Monte-Carlo Simulation mit dem Tool MC-ECO anhand des Beispiels „Verfügbarkeit eines Contact-Center-Services“

Herausgeber

ESG Consulting GmbH
Livry-Gargan-Straße 6
82256 Fürstenfeldbruck

Autor

Dr. Peter Merz
peter.merz@esg-consulting.com

Kontakte

Matthias Reimann
Tel.: +49 (0)89 92161-2802
E-Mail: matthias.reimann@esg-consulting.com

Ulrich Bethäuser
Tel.: +49 (0)89 92161-2517
E-Mail: ulrich.bethaeuser@esg-consulting.com

Inhalt

Abbildungsverzeichnis

Tabellenverzeichnis

Einleitung

Mit diesem Dokument wird nach einer kurzen Einführung in das Thema System-Verfügbarkeit ein Simulations-Szenario zur Ermittlung und Bewertung von durch Störungen in einem Call-Center-System möglicherweise entstehenden finanziellen Schäden beschrieben. Die Call-Center-Software soll auf den Services von „Skype for Business“ (SfB) aufsetzen.
Die Simulationen werden mit dem Tool MC-ECO durchgeführt.
Kennzahlen zur Systemverfügbarkeit sowie die Kostenseite der Nicht-Verfügbarkeit eines Systems werden in dem Artikel „Maschinen-Ausfall Szenario“ beschrieben.

Das Contact-Center System

Das traditionelle Callcenter weicht zunehmend dem digitalen Contact Center, das individuell auf Kundenwünsche reagieren und mit Kunden crossmedial und in Echtzeit agieren kann. Im Vergleich zu einem Call-Center werden beim Contact-Center die unterschiedlichen Kommunikationskanäle (Audio, Video, Konferenzen, Desktop-Sharing, E-Mail, Chat, Messenger oder Social Meida) gebündelt und integriert.
Haben Unternehmen Skype for Business (SfB) bereits im Einsatz oder planen eine Implementierung, können diese durch nützliche Erweiterungen zusätzliche Contact-Center-Funktionalitäten nativ in ihre SfB-Umgebung einbinden. So ist es beispielsweise möglich, sämtliche SfB-Kanäle in Gruppen, beziehungsweise Teams für Chat, Voice und Video zu organisieren sowie Echtzeit-Statistiken und aussagekräftige Reportings über Teamaktivitäten zu erstellen. Dadurch kann nicht nur eine verbesserte Erreichbarkeit für Kundenanfragen sichergestellt werden. Auch die Verfügbarkeit des Services lässt sich durch dynamische Statusanzeigen und intelligentes Routing deutlich erhöhen.
In Abbildung 1 ist eine vereinfachte Architektur des Systems aufgezeichnet. Externe Nutzer und Kunden können sich via PSTN oder VoIP an das Unternehmen wenden. Mitarbeiter können sich remote einwählen und darüber die SfB- und Contact-Center Funktionalitäten nutzen. Das System besteht daher aus einer Demilitarisierten Zone (DMZ), welche nach außen wie nach innen über jeweils eine Firewall abgesichert ist.
Die Anrufe (Voice, Video) gehen im Session Border Controller (SBC) ein. SBC leitet eine Anrufanfrage per SIP an den Edge-Server-Pool weiter. Da mit über 20.000 parallelen Verbindungen gerechnet wird, wird ein Pool aus 2 SBCs aufgesetzt. Bei der unternehmenskritischen Kommunikation sollten alle „Single Points of Failure“ vermieden werden. Viele SBCs lassen sich heute in Hochverfügbarkeits-Konfigurationen betreiben. Dadurch lassen sich zwei SBCs logisch als eine Einheit zusammenschalten. Ein SBC verarbeitet aktiv die aktuellen SIP-Sessions während der andere SBC im Standby-Modus wartet.

SfB:

Ein Pool von Edge-Servern ist notwendig, wenn externe Benutzer, die nicht beim internen Netzwerk der Organisation angemeldet sind, in der Lage sein müssen, mit internen Benutzern zu interagieren. Zu externen Benutzern gehören authentifizierte und anonyme Remotebenutzer, Verbundpartner oder andere mobile Kunden. Durch die Bereitstellung von Edge Server werden auch Mobilitätsdienste aktiviert, die die lync-Funktionalität auf mobilen Geräten unterstützen. Benutzer können mit unterstützten mobilen Geräten (Apple iOS, Android, Windows Phone oder Nokia) Aktionen wie Senden und Empfangen von Chatnachrichten, Anzeigen von Kontakten und Anzeigen der Anwesenheit ausführen. Die Anfrage wird weitergeleitet an den SfB Frontend Server-Pool.

Bei einem Front-End-Pool handelt es sich um einen Satz von Front-End-Servern, die identisch konfiguriert sind und zusammenarbeiten, um Dienste für eine gemeinsame Gruppe von Benutzern bereitzustellen. Ein Pool mit mehreren Servern, auf denen dieselbe Rolle ausgeführt wird, bietet Skalierbarkeit und Failover-Funktionen. Die Frontend-Server stellen u.a. zur Verfügung: Benutzerauthentifizierung und -registrierung, Anwesenheits-informationen und Visitenkartenaustausch, Webkonferenzen, PSTN-Einwahlkonferenzen und A/V-Konferenzen.

Die Back-End-Server sind Datenbankserver mit Microsoft SQL Server, die die Datenbankdienste für den Front-End-Pool bereitstellen. Die Back-End-Server dienen als Sicherungsspeicher für die Benutzer- und Konferenzdaten des Pools und sind die primären Speicher für andere Datenbanken.

Der Mediation-/Vermittlungs-Server übersetzt die Signalübertragung zwischen der internen Enterprise-VoIP-Infrastruktur und einem PSTN-Gateway (Public Switched Telephone Network) oder einem SIP-Stamm (Session Initiation Protocol).

Contact-Center:

Bei Kontaktaufnahme eines Kunden leitet der Mediation-Server wiederum die Signalisierung an den Contact-Center, und zwar an den Connector weiter. Die Anfrage wird an den Trusted Application Server weitergeleitet. Eine vertrauenswürdige Anwendung ist eine Anwendung, die auf Microsoft Unified Communications Managed API (UCMA) 3.0 Core SDK basiert und von SfB Server als vertrauenswürdig eingestuft wird. Über die UCMA können Entwickler kommunikations- und kollaborationsfähige Dienste implementieren.

Die Workflow Engine ist verantwortlich für die Steuerung der Self-Service-Funktionen beim Kundenkontakt und für die Weiterleitung des Kontakts zur Routing Engine mit allen relevanten Informationen (z. B. Skills und Eigenschaften), damit ein passender und verfügbarer Agent für diesen Kontakt gefunden werden kann. Ein spezielles Controller-Plug-In sorgt dafür, dass die Kommunikationskanäle in SfB über die UCMA-Schnittstelle gesteuert werden. Dieser Controller verwaltet auch alle ausstehenden Kontakte in der Warteschlange. Zudem stellt die Workflow Engine die Schnittstellen für Echtzeit-Überwachung von sämtlichen Statistiken zur Verfügung. Damit ist die Administrierbarkeit und Messbarkeit im Contact Center garantiert

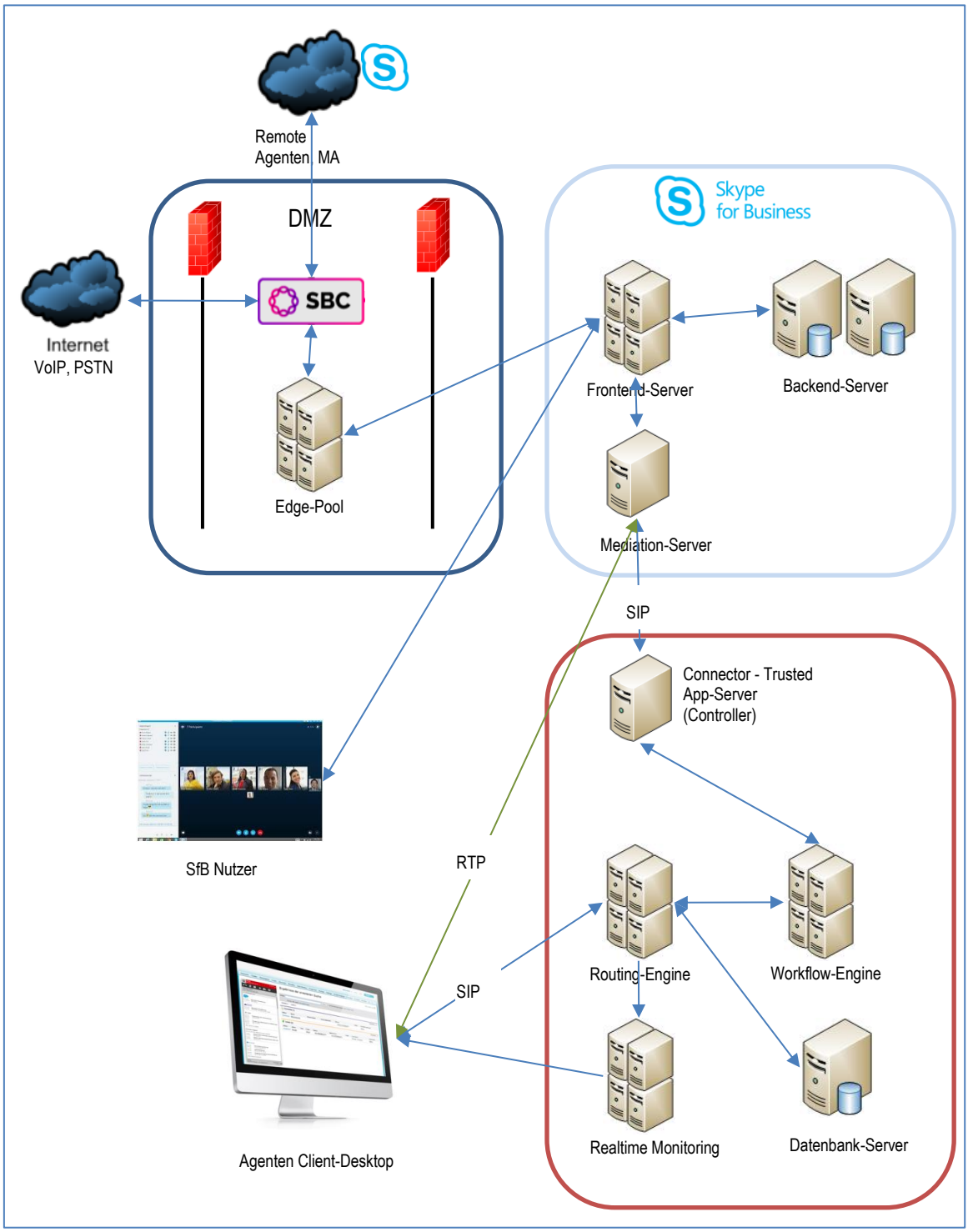

Abbildung 1: vereinfachte Architektur SfB - Contact-Center

Die Routing Engine benötigt Statusinformationen, um Routingentscheidungen treffen zu können. Wenn z. B. ein Agent außerhalb des Contact Center telefoniert, benötigt die Routing Engine diese Information, damit diesem Agenten keine Sprachkontakte zugewiesen werden.

Sobald eine Verbindung zwischen Agent und Kunde aufgebaut ist, werden die Datenströme (Audio, Video) direkt zwischen Client und Media-Server via RTP ausgetauscht.

Ausfallsicherheit

Die technische Aufrechterhaltung, Verfügbarkeit bzw. Ausfallsicherheit der Services ist ein wichtiger Aspekt für eine jede Organisation. Durch Ausfallzeiten können hohe finanzielle Schäden für eine Organisation entstehen. Im Rahmen des Risikomanagements sollte daher auch ein Informationssicherheitsmanagement (ISMS) betrieben werden. Hierbei bewertet eine Organisation die Kombination aus der Eintrittswahrscheinlichkeit und dem potenziellen Auswirkungsschaden eines Ausfalls eines Services. Ist dieser potenzielle Schaden hinreichend hoch, sind Maßnahmen erforderlich, die die Eintrittswahrscheinlichkeit und/oder den Schaden auf ein vertretbares Restrisiko reduzieren.

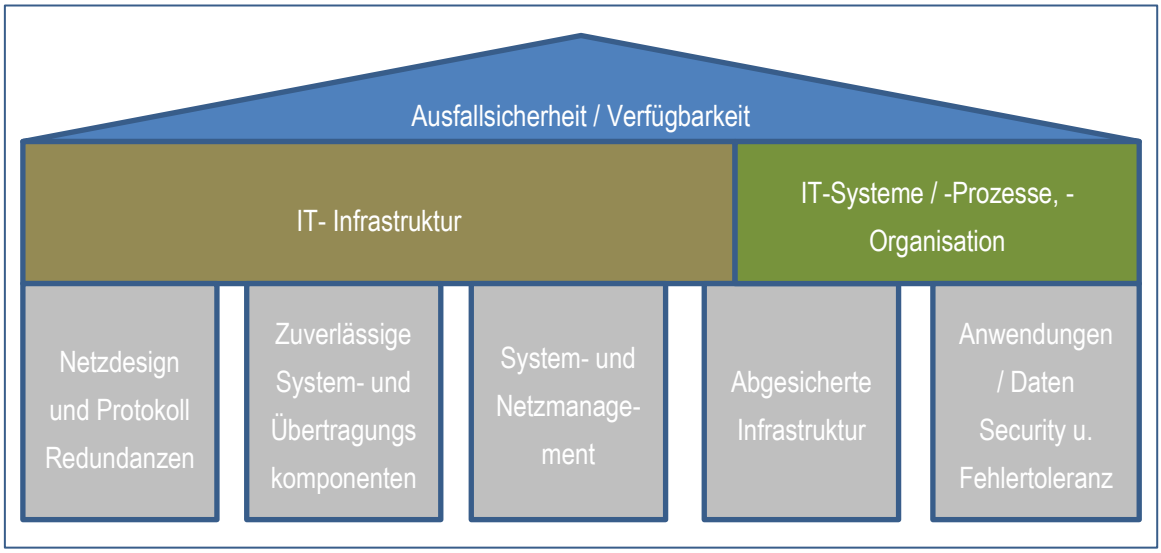

Abbildung 2: Grundsäulen der Ausfallsicherheit von IT-Systemen

Simulation der Systemverfügbarkeit

Das Thema Cyber-Angriffe wird hier bei der Betrachtung der Verfügbarkeit nicht berücksichtigt, obwohl dies natürlich ein wichtiger Aspekt ist. Dieser Aspekt müsste in einem echten Szenario berücksichtigt werden. Weiterhin wird hier nicht die Netzverfügbarkeit des Providers berücksichtigt. In diesem Szenario geht es ausschließlich um die Gesamtverfügbarkeit des Systems bestehend aus Hardware und Software, die für den Betrieb von SfB und den Contact-Center benötigt wird. Switches, Router oder Loadbalancer werden in der Simulation aus Übersichtlichkeitsgründen nicht berücksichtigt. Weiterhin werden hier keine geplanten „Downtimes“ berücksichtigt.

