Elektronische Gesundheitskarte und Telematikinfrastruktur

Feinkonzept

Zero Trust Architektur für die Telematikinfrastruktur

Flexibel, Skalierbar, Zukunftssicher





Version:	1.0.0
Revision:	617681
Stand:	17.03.2023
Status:	freigegeben
Klassifizierung:	öffentlich
Referenzierung:	gemKPT_Zero_Trust

Inhaltsverzeichnis

Hinweis: Aus Gründen der besseren Lesbarkeit wird auf die separate Verwendung von Sprachformen für männlich, weiblich und divers verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

Eine zukunftsorientierte Architektur für die TI 2.0

Als die TI im Jahr 2004 konzipiert und in den darauffolgenden Jahren realisiert wurde, war ihre Architektur State-of-the-Art. Vertrauenswürdige, geschlossene Client-Server-Netzwerke mit dezentraler Datenverarbeitung, Kryptographie auf Basis von Chipkarten und Software-Schnittstellen basierend auf dem Standard SOAP repräsentierten Sicherheit und Skalierbarkeit. Annähernd 20 Jahre später ist es an der Zeit, die damaligen Architekturentscheidungen auf den Prüfstand zu stellen und die derzeitige technologische Lage zu betrachten. Von außen betrachtet stellt die TI 1.0 ein durch VPN abgesichertes, isoliertes Netzwerk dar. Teilnehmer werden durch Identitäten auf Hardware identifiziert und dürfen in Folge an der TI teilnehmen. Die mit dem gewünschten starken Wachstum an Nutzern der TI einhergehenden Änderungen an den Anforderungen bzgl. Skalierbarkeit, Verfügbarkeit, nutzerfreundlicher Sicherheit und mobiler Nutzung, können mit der bisherigen Architektur nicht mehr adäquat erfüllt werden und erfordern neue Ansätze. Dabei ist es ein vordergründiges Ziel, Systeme und Konzepte, die bereits erfolgreich eingeführt wurden, weiterhin nutzen zu können bzw. effiziente Migrationen zu ermöglichen.

Das vorliegende Konzept zeigt auf, wie durch eine moderne Zero Trust Architektur der Schutz von Daten auch bei einem Zugriff über das offene Internet durch eine dynamische Überprüfung von aktuellen Informationen zu Nutzer, Gerät und Kontext der Anfrage gewährleistet werden kann. Es berücksichtigt dabei alle heutigen Faktoren wie andere Sicherheitskonzepte, mobile Zugänge, Identitätsmanagement und die deutlich gestiegenen Erwartungen hinsichtlich der User Experience.

Um bei der Erneuerung der technischen Basis der TI andere Faktoren nicht auszublenden, werden die folgenden kritischen Erfolgsfaktoren adressiert und Lösungsansätze aufgezeigt:

- Einfache Verwendung: Die Teilnahme an der TI sowie der Zugriff und das Teilen von Daten ist von jedem Teilnehmer mit aktuell verfügbaren Endgeräten wie PCs, Smartphones oder Tabletcomputern ohne spezielle Vorrichtungen möglich.
- Niedrigere Zugriffshürden: Applikationen können, ohne Einbußen bei der Sicherheit, sowohl stationär als auch mobil genutzt werden.
- Mitwachsen mit den Verbesserungen bei Sicherheit und Usability der von den Nutzern eingesetzten Endgeräte
- Ganzheitliche Sicherheit: Zugriffe auf Daten und Dienste in der TI werden im Kontext jeder Anfrage nach Prüfung von Sicherheitszustand und Schutzmaßnahmen auf den verwendeten Geräten erlaubt.
- Flexibilität: Ein dynamisches Regelwerk für Zugriffe auf die TI ermöglicht zügige Reaktionen auf aktuelle Entwicklungen und übergreifende Anpassungen ohne Änderung an jedem einzelnen Fachdienst.
- Zukunftssicherheit durch das Anknüpfen an und Weiterentwickeln mit dem Stand der Technik und aktuellen Standards

Das vorliegende Feinkonzept beschreibt die nötigen logischen und technischen Komponenten sowie die organisatorischen Prozesse um die TI als Zero Trust Architektur umzusetzen und die kritischen Erfolgsfaktoren zu erfüllen. Das Konzept wird durch einen Proof of Concept ergänzt, der zeigt, dass die skizzierte Architektur umsetzbar ist und die Anforderungen an Performance und Skalierbarkeit insbesondere aus Sicht des Nutzers erfüllen kann.

Um zu gewährleisten, dass die TI auf Basis der beschriebenen Zero Trust Architektur umgesetzt werden kann, sind folgende weiteren Schritte notwendig:

1. Erstellen eines Stufen- und Migrationsplans sowie einer ersten Kostenschätzung
2. Integration der vorgeschlagenen Architektur in ein TI 2.0 übergreifendes Sicherheits- und Datenschutzkonzept
3. Konkretisierung der Anforderungen an das für die Zugriffskontrolle eingesetzte Regelwerk als Basis für dessen technische Umsetzung
4. Weitere Ausarbeitung der vorgeschlagenen Governance-Prozesse in Abstimmung mit allen beteiligten Stakeholdern
5. Spezifikation der im Konzept beschriebenen Komponenten sowie der Schnittstellen zu anderen Komponenten in der TI
6. Umsetzung der Zero Trust Systemkomponenten
7. Einbringen der Zero Trust Systeme in den Betrieb der TI
8. Ausbau der Föderation der TI für einzelne Sektoren in Abstimmung mit den Entwicklungen der Zero Trust Architektur
9. Anpassen der Zugriffsteuerung der Fachdienste auf Basis der Zero Trust Architektur

Das hier vorliegende Konzept und der dazugehörige Proof-of-Concept zeigen, dass der Zero Trust Ansatz für das Gesundheitssystem funktionieren kann. Die Kapselung eines wesentlichen Teils der Zugriffsteuerung in den Zero-Trust-Komponenten ermöglicht es, den Fokus bei der Entwicklung und Bereitstellen von Fachdiensten auf den fachlichen Mehrwert für die Nutzer der TI zu legen.

1 Einleitung

Die heute verfügbare Telematikinfrastruktur (TI) basiert auf einem zentralen, vertrauenswürdigen Netz, wie in Abbildung 1 dargestellt, in das mittels zugelassener VPN-Komponenten sowohl die Leistungserbringer als auch die Fachdienste eingebunden sind. Auf Seiten der Leistungserbringer wird der Zugang über das zentrale Netz zu den Fachdiensten durch den Betrieb stationärer Konnektoren in Form von Hardwarekomponenten umgesetzt. Diese Architektur bindet den Leistungserbringer an den Ort seines Konnektors und limitiert Skalierung und Flexibilität durch das Bottleneck der VPN-Gateways. In der heutigen Zeit von intensiver Nutzung mobiler Endgeräte von beliebigen Standorten, führt dies zu deutlichen Einschränkungen bei der Nutzerakzeptanz und Usability.

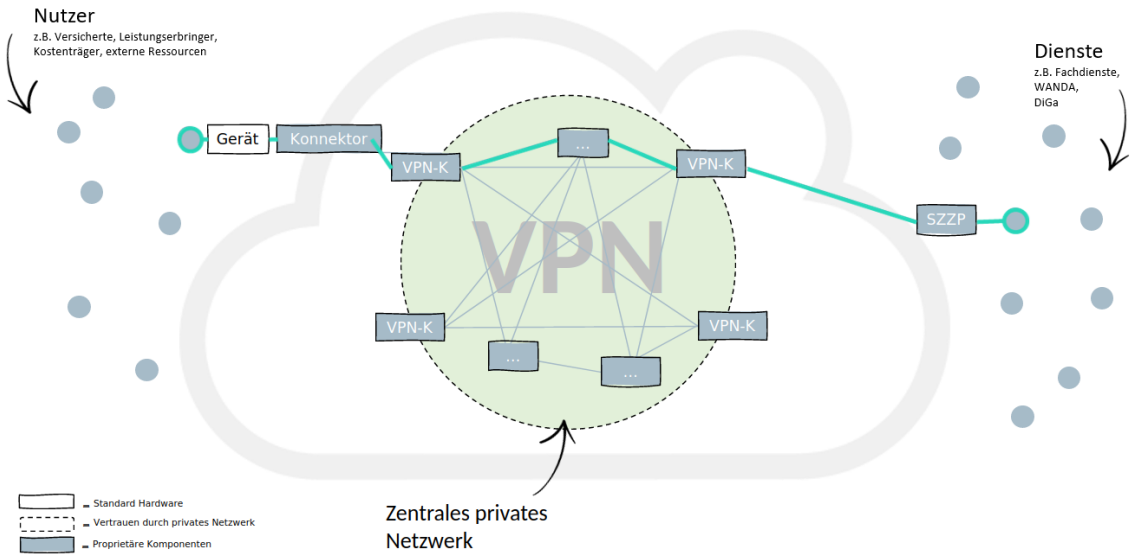

Abbildung 1: Zentrale VPN-Infrastruktur

Neben der VPN-Verbindung sind auch dienstspezifische sichere kryptografische Funktionen Bestandteil der Firmware und Zulassung der Konnektoren. Diese Architektur reduziert die Geschwindigkeit bei der Entwicklung von Fachdiensten, da für Neu- und Weiterentwicklung zuerst einmal Updates der Konnektor-Spezifikationen nötig werden, gefolgt von Softwareentwicklung bei jedem Hersteller von Konnektoren, Interoperabilitätstests und letztlich dem Ausrollen der aktualisierten Firmware auf den Konnektoren bei allen Leistungserbringern. Dazu kommt die limitierte Leistungsfähigkeit der Konnektor-Hardware, welche z.B. in einer Größenbegrenzung für den Austausch von Dokumenten resultiert.

Die Absicherung der Patientendaten in der vom Leistungserbringer selbst betriebenen Infrastruktur ist nicht im Scope der TI 1.0, sondern wird aktuell außerhalb z.B. durch die KBV-Richtlinie betrachtet. Sobald die Tür zur TI per Konnektor geöffnet ist, besteht Netzwerkzugriff auf Fachdienste und zentrale Komponenten der TI. In Verbindung mit der Ende-zu-Ende Verschlüsselung der Daten zwischen den Nutzern einiger Fachdienste können dann beliebige auch potenziell schädliche Daten zwischen den Teilnehmern der Infrastruktur ausgetauscht werden, ohne dass dies durch die Fachdienste verhindert werden kann. Entsprechend hoch sind die Sicherheitsanforderungen an sämtliche

Teilnehmer der TI, Leistungserbringer und Versicherte, sowie ihre Client-Systeme. Das reduziert die mögliche Innovationsgeschwindigkeit innerhalb der TI.

Versicherte spielen in der derzeitigen TI 1.0 eine noch geringe aber stark wachsende Rolle. Die Anbindung der Versicherten erfolgt in Anwendungen wie eRezept oder ePA jedoch nicht wie bei den Leistungserbringern über eine Einbindung in das TI-Netz mittels VPN, sondern es erfolgt eine direkte Kommunikation mit Verschlüsselung auf Transport- (TLS) und ggf. Applikationsebene mit dem jeweiligen Dienst bzw. einem vorgeschalteten Zugangsgateway. Diese Vorgehensweise zeigt die Richtung auf, die auch für die TI 2.0 vorgesehen ist.

Das mit diesem Dokument vorliegende Feinkonzept für eine Zero Trust Architektur (ZTA) der TI soll den im Whitepaper: TI 2.0 - Arena für digitale Medizin von der gematik [gem_Whitepaper_TI2.0] aufgezeigten wesentlichen Anforderungen und Herausforderungen für die TI 2.0 begegnen. Um der sehr hohen Anzahl von Teilnehmern, d.h. jedem Bürger als auch den für die Bürger agierenden Systemen, Zugang zu Diensten in der TI zu ermöglichen, soll - wie in Abbildung 1-2 dargestellt - die Kommunikation zwischen Nutzer und Diensten direkt erfolgen, ohne dass sich Teilnehmer und Dienste in ein zentrales vertrauenswürdiges Netz einwählen müssen. Das geht einher mit einem weitgehenden Verzicht auf TI-spezifische ortsgebundene Zugangshardware, d.h. der Zugriff kann direkt von dem (ggf. mobilen) Endgerät des Nutzers erfolgen - von überall und zu jeder Zeit. Bisher auf dem Konnektor lokalisierte anwendungsspezifische Operationen werden dabei entweder auf das Endgerät oder in neue Dienste der TI 2.0 ausgelagert. Der Verzicht auf die Zugangshardware erhöht nicht nur Mobilität und Nutzungskomfort für den Anwender, sondern senkt auch die Kosten und erhöht die Agilität bei der Entwicklung von Fachdiensten. Mit der Ablösung der Zugangshardware beim Leistungserbringer findet auch eine Ablösung des SZZP (Sicherer Zentraler Zugangspunkt) für die Anbindung der Dienste durch Softwarekomponenten statt.

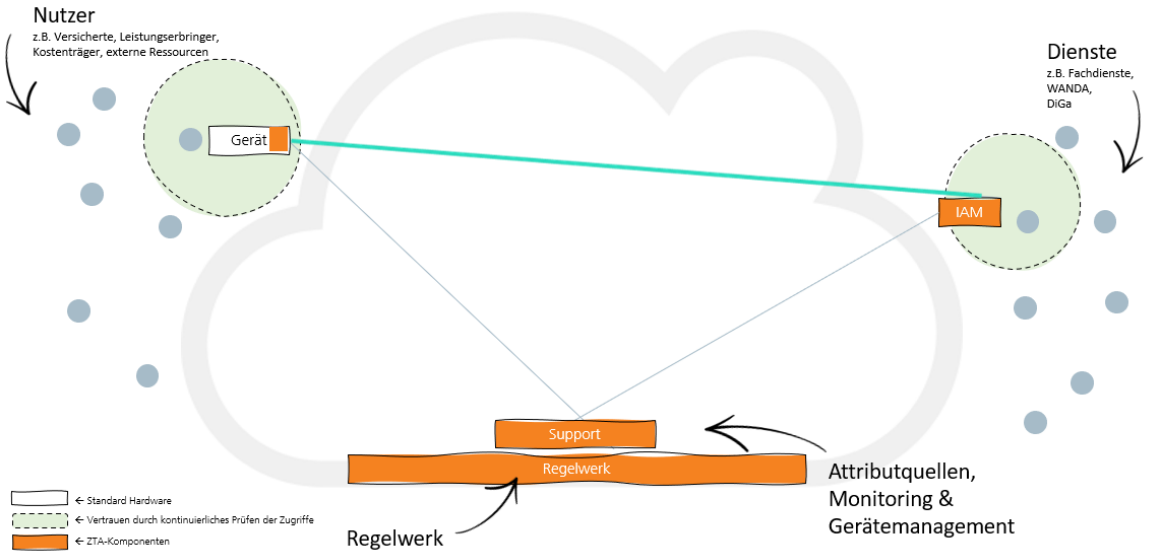

Abbildung 1-2: Zielvorstellung TI 2.0 im Vergleich zu TI 1.0

Das Vertrauen in den zugelassenen Konnektor wird durch die Prüfung von Zugangsgeräten beim Zugriff und eine genauere Prüfung der Zugriffsmodalitäten ersetzt. Sicherheitsvorgaben werden dabei über ein flexibles Regelwerk entsprechend den Anforderungen des jeweiligen Dienstes mittels eines dienstbezogen konfigurierten Identity Access Management durchgesetzt. Dabei werden sowohl Authentisierungsinformationen als auch Informationen über das Endgerät sowie weitere Faktoren in die Zugriffsentscheidung einbezogen. Ebenfalls einbezogen werden Risiken abhängig von der

Kategorie des Zugriffs. Z.B. können für Leistungserbringer höhere Sicherheitsvorgaben beim Zugriff auf die (fremden) Patientendaten bestehen als für den Zugriff des Versicherten auf seine eigenen Daten. Trotz der höheren Granularität der in die Zugriffsentscheidungen einbezogenen Daten, wird auf die Wahrung der Privacy der Teilnehmer geachtet. Eine Verarbeitung oder Speicherung personenbezogener Daten findet nur dort statt, wo es zwingend nötig ist und nur solange es nötig ist. Technische und organisatorische Vorkehrungen verhindern es, sensitive Daten miteinander zu korrelieren und so Nutzerprofile zu erstellen. Wo immer möglich wird eine Assoziation von Kommunikation oder Daten mit einzelnen Anwendern durch Anonymisierung, Pseudonymisierung und Verschlüsselung verhindert oder erschwert.

Das Feinkonzept beschreibt im Folgenden zunächst in Kapitel 2 die relevanten Akteure und Anwendungsfälle sowie die funktionalen und nicht-funktionalen Anforderungen an die Architektur. Basierend darauf werden in Kapitel 3 die grundsätzlichen Zero-Trust-Paradigmen eingeordnet und die Architektur selbst anhand ihrer Komponenten, der verarbeiteten Datenklassen und der relevanten Datenflüsse beschrieben. Für diese konzeptionelle Architektur werden in Kapitel 4 basierend auf dem aktuellen Stand der Technik technische Umsetzungsmöglichkeiten aufgezeigt und damit die tatsächliche Umsetzbarkeit geprüft. Kapitel 5 ergänzt die Governance- und Betriebssicht und zeigt auf, welche Rolle die Akteure einnehmen und wie Betriebsprozesse umgesetzt werden können. In Kapitel 6 wird die definierte Architektur mit Blick auf die Anforderungen evaluiert. Dafür wird einerseits die Umsetzung der definierten Anwendungsfälle aus Nutzersicht beschrieben und andererseits für jede der nicht-funktionalen Anforderungen evaluiert, ob sie mit den technischen und organisatorischen Maßnahmen aus Kapitel 3.5 ausreichend erfüllt ist. Abschließend werden in Kapitel 7 die wichtigsten Erkenntnisse aus dem Konzept zusammengefasst und nächste Schritte auf dem Weg zur Umsetzung identifiziert.

Mit dem Feinkonzept wird ein Proof of Concept erarbeitet, welcher die prinzipielle Machbarkeit und Skalierbarkeit zentraler Teile des Feinkonzepts aufzeigt.

Die Migration von der bestehenden TI 1.0 zur vorgeschlagenen Architektur muss in einem folgenden Konzept betrachtet werden.

2 Anforderungen

Der Umfang des Architekturkonzeptes bestimmt sich aus den Anforderungen an die Architektur. Im Folgenden werden die Akteure (Kapitel 2.1), die Anwendungsfälle (Kapitel 2.2), die für die Anwendungsfälle notwendigen Kernfunktionen im Rahmen funktionaler Anforderungen (Kapitel 2.3), die nicht-funktionalen Anforderungen (Kapitel 2.4) und weitere Annahmen (Kapitel 2.5) beschrieben.

2.1 Akteure

Im Folgenden werden die im Rahmen des Konzeptes betrachteten Akteure der TI beschrieben und in Abbildung 2.1-1 dargestellt. Die Beschreibung basiert auf der Beschreibung der Akteure in der Leistungsbeschreibung [gem_Leistung], sowie dem Glossar [gem_Glossar] der gematik für die TI.

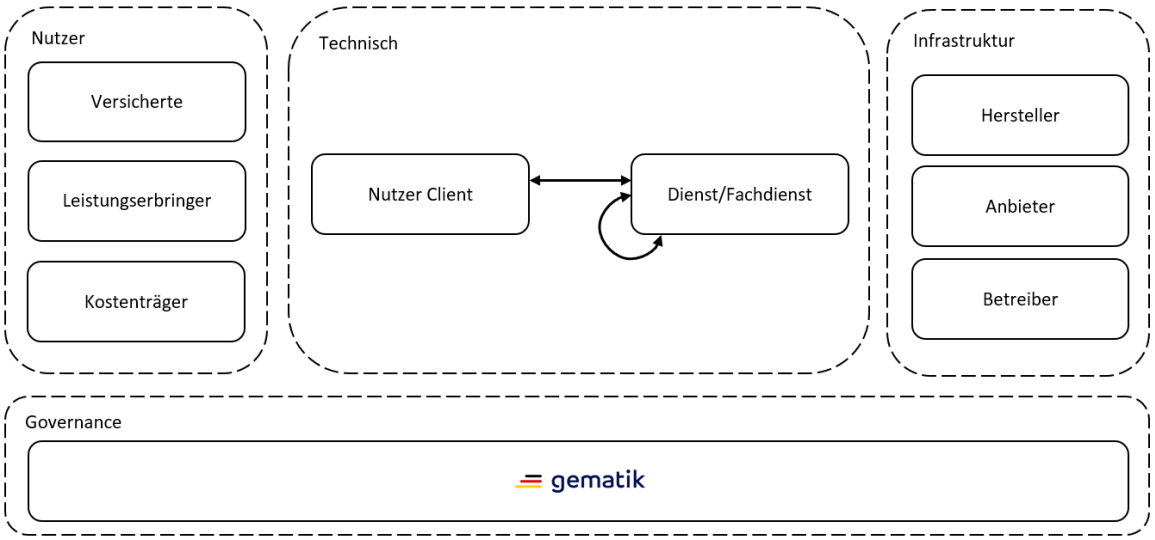

Abbildung 2.1-1: Zentrale Akteure der Architektur

Technisch

Die TI berücksichtigt folgende technische Akteure:

Nutzer Clients sind dezentrale Komponenten, die als Clients mit der TI interagieren (z.B. Smartphone-Apps, Praxisverwaltungssysteme (PVS), Apothekenverwaltungssysteme (AVS) oder Krankenhausinformationssysteme (KIS)). Sie bestehen aus Hard- und Software-Bestandteilen. Als Hardware werden stationäre Endgeräte (z.B. Windows, macOS, Linux PC) und mobile Endgeräte (z.B. Android, iOS-Smartphone) betrachtet.

Dienste/Fachdienste sind zum einen Anwendungs- oder Infrastrukturdienste der TI, bei denen die gematik die Regulierungshoheit innehat. Dazu zählen zum Beispiel die elektronische Patientenakte (ePA), Identity Provider (IDP) oder der National Contact Point eHealth (NCPeH). Zum anderen zählen auch Anwendungsdienste und digitale Gesundheitsanwendungen dazu, die an der TI teilnehmen und nicht direkt unter der Regulierungshoheit der gematik liegen. Das sind z.B. Digitale Gesundheitsanwendungen (DiGA) gem. § 33a SGB V oder weitere Anwendungen für den Datenaustausch in der

Telematikinfrastruktur (WANDA). Dienste können Clients von anderen Diensten sein und so Daten untereinander austauschen.

Nutzer

Die Dienste der TI sollen verschiedenen Nutzergruppen zur Verfügung stehen. Im Fokus stehen dabei die folgenden Akteure:

Versicherte sind z.B. Patienten, die mit einem Client auf einem stationären oder mobilen Endgerät auf Dienste zugreifen.

Leistungserbringer (LE) sind z.B. Ärzte, Physiotherapeuten, Hebammen, Apotheker oder medizinische Fachangestellte, welche Leistungen des Gesundheitswesens für Versicherte erbringen und aus den folgenden Umgebungen auf Dienste zugreifen: Im mobilen Dienst, in kleinerer Institution (bspw. Arztpraxis oder Apotheke), in größeren medizinischen Einrichtungen (z.B. im Krankenhaus).

Kostenträger sind im Kontext der TI die gesetzlichen Krankenversicherungen. Kostenträger greifen vor allem aus stationären Umgebungen auf Dienste zu, um z.B. Informationen zur Abrechnung anderen Nutzern zur Verfügung zu stellen.

Infrastruktur

Im Rahmen des Betreibermodells der TI werden folgende Akteure berücksichtigt:

Anbieter sind Anbieter von Komponenten oder Diensten der TI oder weiteren Anwendungen wie z.B. DiGA. Ein Anbieter von Betriebsleistungen in der TI ist eine Organisation, die Services gegenüber Anwendern oder anderen Servicenehmern anbietet und verantwortet. Ein Anbieter kann seine Services selbst erbringen oder durch Betreiber erbringen lassen, jedoch verbleibt die Serviceverantwortung beim Anbieter selbst.

Betreiber sind Betreiber von Komponenten oder Diensten. Betreiber sind natürliche oder juristische Personen. Unter Betreiben ist das Anschließen an Betriebsmedien (wie z.B. Strom, Netzwerk, Klima), Starten der Betriebsprozesse, Konfigurieren und Überwachen der gewünschten Funktionalität zu verstehen.

Hersteller stellen ein Produkt gemäß den Spezifikationen her und übernehmen die Produkthaftung gemäß den gesetzlichen Vorgaben und den Support gegenüber den Käufern ihrer Produkte. Hersteller unterscheiden sich von Anbietern insbesondere dadurch, dass das verantwortete Produkt keinen IT-Service darstellt, sondern physische Geräte oder Software, welche in der Hoheit der Anwender bzw. in der Hoheit eines Anbieters/Betreibers betrieben werden.

