Ausfallzeit pro Jahr[#□]***
VK 0	Ohne zugesicherte Verfügbarkeit			
VK1	Normale Verfügbarkeit	99,0%	< 8 h	< 88 h
VK 2	Erhöhte Verfügbarkeit	99,9%	< 44 min	< 9 h
VK 3	Hochverfügbarkeit	99,99%	< 5 min	< 53 min
VK 4	Höchstverfügbar	99,999%	< 26 s	< 6 min
VK 5	Verfügbarkeit unter extremen Bedingungen / auch bei höherer Gewalt (Disaster-Tolerant)			
[#□]bei 7 x 24 Std. Betriebszeit				

Tabelle 3: Verfügbarkeitsklassen

Die Betrachtungseinheiten der Klasse VK 0 (ohne zugesicherte Verfügbarkeit) sichern keine Verfügbarkeit zu oder besitzen keine Maßnahmen um Verfügbarkeit zu gewährleisten. In diese Gruppe fallen sehr viele Vorgänge o. ä., bei denen die Datenintegrität nicht essentiell ist. Häufig sind sich die Verantwortungsträger dieses Risikos nicht bewusst. Nur irgendwann trifft das Schadensereignis ein, mit Sicherheit, wie bei der Diskussion über inhärente Eigenschaften oben. Bei Betrachtungseinheiten der Klasse VK 1 (normale Verfügbarkeit) werden längere Ausfallzeiten akzeptiert, aber nach durchaus längerer Zeit erfolgt eine komplette Wiederinstandsetzung und Wiederanlauf. Um dieses gewährleisten zu können, muss die Datenintegrität gegeben sowie weitere Schutzmaßnahmen vorgesehen sein. Die Verfügbarkeit ist ein wichtiges Kriterium, aber nicht das herausragende Funktions- und Qualitätsmerkmal. An die Verfügbarkeit werden bei der Klasse VK2 (erhöhte Verfügbarkeit) deutlich erhöhte Anforderungen gestellt, aber kurze Unterbrechungen sind durchaus akzeptabel. Um diese Qualität erreichen zu können, müssen die entsprechenden Maßnahmen zur Instandsetzung komplett geplant und vorbereitet sein. Die Einheiten der Verfügbarkeitsklasse VK3 (hochverfügbar) stehen nahezu unterbrechungsfrei zur Verfügung. Kurze Ausfälle in der Summe von durchschnittlich 5 Minuten pro Monat werden bei der Nutzung kaum wahrgenommen und beeinträchtigen das Ergebnis nicht. Normale Maßnahmen reichen für so hohe Qualitätsziele nicht aus, sondern müssen mit Automatismen zur Zustandsüberwachung und zur Wiederherstellung in Verbindung mit betriebsbereiten Redundanzen abgesichert werden. So kurze Zeitfenster zur Instandsetzung erlauben keine Prozesse zur Kontrolle und Entscheidungsfindungen von Außen. Anders ausgedrückt, die Zustandsanalyse und die Wiederanlaufsteuerung wurde bei der Konzeption vorausgedacht und erfolgt ohne Ursachenermittlung selbsttätig von Innen durch die Einheit. Ab dieser Klasse wird üblicherweise von Hochverfügbarkeit gesprochen. Die extrem hohe Verfügbarkeit der Klasse VK4 erlaubt nur noch ganz geringe Unterbrechungen, die in der Nutzung

praktisch nicht mehr feststellbar sind: Die Funktionalitäten stehen ununterbrochen zur Verfügung. Diese Anforderung ist nur mit ganz besonders aufwendigen Methoden der Hochverfügbarkeit erreichbar, z. B. mit gleichzeitig für die gleiche Funktion arbeitender Redundanz, also durch Parallelisierung. Bei den Betrachtungseinheiten der letzten Klasse VK5 steht die Verfügbarkeit unter besonderen Bedingungen oder im Falle höherer Gewalt im Vordergrund, man spricht auch von „Disaster Tolerant“ . Die maximale Ausfallzeit kann hier kaum vorherbestimmt werden, weil die extremen Bedingungen nicht bekannt, nicht genau vorstellbar und nicht eingrenzbar sind. Die Einheiten müssen unter „allen möglichen Umständen möglichst lange“ funktionieren, um die Schäden zu minimieren.

Anhang: Verzeichnisse

Abkürzungen und Akronyme

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