1.1 Parametrisierungen der Knoten

Knoten	Verfügbarkeit	Typ
Firewall 1	99,99%	2 Stück als Appliance
SBC	99,99%	2 Stück als Appliance
Edge-Pool	99,0% pro Server	Pool aus 4 Edge-Servern (Parallel-Schaltung)
Firewall 2	99,99%	2 Stück als Appliance
SfB Frontendserver	99,0% pro Server	Pool aus 4 Servern (Parallel-Schaltung)
SfB Backendserver	99,2% pro Server	Pool aus 2 Servern (Parallel-Schaltung)
SfB Mediation-Server	99,0%	1 Server
Trusted App (Connector)	99%	1 Server
Workflow-Engine	99%	Pool aus 4 Servern (Parallel-Schaltung)
Routing-Engine	99%	Pool aus 4 Servern (Parallel-Schaltung)
DB-Server	99,7%	1 Server
Realtime-Monitoring	95%	Pool aus 2 Servern (Parallel-Schaltung)

Tabelle 1: Server-Verfügbarkeiten

Durch einen Gesamtausfall des Systems soll ein geschätzter Schaden von 30.000 EUR pro Stunde entstehen. Davon wären weder Kunden-Anfragen (über den Contact-Center) noch interne Gespräche / Videokommunikation / Konferenzen über SfB möglich. Wenn nur der Contact-Center ausfällt, dann soll ein Gesamtschaden von 20.000 EUR pro Stunde entstehen.

Sobald 1 Knoten bis hin zum Mediation-Server nicht mehr zur Verfügung steht, besteht ein Gesamtausfall. Wenn ein Knoten innerhalb des Contact-Centers ausfällt, dann steht nur der Contact-Center nicht mehr zur Verfügung. Der Ausfall eines Servers sei hierbei unabhängig vom Ausfall eines anderen, d.h. ein jeder Knoten, sei es ein einzelner Server oder ein Pool von Servern als Parallelschaltung kann zu einem Gesamtausfall der Services führen. Weiterhin soll gelten, dass bei einem Ausfall der Firewall-1 (äußere Firewall) ein Schaden von 80.000 EUR pro Stunde entsteht, da neben dem Contact-Center noch mehrere andere Services vom Betrieb der Firewall abhängen. Gleiches gilt für die innere Firewall (Nr. 2), jedoch soll hier beim Ausfall von 1 Stunde „lediglich“ ein Schaden von 60.000 EUR entstehen.

Als Grundlage für die Berechnungen der Ausfallzeiten, Ausfallhäufigkeiten und der damit einhergehende jeweilige Schaden dient das Excel-Sheet „ContactCenter_VerfügbarkeitsRechner.xlsx“. Mit diesem Sheet lassen sich die Serien- und Parallel-Schaltungen der Server berechnen, ausgehend von der Verfügbarkeit eines Servers / eines Pools können damit die Ausfallzeiten (MTTR) berechnet werden. Abhängig von der Reparatur-Zeit eines Ausfalls kann somit weiterhin die Häufigkeit eines Ausfalls pro Jahr berechnet werden.

Gesamtbetriebszeit	8760	Gesamt-MTTR (h)	96,9209904	Schaden bei 1 Stunde Ausfallzeit:		Firewall-1:	80.000 €
						Firewall-2:	60.000 €
Gesamt-Verfügbarkeit	Gesamt-MTTR	Gesamt-Schaden	Nichtverfügbarkeit	1,102305%		SfB-Nutzer:	10.000 €
98,897695%	96,56	1.624.287,96 €	Durchschn. Verfügbark	99,907800%		Agenten (Contact-Center):	20.000 €
98,893596%	96,92		Durchschn. Nichtverf.	0,092200%		Zusammen:	30.000 €
für MTTR << 1							

Komponente	Verfügbarkeit	Nicht-Verfügbarkeit	MTTR (h)	MTBF (h)	λ (Ausfallrate)	MTTR (min)	MTTR (sec)	Schaden / Jahr	Reparatur-zeit (h)	Ausfälle/Jahr	Wahrscheinlichkeit 1es Ausfalls/Jahr	Schaden je Ausfall
Firewall_1	99,990000%	0,010000%	0,88	8759,12	0,000114167	56,06	3.363,84	70.080,00 €	2	0,438000	0,120000%	160.000 €
SBC	99,990000%	0,010000%	0,88	8759,12	0,000114167	56,06	3.363,84	26.280,00 €	2	0,438000	0,120000%	60.000 €
Edge-Pool	99,999999%	0,000001%	0,00	8760,00	0,000114155	0,01	0,34	2,63 €	2	0,000044	0,000012%	60.000 €
Firewall_2	99,990000%	0,010000%	0,88	8759,12	0,000114167	56,06	3.363,84	52.560,00 €	2	0,438000	0,120000%	120.000 €
SfB Frontendserver Pool	99,999999%	0,000001%	0,00	8760,00	0,000114155	0,01	0,34	2,63 €	4	0,000022	0,000006%	120.000 €
SfB Backendserver Pool	99,993600%	0,006400%	0,56	8759,44	0,000114163	35,88	2.152,86	16.819,20 €	4	0,140160	0,038400%	120.000 €
SfB Mediation-Server	99,980000%	0,020000%	1,75	8758,25	0,000114178	112,13	6.727,68	52.560,00 €	3	0,584000	0,160000%	90.000 €
Trusted App (Connector)	99,500000%	0,500000%	43,80	8716,20	0,000114729	2.803,20	168.192,00	876.000,00 €	6	7,300000	2,000000%	120.000 €
CC Workflow Engine Pool	99,999999%	0,000001%	0,00	8760,00	0,000114155	0,01	0,34	1,75 €	8	0,000011	0,000003%	160.000 €
CC Routing Engine Pool	99,999999%	0,000001%	0,00	8760,00	0,000114155	0,01	0,34	1,75 €	8	0,000011	0,000003%	160.000 €
CC DB-Server	99,700000%	0,300000%	26,28	8733,72	0,000114499	1.681,92	100.915,20	525.600,00 €	8	3,285000	0,900000%	160.000 €
CC Realtime-Monitoring	99,750000%	0,250000%	21,90	8738,10	0,000114441	1.401,60	84.096,00	4.380,00 €	10	2,190000	0,600000%	2.000 €

Abbildung 3: Sheet zum Berechnen der Ausfallzeiten, Ausfallhäufigkeiten und Schäden

Anmerkung zur Wahl einer Verteilungsfunktion eines Risikos in MC-ECO:

- Die Häufigkeit eines Ausfalls pro Jahr richtet sich nach der Verfügbarkeit (in %) und der Zeit zur Reparatur bei einem Ausfall
- Ist die Ausfallhäufigkeit kleiner als 1 pro Jahr, dann wird eine Gleichverteilung angenommen
- Ist die Häufigkeit von Ausfällen pro Jahr >= 1, dann wird die Häufigkeitsverteilung benutzt.

1.2 Risiko-Definitionen der Knoten

1.2.1 Firewall 1

Entspricht dem Risiko #1 im Szenario „ContactCenter_Szenario.xlsx“. Da die Firewall-1 auch von weiteren Services benutzt wird, wird hierfür bei einem Ausfall von einem Schaden von 80.000 EUR pro Stunde Ausfall ausgegangen.

- Verfügbarkeit: 99,99%
- MTTR: 0,88 Std., Mittlerer Schaden / Jahr: 70.080 EUR, Reparaturzeit von 2 Stunden
- Ausfälle / Jahr: 0,438
- Verteilungsfunktion: Gleichverteilung mit
  - Eintrittswahrscheinlichkeit: 43,8%
  - Schadensfunktion: Dreiecksverteilung mit 160.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 70.080 / 0,438 EUR), 120.000 EUR Minimalschaden und 200.000 EUR Maximalschaden

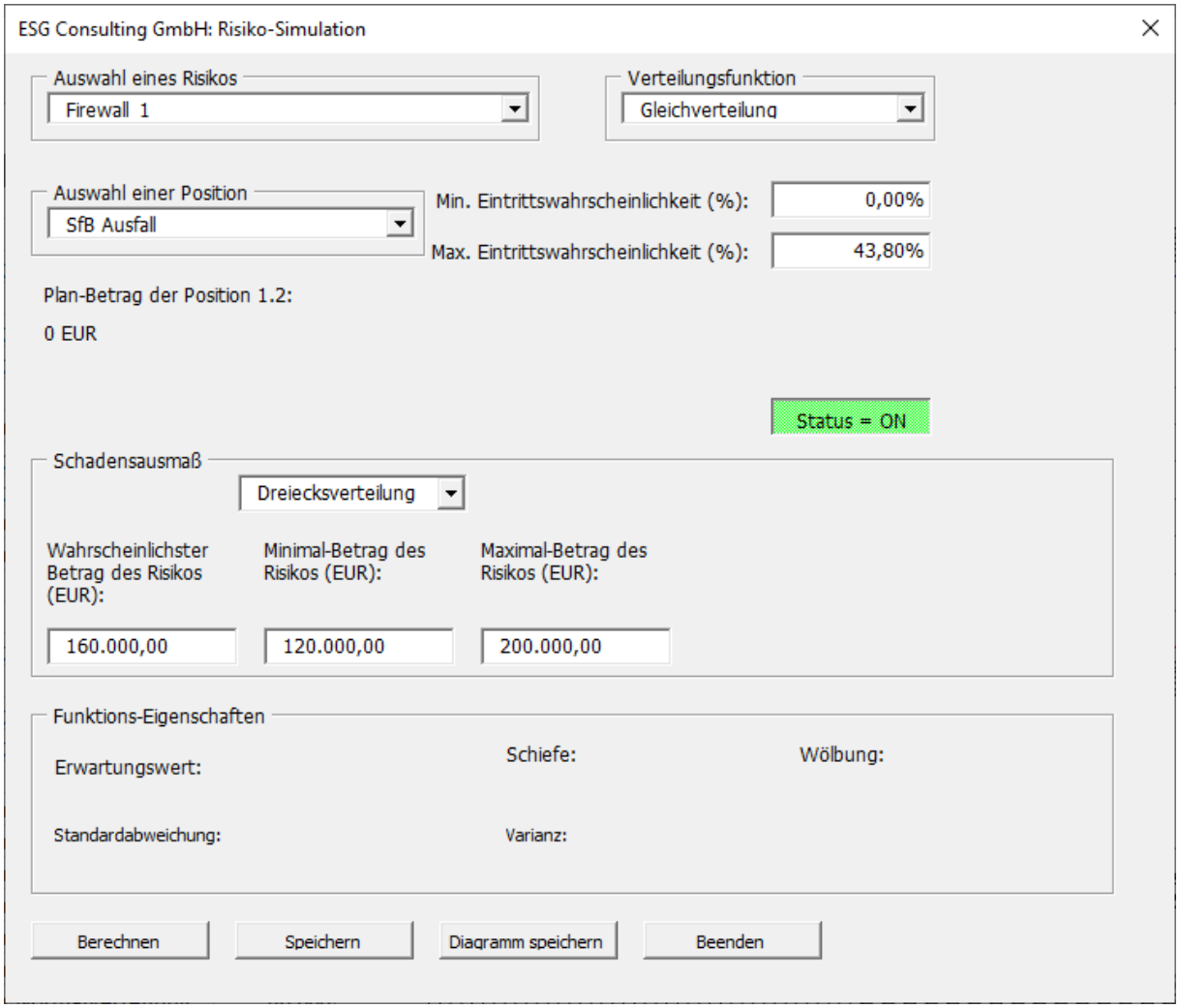

Abbildung 4: Konfiguration des Risikos für Firewall_1

1.2.2 SBC-Appliance

Entspricht dem Risiko #2 im Szenario „ContactCenter_Szenario.xlsx“. Da die SBC-Appliance nur von SfB und dem Contact-Center benutzt wird, wird hierfür bei einem Ausfall von einem Schaden von 30.000 EUR pro Stunde Ausfall ausgegangen.

- Verfügbarkeit: 99,99%
- MTTR: 0,88 Std., Mittlerer Schaden / Jahr: 26.280 EUR, Reparaturzeit von 2 Stunden
- Ausfallhäufigkeit / Jahr: 0,438
- Verteilungsfunktion: Gleichverteilung mit
  - Eintrittswahrscheinlichkeit: 43,8%
  - Schadensfunktion: Dreiecksverteilung mit 60.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 26.280 / 0,438), 40.000 EUR Minimalschaden und 80.000 EUR Maximalschaden

1.2.3 Edge-Pool

Entspricht dem Risiko #3 im Szenario „ContactCenter_Szenario.xlsx“. Da der EDGE-Pool nur von SfB und dem Contact-Center benutzt wird, wird hierfür bei einem Total-Ausfall von einem Schaden von 30.000 EUR pro Stunde Ausfall ausgegangen. Der Pool besteht aus 4 gleichartigen Servern mit einer Verfügbarkeit je Server von 99%.

Bei einem Pool kann man die Häufigkeit mehrerer gleichzeitiger Ausfälle zuerst mittels der Binomialverteilung berechnen.