2.1.1 Governance

Die **gematik** trägt die Gesamtverantwortung für die TI. Sie hat das Betriebs- und Sicherheits-Monitoring der TI und die Betriebsverantwortung der TI inne. Dienste, die sich an die TI angeschlossen werden sollen, müssen im Regelfall zugelassen oder bestätigt werden. Der Zulassungs- bzw. Bestätigungsprozess wird durch die gematik gesteuert.

2.2 Anwendungsfälle

Grundlage für die Anforderungen an das Architekturkonzept der TI sind die im Rahmen der Leistungsbeschreibung [gem_Leistung] aufgeführten Anwendungsfälle. Die Anwendungsfälle werden in Tabelle 2.2-1 kurz beschrieben. Wie die Architektur diese Anwendungsfälle ermöglicht, wird in Kapitel 6- Evaluation beschrieben.

Ref.	Name	Beschreibung
UC1	Lesender Zugriff von stationärem Gerät eines Leistungserbringers	Ein LE greift aus einem (Universitäts-)Krankenhaus mit einem stationären Gerät (PC) auf einen Anwendungsdienst (z.B. ePA) zu, um Daten von Versicherten zu lesen.
UC2	Lesender Zugriff von mobilem Gerät eines Leistungserbringers	Ein LE greift aus einem (Universitäts-)Krankenhaus mit einem mobilen Gerät (Gerät des KH) auf einen Anwendungsdienst (z.B. ePA) zu, um Daten von Versicherten zu lesen.
UC3	Schreibender Zugriff von stationärem Gerät eines Leistungserbringers	Ein LE greift aus seiner Arztpraxis mit einem stationären Gerät (PC) auf einen Anwendungsdienst (z.B. ePA) zu, um Daten von Versicherten zu aktualisieren.
UC4	Schreibender Zugriff von mobilem Gerät eines Leistungserbringers	Ein LE ist unterwegs (z.B. Notarzt) und greift mit seinem mobilen Gerät (Gerät des LE) auf einen Anwendungsdienst (z.B. ePA) zu, um Daten von Versicherten zu aktualisieren.
UC5	Zugriff mit stationärem Gerät eines Versicherten	Ein Versicherter greift mit seinem stationären Gerät (PC) auf einen Anwendungsdienst (z.B. ePA) zu, um seine Daten zu lesen/schreiben.
UC6	Zugriff mit mobilem Gerät eines Versicherten	Ein Versicherter greift mit seinem mobilen Gerät (Smart-Phone) auf einen Anwendungsdienst (z.B. ePA) zu, um seine Daten zu lesen/schreiben.
UC7	Kommunikation zwischen Anwendungsdiensten	Ein Anwendungsdienst (z.B. ePA) kommuniziert mit einem anderen Anwendungsdienst der TI.

Tabelle 2.2-1: Anwendungsfälle

2.3 Funktionale Anforderungen

Die funktionalen Anforderungen beschreiben die zentralen Funktionen der Architektur, die notwendig sind, um die Anwendungsfälle der Architektur zu ermöglichen. Aus den Anwendungsfällen UC1-UC7 wurden die folgenden funktionalen Anforderungen abgeleitet. Abbildung 2.3-1 visualisiert das Ergebnis des Ableitungsprozesses. Innerhalb der Funktionen wird wie folgt unterschieden:

- ZTA-Funktionen: Kernfunktionen, welche benötigt werden, um die ZTA umzusetzen. Die Realisierung der hier identifizierten Funktionen wird im Kapitel 3-Architektur beschrieben.
- Funktionen extern: Funktionen, welche in anderen Konzepten im Detail beschrieben werden. Die Schnittstellen zu bzw. Erwartungen an diese Funktionen werden im Kapitel 3- Architektur beschrieben.

In Tabelle 2.3-1 werden die zentralen Funktionen der Architektur kurz beschrieben.

Scope	Ref.	Name	Beschreibung
ZTA	FA0	Nutzer-Zugriff auf Ressourcen der TI autorisieren	Die Kernfunktion der Architektur besteht in der Autorisierung von Zugriffen der Nutzer auf Ressourcen der TI.
ZTA	FA1	Zugriff gegen Regelwerk prüfen	Die Architektur ermöglicht das Prüfen eines jeden Zugriffs auf Ressourcen gemäß einem definierten Regelwerk, um über die Autorisierung zu entscheiden.
ZTA	FA1.1	Regelwerk erstellen und aktualisieren	Die Architektur ermöglicht das Erstellen und Aktualisieren eines Regelwerks für das Prüfen von Zugriffen auf Ressourcen der TI.
ZTA	FA1.2	Regelwerk überwachen	Die Architektur ermöglicht es genügend Informationen über den Sicherheitszustand des gesamten Systems zu sammeln und auf dieser Grundlage das Regelwerk beständig zu aktualisieren, um damit verlässliche Zugriffsentscheidungen zu treffen. Dazu zählt auch, Informationen über Nutzer-Zugriffe zu erfassen, die es ermöglichen schadhafte Zugriffe zu identifizieren.
ZTA	FA1.3	Regelwerk anpassen	Die Architektur ermöglicht das Anpassen und Aktualisieren des Regelwerks.
ZTA	FA1.4	Attribute des Regelwerks anpassen	Die Architektur erlaubt das automatisierte Aktualisieren ausgewählter Attribute des Regelwerks auf Basis aktueller Zugriffe, um schadhafte Zugriffe zu verhindern.
Extern	FA1.5	Nutzerspezifische Attribute des Regelwerks anpassen	Nutzer können einzelne nutzerspezifische Attribute des Regelwerks anpassen (z.B. vertrauenswürdige Orte oder Zeit-Intervalle für einen Zugriff auf Ressourcen).
ZTA	FA2	Nutzer bei Fachdienst identifizieren	Die Architektur ermöglicht das Identifizieren eines Nutzers bei einem Fachdienst.
Extern	FA2.1	Nutzer authentifizieren	Die Architektur ermöglicht das Authentifizieren von Nutzern vor jedem Zugriff auf Ressourcen der TI.
Extern	FA2.2	Nutzer bei Fachdienst registrieren	Die Authentisierung des Nutzers gegenüber dem Fachdienst erfordert, dass der Nutzer beim Fachdienst einen Account registriert hat.

Scope	Ref.	Name	Beschreibung
Extern	FA2.3	eID-Lifecycle IDP	Das Identifizieren und Authentisieren der Nutzer erfordert, dass diese über ausreichend sichere digitale Identitäten verfügen.
Extern	FA2.4	Vertreterregelung	Eine Vertreterregelung ermöglicht die Vertretung eines Versicherten durch einen anderen Versicherten.
Extern	FA2.5	Single-Sign-On (SSO)	Identitäten gelten fachdienstübergreifend. Nutzer können eine Authentisierung für unterschiedliche Fachdienste verwenden.
ZTA	FA3	Fachdienst gegenüber Nutzer authentisieren	Die Architektur ermöglicht das Authentisieren von Fachdiensten gegenüber dem Nutzer vor dem Zugriff auf Ressourcen des Fachdienstes.
Extern	XFA3.1	Fachdienst bei ZTA registrieren	Die Authentisierung eines Fachdienstes erfordert, dass der Fachdienst zuvor bei der gematik registriert wurde und über Authentisierungsmittel verfügt.
Extern	FA4	Fachdienst attestieren	Fachdienste werden hinsichtlich der Anforderungen an den Fachdienst überprüft. Fachdienste können nur an der TI teilnehmen, wenn sie alle geforderten Eigenschaften aufweisen.
ZTA	FA5	Gerät authentifizieren	Die Architektur ermöglicht das Authentifizieren von Endgeräten der Nutzer vor jedem Zugriff auf Ressourcen der TI.
ZTA	FA5.1	Gerät registrieren	Die Architektur ermöglicht das Registrieren von Endgeräten der Nutzer für den Zugriff auf Ressourcen der TI.
ZTA	FA5.2	Gerät entfernen	Die Architektur ermöglicht das Entfernen von registrierten Endgeräten der Nutzer.
ZTA	FA6	Gerät attestieren	Die Architektur ermöglicht das Prüfen von Endgeräten der Nutzer auf Erfüllung geforderter Sicherheitseigenschaften gemäß Regelwerk.
ZTA	FA6.1	Geräteinformationen erheben	Die Architektur ermöglicht das Erheben von Informationen auf den Endgeräten der Nutzer, welche für die Prüfung geforderter Sicherheitseigenschaften gemäß Regelwerk benötigt werden.

Scope	Ref.	Name	Beschreibung
ZTA	FA7	Autorisierter Zugriff	Die Architektur ermöglicht den autorisierten Zugriff auf Ressourcen der TI, d.h. den autorisierten Austausch von Daten zwischen Client und Diensten/Komponenten.
ZTA	FA7.1	Session-Management	Die Architektur ermöglicht das Management einer Session für den autorisierten Zugriff auf Ressourcen.
ZTA	FA7.2	StepUp-Autorisierung	Die Architektur ermöglicht das Anfordern einer Autorisierung auf einem höheren Vertrauensniveau als das der bereits autorisierten Session.
ZTA	FA8	ZTA-Komponenten überwachen	Die Architektur ermöglicht eine Überwachung des Zustandes von Komponenten in der TI bezüglich Verfügbarkeit, Auslastung und Sicherheitszustand.

Tabelle 2.3-1: Funktionale Anforderungen

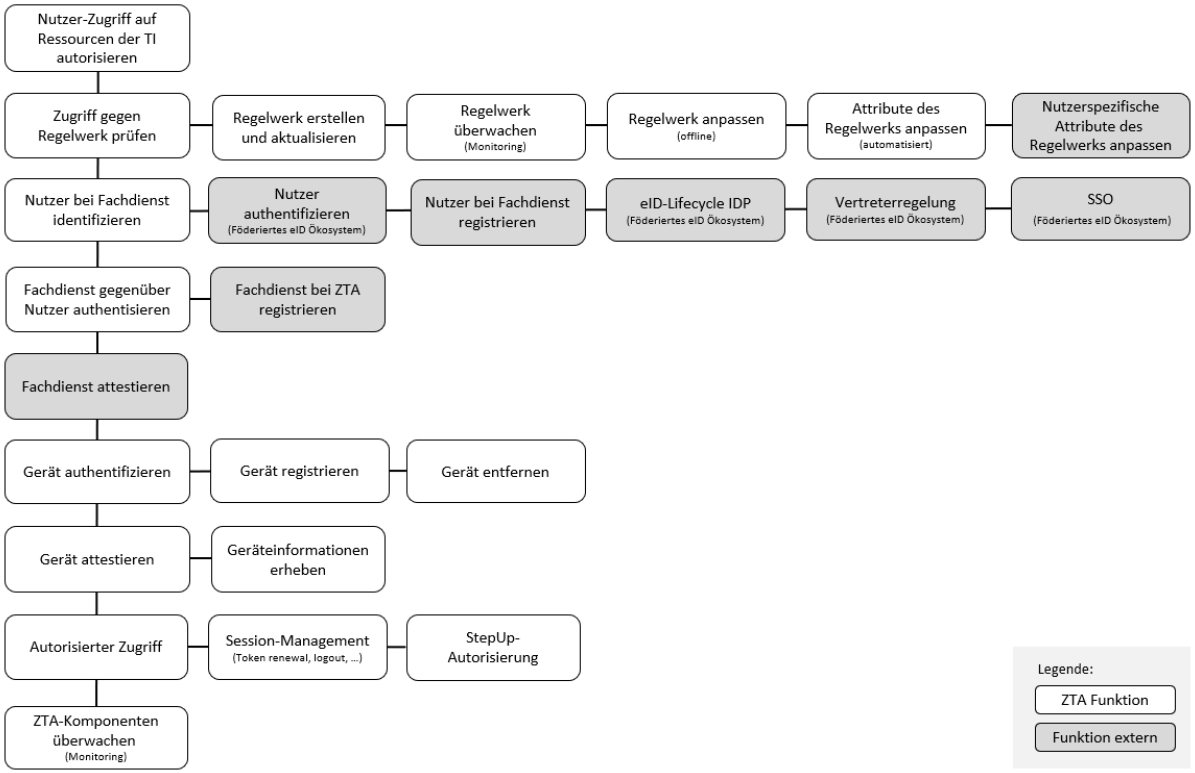

Abbildung 2.3-1: Zentrale Funktionen der Architektur

2.4 Nicht-funktionale Anforderungen

Nicht-funktionale Anforderungen beschreiben Anforderungen an die Qualität, in welcher die geforderten Funktionen erbracht werden sollen. Basierend auf der Leistungsbeschreibung [gem_Leistung] (NFA1-16 aus K*apitel II 6. Anforderungen an eine Zero-Trust-Architektur* und NFA17 aus K*apitel II 4. Zero Trust)* ergeben sich die nicht-funktionalen Anforderungen in Tabelle 2.4-1. Wie die Architektur die nicht-funktionalen Anforderungen erfüllt, wird in Kapitel 6- Evaluation beschrieben.

Ref.	Name	Beschreibung
NFA1	Sichere Kommunikation	Die gesamte Kommunikation innerhalb der ZTA muss durchgehend gesichert sein, unabhängig von Standort und Netzwerkzugehörigkeit.
NFA2	Data (Datenschutz)	Daten sind durchgängig entsprechend ihrem Schutzbedarf gegen unberechtigte Zugriffe (inkl. seitens des Betreibers) zu schützen.
NFA3	Privacy/Datenschutz (Profilbildung)	Es muss verhindert werden, dass ein Akteur unberechtigt individuelle nutzerbezogene Profile über die Nutzung von Diensten anlegt.
NFA4	Data (Just-in-time)	Der Zugriff auf Daten soll möglichst nur für den benötigten Zeitraum (just-in-time-Access) erfolgen, in Abwägung gegen Usability- und Performanceanforderungen.
NFA5	Data (Just-enough)	Der Zugriff auf Daten soll möglichst nur mit den nötigen Privilegien (just-enough-Access) erfolgen, in Abwägung gegen Usability- und Performanceanforderungen.
NFA6	Keine Allmacht (Zugriff auf medizinische Daten)	Kein Akteur darf über alle Mittel verfügen, um systematisch unberechtigt auf schützenswerte medizinische Daten der Versicherten zuzugreifen.
NFA7	Identitäten	Ein Konzept mit föderierten Identitäten muss sich in die Zero-Trust-Architektur integrieren lassen.
NFA8	Performanz	Akteure müssen die Anwendungen der Telematikinfrastruktur performant nutzen können. Konkret muss die Lösung für mehr als 80 Millionen Versicherte und über 200.000 Leistungserbringer praktikabel, d.h. ohne störende Antwort-, Lauf- oder Reaktionszeiten, nutzbar sein.

Ref.	Name	Beschreibung
NFA9	Skalierbarkeit (kurzfristig)	Die Lösung muss skalierbar sein und auch bei kurzfristigen Lastschwankungen performant bleiben.
NFA10	Skalierbarkeit (langfristig)	Langfristig muss die Architektur in der Lage sein, dem absehbaren Wachstum von Anwender- und Leistungserbringergruppen innerhalb und außerhalb Deutschlands, einer großen Anzahl neuer Digital-Health-Anwendungen, Diensten, größeren Datenmengen und möglicher Einsatzszenarien gerecht zu werden.
NFA11	Auswirkungen auf Leistungserbringerprozesse (Erstanwendungen)	Die Lösung muss für die Erstanwendung beim Leistungserbringer einfach und möglichst problemlos integrierbar sein.
NFA12	Auswirkungen auf Leistungserbringerprozesse (Regelnutzung)	Die Lösung muss im Regelnutzungsfall einfach und möglichst problemlos in die Prozesse des Leistungserbringers integrierbar sein.
NFA13	Auswirkungen auf Leistungserbringerprozesse (Geräteregistrierung)	Die Registrierung und Deregistrierung von Geräten muss einfach und möglichst problemlos in die Prozesse des Leistungserbringers integrierbar sein.
NFA14	Usability	Durch die Zero-Trust-Architektur dürfen keine unverhältnismäßigen Zugangshürden (z.B. durch Registrierungs- und Authentisierungsprozesse, Anforderungen an Endgeräte, ...) für die Nutzer entstehen.
NFA15	Standards	Die Lösung soll so weit wie möglich international anerkannte und in ihrem Kern wissenschaftlich geprüfte Standards nutzen, um darauf basierende Produkte einsetzen zu können. Proprietäre Lösungen sind zu vermeiden.
NFA16	Verfügbarkeit	Die Lösung muss die Verfügbarkeit und Stabilität von Anwendungen entsprechend den Anforderungen der Fachdienste gewährleisten.
NFA17	Betreibbarkeit	Die Lösung muss in ihrer Gesamtheit sicher, effizient und mit etablierten Verfahren und Technologien betreibbar sein. Der Betrieb der Lösung muss den gängigen Qualitätskriterien entsprechen.

Tabelle 2.4-1: Nicht-funktionale Anforderungen

2.5 Annahmen

Weitere, dem Architekturkonzept zu Grunde liegende, Annahmen sind in Tabelle 2.5-1 dokumentiert.

Ref.	Name	Beschreibung
A1	Web-API	Fachdienste werden im Regelfall über webbasierte Protokolle (HTTP) angesprochen. Andere Fälle sind denkbar, solange die Zugriffskontrolle webbasiert erfolgt. Diese Ausnahmen werden im vorliegenden Konzept aber nicht behandelt..

Tabelle 2.5-1: Annahmen

3 Architektur

Ein großer Bestandteil des Architekturkonzeptes ist die Beschreibung der wesentlichen Komponenten und die Beschreibung wie diese Komponenten im Zusammenspiel die Kernfunktionen der Architektur erfüllen. Das Architekturkonzept folgt dabei einer Zero-Trust-Strategie. Im Folgenden (Kapitel 3.1) werden diesbezüglich zunächst die wichtigsten Zero-Trust-Paradigmen auf den vorliegenden Kontext der TI abgebildet. Im Anschluss werden die Komponenten (Kapitel 3.2), das verwendete Regelwerk (Kapitel 3.3), die verarbeiteten und gespeicherten Datenklassen (Kapitel 3.4) sowie die Realisierung der Kernfunktionen (Kapitel 3.5) beschrieben.

3.1 Paradigmen

Zu den gängigen, mit einer Zero-Trust-Strategie verbundenen Denkmustern gehören die in der NIST Special Publication 800-207 [NIST_ZeroTrust] in Kapitel 2.1 "Tenets of Zero Trust" erläuterten Paradigmen, die in Tabelle 3.1-1 aufgelistet werden.

Ref.	Name	Beschreibung
P1	Ressourcen	Alle Datenquellen und Rechendienste werden als Ressourcen betrachtet.
P2	Sichere Kommunikation	Die gesamte Kommunikation ist unabhängig vom Standort im Netzwerk gesichert.
P3	Sessions	Der Zugang zu einzelnen Ressourcen wird pro Session gewährt.
P4	Dynamisches Regelwerk	Der Zugang zu Ressourcen wird durch ein dynamisches Regelwerk bestimmt, dass den beobachtbaren Zustand der Nutzeridentität, der Anwendung/des Dienstes und des anfragenden Geräts umfasst und weitere Verhaltens- und Umgebungsattribute einschließen kann.
P5	Monitoring	Das Unternehmen überwacht und misst die Integrität und die Sicherheitslage aller eigenen und zugehörigen Ressourcen.
P6	Dynamische Authentisierung und Autorisierung	Alle Authentifizierungen und Autorisierungen für den Zugriff auf Ressourcen sind dynamisch und werden streng erzwungen, bevor der Zugriff erlaubt wird.
P7	Stetige Verbesserung	Das Unternehmen sammelt so viele Informationen wie möglich über den aktuellen Zustand der Architektur, der Netzinfrastruktur und der Kommunikation und nutzt sie, um seine Sicherheitslage zu verbessern.

Tabelle 3.1-1 Zero Trust Paradigmen gemäß NIST [NIST_ZeroTrust]

Für die Umsetzung der Zero-Trust-Strategie im Rahmen dieses Architekturkonzeptes werden im Folgenden die Paradigmen P1-P7 (Kapitel 3.1.1 - 3.1.7) für die Anwendung innerhalb der TI abgebildet.

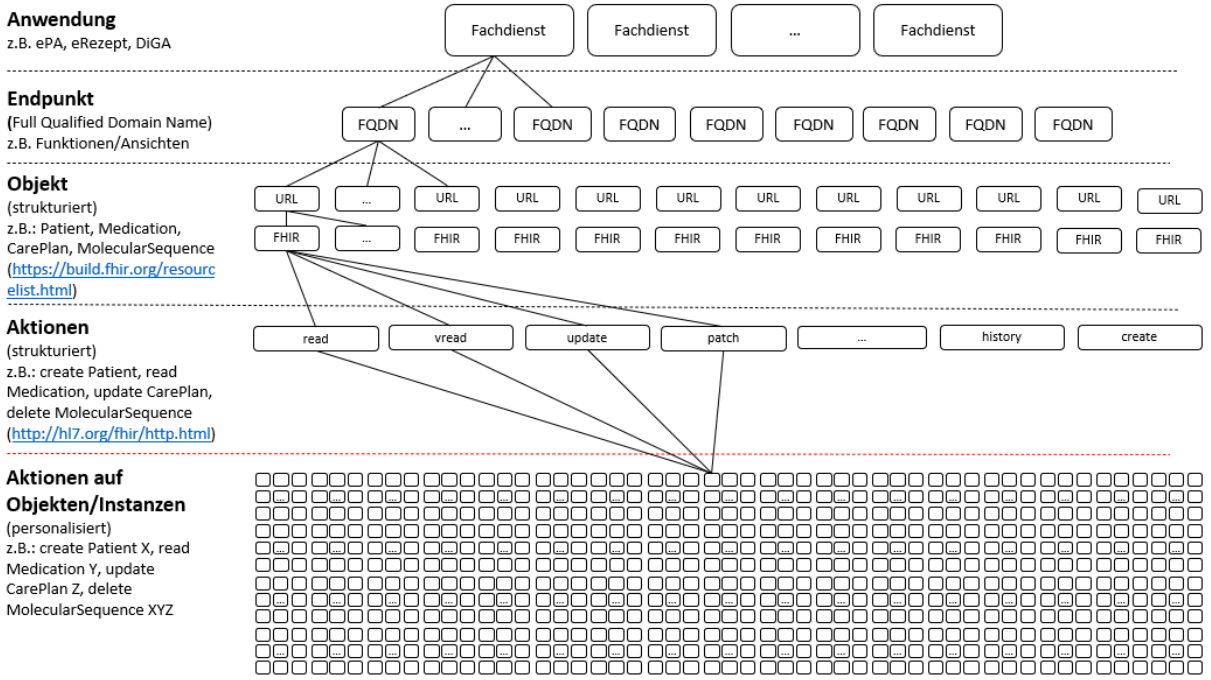

Abbildung 3.1.1-1: Ressourcen im Kontext der TI

3.1.1 Ressourcen

Für die Konzeption einer ZTA ist die Definition dessen, was unter einer Ressource der ZTA verstanden wird, von zentraler Bedeutung. In Abbildung 3.1.1-1 sind unterschiedliche Ebenen hinsichtlich der Granularität einer Ressource im Kontext von Anwendungen der TI dargestellt.

In der obersten Ebene werden Anwendungen in Form von Diensten und Fachdiensten unterschieden. Eine Autorisierung für den Zugriff auf eine Anwendung (z.B. die ePA) kann von einer Autorisierung für den Zugriff auf eine andere Anwendung (z.B. dem E-Rezept) unterschieden werden.

In der darunter liegenden Ebene können für jede Anwendung verschiedene Endpunkte als Ressource hinsichtlich der Autorisierung differenziert werden. Dies kann zum Beispiel anhand des Fully-Qualified Domain Name (FQDN) eines Endpunktes erfolgen.

Die dritte Ebene ermöglicht das Differenzieren zwischen einzelnen Objekten, z.B. FHIR-Ressourcen, welche innerhalb einer Anwendung adressiert werden können. Auf der vierten Ebene wird zusätzlich zur Differenzierung zwischen Objekten auch die angefragte Aktion auf das Objekt, z.B. FHIR-Operationen, unterschieden. Die Unterscheidung von Objekten und Aktionen auf Objekten kann z.B. anhand des URL-Pattern einer Zugriffsanfrage und/oder der HTTP-Methode erfolgen.