- 4 gleichartige Server je Pool
- Verfügbarkeit 1 Servers: 99%
- Gesamt-Verfügbarkeit: 99,999999%
- Bei Gesamtausfall aller 4 Server im Pool: MTTR: 0,34 sec., Mittlerer Schaden / Jahr: 2.63 EUR, Reparaturzeit von 2 Stunden
- Gesamt-Ausfälle / Jahr: 0,00004
- Schaden je Total-Ausfall: 60.000 EUR

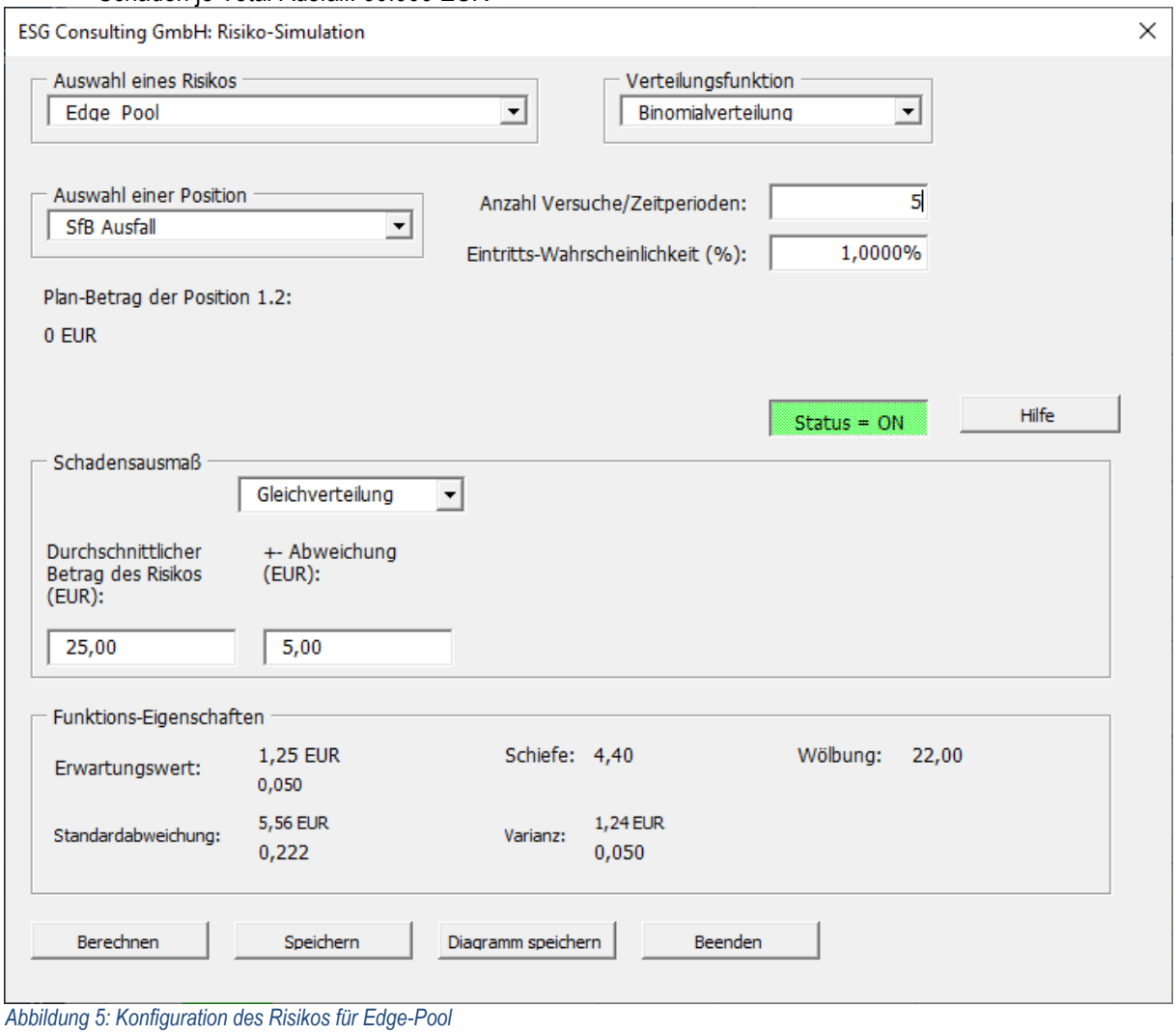

Abbildung 5: Konfiguration des Risikos für Edge-Pool

Der erste Schritt wäre hier:

- Verteilungsfunktion: Binomialverteilung
- Anzahl Versuche: 5 (Pool, bestehend aus 4 gleichartigen Servern)
- Eintrittswahrscheinlichkeit: 1,00% (je Server)
- Schadensfunktion: interessiert in diesem Falle nicht

Dies ergibt eine Verteilung:

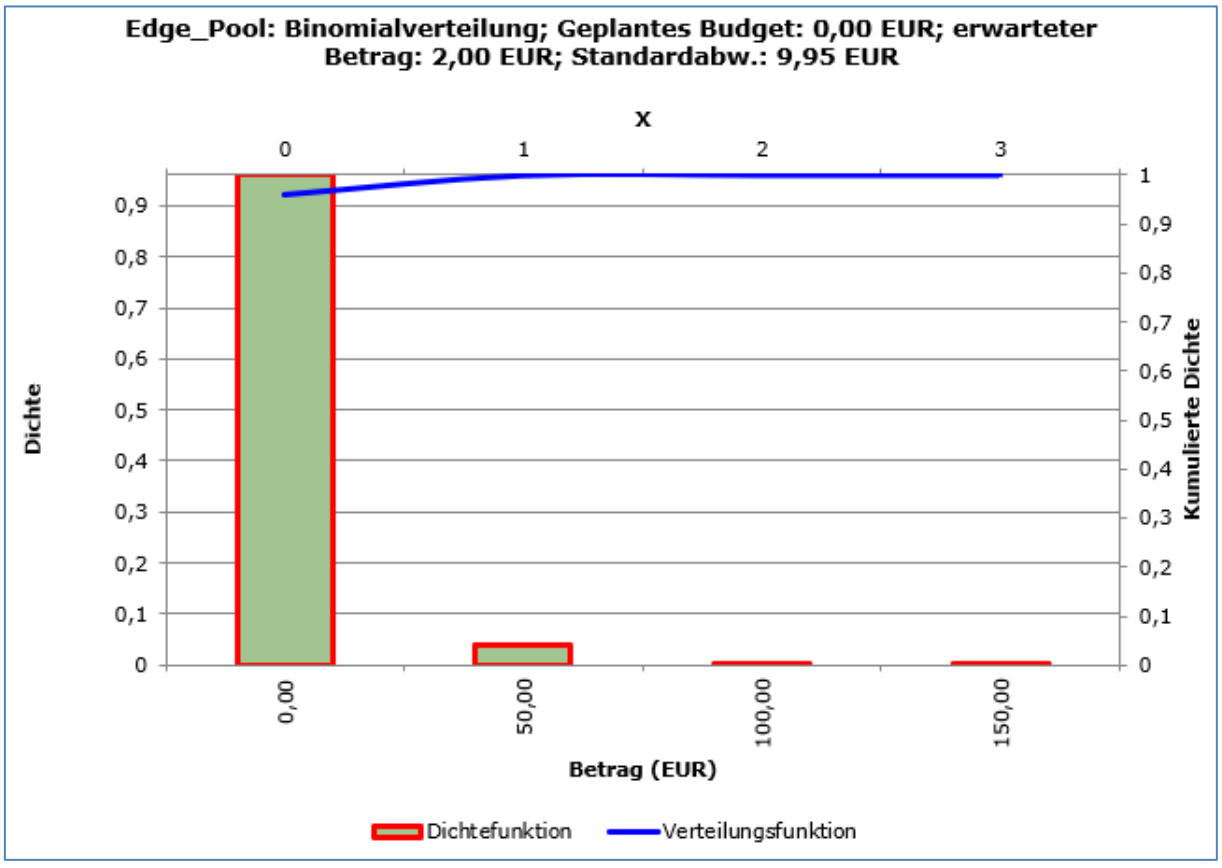

Abbildung 6: Verteilung der gleichzeitigen Ausfälle von Servern im Edge-Pool

Aus dieser Binomial-Berechnung folgt die Häufigkeitsverteilung:

- Ausfall keines Servers: 95,099005%
- Ausfall eines (1) Servers: 4,8029801%
- Gleichzeitiger Ausfall von 2 Servern: 0,0970299%
- Gleichzeitiger Ausfall von 3 Servern: 0,0009801%
- Gleichzeitiger Ausfall von 4 Servern: $< 10^{-7}$ %

Diese Werte können jetzt in einer Diskreten Verteilung benutzt werden. Mit einer diskreten Verteilung können explizit Werte (Schadenhöhen) pro Schätzung (in diesem Fall Ausfallwahrscheinlichkeit) vergeben werden.

- Anzahl Schätzungen: 4

Schätzung	Wahrschein-lichkeit	Schaden
1 (Ausfall keines Servers)	95,099005%	0 EUR
2 (Ausfall eines (1) Servers)	4,8029801%	10.000 EUR (1/6 der gesamten Last kann nicht mehr auf die 3 restlichen Server verteilt werden)
3 (Gleichzeitiger Ausfall von 2 Servern)	0,0970299%	20.000 EUR (ca. 1/3 der Last kann nicht mehr verarbeitet werden)
4 (Gleichzeitiger Ausfall von 3 Servern)	0,0009801%	40.000 EUR (ca. 2/3 der Last kann nicht mehr verarbeitet werden)
5 (Gleichzeitiger Ausfall aller 4 Server)	< 10^{-7} %	Wahrscheinlichkeit ist zu gering, daher wird dieser Fall nicht modelliert

Tabelle 2: Ausfallwahrscheinlichkeiten von Servern im Edge-Pool

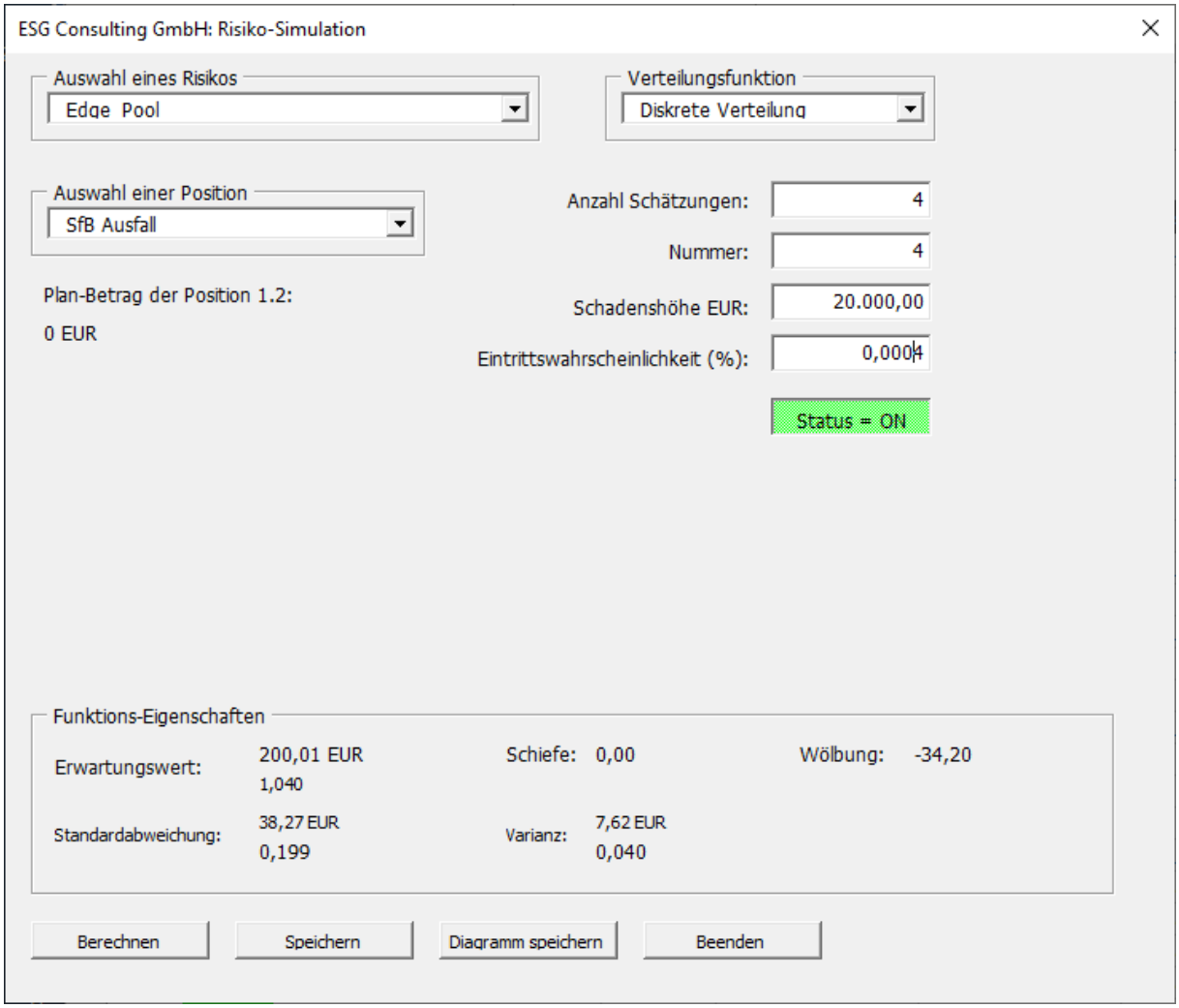

Abbildung 7: Konfiguration des Risikos für Edge-Pool

Die Schadenverteilung ergibt sich für die diskrete Verteilung:

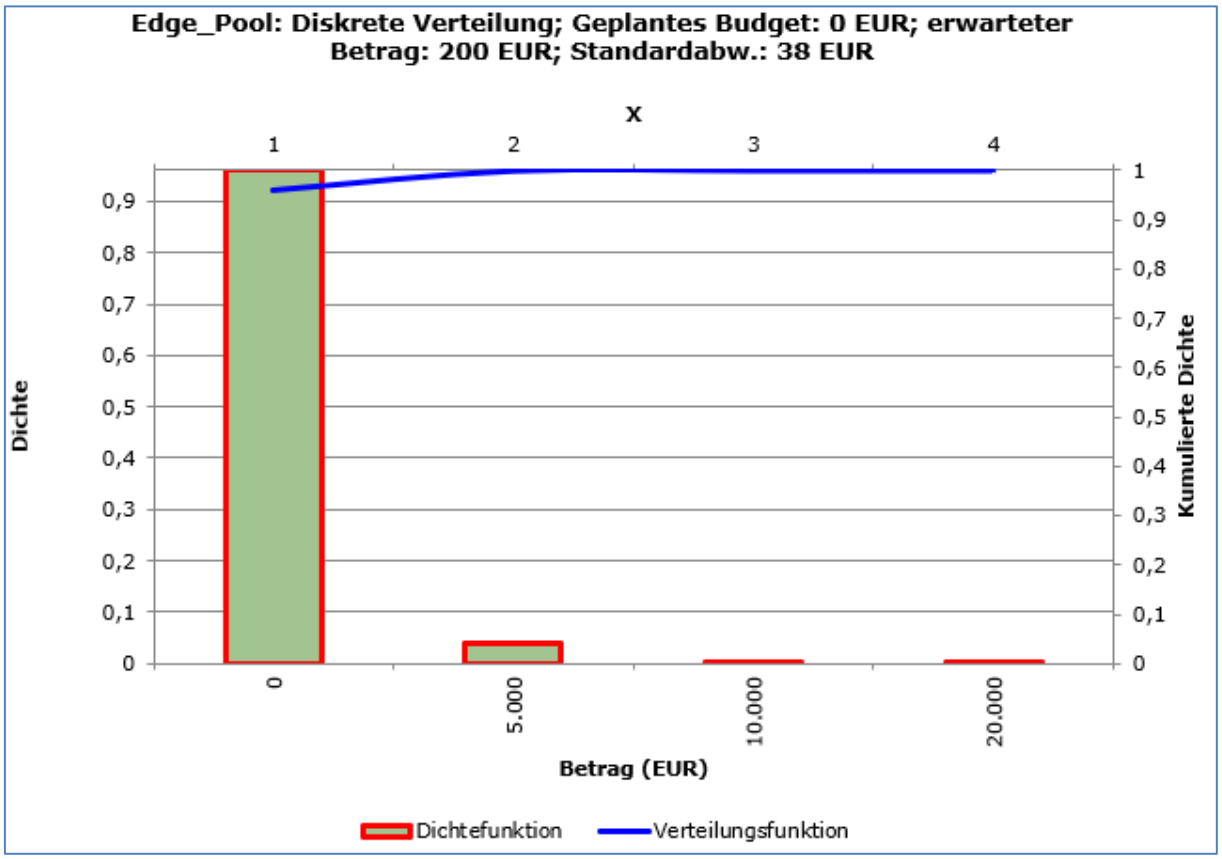

Abbildung 8: Schadenverteilung für den Edge-Pool entsprechend einer diskreten Verteilung

1.2.4 Firewall 2

Entspricht dem Risiko #4 im Szenario „ContactCenter_Szenario.xlsx“. Da die Firewall-2 auch von weiteren Services benutzt wird, wird hierfür bei einem Ausfall von einem Schaden von 60.000 EUR pro Stunde Ausfall ausgegangen.

- Verfügbarkeit: 99,99%
- MTTR: 0,88 Std., Mittlerer Schaden / Jahr: 52.560 EUR, Reparaturzeit von 2 Stunden
- Ausfallhäufigkeit / Jahr: 0,438
- Verteilungsfunktion: Gleichverteilung mit
  - Eintrittswahrscheinlichkeit: 43,8
  - Schadensfunktion: Dreiecksverteilung mit 120.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 52.560 / 0,4382), 100.000 EUR Minimalschaden und 140.000 EUR Maximalschaden

1.2.5 SfB Frontend-Pool

Entspricht dem Risiko #5 im Szenario „ContactCenter_Szenario.xlsx“

Bei einem Pool kann man die Häufigkeit mehrerer gleichzeitiger Ausfälle zuerst mittels der Binomialverteilung berechnen.

- 4 gleichartige Server je Pool
- Verfügbarkeit 1 Servers: 99%
- Gesamt-Verfügbarkeit: 99,9999%
- Bei Gesamtausfall aller 4 Server im Pool: MTTR: 0,34 sec., Mittlerer Schaden / Jahr: 2.63 EUR, Reparaturzeit von 4 Stunden
- Ausfallwahrscheinlichkeit: 0,000006%
- Schaden je Total-Ausfall: 120.000 EUR

Mit der Binomialverteilung und dem Parameter „Anzahl Versuche = 5 (Pool, bestehend aus 4 gleichartigen Servern) und der Eintrittswahrscheinlichkeit: 1,00% (je Server)

Schätzung	Wahrschein-lichkeit	Schaden
1 (Ausfall keines Servers)	95,099005%	0 EUR
2 (Ausfall eines (1) Servers)	4,8029801%	10.000 EUR (1/12 der gesamten Last kann nicht mehr auf die 3 restlichen Server verteilt werden)
3 (Gleichzeitiger Ausfall von 2 Servern)	0,0970299%	20.000 EUR (ca. 1/6 der Last kann nicht mehr verarbeitet werden)
4 (Gleichzeitiger Ausfall von 3 Servern)	0,0009801%	90.000 EUR (ca. 3/4 der Last kann nicht mehr verarbeitet werden)
5 (Gleichzeitiger Ausfall aller 4 Server)	< 10^{-7} %	Wahrscheinlichkeit ist zu gering, daher wird dieser Fall nicht modelliert

Tabelle 3: Ausfallwahrscheinlichkeiten von Servern im SfB-Frontend-Pool

Diese Werte können jetzt in einer Diskreten Verteilung benutzt werden.

1.2.6 SfB Backend-Pool

Entspricht dem Risiko #6 im Szenario „ContactCenter_Szenario.xlsx“

Bei einem Pool kann man die Häufigkeit mehrerer gleichzeitiger Ausfälle zuerst mittels der Binomialverteilung berechnen.

- 2 gleichartige Server je Pool
- Verfügbarkeit 1 Servers: 99,2%
- Gesamt-Verfügbarkeit: 99,999999%
- Bei Gesamtausfall aller 2 Server im Pool: MTTR: 0,56 Std., Mittlerer Schaden / Jahr: 16.819 EUR, Reparaturzeit von 4 Stunden

- Ausfallwahrscheinlichkeit: 0,0348%
- Schaden je Total-Ausfall: 120.000 EUR

Mit der Binomialverteilung und dem Parameter „Anzahl Versuche = 3 (Pool, bestehend aus 2 gleichartigen Servern) und der Eintrittswahrscheinlichkeit: 0,80% (je Server)

Schätzung	Wahrschein-lichkeit	Schaden
1 (Ausfall keines Servers)	98,4064%	0 EUR
2 (Ausfall eines (1) Servers)	1,5872%	40.000 EUR (1/3 der gesamten Last kann nicht mehr auf die 3 restlichen Server verteilt werden)
3 (Gleichzeitiger Ausfall beider Server)	0,064%	120.000 EUR (Totalausfall)

Tabelle 4: Ausfallwahrscheinlichkeiten von Servern im SfB-Backend-Pool

Diese Werte können jetzt in einer Diskreten Verteilung benutzt werden.

1.2.7 SfB Mediation-Server

Entspricht dem Risiko #7 im Szenario „ContactCenter_Szenario.xlsx“.

- Verfügbarkeit: 99,98%
- MTTR: 1,75 Std., Mittlerer Schaden / Jahr: 52.560 EUR, Reparaturzeit von 3 Stunden
- Ausfälle / Jahr: 0,584
- Verteilungsfunktion: Gleichverteilung mit
  - Eintrittswahrscheinlichkeit: 58,4%
  - Schadensfunktion: Dreiecksverteilung mit 90.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 52.560 / 0,584), 70.000 EUR Minimalschaden und 110.000 EUR Maximalschaden

1.2.8 CC Trusted App / Connector

Entspricht dem Risiko #8 im Szenario „ContactCenter_Szenario.xlsx“

- Verfügbarkeit: 99,5%
- MTTR: 43,8 Std., Mittlerer Schaden / Jahr: 876.000 EUR, Reparaturzeit von 6 Stunden
- Ausfallwahrscheinlichkeit / Jahr: 2,0%
- Ausfälle / Jahr: 7,3
- Verteilungsfunktion: Häufigkeitsverteilung mit
  - Häufigkeit: 7,3 ± 10% Abweichung
  - Schadensfunktion: Normalverteilung mit 120.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 876.000 / 7,3) und Standardabweichung: 4%

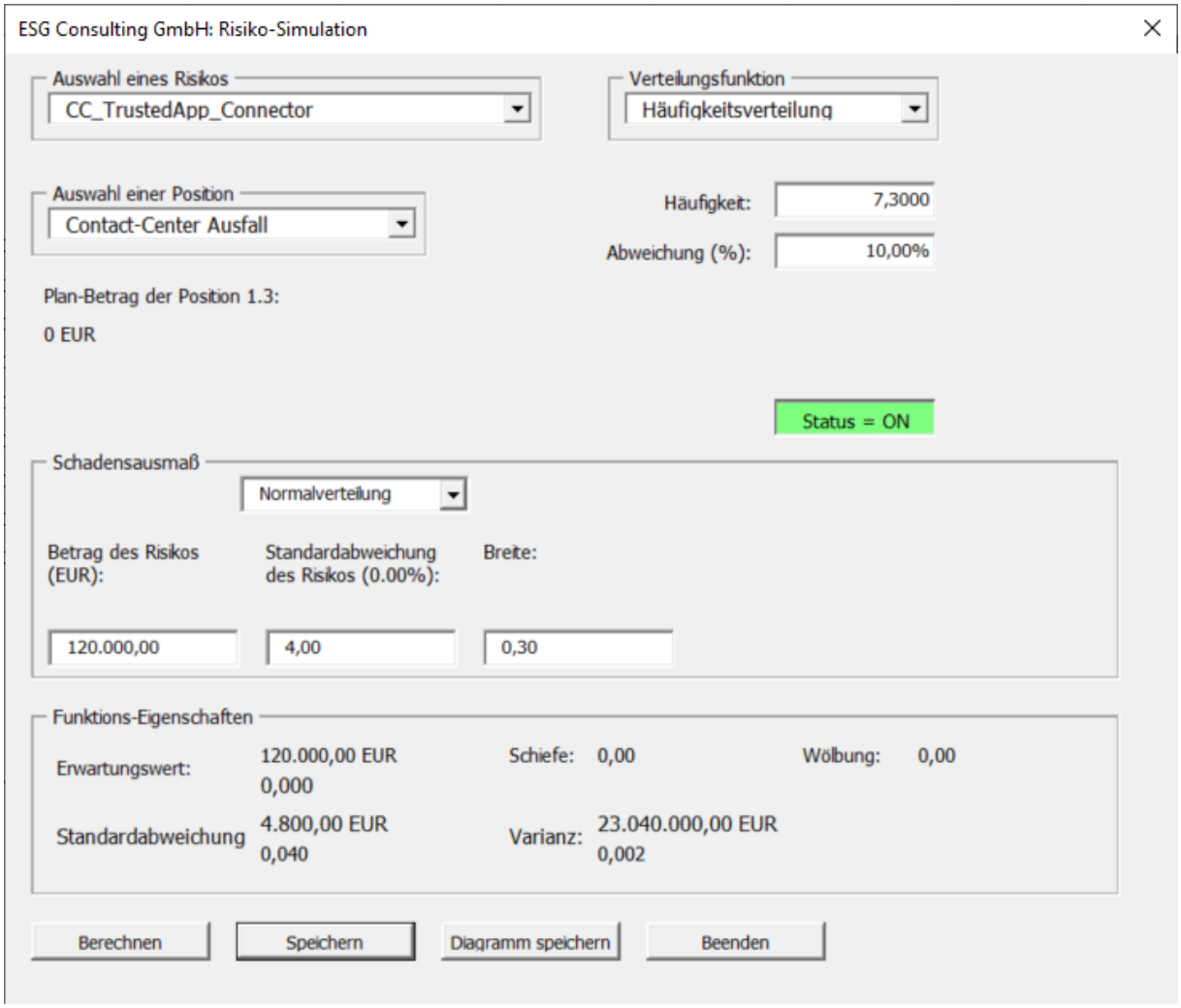

Abbildung 9: Häufigkeitsverteilung für Trusted-App/Connector

1.2.9 CC Workflow Engine

Entspricht dem Risiko #9 im Szenario „ContactCenter_Szenario.xlsx“

Bei einem Pool kann man die Häufigkeit mehrerer gleichzeitiger Ausfälle zuerst mittels der Binomialverteilung berechnen.

- 4 gleichartige Server je Pool
- Verfügbarkeit 1 Servers: 99%
- Gesamt-Verfügbarkeit: 99,999999%
- Bei Gesamtausfall aller 4 Server im Pool: MTTR: 0,34 sec., Mittlerer Schaden / Jahr: 1.75 EUR, Reparaturzeit von 8 Stunden
- Ausfallwahrscheinlichkeit: 0,000003%
- Schaden je Total-Ausfall: 160.000 EUR

Mit der Binomialverteilung und dem Parameter „Anzahl Versuche = 5 (Pool, bestehend aus 4 gleichartigen Servern) und der Eintrittswahrscheinlichkeit: 1,00% (je Server)

Schätzung	Wahrschein-lichkeit	Schaden
1 (Ausfall keines Servers)	95,099005%	0 EUR
2 (Ausfall eines (1) Servers)	4,8029801%	16.000 EUR (1/10 der gesamten Last kann nicht mehr auf die 3 restlichen Server verteilt werden)
3 (Gleichzeitiger Ausfall von 2 Servern)	0,0970299%	40.000 EUR (ca. 1/3 der Last kann nicht mehr verarbeitet werden)
4 (Gleichzeitiger Ausfall von 3 Servern)	0,0009801%	120.000 EUR (ca. 3/4 der Last kann nicht mehr verarbeitet werden)
5 (Gleichzeitiger Ausfall aller 4 Server)	< 10^{-7} %	Wahrscheinlichkeit ist zu gering, daher wird dieser Fall nicht modelliert

Tabelle 5: Ausfallwahrscheinlichkeiten von Servern im CC-Workflow-Engine-Pool

Diese Werte können jetzt in einer Diskreten Verteilung benutzt werden.

1.2.10 CC Routing Engine

Entspricht dem Risiko #10 im Szenario „ContactCenter_Szenario.xlsx“

Bei einem Pool kann man die Häufigkeit mehrerer gleichzeitiger Ausfälle zuerst mittels der Binomialverteilung berechnen.

- 4 gleichartige Server je Pool
- Verfügbarkeit 1 Servers: 99%
- Gesamt-Verfügbarkeit: 99,999999%
- Bei Gesamtausfall aller 4 Server im Pool: MTTR: 0,34 sec., Mittlerer Schaden / Jahr: 1.75 EUR, Reparaturzeit von 8 Stunden
- Ausfallwahrscheinlichkeit: 0,000003%
- Schaden je Total-Ausfall: 160.000 EUR

Mit der Binomialverteilung und dem Parameter „Anzahl Versuche = 5 (Pool, bestehend aus 4 gleichartigen Servern) und der Eintrittswahrscheinlichkeit: 1,00% (je Server)

Schätzung	Wahrschein-lichkeit	Schaden
1 (Ausfall keines Servers)	95,099005%	0 EUR
2 (Ausfall eines (1) Servers)	4,8029801%	16.000 EUR (1/10 der gesamten Last kann nicht mehr auf die 3 restlichen Server verteilt werden)
3 (Gleichzeitiger Ausfall von 2 Servern)	0,0970299%	40.000 EUR (ca. 1/3 der Last kann nicht mehr verarbeitet werden)

4 (Gleichzeitiger Ausfall von 3 Servern)	0,0009801%	120.000 EUR (ca. 3/4 der Last kann nicht mehr verarbeitet werden)
5 (Gleichzeitiger Ausfall aller 4 Server)	< 10^{-7} %	Wahrscheinlichkeit ist zu gering, daher wird dieser Fall nicht modelliert

Tabelle 6: Ausfallwahrscheinlichkeiten von Servern im CC-Routing-Engine-Pool

1.2.11 CC DB-Server

Entspricht dem Risiko #11 im Szenario „ContactCenter_Szenario.xlsx“

Das Contact-Center ist mit 1 DB-Server ausgestattet.