Die unterste Ebene kennzeichnet die Unterscheidung von einzelnen Instanzen von Datenobjekten. Dadurch könnte man z.B. das Anlegen eines neuen Patienten mit Namen X vom Anlegen eines neuen Patienten mit Namen Y unterscheiden.

Im Rahmen des Architekturkonzeptes werden die logischen Entitäten der ersten vier Ebenen als Ressourcen betrachtet, deren Zugriff durch das Regelwerk der ZTA gesteuert werden kann. Der Zugriff auf einzelne, personalisierte Daten, d.h. Daten im Sinne von einzelnen Instanzen auf Ebene 5, gehört jedoch nicht dazu.

Dabei soll es möglich sein, flexibel je nach Zugriffsanfrage und Anwendung, die Zugriffsentscheidung auf unterschiedlichen Ebenen zu treffen und so unterschiedlich granulare Regelwerke durch die ZTA zu unterstützen.

3.1.2 Sichere Kommunikation

Die Kommunikation zwischen Komponenten der ZTA erfolgt über das Internet. Das Architekturkonzept sieht vor, dass jede Kommunikation zwischen den Komponenten der ZTA gemäß dem Stand der Technik verschlüsselt und gegenseitig authentisiert erfolgt. Dies gilt auch wenn Komponenten der ZTA innerhalb organisationsspezifischer Netzwerke kommunizieren. Es werden die gültigen Standards, wie zum Beispiel BSI TR-02102 [BSI_TR02102], herangezogen. Der Transfer von medizinischen Daten erfolgt verschlüsselt zwischen den Entitäten, welche vom Eigentümer der medizinischen Daten (z.B. Patient/Versicherter) für den Zugriff auf diese autorisiert wurden.

3.1.3 Sessions

Das Architekturkonzept sieht vor, dass der Zugang zu Ressourcen auf Basis von Sessions geprüft und gewährt wird. Die mit der Session verknüpften Informationen werden gegen das Regelwerk geprüft.

Eine Session identifiziert im Rahmen des Architekturkonzepts eine etablierte Kommunikationsbeziehung eines Clients mit einem Fachdienst. Diese Session ist dabei

einem einzelnen Nutzer, Gerät und ggf. Umgebung zugeordnet, und kann für mehrere Zugriffe (Requests) auf Ressourcen beim gleichen Fachdienst verwendet werden.

Die Session ist dem Client, PEP und Fachdienst bekannt. Session-spezifische Informationen (wie Identifier) werden ausschließlich während der Session verwendet, und nur für die Lebenszeit einer Session gespeichert. Durch diese clientseitig gespeicherten Session-Informationen allein lassen sich allerdings keine Rückschlüsse auf die hinterlegten Identitäts- und Sicherheitsnachweise ziehen. Die nutzerspezifische Session ist an das Gerät sowie ggf. die Umgebung gebunden und kann nicht in anderen Kontexten verwendet werden.

3.1.4 Dynamisches Regelwerk

Das Architekturkonzept berücksichtigt unterschiedliche Ebenen der Dynamik des Regelwerkes:

- Dynamische Regelauswertung: Das Architekturkonzept sieht ein Regelwerk auf der Basis von Attribute-Based Access Control (ABAC) vor. Dabei können bei der Zugriffsentscheidung Attribute berücksichtigt werden, welche dynamisch zum Zeitpunkt der Zugriffsanfrage erhoben werden (z.B. Zeitpunkt des Zugriffs, Ort des Zugriffs, Sicherheitsattribute des für den Zugriff verwendeten Endgerätes).
- Aktualisierungen von Referenzwerten: Das Regelwerk der ZTA kann für die Bewertung von Attributen Referenzwerte verwenden, die erlaubte (Allowlist) oder verbotene Werte (Blocklist) für ein Attribut festlegen. Diese Referenzwerte können in definierten Prozessen kurz- oder mittelfristig angepasst werden, um auf Monitoring-Ergebnisse oder aktuelle Entwicklungen (z.B. aktuelle Sicherheitsvorfälle) zu reagieren.
- Anpassungen des Regelwerks: Die Regeln im Regelwerk der ZTA sind nicht unveränderlich festgelegt, sondern können über organisatorische Prozesse angepasst, geändert oder entfernt werden. So kann z.B. ein neuer Fachdienst aufgenommen und mittel- und langfristig auf veränderte Anforderungen an die Informationssicherheit reagiert werden.

3.1.5 Monitoring

Das Architekturkonzept sieht drei verschiedene Arten von komponentenübergreifendem Monitoring vor:

- Betriebsmonitoring: Kontinuierliches Monitoring von Informationen über Verfügbarkeit, Auslastung und Performance der Komponenten der TI-Infrastruktur mit dem Ziel, Ausfälle oder Überlastung zu erkennen.
- Security-Monitoring: Kontinuierliches Monitoring, um mögliche Angriffe oder Sicherheitsprobleme zu erkennen. Es umfasst einerseits das Monitoring von Informationen zum Sicherheitsstatus der Komponenten der TI-Infrastruktur zur Identifikation möglicher Schwachstellen und Sicherheitsvorfällen in diesen Komponenten und andererseits das Monitoring von aktuellen Zugriffsanfragen zur Angriffserkennung.
- Monitoring des Regelwerks: Logging und Analyse von Zugriffsentscheidungen, d.h. von zugelassenen und zurückgewiesenen Zugriffen, um die Wirksamkeit des Regelwerkes zu überwachen und Änderungsbedarfe zu erkennen.

Die Monitoring-Informationen werden außerhalb der überwachten Komponenten in der TI-Infrastruktur übergreifend gesammelt, gespeichert und ausgewertet.

Eine Betrachtung des zusätzlich möglichen, internen Komponenten-Monitorings pro Komponente wird im Nachfolgenden nicht vorgenommen.

3.1.6 Dynamische Authentisierung und Autorisierung

Jede Autorisierung eines Zugriffs auf eine Ressource setzt eine vorangegangene Authentisierung des Nutzers voraus. Die Autorisierung ist an den Umfang der Zugriffsanfrage (just-enough) gebunden und zeitlich auf die Dauer der Session begrenzt (just-in-time).

3.1.7 Stetige Verbesserung

Das Architekturkonzept sieht im Rahmen des Monitorings vor, Informationen über den Sicherheitszustand der an der ZTA beteiligten Komponenten, der durch sie geschützten Ressourcen und die erfolgten Zugriffe zu sammeln, um so eine stetige Verbesserung des Regelwerkes und seiner Umsetzung zu ermöglichen.

3.2 Komponenten

Die in diesem Konzept vorgeschlagene Architektur basiert auf den grundlegenden Strategien einer ZTA, wie in NIST SP 800-207 [NIST_ZeroTrust] erläutert. Sie nutzt dafür die gemäß ISO/IEC 29146 [ISO29146] standardisierten und im Rahmen von Zero-Trust etablierten logischen Komponenten. Abbildung 3.2-1 zeigt den konzeptionellen Aufbau der Architektur. Bereits im Rahmen der TI 2.0 konzipierte Komponenten, welche im Rahmen dieses Konzeptes als vorhanden vorausgesetzt werden, sind blau hinterlegt. Neue Komponenten der in diesem Konzept beschrieben ZTA sind orange hinterlegt. Im Folgenden (Kapitel 3.2.1) werden die Komponenten kurz beschrieben. Im Anschluss (Kapitel 3.2.2) wird erläutert, welche Architekturkomponenten an der Realisierung welcher Kernfunktionen beteiligt sind.

3.2.1 Dekomposition

Die Architekturkomponenten können den fünf funktionalen Gruppen zugeordnet werden, die im Folgenden beschrieben sind.

I. Systeme der Nutzer

Die Dienste der TI sollen verschiedenen Nutzergruppen zur Verfügung stehen. Für den Zugriff der Nutzer (Versicherte, Leistungserbringer, Kostenträger sowie andere Fachdienste) auf Dienste der TI werden Systeme benötigt, welche eine sichere Authentisierung und die sichere Verarbeitung von Gesundheitsdaten und deren kryptografischen Schlüsseln ermöglichen.

Fach-Client (FCL)/Trust-Client (TCL)

Nutzer greifen auf Dienste der TI über einen Fach-Client zu, welcher die notwendige Fach-Client spezifische Logik für den Zugriff implementiert. Das können z.B. typische Primärsysteme der Leistungserbringer, aber auch mobile Applikationen der Krankenkassen sein. Für die sichere Nutzung der Dienste müssen Nutzer zum Zeitpunkt des Zugriffs ausreichend sichere Endgeräte nachweisen. Dabei sind Geräte, welche für die Nutzeridentifizierung (Authentisierung) verwendet werden, nicht zwingend identisch mit denen für den Zugriff auf Dienste, d. h. es kann ein separates Gerät mit einer Software zur Bestätigung der Anmeldung zum Einsatz kommen. Der Trust-Client sammelt

Informationen über die Plattformsicherheit des Endgeräts und stellt diese als Basis für Zugriffsentscheidungen bereit. So kann sichergestellt werden, dass das Endgerät die notwendigen Funktionen zum Aufbau sicherer Kommunikationsverbindungen unterstützt und fachspezifische Anforderungen an das Endgerät und seine Ausführungsumgebung erfüllt werden. Zudem nutzt der Trust-Client eine sichere Schlüsselverwaltung nach dem bei den Nutzern verfügbaren Stand der Technik (z.B. Trusted Execution Environment (TEE), Trusted Platform Module (TPM), Secure Element (SE)), um beispielsweise eine Gerätebindung/-registrierung für den Zugriff auf die TI oder die Bestätigung der Plattformsicherheitsattribute zu unterstützen. Diese Schlüsselverwaltung kann über eine Krypto-API auch vom Fach-Client genutzt werden.

Der Trust-Client steht in Form eines SDK oder als eigenständiges Modul für verschiedene Endgeräteklassen zur Verfügung und kann so je nach Nutzerumgebung in einen Fach-Client integriert werden.

1. Alleinstehende Geräte: In Umgebungen, in denen der Nutzer direkt über das Endgerät, auf dem der Fach-Client installiert ist, auf den Fachdienst zugreift, kann der Trust-Client in einfacher Form direkt in die Applikation des Fach-Clients integriert werden.
2. Client-Server-Infrastruktur: In größeren medizinischen Einrichtungen, wie zum Beispiel einem Krankenhaus, kann der Zugriff auf Fachdienste neben alleinstehenden Geräten zusätzlich über eine einrichtungsspezifische Client-Server-Infrastruktur zur Verfügung gestellt werden. Fach- und Trust-Client laufen dann auf einem Fachclientserver der Client-Server-Infrastruktur. Der Trust-Client stellt dort die notwendigen Funktionen zum Aufbau sicherer Kommunikationsverbindungen bereit und sorgt dafür, dass die nötige Information für die fachdienstspezifische Sicherheitsbewertung des Fachclientservers und dessen Ausführungsumgebung zur Verfügung gestellt werden. Endgerät im Sinne einer Gerätebindung/-registrierung für den Zugriff auf die TI durch die medizinische Einrichtung ist dann der Fachclientserver der Client-Server-Infrastruktur. Der sichere Betrieb der Client-Server-Infrastruktur sollte in Verantwortung des Krankenhauses über ein zertifiziertes ISMS des Krankenhauses und geeignete technische Maßnahmen sichergestellt werden. Zusätzlich zur Authentisierung eines registrierten Fachclientservers sollte der Trust-Client eine ggf. pseudonymisierte Gerätekennung des einrichtungsspezifischen Clientgerätes zur Verfügung stellen, welches innerhalb der medizinischen Organisation für den Zugriff auf den Fachclientserver verwendet wurde. Dies ermöglicht der ZTA über ein Monitoring etwaige kompromittierende Zugriffe aus der Umgebung der medizinischen Einrichtung zu erkennen, einem organisationsinternen Terminal zuzuordnen und ggf. die Einrichtung zu informieren oder den Zugriff des spezifischen Terminals selektiv zu blockieren. Die nicht kompromittierten Teile der Client-Server-Infrastruktur könnten durch eine solche granulare Entscheidung weiter für Zugriffe auf Dienste der TI genutzt werden.
3. Proxy-Server-Infrastruktur: Es ist darüber hinaus denkbar, den Trust-Client in medizinischen Einrichtungen als eigenständige Komponente in Form eines Proxy-Servers unabhängig vom Fach-Client zur Verfügung zu stellen. Der Trust-Client läuft dann auf dem Proxy-Server, während Fach-Clients auf verschiedenen dahinterliegenden Geräten installiert sind. Für die Verwaltung dieser, für die Fach-Clients genutzten, internen Endgeräte verfügt die Einrichtung über ein internes Gerätemanagement, welches die Plattformsicherheit der internen Endgeräte attestiert und resultierende Informationen dem Trust-Client auf dem Proxy-Server zur Verfügung stellt. Der Trust-Client stellt dann die notwendigen Funktionen zum Aufbau sicherer Kommunikationsverbindungen vom Fachdienst bis zum Proxy-Server bereit und stellt die für die Bewertung der Zugriffe nötigen Information über die internen Endgeräte der ZTA zur Verfügung. Endgerät im Sinne einer

Gerätebindung/-registrierung für den Zugriff auf die TI durch die medizinische Einrichtung ist dann der Proxy-Server. Der sichere Betrieb der einrichtungsspezifischen Infrastruktur und des internen Gerätemanagements sollte in diesem Fall ebenso in Verantwortung des Krankenhauses über ein zertifiziertes ISMS des Krankenhauses und geeignete organisatorisch-technische Maßnahmen sichergestellt werden. Analog zur Client-Server-Infrastruktur sollte der Trust-Client dabei ggf. pseudonymisierte Gerätekennungen der internen Endgeräte zur Verfügung stellen, um so eine granulare Analyse innerhalb der TI zu ermöglichen.

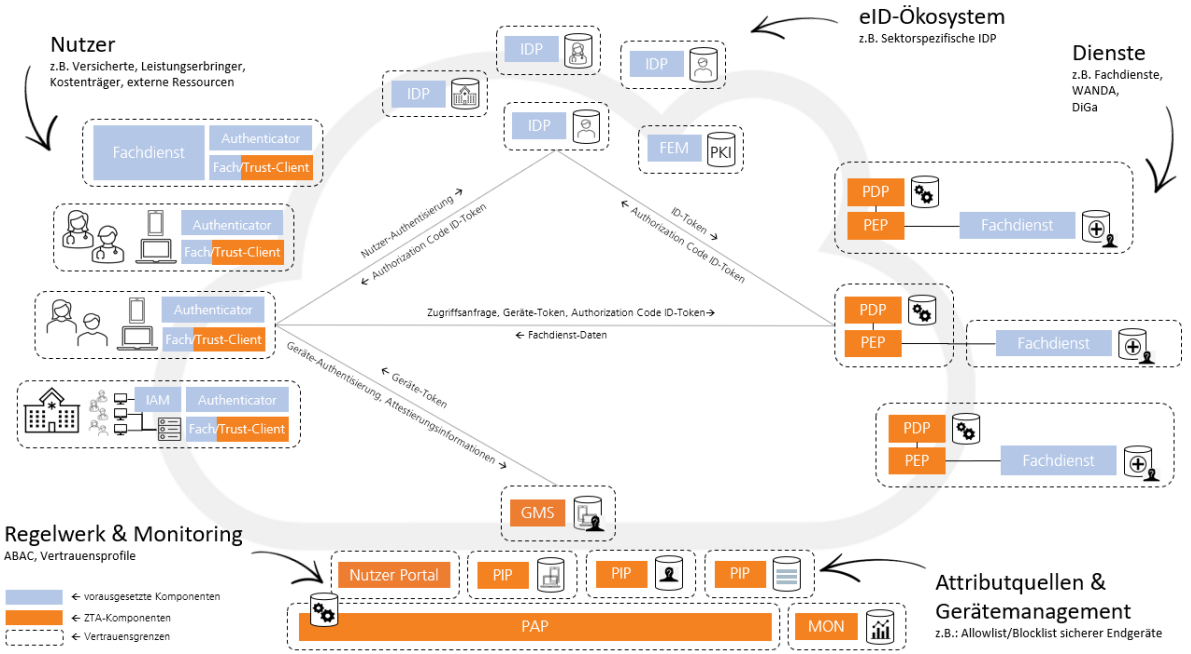

Abbildung 3.2-1: Konzeptioneller Aufbau der Architektur

Im Ergebnis stehen den medizinischen Einrichtungen verschiedene Optionen zur Verfügung, um den Anschluss an die TI in die eigene Architektur und das einrichtungsspezifische Informationssicherheitsmanagement zu integrieren. Während im Rahmen von Option 1 der Großteil der Informationssicherheitsmaßnahmen durch den Trust-Client unterstützt werden kann, steigt mit Option 2 und insbesondere mit Option 3 die Verantwortung für die medizinische Einrichtung und deren ISMS. Medizinische Einrichtungen können so die Aufwände für das Umsetzen von Maßnahmen eines einrichtungsspezifischen ISMS gegen Aufwände für die Integration des Trust-Clients abwägen. Risiken hinsichtlich der Beschränkung des Zugangs zur TI für ggf. kompromittierte Endgeräte oder Server, können für Option 2 und 3 dadurch reduziert werden, dass pseudonymisierte Gerätekennungen für das Monitoring der TI bereitgestellt werden.

Authenticator-Modul (AUM)

Für die sichere Identifizierung und Authentisierung verfügen Nutzer als Teil des IDP-Ökosystems der TI 2.0 über Identitätsmittel. Diese werden mit dem Authenticator-Modul in das jeweilige eID-System verschiedener Identitätsprovider (IDP) der Föderation der TI integriert. Darüber hinaus stellt das Authenticator-Modul Informationen über das für die Identifizierung bzw. Authentisierung genutzte Endgerät und ggf. dessen Ausführungsumgebung zur Verfügung.

Für medizinische Einrichtungen wird hinter dem Authenticator-Modul in der Regel ein organisationsspezifisches Identitäts- und Zugriffsmanagement (IAM) betrieben. Dabei werden organisationsinterne Identitäten der Mitarbeitenden verwendet, welche nicht in der Föderation registriert sein müssen. Stattdessen ist die Organisation mit einer Organisationsidentität in der Föderation der TI registriert. Bei einer Authentisierung eines Zugriffs sollte neben der Organisationsidentität auch die für den Zugriff verwendete organisationsinterne Identität in pseudonymisierter Form als Attribut in das ID-Token mit einfließen und so der ZTA zur Verfügung gestellt werden. Dies ermöglicht der ZTA z.B. etwaige kompromittierende Zugriffe aus der Umgebung der medizinischen Einrichtung zu erkennen, einer organisationsinternen pseudonymen Identität zuzuordnen und diese ggf. zu blockieren. Die Organisationsidentität selbst kann dabei weiter für Zugriffe auf Dienste der TI genutzt werden. Der sichere Betrieb des IAM sollte in Verantwortung der medizinischen Einrichtung über ein zertifiziertes ISMS der medizinischen Einrichtung und geeignete technische Maßnahmen sichergestellt werden.

Im Ergebnis stehen medizinischen Einrichtungen für den Zugang zur TI neben der Verwendung von Identitäten für natürliche Personen (vgl. HBA) auch Organisationsidentitäten (vgl. SMC-B) zur Verfügung. Während im Rahmen der direkten Verwendung von Identitäten für natürliche Personen der Großteil der Informationssicherheitsmaßnahmen hinsichtlich der Identifizierung und Authentisierung durch die Identity Provider der Föderation der TI realisiert wird, steigt mit der Verwendung einer Organisationsidentität die Verantwortung für die medizinische Einrichtung und deren IAM und ISMS. Risiken hinsichtlich der Beschränkung des Zugangs zur TI für ggf. als kompromittiert identifizierte Organisationsidentitäten, können dadurch reduziert werden, dass organisationsinterne Identität in pseudonymisierter Form für das Monitoring der TI bereitgestellt werden.

Nutzerportal

Einige Komponenten der ZTA benötigen bewusste Nutzeraktionen, wie z.B. das Löschen von Geräten aus der Liste registrierter Geräte. In anderen Fällen ist es sinnvoll, dem Nutzer Informationen bereitzustellen, z.B. zur persönlichen Nutzungshistorie, zu Zugriffen auf seine Daten durch Leistungserbringer oder zum Status der für ihn relevanten Fachdienste. Im Sinne einer besseren Transparenz und Nutzbarkeit für den Nutzer, sollen viele dieser Aufgaben in einem Nutzerportal gebündelt werden. Dieses Portal hat keine eigene Funktion innerhalb der ZTA. Es dient nur als Frontend für die Administration oder das Monitoring

spezifischer Komponenten, wie dem Geräte Management Service. Die konkrete Umsetzung wird daher im Feinkonzept nicht betrachtet.

II. Systeme der Dienste der Telematikinfrastruktur und des digitalen Gesundheitswesens

Die TI stellt Dienste und Anwendungen für das Gesundheitswesen zur Verfügung.

Fachdienst

Fachanwendungen sind Anwendungen der TI, wie das Versichertenstammdaten-Management oder die elektronische Patientenakte, mit allen nötigen technischen und organisatorischen Anteilen auf Anwendungsebene. Fachanwendungen können durch einen oder mehrere Fachdienste realisiert werden. Fachdienste können zudem weitere Anwendungen für den Datenaustausch in der Telematikinfrastruktur (WANDA) oder digitale Gesundheitsanwendungen (DiGA) repräsentieren. Fachdienste speichern und verarbeiten unter anderem sensible Nutzerdaten wie Gesundheitsdaten und stellen diese zur Verfügung.

Policy Enforcement Point (PEP)

Jeder Zugriff auf Fachdienste ist ausschließlich nach zuvor erfolgreich erfolgter Autorisierung möglich. Der Policy Enforcement Point setzt dies für jeden einzelnen Zugriff auf einen Fachdienst durch, erlaubt autorisierten Zugriff in der Granularität der ZTA-Ressourcen und schützt vor unautorisiertem Zugriff. Dazu reicht er bei der initialen Autorisierung die vom Client gelieferten Nutzer-, Geräte- und Umgebungsinformationen an den PDP weiter und erwartet von diesem eine Zugriffsentscheidung. Zudem stellt der PEP anonymisierte Informationen über Zugriffsentscheidungen sowie vorverarbeitete Informationen zu auffälligen Zugriffsanfragen für das Monitoring bereit.

Policy Decision Point (PDP)

Der Kern einer Zero-Trust-Architektur besteht in der Entscheidung, ob ein bestimmter Nutzer unter Berücksichtigung der vorliegenden Informationen zu Nutzer, Gerät und Umgebung zu einem bestimmten Zeitpunkt auf eine bestimmte Ressource zugreifen darf. Die Architektur sieht dafür für jeden Fachdienst einen PDP vor, der diese Entscheidung trifft. Voraussetzung dafür sind aktuelle und explizite Richtlinien, die festlegen, wie authentifizierte Nutzer, Dienste, Geräte und Anwendungen miteinander interagieren dürfen. Ein PDP erhält deshalb vom PAP das aktuell gültige Regelwerk des betroffenen Fachdienstes, bestimmt daraus die für den aktuellen Zugriff geltenden Regeln und wertet diese aus. Dafür werden vom PDP alle für die Auswertung notwendigen Informationen ermittelt, soweit sie nicht bereits mit der Autorisierungsanfrage vom Client zur Verfügung gestellt werden.

Das Konzept folgt hinsichtlich der Realisierung von PDP und PEP dem Enterprise-Centric-Implementation-Modell gem. ISO/IEC 29146 [ISO29146] und sieht beide Komponenten innerhalb eines gemeinsames Vertrauensraums vor. Dieser kann realisiert werden durch eine gemeinsame Grenze um Fachdienst und PEP/PDP herum oder als separierte Instanzen (z.B. IaaS-Komponenten), die auf vertrauenswürdige Weise gekoppelt sind.

III. Systeme der eID-Föderation der TI

Die TI 2.0 umfasst ein föderiertes eID-Ökosystem [gem_IDP_Federation] bestehend aus verschiedenen sektorspezifischen IDPs.

Identity Provider (IDP)

Eine der Grundlagen für die Autorisierung eines Zugriffs auf Dienste der TI ist eine erfolgreiche Authentifizierung des Nutzers. IDPs ermöglichen diese Authentifizierung. Sie sind für unterschiedliche Sektoren (z.B. Krankenkassen für Versicherte, Leistungserbringerorganisationen für Ärzte) spezifisch ausgeprägt. IDPs können diese

Aufgabe mit unterschiedlichen Technologien erfüllen (z.B. auf Basis von kartenbasierten Identitätsmitteln, mobilen Endgeräten oder der Integration in eID-Wallets). IDPs registrieren, speichern und verwalten Identitätsdaten.