- Verfügbarkeit: 99,7%
- MTTR: 26,28 Std., Mittlerer Schaden / Jahr: 525.600 EUR, Reparaturzeit von 8 Stunden
- Ausfälle / Jahr: 3,285
- Verteilungsfunktion: Häufigkeitsverteilung mit
  - Häufigkeit: 3,285% ± 10% Abweichung
  - Schadensfunktion: Normalverteilung mit 160.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 8525.600 / 3,285) und Standardabweichung: 4%

1.2.12 CC Realtime-Monitoring

Entspricht dem Risiko #12 im Szenario „ContactCenter_Szenario.xlsx“

Bei einem Pool kann man die Häufigkeit mehrerer gleichzeitiger Ausfälle zuerst mittels der Binomialverteilung berechnen.

- 2 gleichartige Server je Pool
- Verfügbarkeit 1 Servers: 95%
- Gesamt-Verfügbarkeit: 99,75
- Bei Gesamtausfall aller 2 Server im Pool: MTTR: 21,9 Std., Mittlerer Schaden / Jahr: 4.380 EUR, Reparaturzeit von 10 Stunden
- Ausfallhäufigkeit / Jahr: 2,19
- Schaden je Total-Ausfall: 2.000 EUR
- Verteilungsfunktion: Häufigkeitsverteilung mit
  - Häufigkeit: 2,19 ± 10% Abweichung
  - Schadensfunktion: Normalverteilung mit 2.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 4.380 / 2,19) und Standardabweichung: 4%

Diese Werte können jetzt in einer Diskreten Verteilung benutzt werden.

1.3 Simulations-Ergebnisse

Unter Einbeziehung sämtliche in Kap. 1.2 definierten Risiken ergibt sich bei der Berechnung von 25.000 Szenarien folgendes Ergebnis:

Position Kommunikations-Infrastruktur						
Planwert:	0 €	Wahrscheinlichkeit	Differenz (abs)	Differenz (%)	Variations-Koeffizient:	0,08
Bei Konfidenz: 99%	1.902.159 €	99,00%	1.902.159 €		Std-Abweichung:	128.696 €
Wahrscheinlichster Wert:	1.607.611 €	49,74%	1.607.611 €		Schiefe:	0,10
Erwartungswert:	1.610.754 €	50,85%	1.610.754 €		Wölbung:	2,58
Value At Risk:	1.902.159 €	1,00%	1.902.159 €			

			Plan	Erwartungswert	Std-Abweichung
1	1.1	Kommunikations-Infrastruktur	0 €	1.610.754 €	128.696 €
	1.2	SfB Ausfall	0 €	203.547 €	114.256 €
	1.3	Contact-Center Ausfall	0 €	1.407.208 €	58.497 €
2	2.1	Kommunikations-Infrastruktur Szenario_2	0 €	5.113.382 €	146.492 €
	2.2	SfB Ausfall Szenario_2	0 €	203.447 €	113.602 €
	2.3	Contact-Center Ausfall Szenario_2	0 €	4.909.935 €	91.855 €

Abbildung 10: Simulationsergebnis / Erwartungswert für die Kommunikations-Infrastruktur

Der erwartete Schaden pro Jahr ergäbe einen Wert von 1.610.754 EUR. Dieser Wert wird in 50,85% der Fälle unterschritten. Der Value at Risk besagt, dass in 99% der Fälle ein Schaden kleiner als 1.902.159 EUR zu erwarten wäre. Dieser Wert kann zur Festlegung bilanzieller Rücklagen für den Eintritt der Risiken und damit der Risikovorsorge dienen.

Die Wölbung der berechneten Dichtefunktion ist 2,58, entspricht also fast einer Gaußkurve mit einer Wölbung von 3.0. Die Schiefe von 0,10 deutet auf eine sehr leichte rechtsschiefe Verteilung hin. Dies verwundert nicht, da die Komponenten Firewall (1 und 2), SBC und der Mediation-Server mit einer ca. 50%-igen Ausfallwahrscheinlichkeit auf die Szenarien einwirken und bei Eintritt die Schadenshöhen (in Summe, wenn sie alle gleichzeitig eintreten würden, 430.000 EUR) natürlich erhöhen.

Die Server des Contact-Centers verursachen pro Jahr einen Schaden entsprechend dem Erwartungswert in Höhe von 1.407.208 EUR. Dies liegt vor Allem an der Trusted Application inkl. Connector mit 7,3 Ausfällen pro Jahr und einem Schaden von 120.000 EUR pro Ausfall. Aber auch die Verfügbarkeit des CC DB-Servers müsste durch Redundanz erhöht werden.

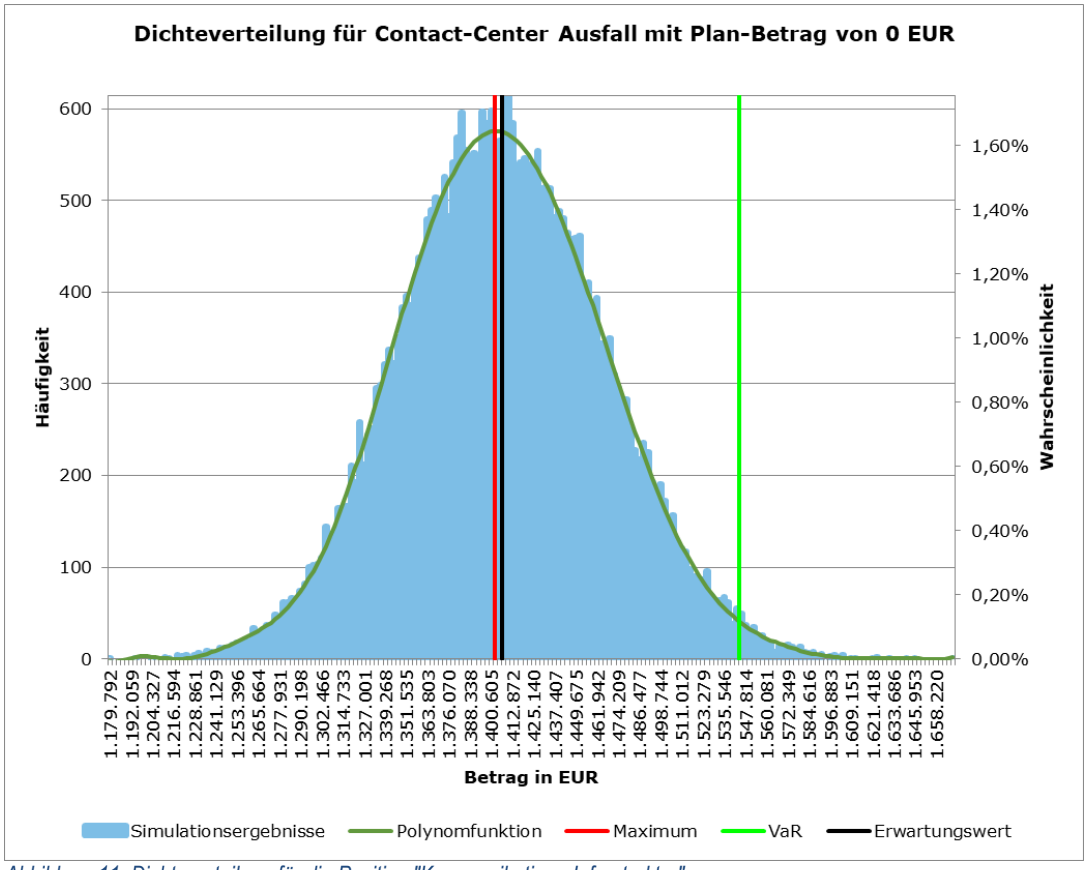

Abbildung 11: Dichteverteilung für die Position "Kommunikations-Infrastruktur"

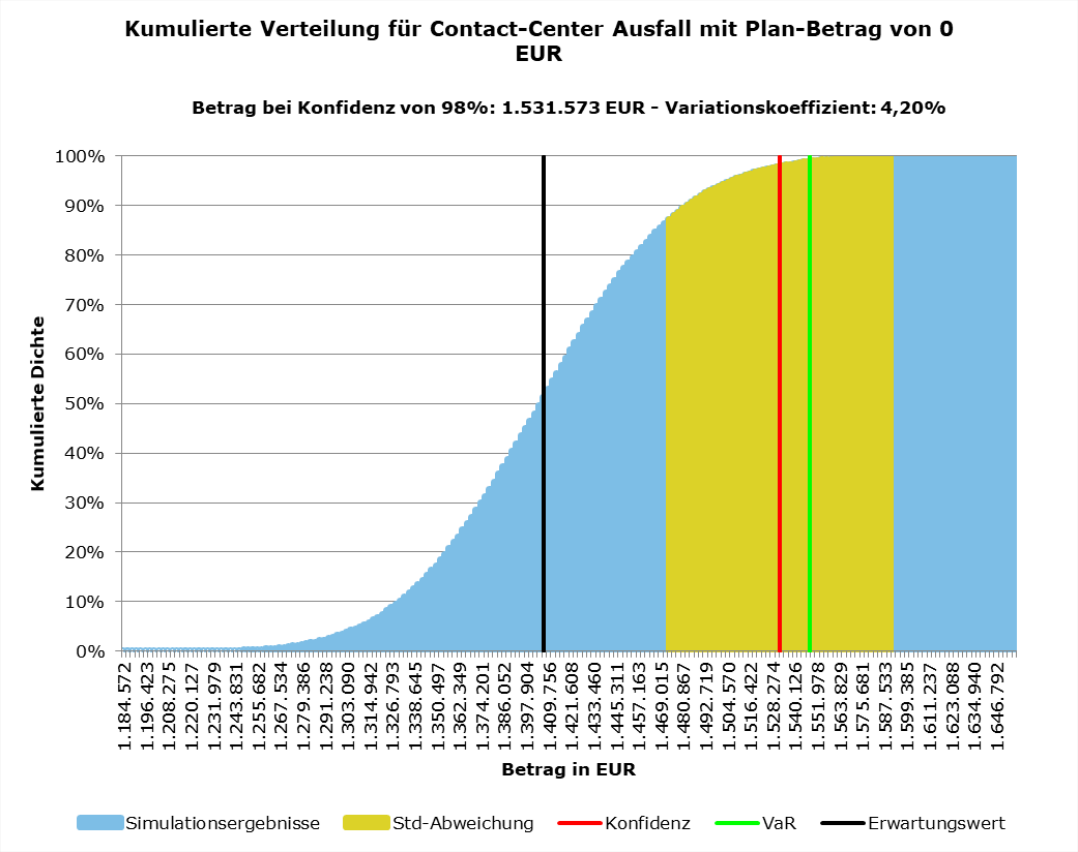

Abbildung 12: Kumulierte Dichteverteilung für die Position "Kommunikations-Infrastruktur"

Modifizierte Version des Szenarios

In dem oberen Beispiel wurde davon ausgegangen, dass das Contact-Center-System (ein Pool von 4 Servern für die Workflow-Engine und die Routing-Engine) redundant ausgelegt ist. In diesem Szenario soll davon ausgegangen werden, dass die Organisation 4 Domänen von Agentengruppen besitzt, und jede Domaine „ihren“ eigenen Server für die Workflow- und die Routing-Engine besitzt. Diese Server sind hier nicht redundant ausgelegt. Der Connector leitet einen Kontakt anhand bestimmter Kennungen an die Workflow-Engine der Domäne, die der jeweiligen Kennung zugeordnet ist. Jede Workflow-Engine wiederum besitzt ihren Routing-Server, an welchen die Informationen für einen Kontakt weitergegeben werden.

Die Konfigurationen, Parametrisierungen und Pools aller anderen Server bleiben bestehen wie in obigem Szenario dargelegt. Für die Simulation und die Risiken des neuen Szenarios werden im MC-ECO-Modell eine zusätzliche Haupt- und 2 darunterliegende Sub-Positionen angelegt. Die Risiken #1 bis #8 sowie #10 und #11 können kopiert werden, müssen lediglich den neuen Sub-Positionen zugewiesen werden.

Knoten	Verfügbarkeit	Typ
Firewall 1	99,99%	2 Stück als Appliance
SBC	99,99%	2 Stück als Appliance
Edge-Pool	99,0% pro Server	Pool aus 4 Edge-Servern (Parallel-Schaltung)
Firewall 2	99,99%	2 Stück als Appliance
SfB Frontendserver	99,0% pro Server	Pool aus 4 Servern (Parallel-Schaltung)
SfB Backendserver	99,2% pro Server	Pool aus 2 Servern (Parallel-Schaltung)
SfB Mediation-Server	99,0%	1 Server
Trusted App (Connector)	99%	1 Server
Workflow-Engine	99%	4 Domänen, 1 Server je Domäne
Routing-Engine	99%	4 Domänen, 1 Server je Domäne
DB-Server	99,75%	1 Server
Realtime-Monitoring	95%	Pool aus 2 Servern (Parallel-Schaltung)

Tabelle 7: Server-Verfügbarkeiten u. Pools für Szenario 2

1.4 CC Workflow-Engine

Entspricht dem Risiko #21 bis #24 im Szenario „ContactCenter_Szenario.xlsx“

Jede Workflow-Engine ist vom gleichen Typ und mit der identischen Basis-Software ausgestattet. Lediglich die Konfigurationen unterscheiden sich, abhängig von der Domäne. Daher können alle 4 Server mit jeweils einem eigenen Risiko und gleicher Parametrisierung versehen werden.