Federation Master (FEM)

Die Public Key Infrastruktur (PKI) des TI-Vertrauensraum wird durch den sogenannten Federation Master verwaltet. Beim Federation Master sind alle Dienste und zentralen Komponenten der TI registriert. Nur dort registrierte Teilnehmer sind berechtigt, die Dienste der Föderation in Anspruch zu nehmen. Der FEM verwaltet und speichert z.B. die öffentlichen Schlüssel der IDPs, der Fachdienste und weiterer Komponenten der ZTA.

IV. Übergreifende Systeme für die Ausführung des Zugriffsmanagements

Für die Autorisierung von Nutzerzugriffen auf Dienste der TI sind Komponenten notwendig, welche weitere Informationen für die Regelauswertung bereitstellen.

Geräte Management Service (GMS)

Die Architektur ermöglicht das Registrieren, Authentisieren und Entfernen von Endgeräten der Nutzer für den Zugriff auf Ressourcen der TI. Diese Funktionalitäten werden durch den GMS realisiert. Endgerät in diesem Sinne ist das Gerät, auf dem der Trust-Client für den gesicherten Zugriff auf die TI läuft. Dabei wird eine Bindung des registrierten Endgerätes an den registrierenden Nutzer berücksichtigt. Beispiele dafür sind die Registrierung eines mobilen Endgerätes durch einen Versicherten oder die Registrierung eines Fachclientservers oder Proxyservers durch eine medizinische Einrichtung. Angriffe auf Dienste der TI von nicht registrierten Endgeräten können von vornherein eingegrenzt werden. Zusätzlich kann für den Zugriff auf Dienste der TI die Nutzerbindung geprüft werden. Je nach Anforderung des Fachdienstes können z.B. Zugriffe auf solche Endgeräte beschränkt werden, die für den authentifizierten Nutzer registriert sind. Zugriffe von Geräten aus medizinischen Einrichtungen könnten für alle registrierten Rollen der TI ermöglicht werden. Der GMS stellt über das Geräte-Token Informationen über das im Rahmen eines Nutzerzugriffs authentisierte Endgerät und dessen Ausführungsumgebung bereit.

Der GMS ist als eine von den sektoralen IDPs der Föderation der TI möglichst unabhängige Komponente geplant, welcher die ZTA neben der Identifizierung und Authentisieren des Nutzers durch den IDP um einen weiteren Sicherheitsanker ergänzt.

Policy Information Point (PIP)

Für die Auswertung einzelner Regeln der Zero-Trust-Architektur sind situationsbezogene Informationen erforderlich. Neben dem IDP, dem GMS und dem Trust-Client, werden diese Informationen von Policy Information Points (PIPs) bereitgestellt. PIPs können Informationen zentral für alle Dienste bereitstellen (fachdienstübergreifender PIP) oder bei einem Fachdienst lokale Informationen für die Regelauswertung bereithalten (fachdienstspezifischer PIP). Ein Beispiel für den ersten Fall sind Informationen über sichere Endgeräteklassen, welche in die Bewertung der Gerätesicherheit und damit in die Zugriffsentscheidung einfließen können. PIPs können auch mit aktuellen Informationen aus dem übergreifenden Monitoring befüllt werden. So können Zugriffsentscheidungen dynamisch auf Basis aktueller Informationen zum Zeitpunkt des Zugriffs getroffen werden.

V. Systeme für die Administration und Verwaltung des Zugriffsmanagements

Policy Administration Point (PAP)

Die Regeln einer Zero-Trust-Architektur müssen erstellt, konfiguriert, getestet und verwaltet werden. Diese Funktion übernimmt der PAP. Das über den PAP verwaltete Regelwerk ermöglicht eine attributbasierte Zugriffskontrolle (ABAC). Dadurch können unterschiedliche Zugriffsanforderungen an verschiedene Rollen wie z.B.

Leistungserbringer oder Versicherte mit unterschiedlichen Anforderungen an deren Endgeräte und Umgebung realisiert werden.

Monitoring (MON)

Das übergreifende Monitoring in der TI2.0 ist dafür zuständig, kontinuierlich den Betrieb und Sicherheitsstatus der TI sowie die Anwendung des Regelwerks zu überwachen. Zu diesem Zweck werden Betriebs- und Sicherheitsinformationen von allen Komponenten der TI-Infrastruktur (d.h. allen Komponenten außer denen auf dem Nutzerendgerät) sowie Informationen zu Zugriffsanfragen und -entscheidungen von den PEPs der Fachdienste erfasst und im Monitoring ausgewertet.
Das Betriebsmonitoring ermöglicht dabei das Erkennen von Ausfällen und die Überwachung von vereinbarten SLAs für die Komponenten der TI.

Als Teil des Security-Monitorings dürfen, ggf. über automatisierte Prozesse, direkt Referenzwerte für bestimmte Attribute (z.B. Blocklisten für verdächtige IP-Adressen) an einem entsprechenden PIP aktualisiert werden. Davon abgesehen sollen Informationen aus dem Security-Monitoring in ein übergreifendes Security Information and Event Management (SIEM) einfließen bzw. einem Security Operations Center (SOC) für die TI zur Verfügung gestellt werden.

Das Monitoring der Anwendung des Regelwerks unterstützt bei der Auswertung von erfolgten Zugriffen und der Bewertung von Effektivität und Angemessenheit des Regelwerks. Darüber hinaus können Informationen gewonnen werden, welche es ermöglichen, die Auswirkung eines veränderten Regelwerkes abzuschätzen, z.B. Informationen über an den Transaktionen der ZTA beteiligte Endgeräteklassen.

3.2.2 Funktionale Abbildung

Jede Kernfunktion der ZTA wird durch das Zusammenspiel mehrerer Komponenten der ZTA realisiert. In Tabelle 3.2-2. ist je Kernfunktion dargestellt, welche Komponenten der ZTA an der Realisierung beteiligt sind. Externe Funktionen sind dabei jedoch nicht aufgeführt. Im Kapitel 3.5 wird das Zusammenspiel der Komponenten für die Realisierung der Kernfunktionen näher erläutert.

FA / Komponente		FCL	TCL	AUM	Nutzerportal	Fachdienst	PEP	PDP	IDP	FEM	PIP	GMS	PAP	MON
FA0	Nutzer-Zugriff auf Ressourcen der TI autorisieren	X	X	X		X	X	X	X	X	X	X		
FA1	Zugriff gegen Regelwerk prüfen						X	X			X			
FA1.1	Regelwerk erstellen und aktualisieren							X			X		X	
FA1.2	Regelwerk überwachen						X							X
FA1.3	Regelwerk anpassen							X			X		X	X
FA1.4	Attribute des Regelwerks anpassen										X			X
FA2	Nutzer bei Fachdienst identifizieren		X	X		X	X	X	X					
FA3	Fachdienst gegenüber Nutzer authentisieren		X				X							
FA5	Gerät authentifizieren		X				X					X		
FA5.1	Gerät registrieren	(X)	X									X		
FA5.2	Gerät entfernen								X			X		
FA6	Gerät attestieren		X									X		
FA6.1	Geräteinformationen erheben		X											
FA7	Autorisierter Zugriff	X	X			X	X	X						
FA7.1	Session-Management		X				X		X					
FA7.2	StepUp-Autorisierung		X	X			X	X	X					
FA8	ZTA-Komponenten überwachen				X	X	X	X	X	X	X	X	X	X

Tabelle 3.2-2: Abbildung der Kernfunktionen auf Komponenten der ZTA

3.3 Regelwerk

Die Umsetzung eines dynamischen Regelwerkes für die Zugriffskontrolle folgt dem Attribute-based Access Control Pattern (ABAC). Die Zugriffsentscheidung basiert dabei auf definierten Attributen, welche der Zugriffsanfrage zugeordnet und mit den Anforderungen einer Ressource verglichen werden können. Identity-based Access Control (IBAC), Pseudonyme-based Access Control (PBAC) und Role-based Access Control (RBAC) sind Sonderfälle von ABAC, bei denen die Attribute Identitäten, pseudonyme Identitäten bzw. Rollen sind. In Abbildung 3.3-1 ist dieses Zusammenwirken grafisch dargestellt.

Jede Zugriffsanfrage erfordert einen authentifizierten Nutzer als Subjekt der Anfrage. Informationen über die Authentisierung und den Nutzer selbst bilden in Form von Attributen die Subjektbeschreibung, die z.B. das Vertrauensniveau der Authentisierung, Rollen des Nutzers in der TI oder den Zeitpunkt der Authentisierung umfassen kann.

Jeder Nutzerzugriff erfolgt mit einem Gerät, welches der Nutzer bedient. Attribute, welche das Gerät beschreiben, bilden die Gerätebeschreibung und können z.B. Informationen über die Registrierung des Gerätes auf den Nutzer, ein Vertrauenslevel für die Sicherheitsfunktionen des registrierten Gerätes oder ggf. Informationen über hinter dem registrierten Endgerät zugreifende Terminals umfassen.

Informationen, welche über die Umgebung des Zugriffs gesammelt werden, werden als Umgebungsbeschreibung berücksichtigt, die z.B. Zeitpunkt oder Ort der Zugriffsanfrage sowie den Adressraum der Zugriffsanfrage enthalten kann. Informationen über die angefragte Ressource (z.B. E-Rezept, medizinisches Dokument, ...) und die damit verbundene Aktion (z.B. lesen, schreiben, ...) können ebenfalls der Zugriffsanfrage entnommen werden und ermöglichen das Ermitteln der über den Zugriff entscheidenden Regeln sowie der dazugehörenden Soll-Attribute bzw. Referenzwerte, bspw. für geforderte Rollen oder Geräteeigenschaften.

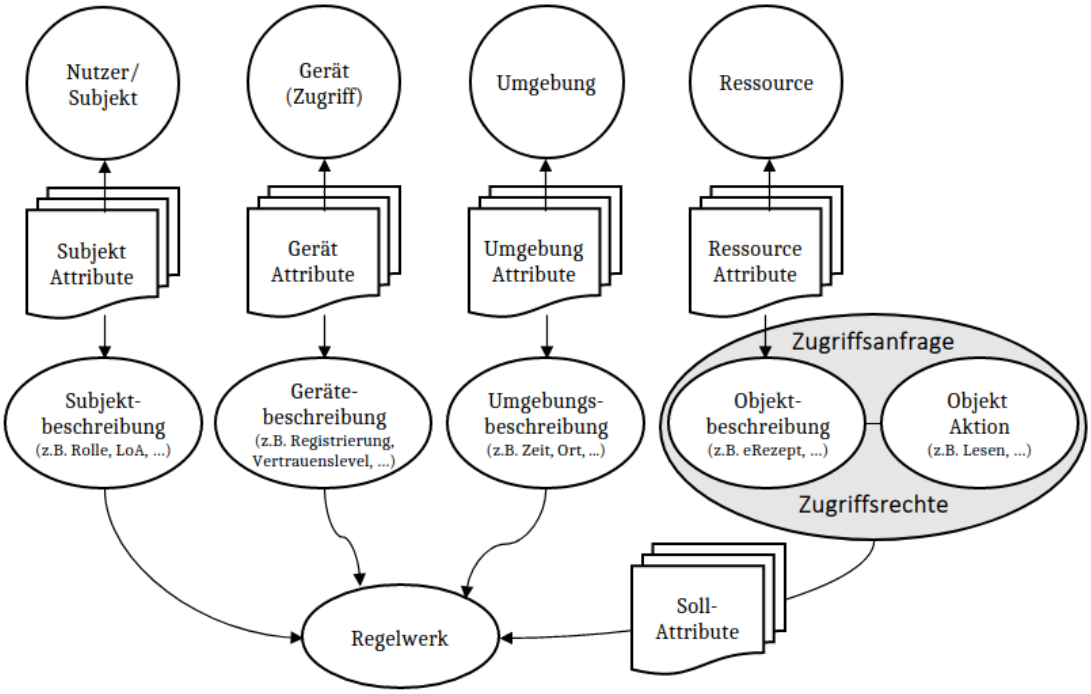

Abbildung 3.3-1: Visualisierung Attribute für ABAC in der ZTA

3.3.1 Mögliche Attribute und Regeln

Auf Basis der zur Verfügung stehenden Attribute können Regeln für den Zugriff auf Ressourcen erstellt werden. In Tabelle 3.3.1-1 sind Beispiele für mögliche Regeln dargestellt. Die Regeln sind den einzelnen Attributklassen aus Abbildung 3.5-1 zugeordnet. Für jede Regel werden die erforderlichen Attribute und deren mögliche Quellen sowie die Soll-Attribute genannt. Die Soll-Attribute selbst sind entweder Teil einer Regel oder sie werden von einem PIP bereitgestellt. Darüber hinaus wird für jede Regel das Ziel hinsichtlich der Informationssicherheit und ein konkretes Beispiel beschrieben.

Attribut-klasse	Regelname	Regel	Attribut → Quelle	Soll-Attribute	Sicherheitsziel	Beispiel
Nutzer	Rollenattribut des Nutzers in der TI	Es wird geprüft, ob der Nutzer die für die Zugriffsanfrage notwendige Rolle nachweisen kann.	Rolle, die der Nutzer im Rahmen der Authentisierung nachgewiesen hat → IDP (ID-Token)	zulässige Rollen für Zugriff	Ausschluss von Zugriffen, für die der Nutzer nicht die richtige Rolle nachweisen kann.	Zugriffe, die nicht den Regelungen des Rollenmodells z.B. gem. §§352 und 361 SGB V entsprechen, werden ausgeschlossen.
Nutzer	Vertrauensniveau (LoA) für Identifizierung/ Authentisierung	Es wird geprüft, ob der Nutzer das für die Zugriffsanfrage notwendige minimale Vertrauensniveau für die Identifizierung/Authentisierung nachweisen kann.	Vertrauensniveau, auf dem die Identifizierung/Authentisierung erfolgt ist (LoA) → IDP (ID-Token)	minimales LoA für Identifizierung / Authentisierung	Ausschluss von Zugriffen, für die der Nutzer nicht ausreichend stark identifiziert/authentisiert ist.	Medizinische Daten dürfen nur nach Identifizierung/ Authentisierung mit dem Vertrauensniveau "gematik-ehealth-loa-high" [gem_Spec_IDP_Sek] eingesehen oder geändert werden.

Attribut-klasse	Regelname	Regel	Attribut → Quelle	Soll-Attribute	Sicherheitsziel	Beispiel
Endgerät	Stärke der Identifizierung/ Authentisierung des Endgeräts (vergleichbar zu LoA)	Es wird geprüft, wie gut das vom Endgerät für die Identifikation/Authentisierung verwendete Merkmal geschützt ist.	Vertrauensniveau der Geräteauthentifizierung → GMS (Geräte-Token)	minimale geforderte Stärke der Geräte-authentifizierung	Ausschluss/Beschränken von Zugriffen von Endgeräten mit nicht hinreichend geschützten Schlüsseln (die z.B. nur auf der Festplatte liegen).	Medizinische Daten dürfen nur geändert werden, wenn der verwendete Schlüssel des zuvor registrierten und erfolgreich authentisierten Endgeräts mit einem Hardware-Sicherheitsmechanismus (z.B. in einem TPM) geschützt ist.
Endgerät	Aktuelles Betriebssystem	Es wird geprüft, ob auf dem Endgerät ein aktuelles Betriebssystem mit eingespielten Updates läuft.	Betriebssystem-Version inkl. Patch-Level → TCL/GMS (Geräte-Token)	Allowlist, Blocklist	Das Gerät verfügt über ein Betriebssystem mit grundsätzlich ausreichenden Sicherheitsfunktionen für den Zugriff.	Zugriffe von Geräten mit veraltetem Betriebssystem wie z.B. Windows XP, Android 8, ... werden ausgeschlossen.

Attribut-klasse	Regelname	Regel	Attribut → Quelle	Soll-Attribute	Sicherheitsziel	Beispiel
Endgerät	Sichere Ausführungsumgebung	Es wird geprüft, ob das Gerät über eine attestierte Ausführungsumgebung (signierter Software-Stack) verfügt und ob diese aktiv ist (nicht gerootet/Jailbreak).	Attestation-Attribute → TCL/GMS (Geräte-Token)	Allowlist, Blocklist	Das Gerät verfügt über eine Ausführungsumgebung, welche die Manipulation der auf dem Endgerät ausgeführten Software (FCL/TCL) erschwert.	Zugriffe von gerooteten Mobiltelefonen werden ausgeschlossen. Zugriffsanfragen von Geräten ohne Secure Boot werden ausgeschlossen.
Endgerät	Antivirus	Es wird geprüft, ob auf dem Endgerät ein zugelassenes und aktuelles Antivirus-Programm läuft.	Attestation-Attribute → TCL/GMS (Geräte-Token)	Allowlist, Blocklist	Das Risiko, dass auf dem Endgerät Schadsoftware installiert ist, welche die Informationssicherheit beeinträchtigt wird, verringert.	Zugriffe werden nur für Geräte mit einem Nachweis für Antiviren-Programme von bekannten Herstellern und mit aktuellen Versionsnummern von Antivirensoftware- und Malwaresignaturdatenbanken zugelassen.

Attribut-klasse	Regelname	Regel	Attribut → Quelle	Soll-Attribute	Sicherheitsziel	Beispiel
Umgebung	Zeitpunkt der Anfrage	Es wird geprüft, ob der Zeitpunkt der Zugriffsanfrage für die Anfrage erlaubt ist.	Zeitpunkt der Anfrage → PEP (aus Zugriffsanfrage)	gültige Zeiträume, ggf. nutzerabhängig und fachdienstabhängig	Ausschluss von Zugriffen, welche zu nicht vertrauenswürdigen Zeitpunkten gestellt werden.	Zugriffe durch einen Nutzer in der Nacht können blockiert werden, wenn der Nutzer in der Regel nicht in der Nacht tätig ist.
Umgebung	Ort der Anfrage	Es wird geprüft, ob der Ort der Zugriffsanfrage für die Anfrage erlaubt ist.	Ort der Anfrage / IP-Adresse → PEP (aus Zugriffsanfrage)	gültige Regionen, ggf. nutzerabhängig und fachdienstabhängig	Ausschluss von Zugriffen, welche von nicht vertrauenswürdigen Orten gestellt werden.	Ausschluss von Zugriffen außerhalb Europas, wenn Dienste dort nicht sinnvoll einsetzbar sind.
Umgebung	IP-Adresse der Anfrage	Es wird geprüft, ob die Zugriffsanfrage von einer IP-Adresse stammt, die zu einer Quell-IP mit schlechter Reputation gehört.	IP-Adresse → PEP (aus Zugriffsanfrage)	Blocklist	Ausschluss von Zugriffen von Adressen, die für sicherheitsproblematische Aktivitäten bekannt sind.	Zugriffe auf Dienste von IP-Adressen, von denen das Monitoring bereits schadhafte Zugriffe verzeichnet hat, oder Adressräume, denen aus anderen Gründen nicht das notwendige Vertrauen entgegengebracht wird, werden blockiert.

Attribut-klasse	Regelname	Regel	Attribut → Quelle	Soll-Attribute	Sicherheitsziel	Beispiel
Umgebung	Impossible Travel	Es wird geprüft, ob die Zugriffsanfrage von einem Ort stammt, der so weit vom Ort der vorangehenden Zugriffsanfrage desselben Nutzers entfernt ist, dass eine Reise im Zeitraum zwischen den beiden Zugriffsanfragen unwahrscheinlich erscheint.	Ort und Zeitpunkt des Zugriffs → PEP (aus Zugriffsanfrage) UND Ort und Zeitpunkt der vorangehenden Zugriffsanfrage → PIP (Daten letzter Zugriff)	Verhältnis Entfernung der Orte und Differenz der Zeit	Ausschluss von nicht plausiblen Zugriffen.	Der letzte Zugriff/Zugriffsversuch eines Nutzers auf einen Dienst erfolgte aus Deutschland. Fünf Minuten später erfolgt ein Zugriff des Nutzers aus Australien. Dieser zweite Zugriff wird blockiert.

Tabelle 3.3.1-1: Beispiele für mögliche Regeln eines ABAC-Regelwerks

3.3.2 Erstellung des Regelwerkes

Für die Erstellung eines Regelwerkes für die Zugriffskontrolle werden zwei Analysen auf unterschiedlichen Abstraktionseben durchgeführt, aus denen sich ein fachdienstspezifisches Regelwerk ergibt: Eine übergreifende Analyse und eine fachdienstspezifische Analyse (siehe Abbildung 3.3.2-1). Beide Analysen bauen auf den notwendigen Schritten für eine Risikoanalyse auf. Grundsätzlich folgt das Regelwerk einem Allowlisting, d.h. Zugriffe auf Ressourcen werden verhindert, solange nicht eine entsprechende Regel im Regelwerk den Zugriff erlaubt.

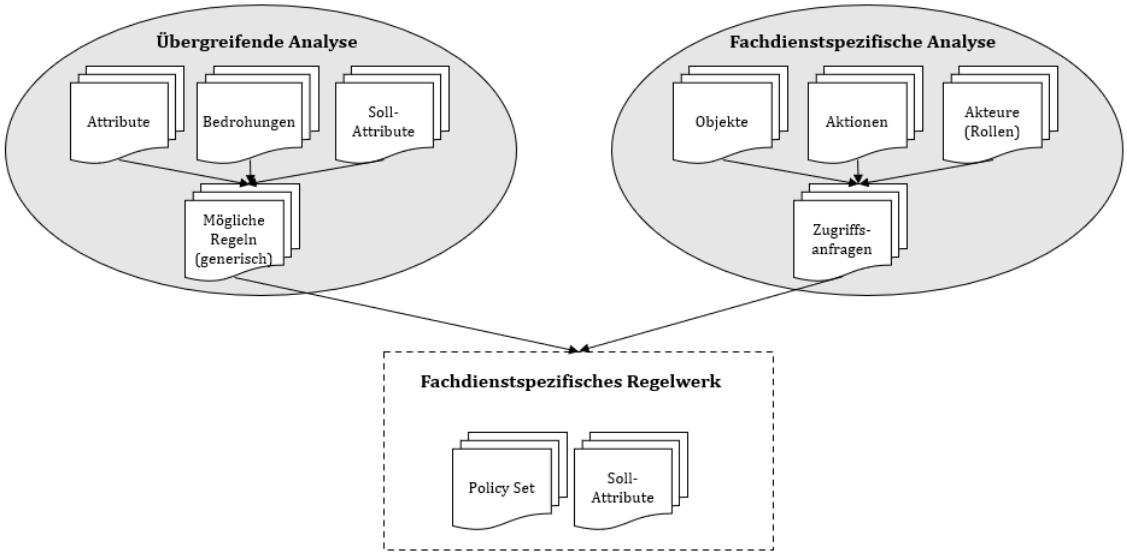

Abbildung 3.3.2-1: Abstraktionsebenen hinsichtlich der Erstellung des Regelwerks

Übergreifende Analyse

Die fachdienstübergreifende Analyse dient dazu, den Werkzeugkasten für die fachdienstspezifische Analyse zu erstellen und umfasst folgende Schritte:

(1) Zunächst werden generelle Bedrohungen und Angriffsszenarien im Zusammenhang mit den Zugriffen auf Ressourcen der TI identifiziert und deren Eintrittswahrscheinlichkeiten abgeschätzt. Dies erfordert das Festlegen auf Kriterien für die nachvollziehbare und vergleichbare Abschätzung der Eintrittswahrscheinlichkeiten von Bedrohungen und Angriffsszenarien.

(2) In einem weiteren Schritt werden Attribute identifiziert, welche Nutzer, Endgerät und Umwelt eines Nutzerzugriffs modellieren und mit dem Stand der Technik erfasst werden können. In diesem Zusammenhang werden auch die Quellen für solche Attribute identifiziert sowie die möglichen Soll-Attribute bzw. Referenzwerte (oder deren Quellen z.B. in Form bestehender Datenbanken) festgelegt.

(3) Im Anschluss werden mögliche Regeln für die Autorisierung eines Nutzerzugriffs auf Basis der identifizierten Attribute entwickelt, welche die Angriffsfläche für einen kompromittierten Nutzerzugriff reduzieren und den identifizierten Bedrohungen entgegenwirken (siehe Beispiele in Tabelle 3.3.1-1). Für jede Regel wird in diesem Schritt auch deren ursächliche Bedrohung bzw. das zugehörige Angriffsszenario dokumentiert, um ein nachhaltiges Management des Regelwerks zu unterstützen. Dabei setzen Regeln entweder Maßnahmen um (z.B. Anfragen von nicht vertrauenswürdigen IP-Adressräumen

blockieren) oder prüfen die Umsetzung von Maßnahmen (z.B. Nachweis über sichere Endgeräte für den Zugriff auf Dienste prüfen). Darüber hinaus wird für jede Regel eingeschätzt, in welchem Maß die Anwendung der Regel die Eintrittswahrscheinlichkeit für die Realisierung der relevanten Bedrohung reduziert.

Fachdienstspezifische Analyse

Im Rahmen der fachspezifischen Analyse wird der Werkzeugkasten aus der übergreifenden Analyse auf einen konkreten Fachdienst angewendet und dadurch ein konkretes Regelwerk für diesen Fachdienst erstellt:

(1) In einem ersten Schritt werden die für diesen Fachdienst spezifischen Ressourcen identifiziert. Dies erfordert das Festlegen auf den Abstraktionsgrad einer Ressource gemäß Kapitel 3.1.1 und das Ermitteln der Ressourcen gemäß diesem Abstraktionsgrad. Hierzu gehört z.B. das Ermitteln der fachdienstspezifischen Objekte, der möglichen Aktionen auf diesen Objekten und der möglichen Akteure bzw. deren Rollen, welche die identifizierten Aktionen auf den Objekten ausführen dürfen.

(2) Im Anschluss werden für jede identifizierte Ressource

- die Anwendbarkeit der in der übergreifenden Analyse identifizierten Bedrohungen geprüft
- die möglichen Schäden bzw. Schadenshöhe durch einen möglicherweise kompromittierten Zugriff auf die jeweilige Ressource ermittelt.

Im Ergebnis ist es möglich, Risiken zu benennen, deren Höhen sich aus der Eintrittswahrscheinlichkeit einer Bedrohung und der Schadenshöhe für eine Ressource ergeben. Hinsichtlich der Bewertung von Schäden und Schadenshöhen ist es erforderlich, einheitliche Schadenskriterien festzulegen.

(3) Die ermittelten Bedrohungen und die damit verbundenen Risiken ermöglichen in einem dritten Schritt die strukturierte Auswahl der passenden Regeln für das Regelwerk des Fachdienstes aus der übergreifenden Analyse und das Festlegen der Soll-Attribute. Darüber hinaus ist es möglich unter Berücksichtigung der mit der Regel verknüpften Wirkung auf die Eintrittswahrscheinlichkeit des Risikos das Restrisiko je Ressource zu ermitteln. Die Risikohöhe kann dabei auch für die Abwägung hinsichtlich der Anwendung einer Regel und dem damit verbundenen Einfluss auf die Usability eines Nutzerzugriffs herangezogen werden.

3.4 Daten

Für die Realisierung der Kernfunktionen werden zwischen den Komponenten der ZTA Daten ausgetauscht und Daten auf den Komponenten gespeichert. Dafür werden die im Folgenden (Kapitel 3.4.1) dargestellten Datenklassen berücksichtigt. Im Anschluss (Kapitel 3.4.2) wird erläutert, welche Komponenten welche Datenklassen verarbeiten oder speichern.

3.4.1 Datenklassen

Im Rahmen dieses Konzeptes werden aufgrund ihres unterschiedlichen Schutzbedarfes folgende Datenklassen unterschieden (siehe Tabelle 3.4.1-1). Die Tabelle ist absteigend sortiert nach Schutzbedarf, d.h. DS1 hat den höchsten Schutzbedarf.

Ref.	Name	Beschreibung	Beispiele
DS1	Personenbezogene medizinische Daten	Informationen, die sich auf die Gesundheit eines Individuums beziehen. Gemäß Art. 4 DSGVO [EU_DSGVO] sind dies personenbezogene Daten, die sich auf die körperliche oder geistige Gesundheit einer natürlichen Person, einschließlich der Erbringung von Gesundheitsdienstleistungen, beziehen und aus denen Informationen über deren Gesundheitszustand hervorgehen.	Rezepte, medizinische Messwerte, Diagnosen, Notfalldatensatz, Metadaten wie Zugriffsanfragen aus denen die Erbringung bestimmter Gesundheitsleistungen hervorgeht
DS2	Personenbezogene Daten	Informationen ohne Gesundheitsbezug, die sich gemäß Art. 4 DSGVO [EU_DSGVO] auf eine identifizierte oder identifizierbare natürliche Person beziehen.	Versichertenstammdaten, ID-Token, Identifier eines Nutzer-Gerätes
DS3	Daten für die Zugriffssteuerung ohne Personenbezug	Informationen ohne Personenbezug, welche für die Steuerung des Zugriffs auf Ressourcen verwendet werden.	Regelwerk, Attribute ohne Personenbezug
DS4	Interne Informationen zum Betrieb	Nicht-öffentliche Informationen ohne Personenbezug über den Sicherheitszustand der ZTA und über erfolgte Zugriffe.	Protokolle, Monitoring-Daten
DS5	Öffentliche Informationen	Öffentliche Informationen über die ZTA ohne Personenbezug.	Verfügbarkeit der Fachdienste, Liste der Fachdienste

Tabelle 3.4.1-1: Datenklassen

3.4.2 Verarbeiten und Speichern von Datenklassen

Jede der Komponenten in der ZTA verarbeitet oder speichert Daten. Tabelle 3.4.2-1 stellt hierfür pro Komponente dar, welche Datenklassen dort verarbeitet/gespeichert werden. Dabei werden Daten der entsprechenden Datenklasse nur aufgeführt, wenn diese in den entsprechenden Komponenten unverschlüsselt verarbeitet (V) oder persistent gespeichert werden. Je nach Art und Schutzbedarf der Daten muss die Verarbeitung in einer vertrauenswürdigen Ausführungsumgebung erfolgen. Persistente Speicherung bedeutet hier, dass die Daten nach einem Neustart der entsprechenden Komponente noch vorhanden sind. Je nach Art und Schutzbedarf der Daten muss diese Speicherung vertraulich und daher verschlüsselt (S**) oder ausschließlich integritätsgesichert (S*) erfolgen. Da Daten der Klasse DS5 als öffentliche Daten und ohne Bezug zur

Zugriffsteuerung nicht im Fokus der ZTA stehen, werden sie in der Tabelle nicht betrachtet.

Datenklasse/ Komponente		FCL	TCL	AUM	Nutzerportal	Fachdienst	PEP	PDP	IDP	FEM	PIP	GMS	PAP	MON
DS1	Personenbezogene medizinische Daten	V/S**	V	--	--	V/S**	V	V	V	--	--	--	--	--
DS2	Personenbezogene Daten	V/S**	V/S**	V/S**	V	V/S**	V/S**	V	V/S**	--	(V/S**)	V/S**	--	(V/S**)
DS3	Daten für die Zugriffssteuerung ohne Personenbezug	--	--	--	--	--	V	V	--	V/S*	V/S*	V/S*	V/S*	V/S*
DS4	Interne Informationen zum Betrieb	--	--	--	V	V	V	V	V	V	V	V	V	V/S*

Tabelle 3.4.2-1: Verarbeiten (V) und persistentes Speichern (S*/S) von Datenklassen**

Fach-Client (FCL)

Der Fach-Client ermöglicht den Austausch von personenbezogenen medizinischen Daten (DS1) und personenbezogenen Daten (DS2) zwischen dem Nutzer und dem jeweiligen Fachdienst. Je nach Fachanwendung kann der Fach-Client auch personenbezogene medizinische Daten (DS1) oder personenbezogene Daten (DS2) für die weitere Verarbeitung persistent speichern, solange dies den berufsständischen bzw. allgemeinen Datenschutzanforderungen gemäß geschieht. Eine Verarbeitung oder Speicherung von Daten für die Zugriffssteuerung (DS3) oder von Daten zum Betrieb der TI (DS4) ist nicht vorgesehen.

Trust-Client (TCL)

Der Trust-Client ermöglicht für den Fach-Client den Austausch von Daten mit dem Fachdienst und interagiert mit anderen ZTA-Komponenten, wenn das für einen Zugriff auf die entsprechenden Ressourcen benötigt wird. Dabei verarbeitet der TCL personenbezogene medizinische Daten (DS1), ohne sie jedoch persistent zu speichern. Personenbezogene Daten (DS2) werden durch den TCL verarbeitet, z.B. das Geräte-Token mit den enthaltenen IDs für Nutzer und Gerät. Ein Teil der personenbezogenen Daten (DS2) wird auch persistent gespeichert, z.B. die ID des Gerätes (basierend auf einem asymmetrischen Schlüsselpaar), das im Rahmen der Geräteregistrierung beim GMS hinterlegt wird und danach für die Identifizierung und Authentifizierung des Gerätes dient. Eine Verarbeitung oder Speicherung von Daten für die Zugriffssteuerung (DS3) oder von Daten zum Betrieb der TI (DS4) ist nicht vorgesehen.

Authenticator-Modul (AUM)

Das Authenticator-Modul verarbeitet personenbezogene Daten (DS2) für die Authentifizierung des Nutzers beim IDP. Je nach Realisierung des sektoralen IDPs kann das AUM solche Daten auch persistent speichern, z.B. als Schlüsselmaterial im Secure Element des Endgerätes oder in Form von Identifiern, welche das AUM mit der Identität des Nutzers verbinden.

Nutzerportal

Das Nutzerportal bietet dem Nutzer eine Möglichkeit, bestimmte Funktionalitäten der ZTA-Dienste über eine einzige Oberfläche zu nutzen, um z.B. einen Überblick über seine beim GMS registrierten Geräte zu bekommen oder benutzerspezifische Werte für Soll-Attribute im Regelwerk zu definieren. Im Rahmen dieser Aufgaben erhält das Nutzerportal personenbezogene Daten (DS2) zu dem Nutzer, um sie in der Nutzer-Oberfläche anzuzeigen. Das Nutzerportal hat keinen persistenten Speicher und verarbeitet auch keine Daten für die Zugriffssteuerung (DS3). Wenn das Nutzerportal als zentrale Komponente

(z.B. als Webserver) umgesetzt wird, stellt sie Daten zu ihrem Betrieb (DS4) für das übergreifende Monitoring zur Verfügung.

Fachdienst

Die Fachdienste dienen in den meisten Fällen dazu, für den Nutzer personenbezogene medizinische Daten (DS1) und personenbezogene Daten (DS2) zu verarbeiten und (in verschlüsselter Form) zu speichern. Für das übergreifende Monitoring (MON) stellt der Fachdienst Daten über den Betrieb des Fachdienstes (DS4) zur Verfügung, die jedoch für die im Konzept beschriebenen ZTA-Funktionen nicht persistent gespeichert werden müssen.

PEP

Die PEPs der Fachdienste dienen als TLS-Endpunkte und sind dafür verantwortlich, dass für jeden Zugriff die Erfüllung des Regelwerks durch den PDP geprüft und die Entscheidung des PDPs durchgesetzt wird. Zu diesem Zweck verarbeiten die PEPs die Zugriffsanfragen (inkl. der angefragten URLs), die personenbezogene medizinische Daten (DS1) enthalten/darstellen können. Zudem verarbeiten die PEPs personenbezogene Daten (DS2), z.B. in Form der ID-Token, sowie Daten für die Zugriffssteuerung ohne Personenbezug (DS3) wie z.B. die öffentlichen Schlüssel, die für die Prüfung der Signaturen an ID- und Geräte-Token verwendet werden. Für das übergreifende Monitoring (MON) stellt der PEP Daten über den Betrieb des PEPs sowie über erhaltene Zugriffsanfragen (DS4) zur Verfügung. Eine Speicherung von Daten (DS2) im PEP ist nur kurzzeitig im Rahmen des Session-Managements und für eine Aggregation von Informationen zu erfolgten Zugriffen für das Monitoring vorgesehen.

PDP

Die PDPs treffen Entscheidungen über die Gültigkeit angefragter Zugriffe basierend auf dem Regelwerk. Zu diesem Zweck verarbeiten sie Informationen zur angefragten Ressource, die personenbezogene medizinische Daten (DS1) enthalten/darstellen kann. Personenbezogene Daten (DS2) werden z.B. in Form von IP-Adressen verarbeitet. Für die Entscheidung benötigt ein PDP die Daten zur Zugriffssteuerung ohne Personenbezug (DS3) wie die für den entsprechenden Fachdienst geltenden Regeln des Regelwerks oder Informationen aus PIPs. Für das übergreifende Monitoring (MON) stellt der PDP Daten über seinen Betrieb (DS4) zur Verfügung. Eine persistente Speicherung von Daten im PDP ist nicht vorgesehen.

IDP

Die sektoralen IDPs sind verantwortlich für die Identifizierung und Authentisierung der Nutzer. Zu diesem Zweck erhalten und verarbeiten IDPs Informationen über die angefragten Fachdienste und Ressourcen, was personenbezogene medizinische Daten (DS1) darstellen kann. Diese werden jedoch nicht persistent gespeichert, um eine Profilbildung über das Nutzerverhalten am IDP zu verhindern. Der IDP verarbeitet und speichert personenbezogene Daten (DS2), um damit die Identifizierung und Authentifizierung der Nutzer vorzunehmen. Am IDP ist keine Verarbeitung von Daten für die Zugriffssteuerung ohne Nutzerbezug (DS3) vorgesehen, da Datenverarbeitung am IDP in der Regel im Kontext eines Nutzers geschieht. Für das übergreifende Monitoring (MON) stellt der IDP Daten über seinen Betrieb (DS4) zur Verfügung, die jedoch für die im Konzept beschriebenen ZTA-Funktionen nicht persistent gespeichert werden müssen.

FEM

Der Federation Master verarbeitet und speichert keine personenbezogenen (medizinischen) Daten (DS1, DS2). Er speichert aber Daten für die Zugriffssteuerung ohne Personenbezug (DS3), insbesondere die öffentlichen Schlüssel/Zertifikate für IDPs, GMS und Fachdienste und stellt diese für die TI zur Verfügung. Für das übergreifende Monitoring

(MON) stellt der FEM Daten über seinen Betrieb (DS4) zur Verfügung, die jedoch für die im Konzept beschriebenen ZTA-Funktionen nicht persistent gespeichert werden müssen.

PIP

Die Policy Information Points speichern Daten für die Zugriffssteuerung ohne Personenbezug (DS3), insbesondere Referenzwerte für Attribute, und stellen diese den PDPs der Fachdienste zur Verfügung. PIPs verarbeiten und speichern keine personenbezogenen medizinischen Daten (DS1). Die Verarbeitung von personenbezogenen Daten (DS2) kann im Rahmen des Feinkonzeptes nicht ausgeschlossen werden, da sie von den konkreten Attributen des Regelwerks abhängt. Wenn personenbezogene Daten, wie z.B. IP-Adressen, als Referenzwerte für das Regelwerk verwendet werden sollen, muss für den Einzelfall geprüft werden, ob eine Anonymisierung oder Pseudonymisierung möglich ist oder wie durch weitere Schutzmaßnahmen eine datenschutzkonforme Verarbeitung gewährleistet werden kann. Für das übergreifende Monitoring (MON) stellen PIPs Daten über ihren Betrieb (DS4) zur Verfügung, die jedoch für die im Konzept beschriebenen ZTA-Funktionen nicht persistent gespeichert werden müssen.

GMS

Der GMS verarbeitet und speichert personenbezogene Daten (DS2) im Rahmen der Registrierung bzw. Deregistrierung von Geräten und deren Assoziation zu Nutzer-IDs. Er erhält jedoch keine Informationen über verwendete Fachdienste oder andere personenbezogene medizinische Daten (DS1). Der GMS verarbeitet und speichert Daten für die Zugriffssteuerung ohne Personenbezug (DS3), z.B. in Form von Vertrauensankern für die Prüfung verwendeter Hardware-Sicherheitsanker. Für das übergreifende Monitoring (MON) stellt der GMS Daten über seinen Betrieb (DS4) zur Verfügung, die jedoch für die im Konzept beschriebenen ZTA-Funktionen nicht persistent gespeichert werden müssen.

PAP

Der PAP speichert und verarbeitet mit dem Regelwerk Daten zur Zugriffssteuerung (DS3). Für das übergreifende Monitoring (MON) stellt er Daten über seinen Betrieb (DS4) zur Verfügung, die jedoch für die im Konzept beschriebenen ZTA-Funktionen nicht persistent gespeichert werden müssen. Er verarbeitet und speichert keinerlei personenbezogene (medizinische) Daten (DS1, DS2).

MON

Das übergreifende Monitoring (MON) sammelt von allen TI-Komponenten Informationen über deren Betrieb und Sicherheit (DS4) und erhält nutzerunabhängige Attributwerte und Zugriffsentscheidungen (DS3) von den PEPs der Fachdienste. Diese Informationen werden für spätere Auswertungen geloggt und integritätsgeschützt gespeichert. Bei der Spezifikation der Nutzung der Schnittstellen des Monitorings durch andere Dienste muss darauf geachtet werden, dass keine personenbezogenen medizinischen Informationen (DS1) an die Monitoring-Komponente übergeben werden. Die korrekte Umsetzung der Spezifikation wird durch Produktzulassungen sichergestellt. Eine Verarbeitung von personenbezogenen Daten (DS2) kann im Rahmen des Feinkonzeptes nicht ausgeschlossen werden, da sie von den konkreten Attributen des Regelwerks abhängt. Wenn personenbezogene Daten, wie z.B. IP-Adressen, im Rahmen des Security-Monitorings für ein automatisiertes Anpassen von Referenzwerten für das Regelwerk verarbeitet werden sollen, muss für den Einzelfall geprüft werden, ob eine Anonymisierung oder Pseudonymisierung möglich ist oder wie durch weitere Schutzmaßnahmen eine datenschutzkonforme Verarbeitung gewährleistet werden kann.

3.5 Funktionen und Datenflüsse

In diesem Kapitel erfolgt die nähere Beschreibung der Kommunikationsflüsse in der ZTA. Je Kernfunktion wird dargestellt, wie einzelne Komponenten für die Realisierung der Funktion miteinander interagieren, welche Protokolle dafür eingesetzt werden und welche Rollen einzelne Komponenten innerhalb der Protokolle einnehmen.

In Abbildung 3.5-1 ist die Interaktion zwischen Komponenten der Nutzer und Komponenten der ZTA für die Autorisierung von Nutzerzugriffen im Überblick dargestellt. In den folgenden Kapiteln werden die Interaktionen zu den Kernfunktionen, wie in Kapitel 2 definiert, detaillierter dargestellt. Kapitel 3.5.1 beleuchtet die Identifizierung und Authentisierung der Nutzer. Im Anschluss werden in Kapitel 3.5.2 die Funktionen im Zusammenhang mit der Authentifizierung und Attestierung der Endgeräte der Nutzer betrachtet sowie in Kapitel 3.5.3 die Funktionen hinsichtlich der Authentifizierung und Attestierung der Fachdienste. In Kapitel 3.5.4 werden die Funktionen für das Erstellen, Prüfen, Überwachen und Anpassen des Regelwerks beschrieben. Kapitel 3.5.5 erläutert die Interaktion im Zusammenhang mit dem Zugriffs- und Session Management und in Kapitel 3.5.6 wird das Betriebsmonitoring der ZTA-Komponenten erläutert. Abschließend wird in Kapitel 3.5.7 der Daten- und Kommunikationsfluss für die Autorisierung von Nutzerzugriffen detaillierter dargestellt und erläutert.

Zur vereinfachten Darstellung werden Fach- und Trust-Client hier und in den folgenden Kapiteln als eine integrierte Komponente analog zu Option 1 (Alleinstehende Geräte) und 2 (Client-Server Infrastruktur) in Kapitel 3.2.1 betrachtet. Zusätzlich werden die beim Federation Master hinterlegten Informationen ausgeklammert, beziehungsweise als bereits bei den Akteuren vorliegend angenommen. Ebenso wird vereinfachend angenommen, dass sich Authenticator-Modul und Fach- und Trust-Client auf dem gleichen Gerät befindet.

In der ZTA werden alle Zugriffe auf Fachdienste anhand eines Regelwerks geprüft (9, 16). Dafür müssen im Rahmen des Nutzerzugriffs die notwendigen Attribute zu Subjekt, Gerät, Umgebung und Ressource erfasst werden. Übergreifend kann der Ablauf für einen Nutzerzugriff, wie in Abbildung 3.5-1 skizziert, in drei Teile unterteilt werden: Die Autorisierung des Geräts (2-6), die Authentifizierung des Nutzers (8-14) und der Zugriff auf die Ressource mit den beiden vorliegenden Nachweisen (15-17).

Die Autorisierung des Geräts erfolgt direkt nach Start der Anwendung (1), ohne Interaktion mit dem Nutzer.

Die zur Autorisierung und Attestierung des Geräts notwendigen Schritte sind von dem Fach/Trust-Client gegen das Gerätemanagement-Service (GMS) gerichtet und folgen dem OAuth2 Standard [RFC6740_OAuth2].. Vorausgehend, und hier nicht abgebildet, ist die Registrierung des Geräts beim GMS (siehe Abschnitt 3.5.2). Durch diese steht jedem Gerät eine hardwaregebundene Geräte-ID auf Basis eines Geräte-Schlüsselpaars zur Verfügung, welche das Gerät eindeutig identifiziert, und es an einen Nutzer bindet. Für die Autorisierung des Geräts baut der Client eine mTLS gesicherte Verbindung zum GMS auf und übermittelt dabei durch das verwendete Zertifikat implizit seine Geräte-ID (2). Der GMS prüft nun durch geeignete Maßnahmen welche Eigenschaften des Geräts durch ihn bestätigt werden können (4, 5). Zusätzlich hinterlegt der GMS die auf das Gerät registrierte Nutzer-Identität, welche zusammen mit dem Gerät verwendet werden darf. (5). Dieses Token erhält der Client und kann es beim PEP als Nachweis für die Attribute seines Geräts verwenden (6). Das Token ist an das TLS Client Zertifikat gebunden und kann so nur von dem tatsächlich attestierten Gerät verwendet werden.

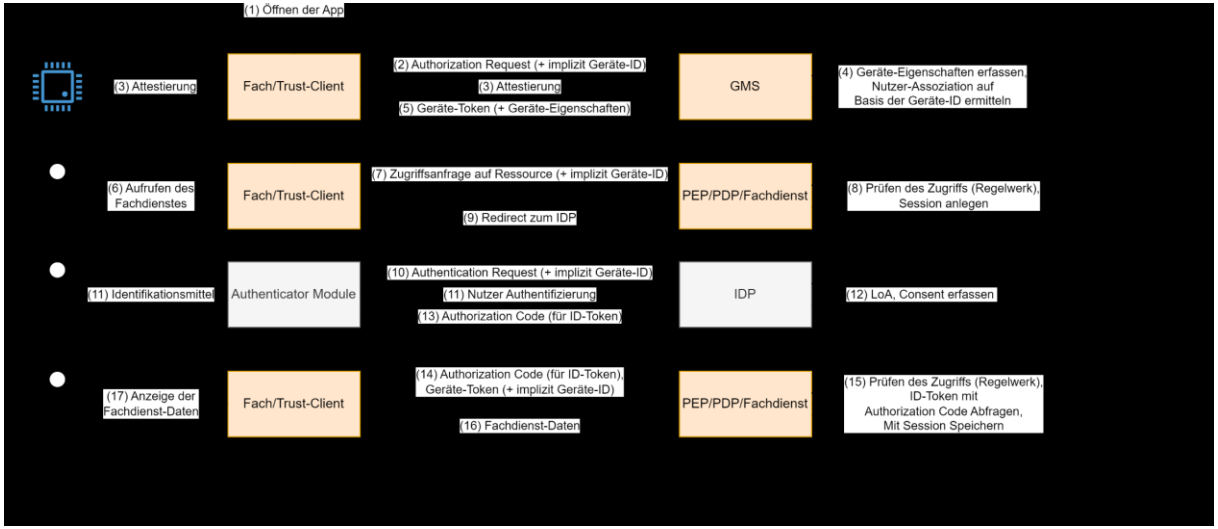

Abbildung 3.5-1: Schematische Übersicht des Daten- und Kommunikationsflusses (vereinfacht)

Erst durch Interaktion des bereits registrierten Nutzers mit dem Fach-Client (7), startet der Authentifizierungsvorgang des Nutzers per OpenID Connect (OIDC) [OpenID-Connect] am IDP. Dazu ruft der Client zuerst die angefragte Ressource auf (8), woraufhin der PEP einen OIDC Authentication Request gegen den IDP mit entsprechenden Scopes formuliert (9). Dieser wird an den Client zurückgegeben (10), welcher wiederum die Authentifizierung an das Authenticator-Modul delegiert. Nach entsprechender Authentifizierung erhält der Client einen Authorization Code (11-14), mit welchem der PEP nach Authentisierung beim IDP ein ID-Token des bereits registrierten Nutzers abfragen kann (15, 16).