Jede Maschine hat eine Verfügbarkeit von 99%:

- Verfügbarkeit: 99,0%
- MTTR: 87,6 Std., Mittlerer Schaden / Jahr: 438.000 EUR, Reparaturzeit von 8 Stunden
- Wahrscheinlichkeit 1 Ausfalls / Jahr: 3,0%
- Ausfälle / Jahr: 10,95
- Verteilungsfunktion: Häufigkeitsverteilung mit
  - Häufigkeit: 10,95 ± 10% Abweichung
  - Schadensfunktion: Normalverteilung mit 40.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 438.000 / 10,95) und Standardabweichung: 4%

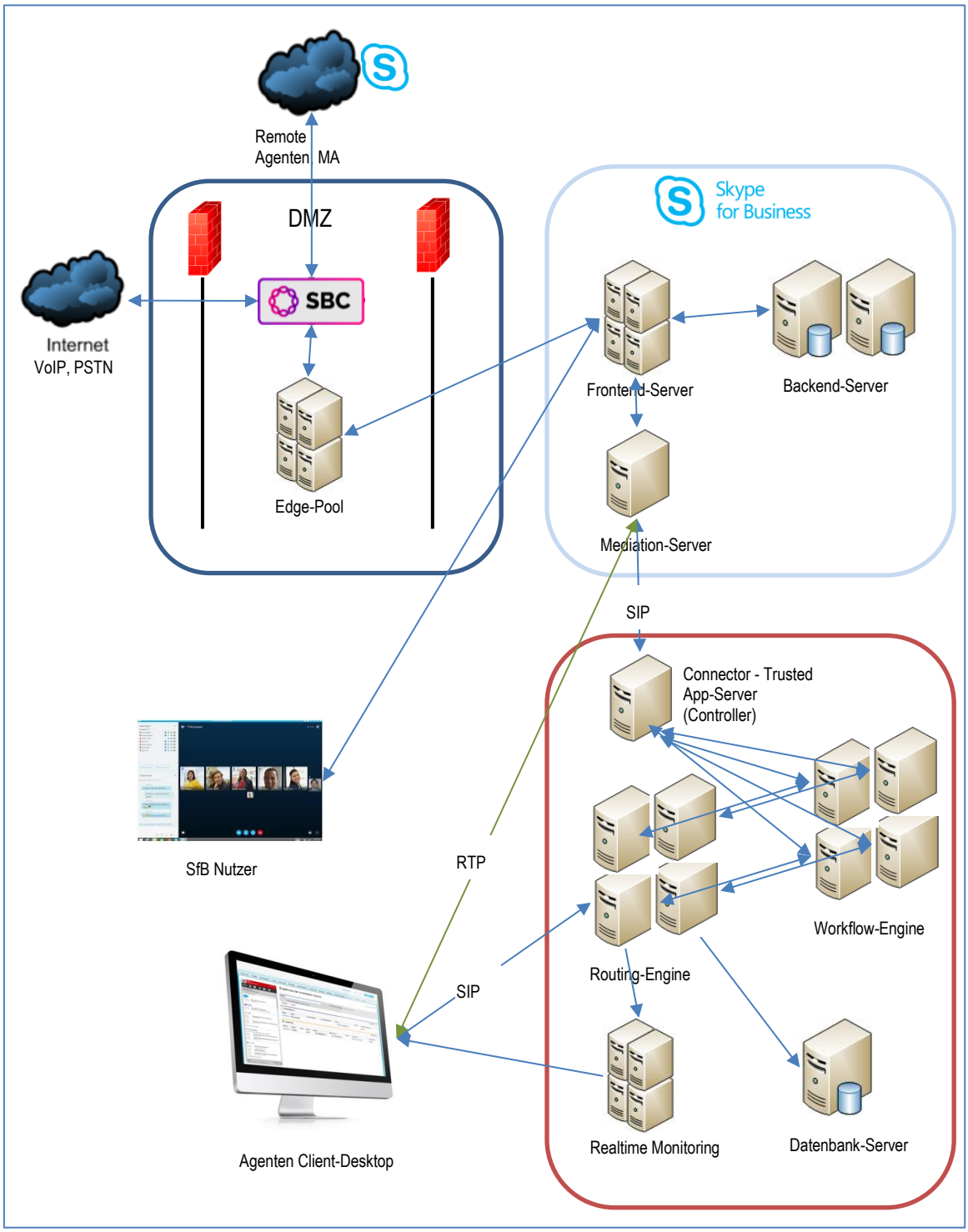

Abbildung 13: modifizierte vereinfachte Architektur SfB - Contact-Center für Szenario 2

1.5 CC Routing-Engine

Entspricht dem Risiko #25 bis #28 im Szenario „ContactCenter_Szenario.xlsx“

Jede Routing -Engine ist vom gleichen Typ und mit der identischen Basis-Software ausgestattet. Lediglich die Konfigurationen unterscheiden sich, abhängig von der Domäne. Daher können alle 4 Server mit jeweils einem eigenen Risiko und gleicher Parametrisierung versehen werden.
Jede Maschine hat eine Verfügbarkeit von 99%:

- Verfügbarkeit: 99,0%
- MTTR: 87,6 Std., Mittlerer Schaden / Jahr: 438.000 EUR, Reparaturzeit von 8 Stunden
- Ausfälle / Jahr: 10,95
- Verteilungsfunktion: Häufigkeitsverteilung mit
  - Häufigkeit: 10,95 ± 10% Abweichung
  - Schadensfunktion: Normalverteilung mit 40.000 EUR als wahrscheinlichstem Schaden (Mittlerer Schaden/Jahr / Ausfälle/Jahr = 438.000 / 10,95) und Standardabweichung: 4%

1.6 Simulations-Ergebnisse

Unter Einbeziehung sämtliche in Kap. 1.2 definierten und in Kap. 1.4 und Kap 1.5 angepassten Risiken ergibt sich bei der Berechnung von 25.000 Szenarien folgendes Ergebnis für das Szenario 2, Hauptposition „Kommunikations-Infrastruktur Szenario_2“:

Position Contact-Center Ausfall						
Planwert:	0 €	Wahrscheinlichkeit	Differenz (abs)	Differenz (%)	Variations-Koeffizient:	0,04
Bei Konfidenz: 98%	1.527.909 €	98,00%	1.527.909 €		Std-Abweichung:	58.422 €
Wahrscheinlichster Wert:	1.404.807 €	48,73%	1.404.807 €		Schiefe:	0,07
Erwartungswert:	1.407.629 €	51,63%	1.407.629 €		Wölbung:	2,88
Value At Risk:	1.544.143 €	1,00%	1.544.143 €			

			Plan	Erwartungswert	Std-Abweichung
1	1.1	Kommunikations-Infrastruktur	0 €	1.610.487 €	128.059 €
	1.2	SfB Ausfall	0 €	202.858 €	113.568 €
	1.3	Contact-Center Ausfall	0 €	1.407.629 €	58.422 €
2	2.1	Kommunikations-Infrastruktur Szenario_2	0 €	5.114.095 €	145.806 €
	2.2	SfB Ausfall Szenario_2	0 €	203.317 €	113.965 €
	2.3	Contact-Center Ausfall Szenario_2	0 €	4.910.777 €	91.566 €

Abbildung 14: Simulationsergebnis / Erwartungswert für die Kommunikations-Infrastruktur, Szenario 2

Der erwartete Schaden pro Jahr ergäbe einen Wert von 5.114.095 EUR (verglichen mit Szenario 1 in Höhe von 1.610.754 EUR). Dieser Wert wird in 51,58% der Fälle unterschritten. Der Value at Risk besagt, dass in 99% der Fälle ein Schaden kleiner als 5.446.436 EUR zu erwarten wäre. Dieser Wert kann zur Festlegung bilanzieller Rücklagen für den Eintritt der Risiken und damit der Risikovorsorge dienen.
Der erwartete Schaden ist in diesem Szenario natürlich höher, da die Workflow- und die Routing-Engine nicht redundant ausgelegt wurden und 1 Serverausfall zum Gesamtausfall einer gesamten Domäne führt.
Auch hier zeigt die Dichtefunktion mit einer Wölbung von 2,77 eine fast ideale Gaußverteilung mit einer Rechts-Schiefe von 0,008.
Der hohe Erwartungswert des Schadens ist hauptsächlich durch die nur 99%-ige Ausfallsicherheit der für die jeweilige Domäne verfügbaren Workflow- und Routing-Server bedingt. Eine jeweilige Redundanz der beiden Server-Typen würde den Erwartungswert erheblich reduzieren.

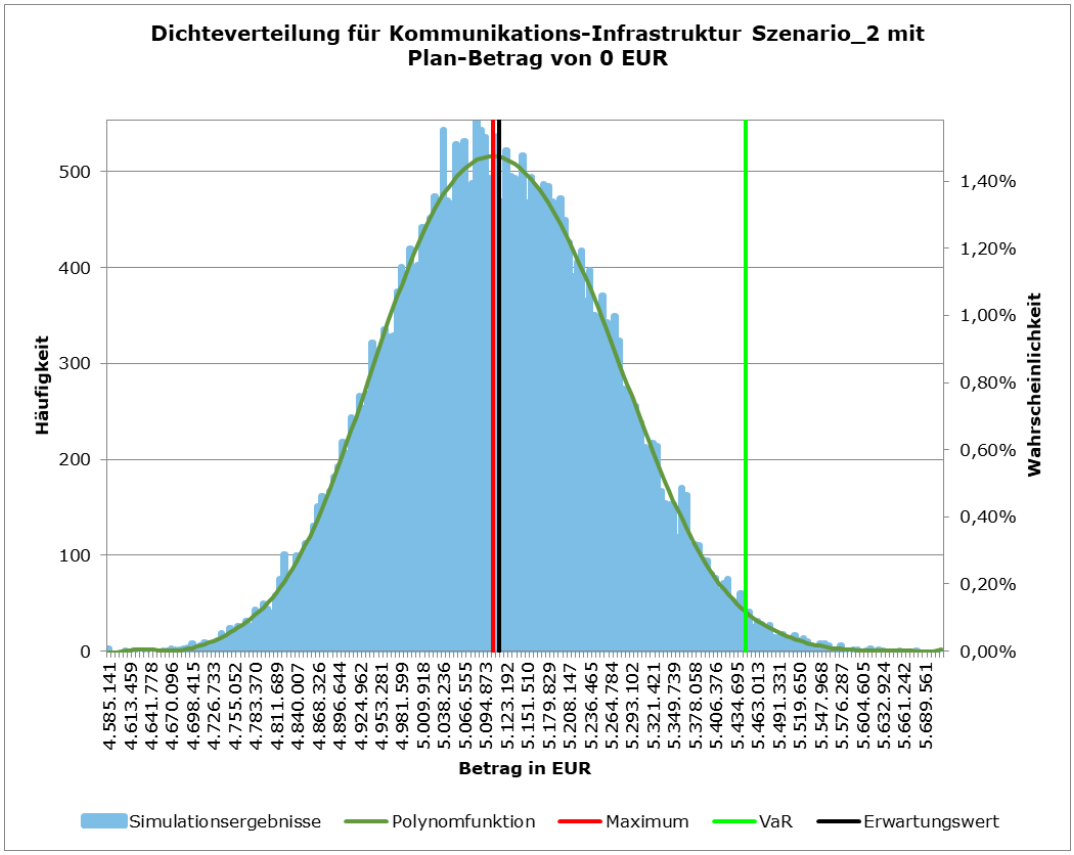

Abbildung 15: Dichteverteilung für die Position "Kommunikations-Infrastruktur Szenario 2"

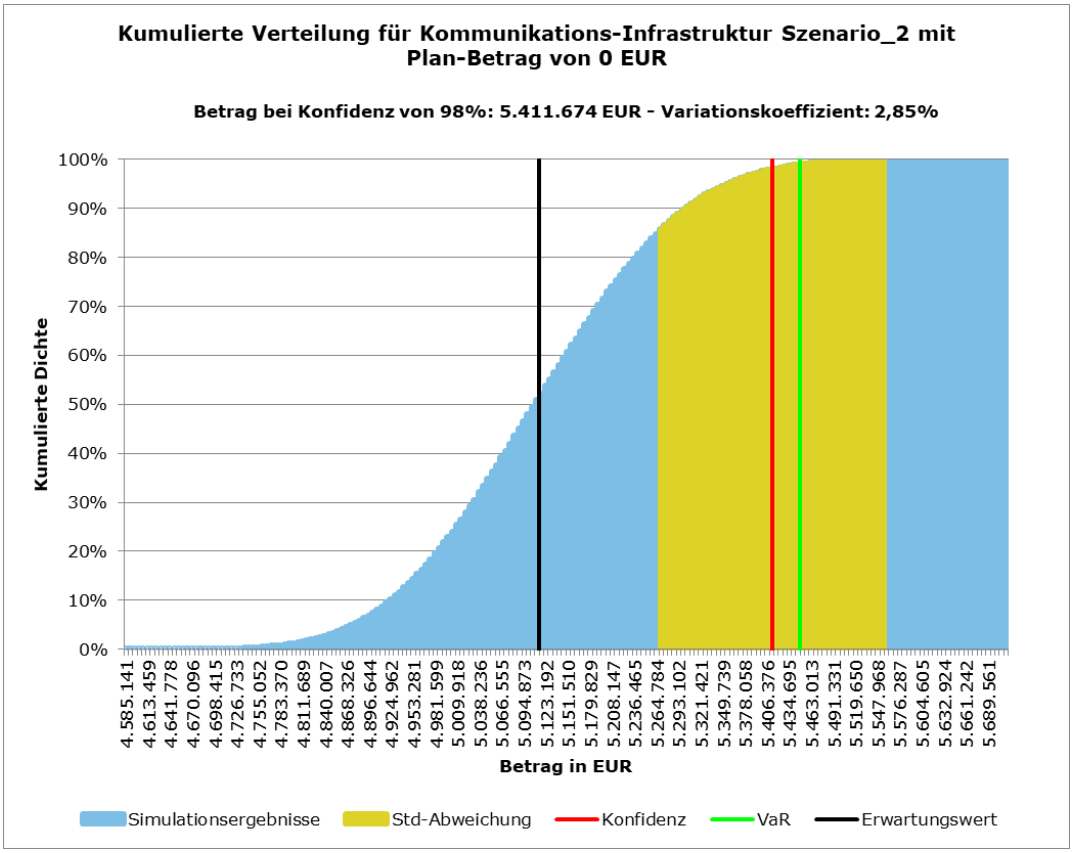

Abbildung 16: Kumulierte Dichteverteilung für die Position "Kommunikations-Infrastruktur Szenario 2"

Abkürzungen

Abkürzung	Beschreibung
DMZ	Demilitarisierte Zone
ISMS	Informationssicherheits-Managementsystem
MTTR	Mean Time to Repair
PSTN	Public Switched Telephony Network
SBC	Session Border Controller
SfB	Skype for Business
VoIP	Voice-over-IP-Telefonie

Tabelle 8: Abkürzungen

Anhang

Eine Einführung in das Thema Hochverfügbarkeit

Diese Einführung ist dem Artikel von Andrea Held entnommen:
https://www.informatik-aktuell.de/betrieb/verfuegbarkeit/hochverfuegbarkeit-und-downtime-metriken.html

Ein Service wird als verfügbar bezeichnet, wenn er in der Lage ist, die Aufgaben zu erfüllen, für die er vorgesehen ist. Als Verfügbarkeit wird aber auch die Wahrscheinlichkeit bezeichnet, dass ein System innerhalb eines spezifizierten Zeitraums funktionstüchtig (verfügbar) ist. Die Verfügbarkeit wird als Verhältnis aus Downtime und Uptime eines Systems bemessen:

Verfügbarkeit = Uptime / (Downtime + Uptime)

Ein System gilt als hochverfügbar, wenn eine Anwendung auch im Fehlerfall weiterhin verfügbar ist und ohne unmittelbaren menschlichen Eingriff weiter genutzt werden kann. Der Anwender sollte keine oder nur eine kurze Unterbrechung wahrnehmen. Hochverfügbarkeit (abgekürzt auch HA, abgeleitet von engl. High Availability) bezeichnet also die Fähigkeit eines Systems, bei Ausfall einer seiner Komponenten einen uneingeschränkten Betrieb zu gewährleisten.