Mit dem Geräte-Token und der Information, mit der der PEP das ID-Token abrufen kann, ruft der Client nun erneut die Ressource auf (15). Die Nachweise werden für zukünftige Anfragen mit der Session verknüpft gespeichert (16). Diese Session ist ebenfalls an dasselbe TLS Client Zertifikat gebunden und kann so nur von dem authentifizierten Gerät verwendet werden. Nachdem alle Anforderungen zum Nachweis eines gültigen Geräts und einer gültigen Nutzeridentität erfüllt sind, prüft der Fachdienst, ob der Nutzer eine Berechtigung für den Zugriff auf die angefragten Nutzer-Daten hat und der Client erhält im positiven Fall Zugriff auf die angefragten Fachdienst-Daten (17).

Zukünftige Anfragen müssen nun nur noch auf die bestehende Session verweisen (in der Regel unter Nutzung von TLS-gebundenen Access- und Refresh-Token), um die gleichen Nachweise nutzen zu können.

3.5.1 Nutzer authentisieren und identifizieren

Für die Identifizierung und Authentifizierung von Nutzern, d.h. Versicherten, Leistungserbringern aber auch Fachdiensten selbst, integriert die ZTA die bereits in Spezifikation befindliche Föderation der TI [gem_IDP_Federation] der TI. Dabei wird insbesondere die Spezifikation des Sektoralen Identity Providers [gem_Spec_IDP_Sek] sowie der in diesem Zusammenhang relevante Standard OpenID Connect [OpenID-Connect] berücksichtigt.

FA2 Nutzer bei Fachdienst identifizieren

Die Funktion FA2 hat zum Ziel, die für einen Nutzerzugriff notwendigen Identitätsattribute eines Nutzers zur Verfügung zu stellen. Dazu gehören sowohl Identitätsattribute, welche für die Registrierung des Nutzers bei einem Fachdienst notwendig sind (z.B. Vorname, Name, medizinische Einrichtung, KVNR, ...), als auch Attribute hinsichtlich der TI-spezifischen Rolle des Nutzers oder der gemäß OpenID Connect spezifizierte Pairwise Pseudonymous Identifier [OpenID-Connect] für die Abbildung der Identität des Nutzers auf einen konkreten Nutzer-Account beim Fachdienst. Diese Funktion wird über einen IDP der Föderation der TI auf Basis der Authentifizierung des Nutzers durch diesen IDP realisiert.

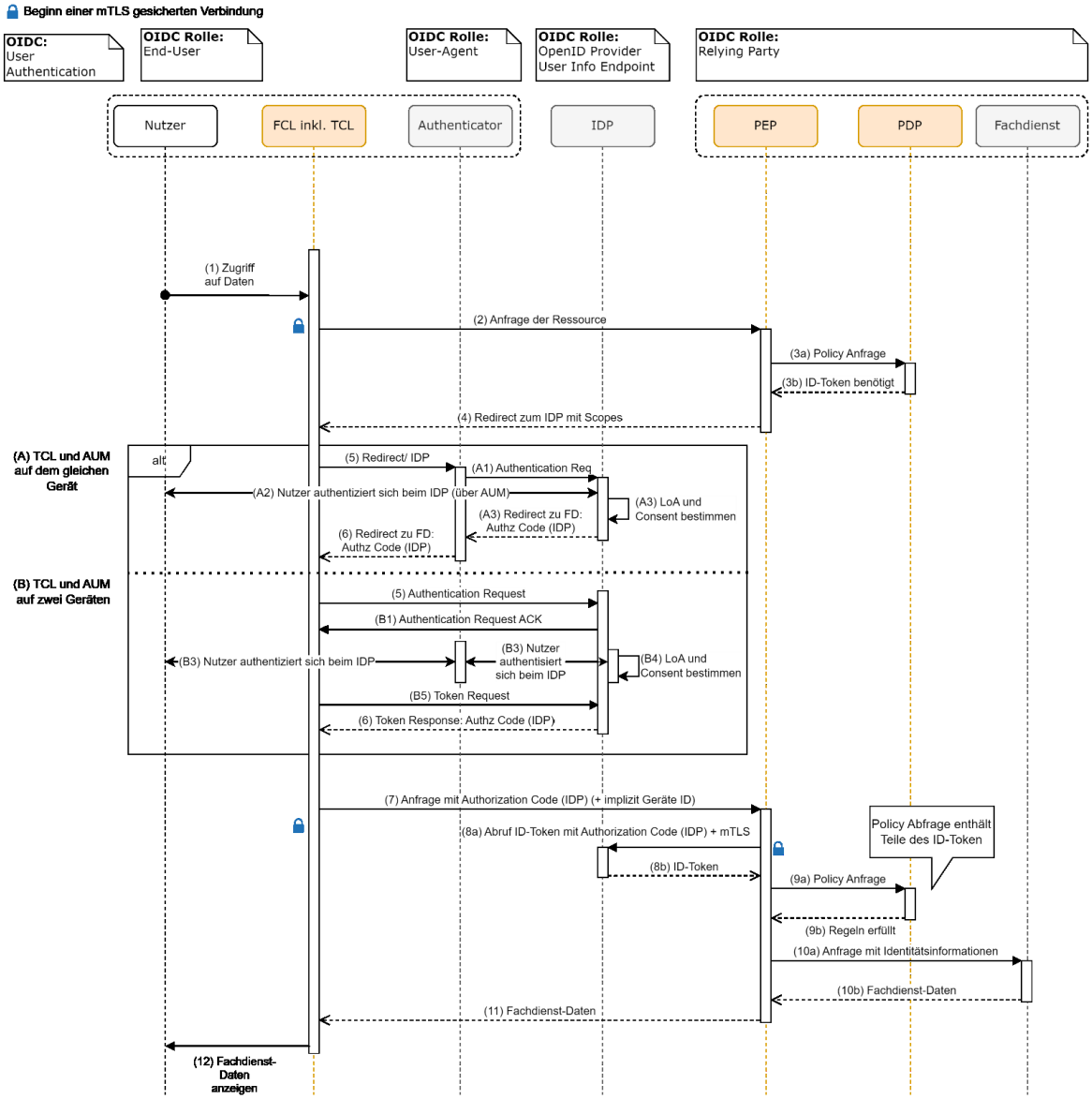

Abbildung 3.5.1-1: Authentifizierung des Nutzers per OIDC am PEP und Fachdienst

Der Datenfluss ist in Abbildung 3.5.1-1 abgebildet und beschreibt die Identifizierung des Nutzers auf Basis von OIDC in der ZTA. Die Identifizierung des Nutzers erfolgt erst durch eine Zugriffsanfrage des Nutzers auf eine Ressource (1,2). Diese erste Anfrage wird aufgrund des fehlenden ID-Tokens (3a, 3b) abgelehnt, und an den IDP weiterverwiesen (4). Dabei übergibt der PEP ebenfalls alle notwendigen Scopes, die für die Anfrage des Nutzers notwendig sind. Anschließend sind zwei Datenflüsse möglich, die mit A und B gekennzeichnet sind. Diese werden weiter unten genauer erklärt. Sie beginnen beide mit der Übergabe des Authentication Requests (5) und enden mit dem Erhalt des Authorization Codes (6). Dieser wird anschließend an den PEP übergeben (7), welcher wiederum das ID-Token abruft (8a, 8b). Dabei muss sich der PEP mittels mTLS gegenüber dem IDP authentifizieren. Nach einer erfolgreichen Policy-Abfrage (9a, 9b), wird die Anfrage an den Fachdienst weitergeleitet (10a). Dabei werden die Informationen aus dem ID-Token dem Fachdienst zugänglich gemacht, um eine fachdienstspezifische Verwendung der Identitätsattribute zu ermöglichen. Anschließend werden die Fachdienst-Daten über den PEP und Client dem Nutzer angezeigt (10b, 11, 12).

Die beiden angesprochenen Datenflüsse A und B beschreiben die Identifizierung des Nutzers gegenüber dem IDP genauer. Der Kommunikationsfluss hängt hier davon ab, ob der Fach-Client inkl. Trust-Client und das Authenticator-Modul auf dem gleichen Gerät sind oder nicht. Dementsprechend implementiert Datenfluss A den OIDC-Authorization-Code-Flow, bei dem der TCL direkt mit dem AUM auf demselben Gerät interagiert. Bei Datenfluss B hingegen befinden sich Trust-Client und Authenticator-Modul auf getrennten Geräten und der IDP stellt über geeignete Maßnahmen eine Verknüpfung der beiden Prozesse sicher. Im Diagramm ist dies beispielhaft über OIDC Client Initiated Backchannel Authentication dargestellt [OpenID-Connect_CIBA]. Dabei stellt der Fach-/Trust-Client direkt eine Anfrage an den IDP (5, B1), der wiederum das AUM kontaktiert (B2-B4). Der Fach-/Trust-Client fragt im Hintergrund regelmäßig den aktuellen Status der Authentifizierung an (B6), und erhält nach einer erfolgreichen Authentifizierung ein entsprechendes Token (6). Bei beiden muss sich der Nutzer gegenüber dem IDP über das AUM auf einem separaten Gerät identifizieren (A2, B4) und seine Einwilligung in die Datenweitergabe hinterlegen (A3, B4).

Der Datenfluss kann ebenfalls mittels Pushed-Authorization Requests (PAR) gestaltet werden [RFC9126_OAuth2_PAR]. Dabei kontaktiert der PEP in zwischen Schritt 3b und 4 zusätzlich den IDP, und übergibt in Schritt 4 ausschließlich eine Referenz auf die von ihm zuvor gestellte Anfrage an den Fach-/Trust-Client. Die übrigen Datenflüsse bleiben gleich.

FA2.1 Nutzer authentifizieren (extern)

Die Funktion FA2.1 dient der Authentifizierung eines Nutzers gegenüber einem für diesen Nutzer zuständigen sektoralen IDP der Föderation der TI. Darüber hinaus ist diese Funktion für das Erbringen der Nutzerzustimmung für die Weitergabe von Identitätsattributen an einen Fachdienst zuständig. Der Authentifizierungsmechanismus und das für die Authentifizierung notwendig beim Nutzer verfügbare Authentifizierungsmittel (z.B. mobiles Endgerät, eGK, HBA, SMC-B, ...) ist dabei abhängig von der Implementierung des sektoralen IDP. Schnittstelle für die Authentisierung des Nutzers mit Hilfe des Authentisierungsmittels ist dabei das AUM. Die Authentifizierung selbst erfolgt auf einem definierten Vertrauensniveau hinsichtlich der Widerstandsfähigkeit der Prozesse gegen Angriffe und Manipulation des Authentifizierungsprozesses (z.B. gematik-ehealth-loa-substantial, gematik-ehealth-loa-high) gemäß Spezifikation des Sektoraler Identity Provider [gemSpec_IDP_Sek]. Das Vertrauensniveau der Authentifizierung wird als ein Inhalt des ID-Token vom IDP für die Berücksichtigung in der ZTA zur Verfügung gestellt.

FA2.2 Nutzer bei Fachdienst registrieren (extern)

Die Funktion FA2.2 umfasst das Anlegen eines Nutzer-Accounts bei einem Fachdienst. Die dafür notwendigen Identitätsattribute des Nutzers werden über die Funktion FA2 in Form eines ID-Token bereitgestellt. Die Umsetzung dieser Funktion obliegt darüber hinaus dem Fachdienst.

FA2.3 eID-Lifecycle IDP (extern)

Die Funktion FA2.3 umfasst die für die Verwaltung einer elektronischen Identität innerhalb der TI notwendige Funktionalität (z.B. Registrieren beim IDP, Aktualisieren von Identitätsattributen, Sperren und Entsperren von Identitätsmitteln, Löschen, usw.) Dies Umsetzung dieser Funktionalität obliegt dem IDP der Föderation der TI.

FA2.4 Vertreterregelung (extern)

Durch fortschreitende Digitalisierung der Gesundheitsdienste tritt auch die Digitalisierung der Vertretung eines Nutzers der TI durch einen anderen Nutzer der TI in den Vordergrund, z.B. bei:

- Eltern in Vertretung für Ihre, ggf. minderjährigen, Kinder oder

- Kinder in Vertretung für Ihre, durch ihre hohen Alter ggf. hilfsbedürftigen, Eltern oder
- Pfleger in Vertretung für in Pflege befindlicher Nutzer,
- aber auch Dienste im Auftrag eines Nutzers.

Die Funktion FA2.4 umfasst diese Funktionalität. Theoretisch wäre eine Umsetzung in Form einer Impersonifizierung des Nutzers oder der Delegation von Rechten des Nutzers möglich [RFC8693_OAuth2_Token_Exchange]. Für die TI wird jedoch eine Umsetzung in Form einer Delegation von Rechten des Nutzers empfohlen, so dass für relevante Komponenten der ZTA erkennbar bleibt, welche Identität sich tatsächlich hinter einem Nutzerzugriff verbirgt und es möglich bleibt das vom Nutzer genutzte Gerät auch im Falle einer Vertretung der Identität des Nutzers zuzuordnen.

Grundsätzlich sind hinsichtlich der Realisierung der Funktion mehrere Lösungen denkbar:

- Option 1: IDPs der Föderation der TI ermöglichen das Hinterlegen von Vertreterbeziehungen
- Option 2: Fachdienste ermöglichen das Hinterlegen von Vertreterbeziehungen
- Option 3: Ein unabhängiger Dienst ermöglicht das Hinterlegen von Vertreterbeziehungen

Option 1 nutzt die bestehende Vertrauensbeziehung zwischen Nutzer und sektoralem IDP, wirkt sich aber ungünstig auf NFA2 (Keine Allmacht) aus, da sich der IDP durch das Anlegen eines kompromittierten Vertreters und einer entsprechend kompromittierten Vertreterbeziehung Zugang zu einem Nutzerkonto eines Fachdienstes verschaffen könnte. Darüber hinaus erfordert das Anlegen von Vertreterbeziehungen, an denen Identitäten von verschiedenen Sektoraler Identity Providern beteiligt sind, ein Protokoll, welches diese Kooperation sicher ermöglicht.

Option 2 nutzt die bestehende Vertrauensbeziehung zwischen Nutzer und Fachdienst. Allerdings können bei einem Fachdienst nur Vertreterbeziehungen für diesen Fachdienst hinterlegt werden. Eine Generalvertretung müsste also bei jedem Fachdienst einzeln angelegt werden. Darüber hinaus könnten nur bei dem Fachdienst registrierte Identitäten als Vertreter eingerichtet werden.

Option 3 erfordert einen neuen Vertretungsdienst. Als unabhängiger Dienst könnte er Vertreterbeziehungen Identity-Provider- und Fachdienst-übergreifend verwalten sowie Dienste, welche im Auftrag von Nutzern agieren, als Vertreter ermöglichen.

Mit Option 1 und 2 kann eine umfassende Vertreterregelung aus den voran genannten Gründen nicht oder nur mit Nachteilen hinsichtlich NFA2 realisiert werden. Im Rahmen dieses Konzeptes wird deshalb eine Realisierung der Funktion gemäß Option 3 empfohlen. Ein unabhängiger Vertretungsdienst könnte Nutzern zentral über das Nutzerportal zugänglich gemacht werden. Für die Integration der Vertreterbeziehungen in die Evaluierung von Nutzerzugriffen durch die PDPs müsste der Vertretungsdienst als PIP ausgelegt sein.

3.5.2 Geräte der Nutzer authentisieren und attestierten

Die Architektur ermöglicht das Authentisieren und Attestieren von Endgeräten der Nutzer, welche für den Zugriff auf Ressourcen der TI genutzt werden. Dabei soll die Authentisierung und Attestierung der Endgeräte einen weiteren von den IDPs unabhängigen Sicherheitsanker der ZTA darstellen, der die Identifizierung und Authentisieren des Nutzers durch den IDP ergänzt. Zusätzlich zum Nachweis über die Identität des Nutzers (ID-Token) soll ein Nachweis über den Besitz eines auf diesen Nutzer registrierten Endgerätes (Geräte-Token) erbracht werden. So kann verhindert werden,

dass die Ausstellung eines ID-Tokens ausreicht, um eine erfolgreiche Autorisierung am Fachdienst vorzunehmen, wodurch dem IDP als Aussteller des ID-Token eine besondere Allmachtstellung in der Architektur zukäme. Für das Geräte-Token wird nachfolgend immer der OAuth2 Authorization Code Flow betrachtet [RFC6740_OAuth2].

Voraussetzung für die Zuordnung dieser beiden Nachweise zueinander ist eine erfolgreiche Verknüpfung des registrierten Geräts mit der Identität des Nutzers (Geräte-Nutzer-Assoziation). Diese beruht auf der Verwendung von zwei eindeutigen Identifiern:

- **UUID_Nutzer**: Eindeutiger unveränderlicher Identifier einer Nutzer-Identität in der TI, die sowohl dem IDP als auch dem GMS im Rahmen der Nutzer-/Geräte-Registrierung bekannt wird und vom Fachdienst zur Adressierung des Nutzeraccount benutzt wird
- **UUID_Gerät**: Eindeutiger Identifier des Geräts, der aus dem selbst-signierten Zertifikat eines auf dem Endgerät generierten asymmetrischen Schlüsselpaares abgeleitet wird (z.B. Zertifikat-Fingerprint)

Die Assoziation von Nutzer- und Geräteidentitäten erfolgt im Rahmen der Geräteregistrierung (FA5.1 - Gerät registrieren). Alle gültigen UUID_Nutzer <-> UUID_Gerät-Kombinationen sind beim GMS gespeichert und können dort verwaltet werden. Das Geräte-Token ist somit nicht fachdienstspezifisch, sondern kann für alle Fachdienste verwendet werden.

Falls der Fach-Client von mehreren Personen verwendet wird, erhält jeder Nutzer ein eigenes Schlüsselpaar. Daher ist immer nur genau ein Nutzer einem Schlüsselpaar beziehungsweise UUID_Gerät zugeordnet.

Das folgende Datenflussdiagramm zeigt den Kommunikationsfluss für einen Zugriff auf eine Ressource des Fachdienstes basierend auf dieser Assoziation:

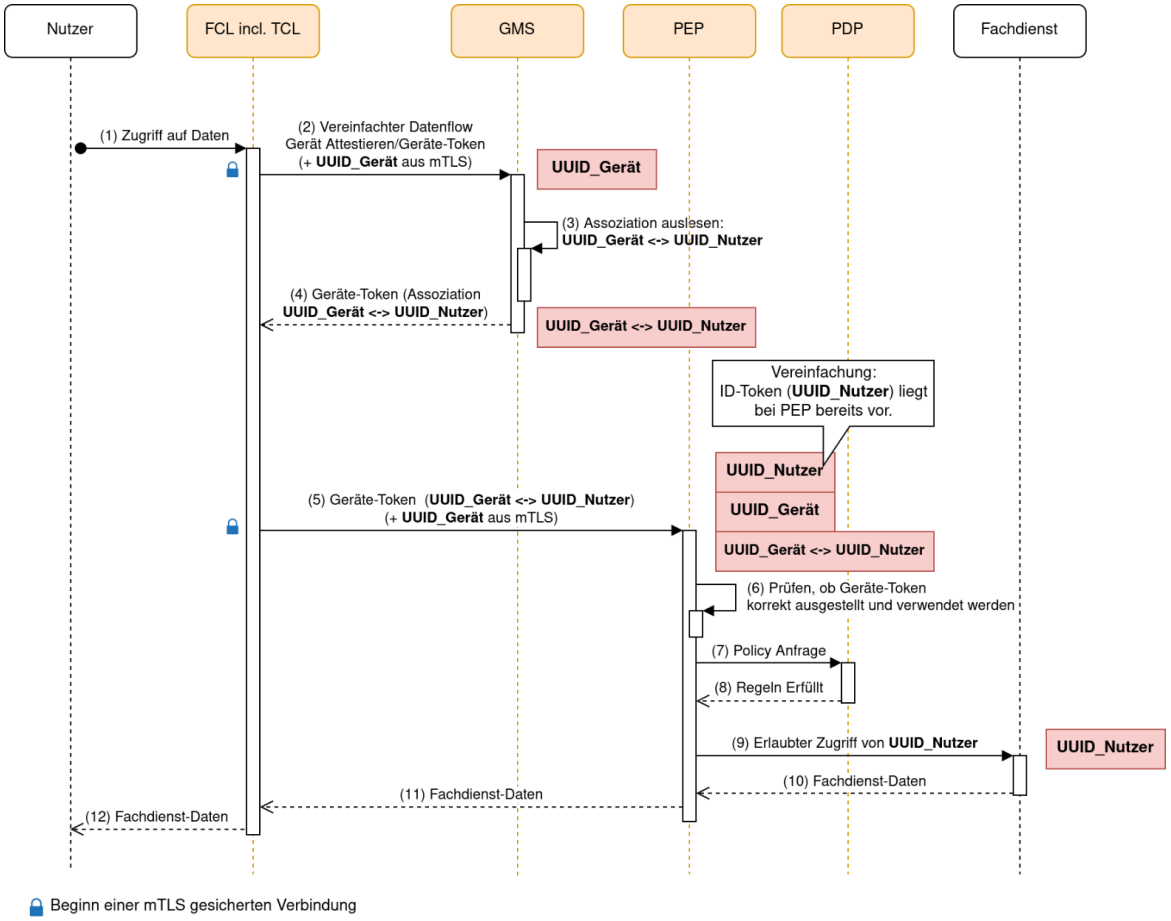

Abbildung 3.5.2-1: Übersicht über Geräte-Nutzer-Assoziation bei Zugriff

Die UUID_Nutzer ist beim IDP gespeichert und wird nach erfolgreicher Nutzer-Authentifizierung durch ein gültiges ID-Token nachgewiesen (FA2 Nutzer authentisieren, im Bild nur durch Vorhandensein des ID-Tokens beim PEP abgebildet). Die Authentifizierung des Nutzer-Gerätes (FA5 Gerät authentisieren) erfolgt beim GMS und führt zur Ausstellung eines Geräte-Tokens, in dem unter anderem die gültige Geräte-Nutzer-Assoziation (UUID-Gerät <-> UUID_Nutzer) enthalten ist.

Das ID-Token und das Geräte-Token werden als Teil der Zugriffsanfrage an den PEP des Fachdienstes übertragen. Dieser überprüft zusammen mit dem PDP die Identitätsnachweise und ihre Assoziation und lässt nur bei Übereinstimmung einen Zugriff auf den Fachdienst zu. Der PEP ist dabei nur für die Prüfung von technischen Anforderungen zuständig, unter anderem ob die Token korrekt signiert wurden oder ob bei einem zertifikatsgebundenen Token auch eine Verbindung mit dem korrekten Zertifikat vorliegt [RFC8705_OAuth2_mTLS]. Die inhaltliche Prüfung erfolgt durch den PDP. Die Geräte-Token haben eine begrenzte Gültigkeit und müssen nach Ablauf ihrer Lebensdauer erneut angefordert werden.

Ein Gerät (UUID_Gerät) kann auch auf mehrere Nutzer (UUID_Nutzer) registriert werden, wenn mehrere Benutzer ein Gerät gemeinsam nutzen. Alternativ kann auch durch lokale Systeme (z.B. bei Kontentrennung im OS) oder Fach-Client spezifische Zuordnung (z.B. bei Kontentrennung im FCL) jeweils eine eigene UUID_Gerät erzeugt werden.

App spezifische gerätegebundene Identität

Sollte es aufgrund von Regeln der Geräte-Plattform nicht möglich sein, dass Anwendungen auf einem Gerät Zugriff auf eine UUID_Gerät und das dahinterliegende Schlüsselpaar erhalten, muss jede Anwendung ihre eigene UUID_App_Gerät erzeugen. Für die Registrierung mehrerer UUID_App_Gerät auf einem Gerät sollten vereinfachte Registrierungs- und Managementprozesse (u.a. im Nutzerportal) vorgesehen werden. Eine erste Betrachtung der aktuellen technischen Möglichkeiten auf marktüblichen Plattformen ist in Anhang 1 zu finden. In allen vorstehenden und folgenden Betrachtungen zur Geräteregistrierung ist UUID_Gerät stellvertretend für UUID_Gerät oder UUID_App_Gerät anzusehen.

Pairwise-Identifier des Nutzers

Die OIDC-Spezifikation sieht die Nutzung eines *Pairwise-Identifiers* als OIDC-Client-spezifische Nutzeridentität vor [OpenID-Connect]. In diesem Fall erhält jeder PEP bzw. Fachdienst eine fachdienstspezifische Nutzerkennung. Für den für *Pairwise-Identifier* muss ein zusätzlicher Identifier eingeführt werden: UUID_FD_Nutzer. Um die fachdienstübergreifende Geräteregistrierung weiterhin zu ermöglichen, müssen UUID_FD_Nutzer und UUID_Nutzer für den Fachdienst zuordenbar sein. Beispielsweise könnte UUID_FD_Nutzer zusammen mit der OIDC-Client-ID aus UUID_Nutzer abgeleitet werden. Im folgenden Diagramm (Abb. 3.5.2-2*)* ist nun einmalig die Verwendung von UUID_FD_Nutzer und UUID_FD_Gerät dargestellt.

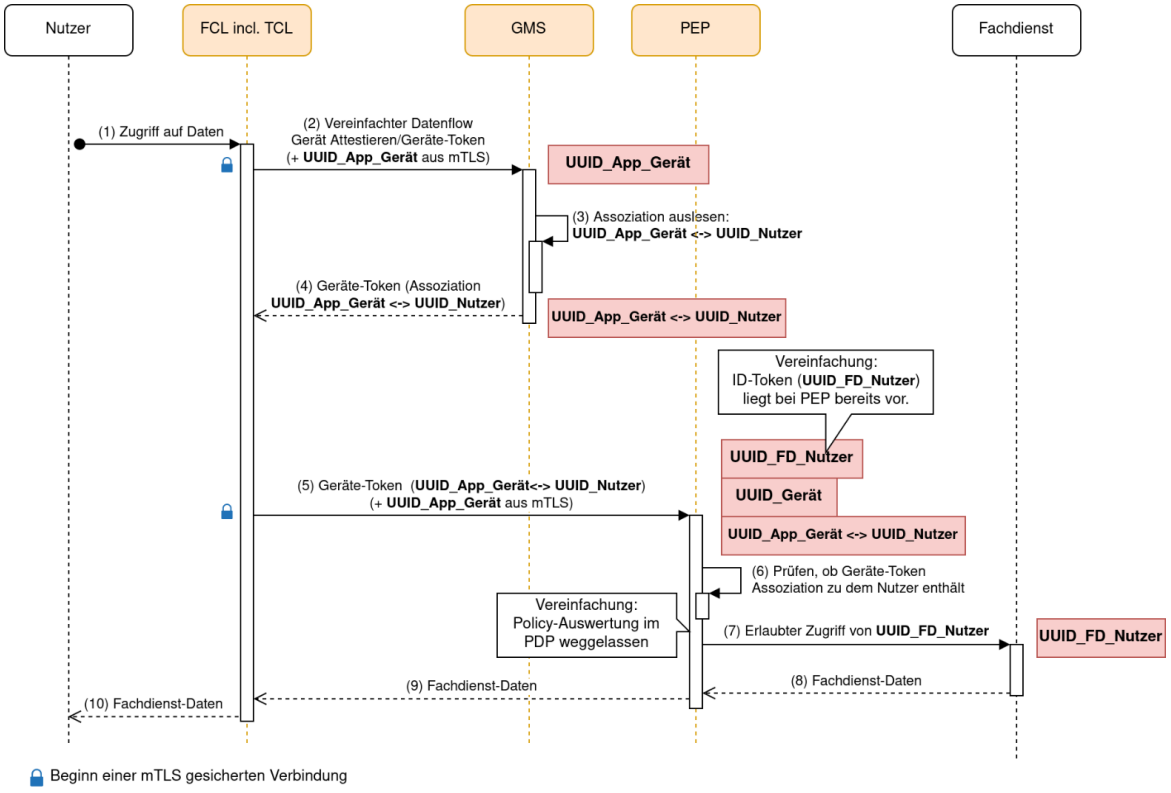

Abbildung 3.5.2-2: Übersicht über Geräte-Nutzer-Assoziation bei Zugriff mit erweiterter Nomenklatur

Für eine konkrete Ausgestaltung der UUID_Nutzer, UUID_FD_Nutzer, UUID_Gerät und UUID_App_Gerät, sowie deren Ableitung, sind tiefergehende Analysen notwendig. Idealerweise sollte UUID_Nutzer bzw. UUID_FD_Nutzer so gewählt werden, dass diese keine direkten Rückschlüsse auf die dahinterliegende Person zulassen. Dies würde unter

anderem ermöglichen, dass der GMS keine personenbezogenen Informationen persistent speichern muss.

Geräte-Token Refresh

Der Refresh des Geräte-Token erfolgt ausschließlich durch den Trust-Client. Auf Basis der im Token enthaltenen Lebensdauer kann der Trust-Client periodisch ein neues Geräte-Token anfragen und beim nächsten Ressourcen-Zugriff dem PEP übergeben.

Hier kann durch Refresh-Token und die Lebensdauer des Geräte-Tokens an sich eine Abwägung zwischen verschiedenen Schutzzielen und Usability getroffen werden. So ermöglichen kurze Laufzeiten ein schnelles Sperren von Geräten, oder das schnelle Reagieren auf Veränderungen am System. Lange Lebensdauern erhöhen die Performance. Sie erfordern weniger Abfragen und aufwändige Remote-Attestation Vorgänge müssen seltener durchlaufen werden.

FA5 - Gerät authentisieren

Ein Nutzer-Endgerät wird durch den eindeutigen Geräte-Identifier UUID_Gerät identifiziert und kann sich durch den Besitz des privaten Schlüssels aus einem asymmetrischen Schlüsselpaar in der ZTA authentisieren. Die UUID_Gerät ergibt sich direkt aus dem öffentlichen Schlüssel des Schlüsselpaares und wird im Rahmen des Registrierungsprozesses beim GMS hinterlegt. Den privaten Schlüssel und das Zertifikat verwendet der Trust-Client beim TLS-Verbindungsaufbau zu seinen Kommunikationspartnern, um seine Identität nachzuweisen.

Beim GMS ist hierbei das Zertifikat allein für die Authentifizierung ausreichend, da sich die UUID_Gerät daraus ergibt und im Rahmen des Registrierungsprozesses beim GMS abgespeichert wurde.

Zur Umsetzung der hier gewünschten Funktionalität wurde ein selbst-signiertes Zertifikat als ausreichend erachtet, da das Zertifikat ausschließlich für die Übermittlung des öffentlichen Schlüssels dient, um über die standardisierten mTLS-Nachrichten den Nachweis über den Besitz des dazugehörigen privaten Schlüssels zu erbringen. Das erwartete Zertifikat bzw. der öffentliche Schlüssel wird über das Geräte-Token übermittelt und beim Fachdiensten (bzw. dessen PEP) mit dem für den Aufbau der TLS-Verbindung verwendeten Zertifikat abgeglichen, womit eine Art Zertifikats-Pinning bzw. Publik-Key-Pinning umgesetzt wird [RFC8705_OAuth2_mTLS]. Ein von einer PKI ausgestelltes Zertifikat bietet für diese beschriebenen Funktionalitäten keinen Mehrwert.

Eine detaillierte Betrachtung der Möglichkeiten ist in Anhang 1 zu finden.

FA5.1 - Gerät registrieren

Bei der Registrierung eines Geräts wird die Nutzeridentität (UUID_Nutzer) mit der Geräte-Identität (UUID_Gerät) verknüpft und diese Assoziation beim GMS hinterlegt.

Das folgende Datenflussdiagramm (Abbildung 3.5.2-3) zeigt die Registrierung eines Geräts am GMS. Direkt nach Start der Anwendung wird die hardwaregebundene Geräte-ID erzeugt (2). Nachdem der Nutzer die Geräte-Registrierung initiiert hat (3), authentisiert sich der TCL implizit per mTLS am GMS, und startet die Registrierung (4). Anschließend fordert der GMS den TCL auf, einen entsprechenden Nutzernachweis vorzulegen (5). Dies wird durch Identifizierungsmittel erreicht, die möglichst unabhängig von den sektoralen IDP sind (z.B. Identifizierungsmittel, welche auch für die Erstidentifizierung bei einem sektoralen IDP zugelassen sind). Dabei ist notwendig, dass der Nachweis durch die mTLS gesicherte Verbindung am GMS eingereicht wird (7), um den Nachweis sicher mit der Geräte-ID verbinden zu können. Dieser Nachweis wird am GMS hinterlegt (7), und dem Gerät und Nutzer bestätigt (9, 10).

Für die sichere Implementierung müssen einige Besonderheiten beachtet werden:

- **UUID_Nutzer** muss von einer unabhängigen dritten Stelle vergeben und bestätigt werden. Dies wäre unter anderem direkt durch das Authentifizierungsmerkmal (eGK, HBA) oder einen vertrauenswürdigen dritten Dienst möglich.
- Für die Identifizierung des Nutzers beim FD muss entsprechend diese beim GMS registrierte UUID_Nutzer verwendet werden. Dies kann entweder durch eine direkte Nutzung der UUID_Nutzer oder die Verwendung einer davon abgeleiteten UUID_FD_Nutzer als Pairwise-Identifier erfolgen. Dadurch wird sichergestellt, dass der IDP die Gerätebindung nicht umgehen kann.

Zudem muss sichergestellt werden, dass die Authentifizierung des Nutzers resistent gegen Angriffe ist, wie beispielsweise Phishing. Das ist insbesondere herausfordernd, wenn das Authentisierungsmittel nicht direkt mit dem Trust-Client kommunizieren kann.

Die Registrierung wird direkt vom Trust-Client durchgeführt, kann allerdings vom Fach-Client angestoßen werden, sofern der Trust-Client keine eigene Bedienoberfläche hat, oder dessen Bedienoberfläche direkt in den Fach-Client integriert ist.

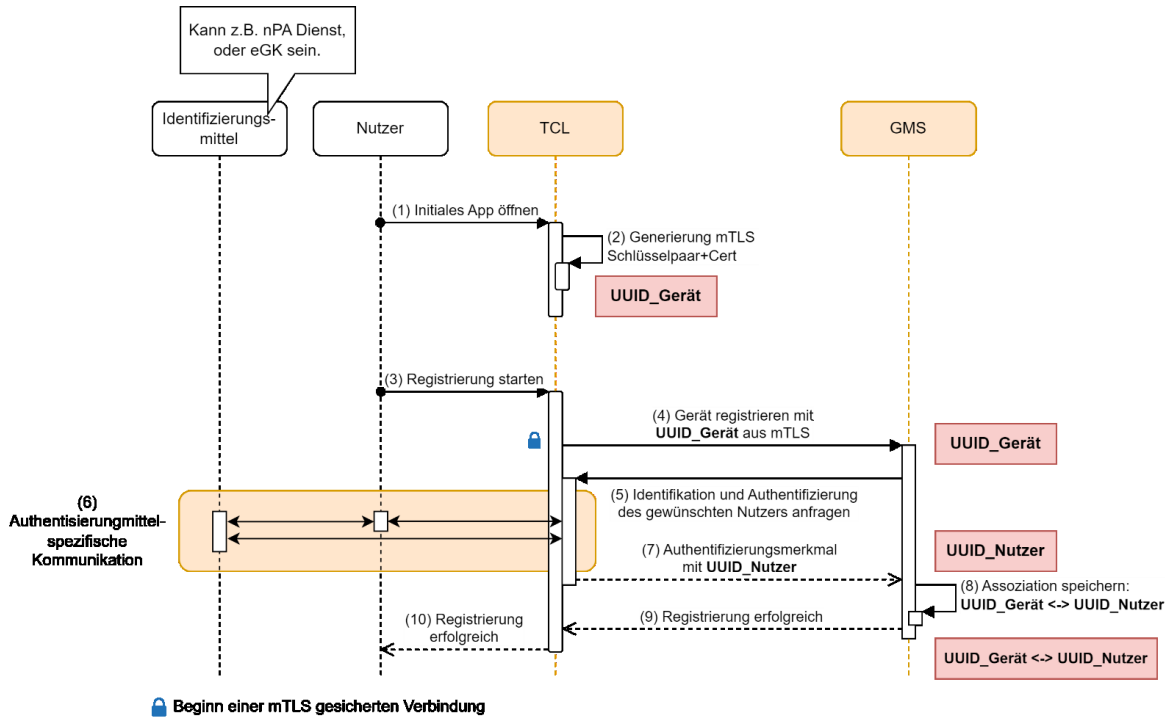

Abbildung 3.5.2-3 Registrierung eines Geräts beim GMS

FA5.2 - Gerät entfernen

Die Geräte-Nutzer-Assoziation (UUID_Gerät <-> UUID_Nutzer) ist nur am GMS und im von ihm ausgestellten Geräte-Token hinterlegt. Geräte-Token verlieren automatisch ihre Gültigkeit, nachdem ihre Lebensdauer überschritten ist. Dementsprechend kann die Gerätebindung durch eine einzelne Operation am GMS entfernt werden, welche spätestens nach Ablauf der maximalen Lebensdauer eines Geräte-Token wirksam wird.

Diese Operation muss ohne den Zugriff auf das registrierte Gerät möglich sein, um nicht mehr zugängliche Geräte ebenfalls entfernen zu können. Das Entfernen von Geräten ist weniger sicherheitskritisch als die initiale Registrierung. Daher können hier verschiedene Methoden zur Authentifizierung eingesetzt werden. Hier ist beispielhaft eine Authentifizierung durch den IDP per OIDC dargestellt [OpenID-Connect]. Es ist aber auch möglich das Authentisierungsmittel für die Geräteregistrierung zu verwenden, oder Informationen mit Gerätebezug zu verwenden. So könnte bei der Geräteregistrierung ein

Lösch-Code mit ausgestellt werden. Weiterhin könnte die Löschung auch mit einem anderen authentisierten Gerät möglich sein, evtl. über das Nutzer-Portal.

Das nachfolgende Datenflussdiagramm (Abbildung 3.5.2-4) zeigt beispielhaft den Kommunikationsfluss für das Entfernen eines registrierten Gerätes (basierend auf der UUID_Gerät) mittels OIDC gegen den IDP.

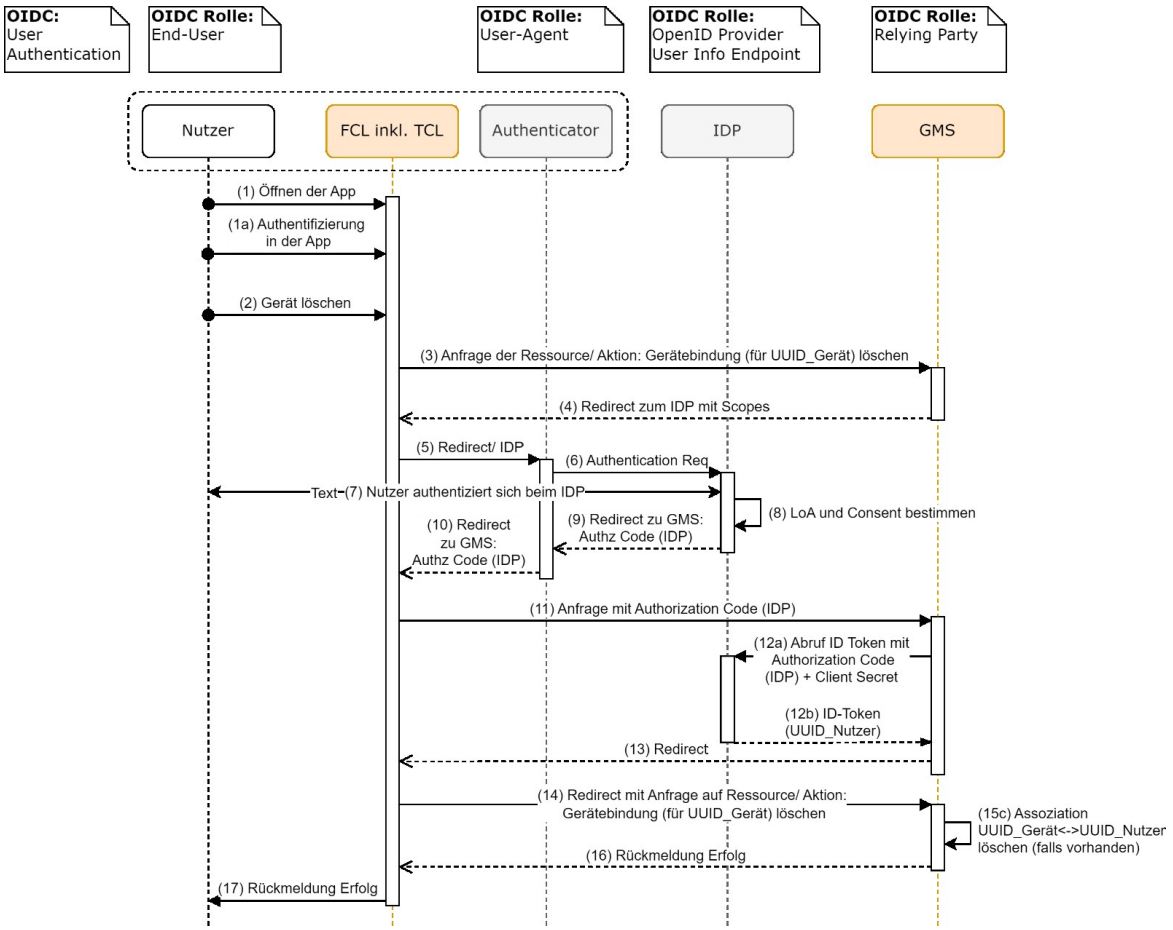

Abbildung 3.5.2-4 Entfernen einer Nutzer-Geräte-Assoziation beim GMS

FA6 - Gerät attestieren

Informationen über den Sicherheitszustand des Gerätes werden durch den GMS geprüft und zusammen mit der Geräte-Nutzer-Assoziation in einem Geräte-Token bestätigt. Im Wesentlichen findet dabei ein OAuth2 Authorization Code Flow mit PKCE [RFC6749_OAuth2, RFC7636_OAuth2_PKCE] statt. Der Fach-/Trust-Client kann in dem initialen Authentication Request (2) durch Scopes den GMS darauf hinweisen, welche Nachweise beziehungsweise Claims der Client als notwendig erachtet. Dies kann durch eine vom Fach-Client lokal vorgehaltene Liste erfolgen. Anschließend fordert der GMS vom Trust-Client Informationen über das Nutzerendgerät an, das dieser mithilfe von plattform-spezifischen Frameworks und Diensten zur Verfügung stellt (3). Der GMS überprüft die erhaltenen Attestierungs-Informationen mit dem für den Gerätetyp anzuwendenden Verifizierungsmechanismus (4,5). Je nach verwendetem Framework kann dafür eine Interaktion mit Backend-Diensten des Framework-Anbieters (z.B. Google, Apple, Microsoft) nötig sein (4). Mithilfe eines Authorization Codes (6) können die verifizierten Informationen dann über das Geräte-Token vom GMS angefordert werden (7). Das Geräte-Token enthält einerseits die beim GMS hinterlegte Geräte-Nutzer-Assoziation und

andererseits die verifizierten Informationen über das Endgerät in für die TI standardisierter Form (8).

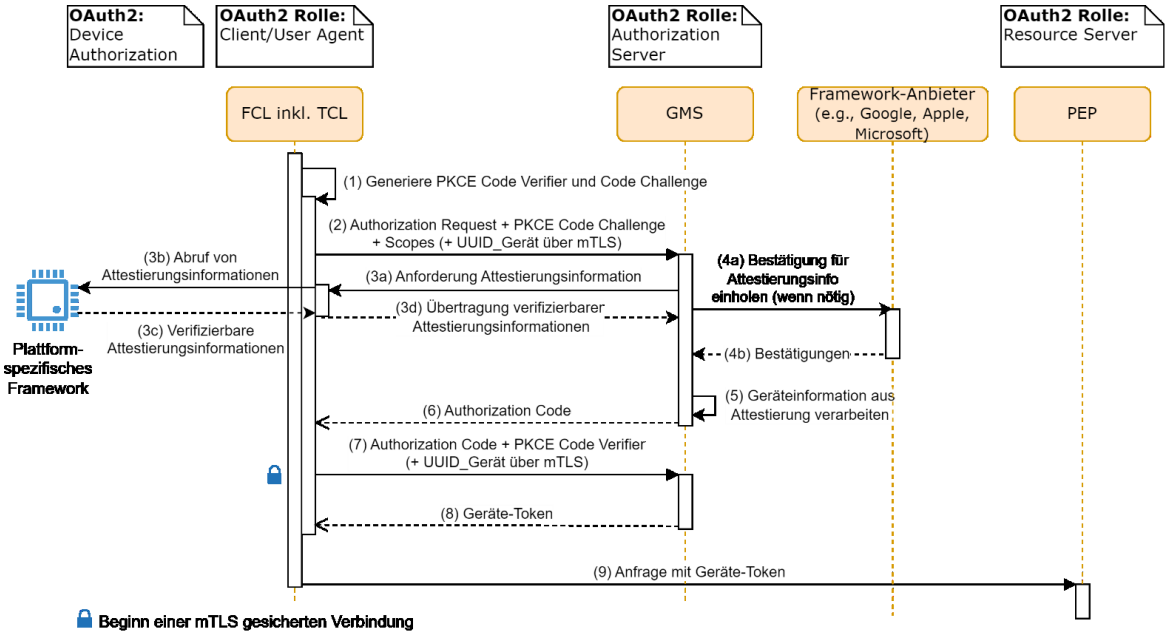

Abbildung 3.5.2-5 Attestierung von Geräte-Eigenschaften

FA6.1 - Geräteinformationen erheben

Die für die Bewertung des Gerätezustands nötigen Informationen werden auf dem jeweiligen Endgerät des Nutzers erhoben. Um zu verhindern, dass ein kompromittiertes Endgerät die Informationen über den Gerätezustand manipulieren kann, müssen die Informationen durch eine vertrauenswürdige Komponente auf dem Gerät zur Verfügung gestellt oder bestätigt werden. Hierfür können beispielsweise TPMs oder Trusted Execution Environments verwendet werden. Da die Nutzung der TI mit den im Alltag bereits genutzten Endgeräten der Nutzer möglich sein soll (besonders bei den Versicherten), sollten für die Erhebung der Geräteinformationen auf den Endgeräten die dort bereits verfügbaren Frameworks bzw. Lösungen verwendet werden. Diese sind in der Regel abhängig vom verwendeten Betriebssystem und benötigen bisher meist die Unterstützung durch den entsprechenden Anbieter. So liefert beispielsweise die Google Play Integrity API(https://developer.android.com/google/play/integrity/overview) mithilfe von Key Attestation über einen Google Backend-Server Informationen über die Integrität einer verwendeten Anwendung und eine Bewertung der Geräteintegrität. Auch das Device Check Framework(https://developer.apple.com/documentation/devicecheck) von Apple bietet Nachweise über die Integrität von Apps und die Identität des Geräts. Unter Windows stehen mehrere Möglichkeiten zur Verfügung, um Geräteinformationen zu erheben. Einerseits besteht die Möglichkeit, direkt mit dem TPM als Root-of-Trust zu interagieren und den Systemzustand zu attestieren. Ebenfalls steht die Windows Device Health Attestation API(https://learn.microsoft.com/en-us/windows/security/threat-protection/protect-high-value-assets-by-controlling-the-health-of-windows-10-based-devices) zu Verfügung, die mehrere Komponenten in einer vereinheitlichen Schnittstelle für eine umfassendere Systemattestierung bietet. Diese ist allerdings möglicherweise nicht in allen Systemkonfigurationen verfügbar (bspw. nur AD-Joined).

3.5.3 Fachdienst authentisieren und attestieren

FA3 - Fachdienst gegenüber Nutzer authentisieren

Das Identifizieren und Authentifzieren eines Fachdienstes durch den Nutzer bzw. seinen Client erfolgt über das TLS-Zertifikate des Fachdienstes bzw. seines PEP. Diese Zertifikate werden im Rahmen des TLS-Verbindungsaufbaus an den Trust-Client übertragen. Der Trust-Client prüft gemäß TLS-Standard, ob der Fachdienst im Besitz des zugehörigen privaten Schlüssels ist (Abschnitt 4.4.3. in für TLS1.3 [RFC8446-TLS]) und validiert das Zertifikat nach anerkannten Standards (z.B. [RFC5280-Certs]) gegen eine für die TI zugelassene CA. Der Fachdienst verwendet OCSP-Stapling (wie in Abschnitt 4.4.2.1 in [RFC8446-TLS] beschrieben) und der Trust-Client prüft anhand der übermittelten Informationen, dass das Zertifikat nicht gesperrt oder widerrufen wurde. Falls eine der Prüfungen fehlschlägt, verweigert der Trust-Client den Zugriff auf den Fachdienst und informiert den Endnutzer über das vorliegende Problem.