1.7 Hochverfügbare Architekturen

Hochverfügbare Architekturen verfügen meist über folgende Eigenschaften:

- Toleranz und Transparenz gegenüber Fehlern
- Präventive Build-In-Funktionalitäten
- Proaktives Monitoring und schnelle Fehlererkennung
- Schnelle Wiederherstellungsmöglichkeiten
- Automatisierte Wiederherstellung ohne administrative Eingriffe
- Kein oder geringer Datenverlust

Um geeignete Maßnahmen treffen zu können, sollte man vorab mögliche Systemausfälle und deren Gründe kategorisieren. Denn: Je nach Typ sind unter Umständen grundlegend andere Verfahren zur Vermeidung eines Ausfalls geeignet.

1.8 Fehlertolerante Systeme

Fehlertolerante Systeme müssen auf nahezu alle erdenklichen Fehlerursachen reagieren können. Erreicht wird dies durch den Einsatz intelligenter Software in Kombination mit einer Eliminierung von Single Points of Failure (SPOF). Ein SPOF ist eine Komponente, die im System nur einmal vorhanden und für die korrekte, sichere und zuverlässige Funktionsfähigkeit des Gesamtsystems zwingend erforderlich ist. Nicht nur Speicher- und Rechnerkomponenten, auch das Design des Netzwerkes und der Speichertechnik können einen SPOF darstellen: Oft ist der Backend-Server zwar durch ein Cluster aus miteinander vernetzten Rechnerknoten abgesichert, so dass

bei einem Ausfall eines Rechners ein anderer dessen Aufgabe übernehmen kann. Allerdings hängen die Clusterknoten manchmal am selben Netzwerkswitch. Fällt dieser aus, nützt auch der Cluster als Schutzmaßnahme nichts: Der Service ist nicht verfügbar.

Redundanz ist hier die Lösung: Korrekt konfigurierte redundante - also mehrfach ausgelegte - Komponenten stellen sicher, dass bei einem Ausfall einer einzelnen Komponente andere des gleichen Typs die Aufgaben übernehmen können. Jede Hardwarekomponente in einem Computersystem - wie beispielsweise Storage-Adapter, Netzwerkkarten, interner Festspeicher oder auch das Netzteil des Rechners - sollte deshalb mindestens zweimal vorhanden sein. Doch kann der Rechner als Ganzes ausfallen. Bei erhöhten Verfügbarkeitsanforderungen muss daher auch die gesamte Rechnerhardware redundant – z. B. in Form eines Standby Systems – gehalten werden. Fällt ein Rechner aus, übernimmt der andere.

Redundant ausgelegte Rechnersysteme nutzen jedoch wenig, wenn eine andere Komponente einen SPOF bildet. So ist die Stromversorgung ein wichtiger Punkt, der gerne vergessen wird. Das Datacenter von IXEurope zum Beispiel verfügte im Jahr 2004 über mehrfach redundante Anbindungen von bis zu einem Dutzend unabhängigen Carrier und gewährleistete damit eine 99,999-prozentige Stromversorgung. Im Falle eines partiellen Netzausfalles kann der Kunde in solchen Konfigurationen innerhalb kurzer Zeit auf einen alternativen Carrier umgeschaltet werden.

1.8.1 Hardwar- Fehlertoleranz

Fehlertoleranz kann durch Hardware unterstützt werden. Ein gutes Beispiel ist die Fehlertoleranz eines RAID 1 Storage. Hier werden Daten über Plattenspiegelung redundant, also "mehrfach" auf physisch unterschiedlichen Platten gespeichert. Fällt eine Festplatte aus, so hat dies zunächst keinen Einfluss auf die Verfügbarkeit des Gesamtsystems. Der Festplattenspiegel fängt diesen Fehler auf. Der Ausfall bleibt für den Benutzer in der Regel transparent, er wird für ihn also nicht sichtbar.

1.8.2 Software-Fehlertoleranz

Software-Fehlertoleranz wird häufig mit den bereits genannten Methoden der Hardware-Redundanz kombiniert. Damit eine Anwendung auch nach einem Failover korrekt weiterverarbeiten kann, muss diese in der Regel auf eine Fehlersituation reagieren können. So entsteht bei einem Failover in einem Cluster ein zumindest vorübergehender Verbindungsverlust. In der Regel geht dies mit einem Verlust offener Transaktionen einher. Eine Anwendung muss also in der Lage sein, dies zu erkennen und gegebenenfalls zurückgerollte Verarbeitungsschritte nach einem Wechsel der Verbindung auf dem Ersatzsystem nachzuholen.

1.8.3 Hybride Verfahren

Im Rahmen von Clustersystemen, bei denen ein ganzes Rechnersystem redundant vorgehalten wird, wird Fehlertoleranz in einer Kombination aus Hardware und intelligenter Software bereitgestellt. Hier wird zum einen die hardwareseitige, redundante Auslegung eines Gesamtsystems genutzt, andererseits spielen aber auch Softwarekomponenten eine herausragende Rolle. So ist die Clustersoftware für die Erkennung eines Systemausfalls, die nachfolgende Fehlerbehandlung und erneute Bereitstellung der Dienste verantwortlich. Durch

eine automatische Wiederherstellung der Betriebsfähigkeit eines Systems wird hier die Systemverfügbarkeit ohne menschlichen Eingriff wiederhergestellt. Idealerweise sollten hierbei kein Daten- oder Verbindungsverluste entstehen.

1.9 Downtime: Ursachen und Kategorien

Eine **geplante Downtime** ist ein Zeitraum, in dem das System aufgrund geplanter Tätigkeiten nicht verfügbar ist. Geplante Tätigkeiten sind z. B. Wartungsarbeiten, Ausbau der Hardware oder auch das Einspielen von Software Upgrades und Patches. Gelegentliche Failover Tests zum Überprüfen der Übernahme im Fehlerfall fallen ebenso darunter wie die Aktualisierung von Applikationen.

Als **ungeplante Downtime** wird eine Zeitspanne bezeichnet, in der das System aufgrund ungeplanter Ereignisse nicht verfügbar ist. Die Ursache kann in einem Ausfall von Hardware (Festspeicher, Adapter, Netzteil, Netzwerkkomponenten usw.), fehlerhafter Software (unbehandelte Fehler/Ausnahmen) Datenkorruptionen oder auch Bedienerfehlern (Stammdaten gelöscht, Not-Aus-Schalter gedrückt) liegen. Die Anlässe für ungeplante Ausfälle sind ebenso vielfältig wie die Komponenten – gleich ob technischer oder menschlicher Natur – die an einem solchen System zusammenwirken.

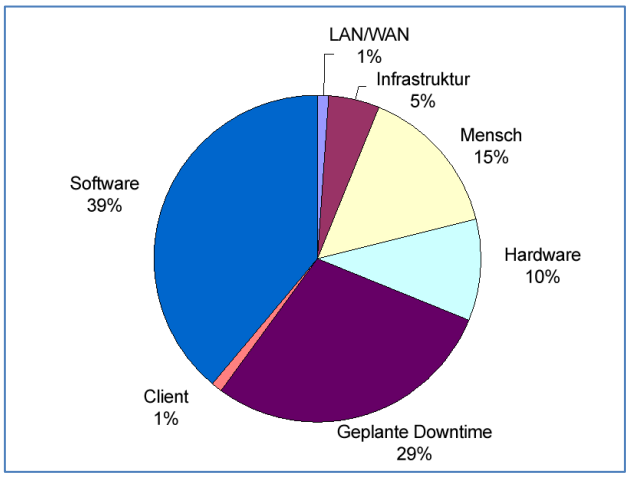

Abbildung 17: Ursachen für Ausfallzeiten. © IEEE Computer

Hochverfügbarkeit erfordert auf jeden Fall Vermeidung und Begrenzung ungeplanter Ausfallzeiten. Ungeplante Ausfallzeiten können durch Vorsorge, wenn auch nicht vermieden, so doch zumindest erheblich reduziert werden:

- Defekte Hardware: Einsatz redundanter Hardware
- Fehlerhafte Software: Verwendung aktueller Patches
- Bedienungsfehler: Aktuelles und intaktes Backup, besser noch Standby System mit Delay in der Replikation
- Stromausfälle: Große Rechenzentren verfügen über mindestens zwei Strom-Lieferanten, getrennte Stromnetze (redundante Lieferanten nutzen schließlich nichts, wenn der Bagger ein Kabel durchtrennt) sowie eine unterbrechungsfreie Stromversorgung

- Netzwerkprobleme: Redundante Auslegung der Netzwerkkomponenten
- Katastrophen / Desaster (Überflutung, Erdbeben, Bombenanschläge): Verwendung von Ausweichrechenzentren, Geo-Cluster, Geo-Spiegelung, Replikation

Zusätzlich gilt, dass geplante Ausfallzeiten vermieden oder zumindest begrenzt werden. Wartungsarbeiten und Upgrades sind jedoch unter bestimmten Umständen zwingend notwendig. Den erforderlichen Zeitraum kann man nur schwer verringern. Jedoch kann man auf Ausweichszenarien zurückgreifen. Nur ein paar Beispiele:

- Hardware Upgrades: Verwendung von Hot Plug Hardware
- Software Upgrades: Rolling Upgrades über mehrere Clusterknoten hinweg
- Wartungsarbeiten: Switchover auf einen anderen Knoten im Cluster

Es handelt sich bei den oben genannten Maßnahmen nur um Beispiele. Tatsächlich ist die Fülle möglicher Vorkehrungen enorm. Die Kosten variieren stark.

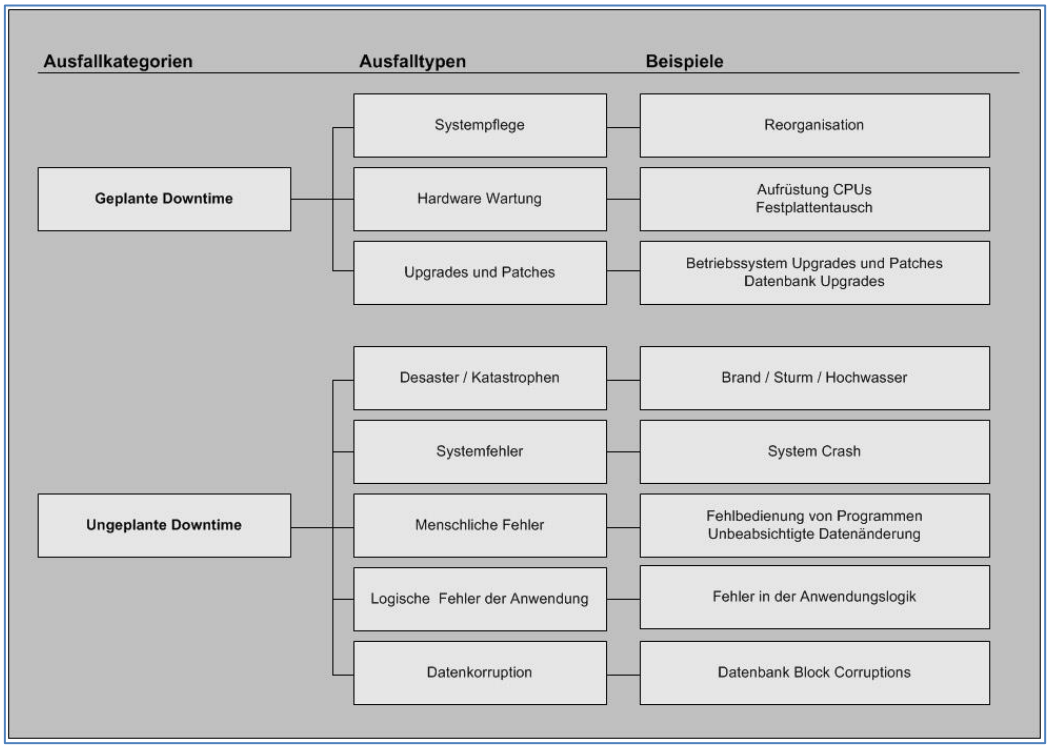

Abbildung 18: Ausfallkategorien und Ausfalltypen

1.10 Verfügbarkeitsklassen

Die Verfügbarkeit von Computersystemen wird meist in "Dauer Downtime pro Jahr" gemessen und in Prozent angegeben. Die Harvard Research Group (HRG) teilt Hochverfügbarkeit in ihrer Availability Environment Classification (kurz: AEC) in 6 Klassen ein:

- Conventional (AEC-0): Funktion kann unterbrochen werden, Datenintegrität ist nicht essentiell
- Highly Reliable (AEC-1): Funktion kann unterbrochen werden, Datenintegrität muss jedoch gewährleistet sein

- High Availability (AEC-2): Funktion darf nur innerhalb festgelegter Zeiten bzw. zur Hauptbetriebszeit minimal unterbrochen werden
- Fault Resilient (AEC-3): Funktion muss innerhalb festgelegter Zeiten bzw. während der Hauptbetriebszeit ununterbrochen aufrechterhalten werden
- Fault Tolerant (AEC-4): Funktion muss ununterbrochen aufrechterhalten werden, 24*7 Betrieb (24 Stunden, 7 Tage die Woche) muss gewährleistet sein
- Disaster Tolerant (AEC-5): Funktion muss unter allen Umständen verfügbar sein

Verfügbarkeits-Klasse	Bezeichnung	Verfügbarkeit in %	Downtime pro Jahr
2	Stabil	99,0	3,7 Tage
3	Verfügbar	99,9	8,8 Stunden
4	Hochverfügbar	99,99	52,2 Minuten
5	Fehlerunempfindlich	99,999	5,3 Minuten
6	Fehlertolerant	99,9999	32 Sekunden
7	Fehlerresistent	99,99999	3 Sekunden

Tabelle 9: Verfügbarkeitsklassen:

1.11 Ausfallkosten und Aufwände für Verfügbarkeit

Das Geschäft mit der Angst vor einem möglichen Systemausfall boomte während der Jahre, als Internet-Zeitalter und New Economy enorme Gewinne zu versprechen schienen. 24 Stunden Verfügbarkeit an sieben Tagen der Woche schien ein absolutes Muss zu sein. Dabei darf bezweifelt werden, dass ein solches Maß an Verfügbarkeit generell sinnvoll ist. Die finanziellen Mittel, die für Hardware und Personal zur Realisierung einer solchen Zielvorgabe nötig sind, sind nicht zu unterschätzen.

Mittlerweile richten sich Unternehmen wieder mehr an den tatsächlichen Erfordernissen Ihres Geschäftsfeldes aus. Die Zeiten, in denen jeder noch so kleine Geschäftszweig eine 99,999-prozentige Verfügbarkeit zu realisieren suchte, sind vorbei.

Die Kosten, die aufgrund der Nichtverfügbarkeit eines Dienstes entstehen, hängen sehr stark von der jeweiligen Anwendung und vom Geschäftsfeld ab. Für ein Unternehmen ist es wichtig, diese Auswirkungen zu kennen. Nur mit dem Wissen um diese Wirkungen lassen sich die erforderlichen und geeigneten Gegenmaßnahmen planen und ergreifen. Dabei gilt es, die Kosten für verschieden lange Ausfallzeiten zu betrachten, da diese mit fortschreitender Zeit oftmals progressiv ansteigen. Im nächsten Schritt sollte man sich ein klares Bild über den finanziellen Aufwand für Hochverfügbarkeitsmaßnahmen verschaffen. Hierzu zählen redundante Hardware und Infrastruktur, zusätzliches IT-Personal und/oder dessen Weiterbildung, Systemmanagementlösungen, um Probleme frühzeitig zu erkennen und darauf reagieren zu können, genaueres Performancemanagement und Kapazitätsplanungen, verbesserte Abstimmung von Problem- und Change-Management-Prozessen sowie

Vereinbarung von Dienstleistungen bei Zulieferern mit einem bestimmten Leistungsniveau in Form von SLAs (Service Level Agreement).
Wie auch die ausfallbedingten Kosten sind die Aufwendungen zur Verhinderung eines Ausfalls und zur schnellen Wiederherstellung zeitabhängig: Je höher die angestrebte Verfügbarkeit ist und je kürzer die Wiederherstellzeit sein darf, umso höher sind die dafür erforderlichen Investitionen. Hierbei ist es wichtig zu wissen, dass die Kosten ab einem gewissen Punkt schneller ansteigen als der dadurch erreichte Zugewinn an Verfügbarkeit beziehungsweise an Schnelligkeit bei ihrer Wiederherstellung.
Daher gilt es, beide Kostenarten abzuwägen. Die Aufwendungen zur Verhinderung des Ausfalls sollten die Kosten des Ausfalls nicht übersteigen. Die Wahl der geeigneten Technologien zur Verfügbarkeitserhöhung und schneller Wiederherstellung richtet sich dabei nach den Geschäftserfordernissen. Je höher die erwarteten Kosten eines Ausfalls sind, umso größer dürfen potenziell auch die Gegenmaßnahmen sein.

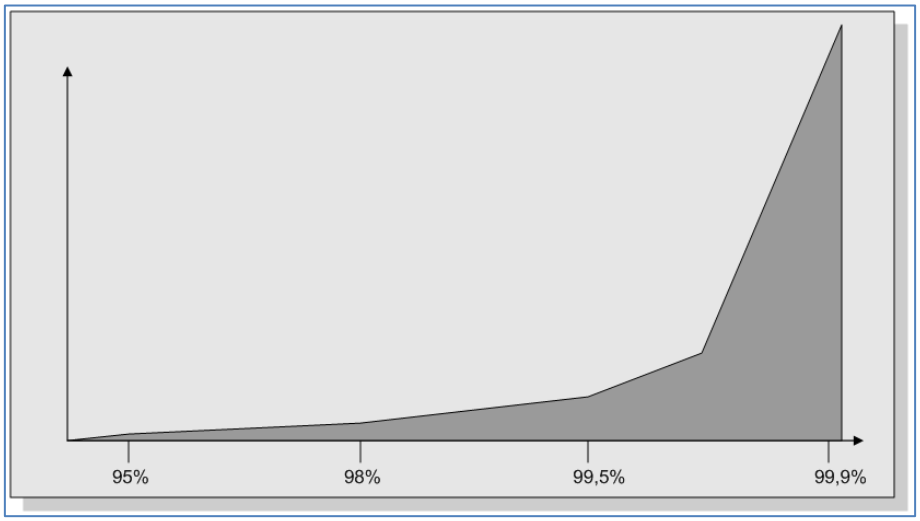

Abbildung 19: Kostensteigerungskurve nach Verfügbarkeit (Quelle: Gartner Research)

Es ist nicht immer einfach, die Kosten für Ausfallzeiten abzuschätzen. Verärgerte Kunden und Imageverluste lassen sich nicht unmittelbar messen. Die Kosten für eine HA-Lösung dagegen lassen sich klar berechnen. Generell kann man sagen, dass die Systemkosten mit steigender Verfügbarkeit weit überproportional zunehmen. Für etwa 20.000 Euro reine Hardwarekosten sind Systeme realisierbar, für die mittlere Ausfallzeiten von ca. 10 Stunden im Jahr akzeptabel sind. Um eine jährliche Downtime von maximal 1 Stunde zu erreichen, sind dagegen Anschaffungskosten von mindestens etwa 1 Mio. Euro erforderlich.
Man sollte also eine realistische Kosten-Nutzen-Analyse erstellen, bevor Investitionen getätigt werden. Dabei spielen die Hardwarekosten, aber auch Gelder für Softwarelizenzen eine Rolle. Maßgeblich sind jedoch die Betriebskosten, die aufgrund der Notwendigkeit einer oft beträchtlichen Qualifikation der Mitarbeiter erheblich höher sind.

Vor der Entwicklung eines Hochverfügbarkeitskonzeptes sollte zusammen mit den Endbenutzern präzise spezifiziert werden, welche Dienste hochverfügbar sein sollen. Über eine Systemanalyse der betreffenden Dienste sollte die gesamte Infrastruktur erfasst und potenzielle Fehlerquellen identifiziert werden.

1.12 Verfahren zur Kostenanalyse

Den Realisierungskosten einer geplanten Verfügbarkeitsstrategie sollte man – wie eingangs angerissen – die Kosten für Ausfallzeiten gegenüberstellen. Hierzu zählen die anteiligen Festkosten des Geschäftsbereiches pro Stunde, Kosten pro Mitarbeiter sowie der Verlust durch verlorene Einnahmen.

Viele Unternehmen kennen die Kosten ihrer Downtimes nicht. Doch im Grunde sind sie eine wichtige Entscheidungsgrundlage bei der Auswahl der Verfügbarkeitsstrategie. Relevant für die Kostenanalyse sind unter anderem folgende Fragestellungen:

- Betroffene Anwendungen?
- Entstehende Einnahmeverluste je Ausfall-Minute?
- Entstehende Produktionsausfälle entstehen je Ausfall-Minute?
- Kosten für Personal, Räumlichkeiten usw. je Minute?
- Länge der Ausfallzeit?
- Verlust von Kunden?
- Mögliche Rechtskosten und Regressansprüche von Kunden und Geschäftspartnern?
- Entstehen Datenverluste? Falls ja: Mit welchen Auswirkungen?

In der folgenden Tabelle sind exemplarisch die Ausfallkosten pro Stunde ausgewählter Geschäftsbereiche aufgeführt.

Geschäftsfeld	Kosten je Stunde Ausfallzeit
Airline Reservation Centers	67.000$ – 112.000$
Bank ATM Service Fees	12.000$ – 17-000$
Brokerage House (Stock)	5.600.000$ – 7.300.000 $
Catalog Sales Centers	60.000$ – 120.000$
Package Shipping	28.000$ – 32.000$
Credit Card / Sales Authorization	2.200.000$ – 3.100.000$
Pay-per-View Television	67.000$ – 112.000$
Teleticket Sales	56.000$ – 82.000$

Tabelle 10: Kosten je Stunde Ausfallzeit (Quelle Contingency Planning Research, 2001)

1.13 Die Bedarfsanalyse

Kern jeder Disaster Recovery-Strategie ist die richtige IT-Lösung. Da die Kosten der einzelnen Lösungen und Technologien stark variieren, sollte immer genau abgewogen werden, wie lange ein Unternehmen im Notfall auf seine Systeme verzichten kann, ohne dauerhaft Schaden zu nehmen. Die Bedarfsanalyse gliedert sich in

Betrachtungen des Bestandes, möglicher Gefahren für die IT-Umgebung, der durch Ausfälle verursachten Risiken sowie der Kosten. Sie bildet die Grundlage einer abschließenden Bewertung.

1.13.1 Bestandsanalyse

In der Bestandsanalyse wird erfasst, welche Anwendungen wo im Unternehmen implementiert sind, auf welchen Plattformen diese laufen, und welche Ressourcen benötigt werden, um ihren reibungslosen Betrieb zu gewährleisten. Diese Verfahrensweise garantiert, dass wirklich alle Organisationsebenen und Abteilungen eines Unternehmens bzw. des betreffenden Geschäftsbereiches vom High End-Server über die Desktops der Mitarbeiter bis hin zu den Laptops des Außendienstes in die Disaster Recovery-Strategie miteinbezogen werden.

1.13.2 Gefahrenanalyse

Die IT-Umgebung sollte genau auf alle potenziellen Schwachstellen hin untersucht werden. Eine individuelle Analyse der Worst-Case Szenarien im eigenen Unternehmen ist hilfreich.

1.13.3 Risikoanalyse

Der Einfluss eines Systemausfalls auf das Weiterlaufen der Geschäftsprozesse ist immens wichtig. Es macht selbstverständlich einen großen Unterschied, ob eine Anwendung, die nur unregelmäßig benötigt wird und nur zeitunkritische Prozesse unterstützt, oder ob eine geschäftskritische Anwendung wie das Kassensystem eines Kaufhauses ausfällt. Wurden Risiken für das Unternehmen und dessen IT-Systeme identifiziert, müssen auch die individuellen Risiken für jede einzelne Anwendung analysiert werden. Hieraus lässt sich ableiten, welche Geschäftsprozesse besser gesichert werden müssen und welche bereits mit minimalem Aufwand ausreichend geschützt werden können.

1.13.4 Kostenanalyse

Aus den gewonnenen Erkenntnissen über den individuellen Sicherheitsbedarf eines Unternehmens lässt sich berechnen, welche Kosten im Rahmen verschiedener Disaster Recovery-Strategien einzelner Anwendungen entstehen können.

1.13.5 Break-even-Analyse

Die durch die angenommene und geschätzte Ausfallwahrscheinlichkeit und -dauer gewichteten Kosten für einen Systemausfall müssen abschließend den Kosten einzelner Maßnahmen zur Absicherung gegen Systemausfälle gegenübergestellt werden. So kann der Break-even-Point zwischen jährlichen, durch Systemausfälle verursachten Kosten und jährlichen Kosten zur Sicherung der Systemlandschaft gegen Ausfälle identifiziert werden. Neben diesem können dann auch weiche Faktoren wie potenzielle Imageverluste und Verlust der Kundenbindung zur Bewertung herangezogen werden.

Wichtig ist nicht nur, wie viel Ausfallzeit ein Unternehmen verkraften kann, sondern ebenso, wie hoch der Datenverlust maximal sein darf. Manche Unternehmen kommen mit einem einfachen Backup aus. Wird damit täglich gesichert, gehen maximal die Änderungen des letzten Arbeitstages vor dem Ausfall verloren. Banken und Finanzdienstleister dagegen müssen in einigen Geschäftsfeldern Datenverluste gänzlich ausschließen können. Hier werden häufig Technologien wie synchrone Replikation über große Reichweiten hinweg eingesetzt. Dies bietet

nicht nur Schutz vor Ausfällen einer einzelnen Speicherkomponente. Auch bei Verlust eines ganzen Rechenzentrums sind alle Daten auf dem entfernten Spiegel gesichert.

Es gibt spezielle Anforderungen in manchen Geschäftszweigen, die dazu führen, dass in diesen Umgebungen das Gesamtsystem heruntergefahren wird, sobald das Spiegelsystem nicht mehr verfügbar ist. Das klingt zunächst merkwürdig, hat jedoch einen Sinn: Ein möglicher Datenverlust würde schwerer wiegen und könnte teurer werden als eine vorübergehende Downtime.

1.13.6 Abschließende Bewertung

Zwei Faktoren bilden wichtige Eckpunkte der Bewertung: Die Kenngröße Recovery Point Objective (RPO) für den möglichen Datenverlust sowie Recovery Time Objective (RTO) für die Ausfallzeit. Letztere ist identisch mit der MTTR (s. o.).

Auch innerhalb eines Unternehmens können unterschiedliche Anforderungen an RPO und MTTR bestehen. Ein Datei- und Print-Server muss im Allgemeinen lange nicht so gut vor Systemausfällen und Datenverlust geschützt sein wie ein Online-Bestellsystem. RPO und MTTR können auch völlig unterschiedlich gewichtet werden. Ein Kassensystem kann sich keinen Datenverlust erlauben und erfordert daher einen sehr geringen RPO-Wert, muss aber vielleicht nicht zwingend sofort nach einem Ausfall wieder in Betrieb gehen, wenn etwa außerhalb der Öffnungszeiten nicht auf die Daten zugegriffen wird. In diesem Fall kann die MTTR recht hoch sein, ohne dass dies Auswirkungen auf die Geschäfte des Unternehmens hat.

Auf der anderen Seite ist ein Webserver gegenüber Datenverlusten relativ tolerant, muss aber konstant verfügbar sein. Daraus folgt für die Anwendung ein relativ hoher RPO, während die MTTR sehr niedrig sein muss.

Die abschließende Bewertung sollte auch eine Prüfung der Effizienz bereits vorhandener Hochverfügbarkeits- und Disaster Recovery-Lösungen einschließen, deren Ergebnisse mit den Anforderungen verglichen werden. So lässt sich genau feststellen, wo Verbesserungen gemacht werden können. Oft können Neuinvestitionen in kostspielige Hardware durch eine effiziente Nutzung der bestehenden Ressourcen vermieden werden.

1.13.7 Resümee

Um das Maß der Verfügbarkeit eines Systems zu bestimmen, können verschiedene Kennzahlen, insbesondere aber die Mean Time To Repair sowie die Mean Time Between Failures herangezogen werden. Mittels mathematischer Modelle lassen sich zudem Vorhersagen bezüglich der Verfügbarkeit eines Systems aufgrund von dessen Verschaltung treffen. Dazu werden Formeln herangezogen, die für die Serien- und Parallelschaltung gelten.

Im Rahmen einer Analyse zur Bestimmung der Verfügbarkeitsmaßnahmen sollte den Aufwänden für die Umsetzung einer Verfügbarkeitslösung die Ausfallkosten gegenübergestellt werden. Dazu ist es notwendig, über grundlegende Informationen zum Unternehmen bzw. dem Unternehmensbereich einer Anwendung zu verfügen. Neben den Personalkosten spielen auch Geschäftszeiten und mittlere Einnahmen eine wichtige Rolle zur Berechnung der Ausfallkosten.

Bevor eine Verfügbarkeitsstrategie geplant werden kann, sind die Eckpunkte in Form einer Bedarfsanalyse zu eruieren. Dazu wird zunächst der Ist-Zustand erfasst. Eine Analyse potenzieller Gefahren, der damit für das eigene

Geschäft verbundenen Risiken sowie eine Analyse der Kosten zur Verhütung derselben bildet die Grundlage für eine Bewertung des Break-even. Die zu bestimmenden Maßnahmen zur Sicherung der Verfügbarkeit adressieren dabei u. U. mehrere Kategorien bzw. Typen von Downtimes.