FA4 - Fachdienst attestieren (extern)

Dienste, die an die TI angeschlossen werden sollen, werden initial von der gematik in einem Zulassungs- bzw. Bestätigungsprozess gegen die Anforderungen der TI geprüft. Zu den Anforderungen kann es gehören, dass der Dienst mithilfe von Attestierung auch im laufenden Betrieb geprüft wird. Hierbei werden Eigenschaften der laufenden Fachdienst-Instanz überprüft, um ihre Vertrauenswürdigkeit im Betrieb zu bewerten. Zu den überprüften Eigenschaften können die Integrität und Aktualität der Software, die Korrektheit der Systemkonfiguration oder die Lokalisierung der Instanz in einer zugelassenen Rechenzentrumsumgebung gehören. Die Integritätsprüfung kann mittels eines TPMs mit Measurements des gesamten SW-Stacks oder mittels der Remote Attestation Features von Confidential Computing Technologien wie AMD SEV-SNP oder Intel TDX umgesetzt werden. Die Attestierung kann implizit durch das Binden der verwendeten kryptografischen Schlüssel an den Zustand des Software-Stacks oder explizit durch eine Attestierung gegenüber einem zentralen Dienst in der TI erfolgen.

Die Konzeptionierung und Umsetzung der Attestierungsfunktion für Fachdienste ist nicht Teil dieses Projektes. Das vorliegende Feinkonzept beschränkt sich deshalb auf das Prüfen der fachdienstspezifischen TLS-Zertifikate.

3.5.4 Regelwerk prüfen, erstellen, überwachen und anpassen

Das Erstellen, Verwalten und Anwenden von Zugriffsregeln innerhalb einer Zero-Trust-Architektur erfordert ein definiertes Schema für das Beschreiben der Zugriffsregeln. Der offene Standard eXtensible Access Control Markup Language (XACML) [OASIS_XACML] der Organisation OASIS (unter anderem vertreten durch Cisco Systems, IBM, Microsoft, Oracle und Red Hat) beschreibt ein solches Schema, in Form einer auf XML basierenden Beschreibungssprache für ein Regelwerk sowie einem Datenflussmodell. Dabei orientiert sich das Datenflussmodell an den logischen Komponenten des Standards ISO/IEC 29146 [ISO29146], welche auch in diesem Konzept berücksichtigt werden.

In Abbildung 3.5.4-1 ist der Aufbau eines Regelwerkes gemäß XACML 3.0 grafisch dargestellt.

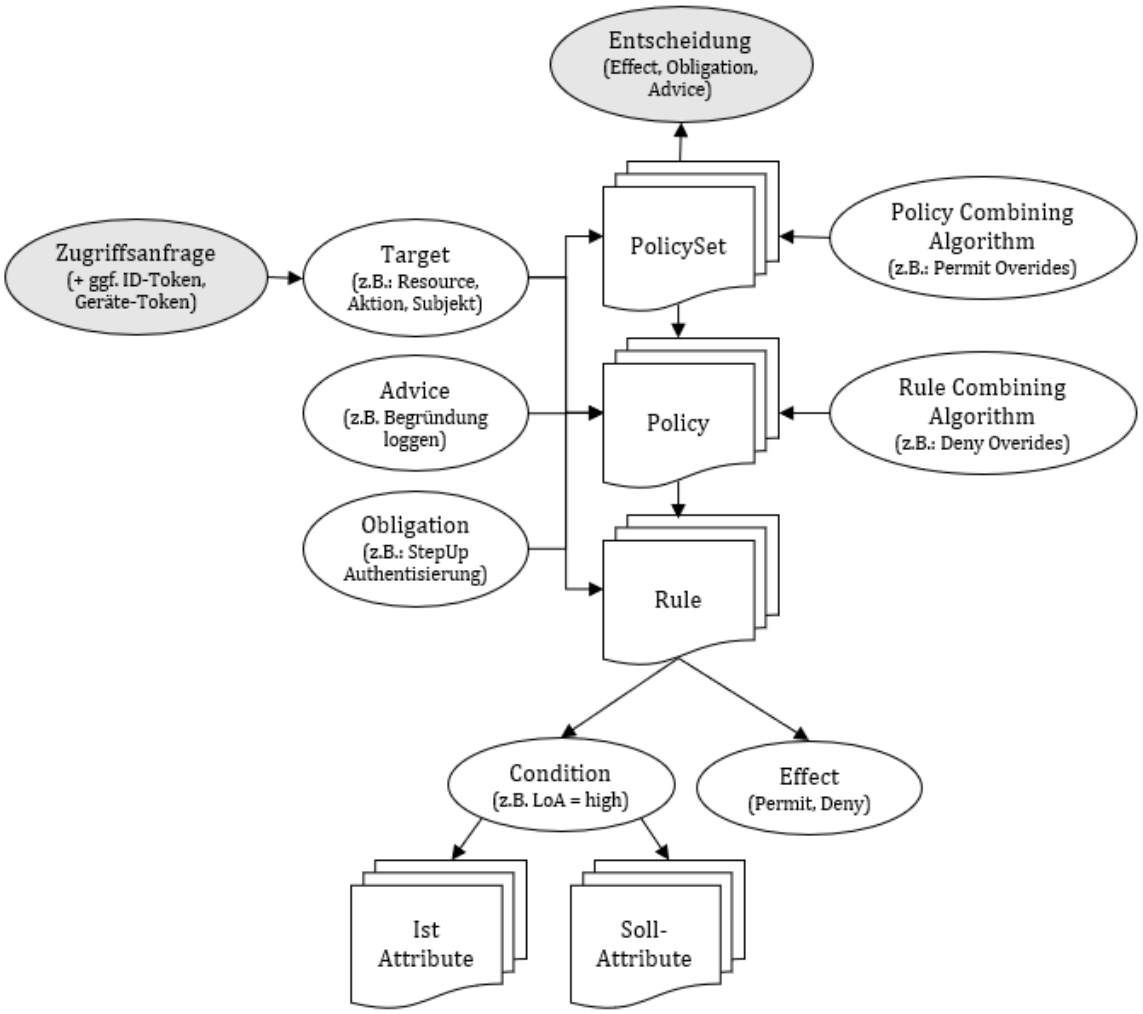

Abbildung 3.5.4-1: Modell Regelwerk XACML 3.0

XACML definiert eine hierarchische Struktur für das Verwalten von Regeln. Eine Regel (Rule) ist die elementare Einheit, in der eine Zugriffsentscheidung getroffen werden kann.

Eine Regel ist immer einer Policy zugeordnet. Policies können darüber hinaus in PolicySets gruppiert werden. Dabei können PolicySets wiederum PolicySets zugeordnet werden, wodurch sich Regelwerke beliebig fein untergliedern lassen. Für welche Zugriffsanfragen PolicySets, Policies und Regeln Anwendung finden wird in jeder dieser Komponenten durch die Beschreibung eines Targets in Form eines logischen Ausdrucks über die Attribute der Zugriffsanfrage definiert.

Jede Regel kann eine Bedingung (Condition) in Form eines boolschen Ausdrucks enthalten, welcher den Anwendungsbereich der Regel in Abhängigkeit von weiteren Attributen (Ist-Attribute, Soll-Attribute) weiter einschränkt. XACML stellt eine lange Liste an Funktionen zur Verfügung, um Attribute zu bearbeiten oder mit anderen Attributen zu vergleichen.

Jede Regel erzielt eine Wirkung (Effect) als Teilentscheidung hinsichtlich der Zugriffsanfrage. Die Wirkung der Regel gibt die beabsichtigte Konsequenz (Permit, Deny) an, die der Verfasser der Regel bei einer positiven Auswertung (true) der Bedingung der Regel beabsichtigt.

Konflikte im Rahmen von unterschiedlichen Zugriffsentscheidungen verschiedener Regeln einer Policy werden durch einen definierten Algorithmus (Rule Combining Algorithm) der

Policy aufgelöst. PolicySets verfügen über einen analogen Algorithmus (Policy Combining Algorithm) zum Auflösen von Konflikten von Zugriffsentscheidungen darunterliegender Elemente.

Für PolicySets, Policies und Regeln können darüber hinaus verpflichtende (Obligation) oder empfohlene (Advice) Ausführungsanweisungen definiert werden, welche in Abhängigkeit von der Zugriffsentscheidung zusammen mit dieser an den PEP übermittelt werden.

XACML wurde im Jahr 2001 als Policy-Sprache für ABAC entwickelt und im Jahr 2013 in der Version 3.0 veröffentlich. 2017 wurde der Standard letztmalig aktualisiert.

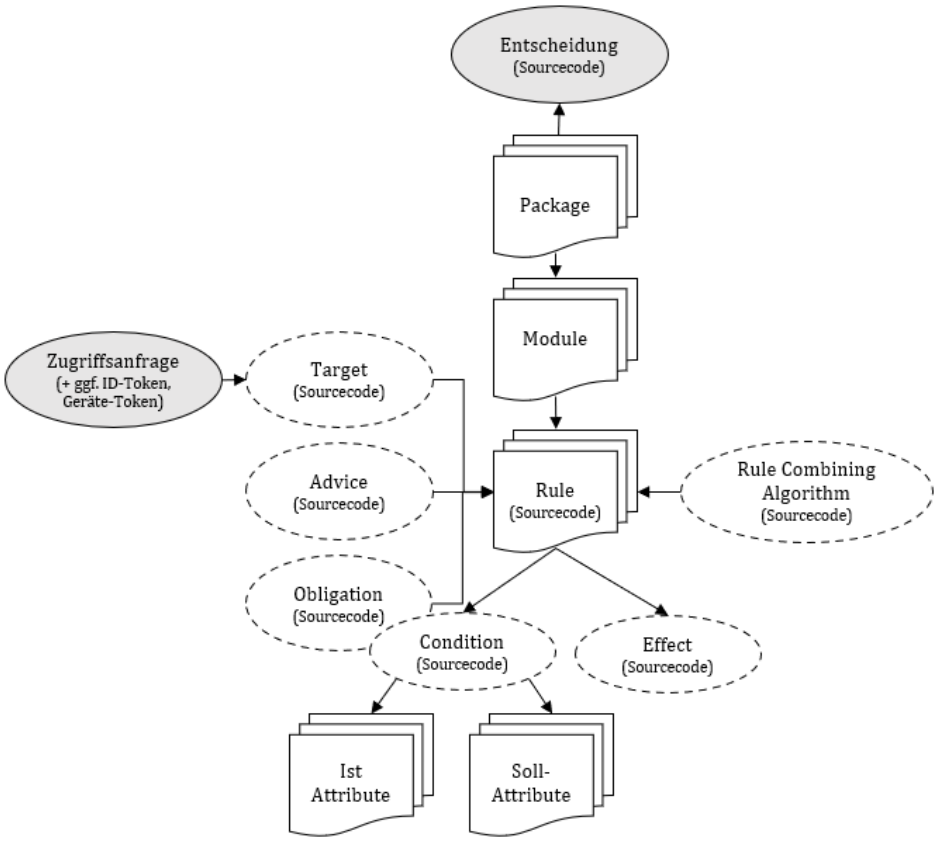

Abbildung 3.5.4-2: Vergleich Modell Regelwerk Rego / XACML 3.0

Eine weitere Regelbeschreibungssprache ist Rego [OPA_Rego]. Rego wurde von der Firma Styra Inc. entwickelt und 2018 zusammen mit der Regelauswertungsengine Open Policy Agent (OPA)(https://www.openpolicyagent.org/) als Open Source Projekt in die Cloud Native Computing Foundation (CNCF) aufgenommen. Rego selbst ist jedoch nicht standardisiert. Rego ermöglicht das Beschreiben von Regeln in Form einer Programmiersprache. Dadurch ergibt sich ein gegenüber XACML erhöhter Grad an Flexibilität und Ausdrucksstärke.

In Abbildung 3.5.4-2 ist der Aufbau eines Regelwerkes mit Rego anhand der Artefakte von XACML grafisch dargestellt.

Grundsätzlich lässt sich in Rego eine ähnliche Struktur des Regelwerks abbilden, wie dies in XACML möglich ist. Regeln können in Packages und Modulen strukturiert werden. Elemente wie Target, Advice, Obligation, Condition, Effect und Rule Combining Algorithm sind in Rego nicht direkt definiert, können aber über die Logik der Programmiersprache in den Regeln abgebildet werden. Darüber hinaus ist es so möglich, das Modell des Regelwerks weiter an domainspezifische Bedürfnisse anzupassen. Objekte, welche im Rahmen der Zugriffsanfrage an den PEP übergeben werden oder als Entscheidung an den PEP zurückgegeben werden, können zum Beispiel flexibel definiert werden. Ebenso können

komplexe Bedingungen durch den erhöhten Freiheitsgrad in Rego einfacher beschrieben werden als in XACML.

XACML oder Rego eröffnen grundsätzlich die Möglichkeit zur Umsetzung des Regelwerkes. Vor einer Umsetzung des Regelwerkes und der damit verbundenen Komponenten sollte jedoch noch ein tiefergehender Vergleich der vorgestellten Schemas mit weiteren Open-Source-Optionen wie z.B. NGAC [NIST_ABAC-Comparison] erfolgen.

FA1 - Zugriff gegen Regelwerk prüfen

In Abbildung 3.5.4-3 ist der Datenfluss für das Prüfen eines Zugriffs gegen das Regelwerk dargestellt.

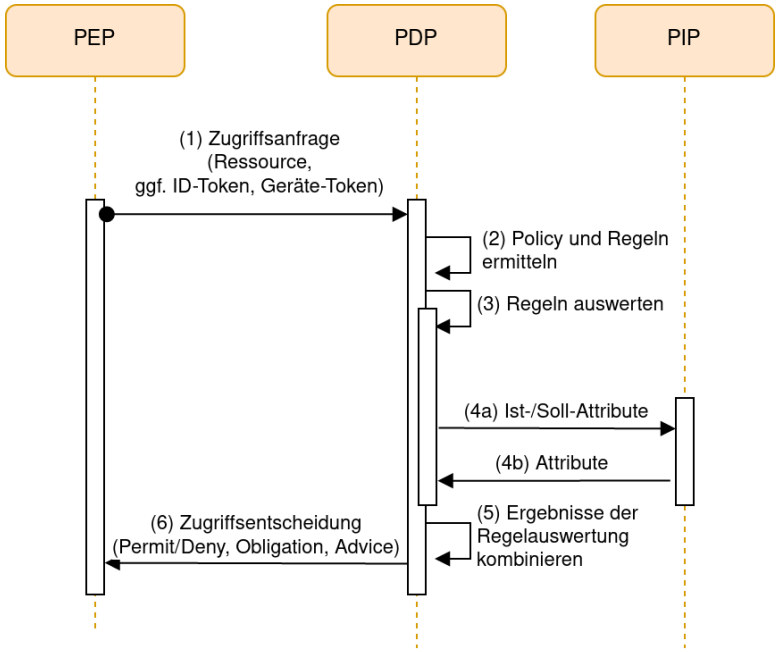

Abbildung 3.5.4-3: Datenfluss für das Prüfen eines Zugriffs gegen das Regelwerk

Zugriffsanfragen auf einen Fachdienst leitet der PEP zunächst an den PDP weiter (1). Eine Zugriffsanfrage enthält immer Informationen über die angefragte Ressource, d.h. das angefragte Objekt und die intendierte Aktion. Darüber hinaus werden, sofern vorhanden, ID-Token und Geräte-Token zusammen mit der Zugriffsanfrage an den PDP weitergeleitet.

Der PDP ermittelt anhand der Informationen aus der Zugriffsanfrage die relevanten Policies und Regeln (2) und wertet diese detaillierter aus (3). Attribute, welche für das Auswerten von Bedingungen der Regeln erforderlich sind, werden dabei entweder der Zugriffsanfrage entnommen oder bei einem PIP angefragt (4a, 4b).

Der PDP kombiniert die Ergebnisse aller ausgewerteten Regeln zu einer Zugriffsentscheidung (5) und sendet diese zusammen mit möglichen weiteren Handlungsanweisungen an den PEP (6) zurück.

FA1.1 - Regelwerk erstellen und aktualisieren

Die Regeln für das Regelwerk der TI werden vor ihrer Anwendung in der ZTA durch einen geeigneten Prozess definiert und abgestimmt (siehe Kapitel 5). Um die Diskussion aller Beteiligten über die angedachten Regeln zu ermöglichen, werden die Regeln dabei in natürlicher Sprache oder einer geeigneten leicht verständlichen Form wie z.B. einer Tabelle oder low-code formuliert. Das könnte beispielsweise wie folgt aussehen:

Option 1: Natürliche Sprache
Der Vorteil an der Verwendung von natürlicher Sprache ist, dass diese grundsätzlich von jedem Diskussionsteilnehmenden verstanden werden kann und kein tiefergehendes technisches Verständnis der Beteiligten erfordert. Die Herausforderung liegt jedoch darin, in natürlicher Sprache die Regeln klar und ohne Raum für Interpretation zu definieren, ggf. komplexe Zusammenhänge sauber aufzuschreiben und danach eine geeignete Übersetzung in eine maschinenlesbare Sprache zu finden. Falls dieser Weg gewählt wird, ist die Verwendung von Templates für die Regelsätze zu empfehlen.

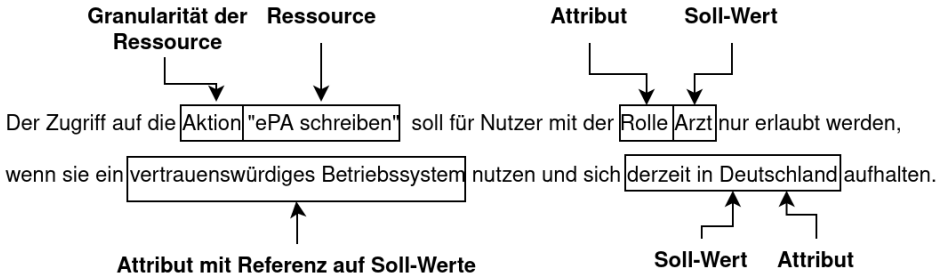

Abbildung 3.5.4-4: Beispiel für eine Regel in natürlicher Sprache

Option 2: Tabellenform

Eine Tabelle erlaubt ein strukturiertes Vorgehen und ein konsequentes Durchgehen/Diskutieren aller möglichen Attribute pro Regel. Sie schränkt jedoch die Flexibilität stark ein, da eine Abbildung komplexerer Regel-Kombinationen (z.B. Verknüpfung mehrerer Zeilen) nur schwierig möglich ist.

Granularität der Ressource	Ressource	Nutzerrolle(n)	Vertrauenswürdiges Betriebssystem	Ort der Anfrage	Zeitpunkt der Anfrage	...
Aktion	ePA schreiben	Arzt	ja	DE	--	
...						

Tabelle 3.5.4-5: Beispiel für eine Regel in Tabellenform

Option 3: Low-Code

Low Code beschreibt die Verwendung einer visuellen Gestaltungsoberfläche anstelle von klassischen textbasierten Programmiersprachen. Mithilfe von Low-Code kann eine Darstellung der Regeln geschaffen werden, die nahe an der gewünschten Sprache der Policy-Engine ist und so die technische Abbildung des Regelwerkes vereinfacht. Im Rahmen der Marktanalyse für die zu verwendende Policy-Sprache sollte untersucht werden, ob es bereits geeignete graphische Policy Editoren für die jeweilige Sprache gibt oder wie viel Aufwand es wäre, ein solches Tooling zu entwickeln.

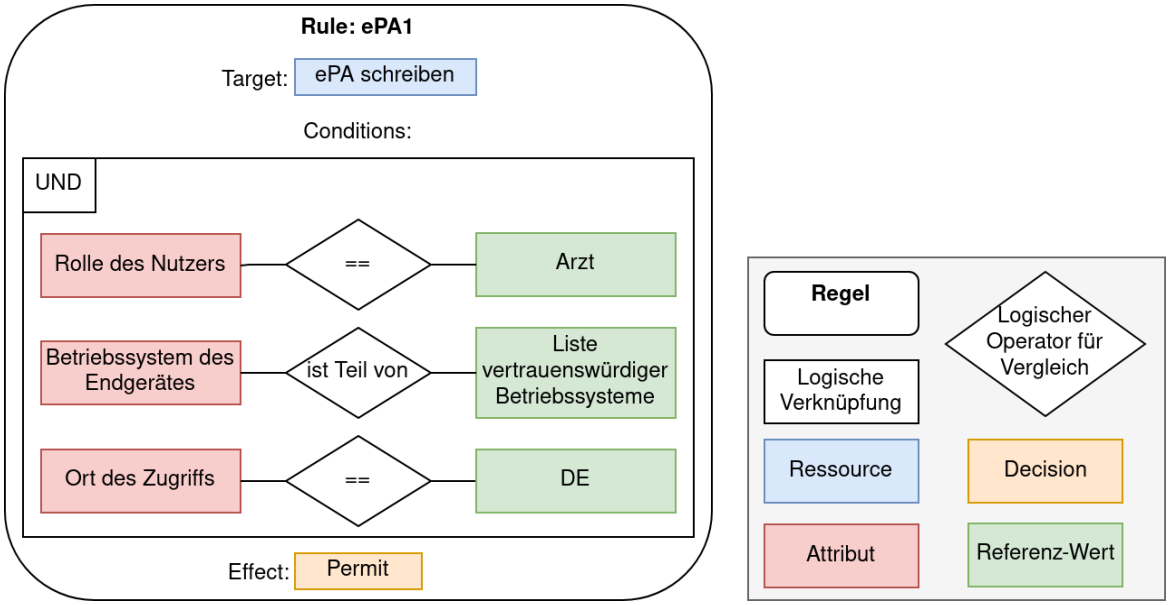

Abbildung 3.5.4-5: Beispiel für eine Regel in Low-Code

Die gewünschte Darstellungsform sollte in Absprache mit allen an der Regelsetzung Beteiligten festgelegt werden. Wichtig ist es, dass die Regeln klar formuliert und ohne Interpretationsspielraum dokumentiert werden können.

Für die Verwendung der Regeln in der TI müssen sie aus der gewählten Darstellungsform in die Sprache der Policy-Engine (z.B. XACML oder Rego) übersetzt werden. Diese Übersetzung kann entweder manuell erfolgen oder auch automatisiert werden, wenn es eine eindeutige Interpretation/Übersetzungsmethodik von der ursprünglich verwendeten Form auf die maschinenlesbaren Regeln gibt.

Nach der Übersetzung sollte die Übereinstimmung der maschinenlesbaren Version mit dem in natürlicher Sprache verfassten Regelwerk geprüft werden. Mögliche Methoden sind z.B.:

- das Testen mit definierten Beispielfällen: Es werden verschiedene Input-Daten für die möglichen Attribute sowie deren erwartete positive oder negative Auswertungsergebnisse definiert und das generierte Regelwerk darauf angewendet/dagegen getestet
- das Rückübersetzen der Policy in eine natürliche Sprache/das Ursprungsformat und der Abgleich, ob die initiale und die rückübersetzte Regel übereinstimmen

Zudem sollte ggf. Tooling zum Einsatz kommen, welches das fertig definierte Regelwerk auf Widerspruchsfreiheit und Konsistenz prüft.
Im Rahmen der Marktanalyse für die zu verwendende Policy-Sprache sollte untersucht werden, ob es zu der jeweiligen Sprache bereits Tooling für die Verifikation oder das Testen gibt (ggf. angelehnt an bestehenden Verifizierungs- und Testmethoden für Policies/Modelle [NIST_PolicyVerification]).

Eine erfolgreiche Prüfung des übersetzten Regelwerks und der darin enthaltenen PolicySets (bzw. Packages) wird vor dem Einbringen in die Komponenten der ZTA durch kryptographische Signaturen bestätigt. Das Regelwerk wird dabei basierend auf der Gültigkeit für die Fachdienste in der TI in verschiedene PolicySets unterteilt und jedes dieser PolicySets wird nach erfolgreicher Prüfung einzeln signiert. Das erlaubt es, den ZTA-Komponenten mit den entsprechenden öffentlichen Schlüsseln die Integrität und Authentizität der erhaltenen Policy-Sets zu prüfen.

Das gesamte Regelwerk wird (in der Sprache der Policy-Engine) am PAP in die TI eingebracht, indem die signierten PolicySets dort z.B. über Konfigurationsdateien zur

Verfügung gestellt werden. PDPs sind dazu verpflichtet, in regelmäßigen Abständen (z.B. täglich) beim PAP die aktuelle Version der für ihren Fachdienst gültigen Teile des Regelwerks abzufragen. Sollte die damit gegebene Updatefrequenz für kurzfristige Updates im Emergency Fall nicht ausreichen, so müsste ein zusätzlicher Weg geschaffen werden, über den ein sofortiges Update der PDP bewirkt werden kann (z.B. einen Notification-Mechanismus, über den der PAP den PDP kurzfristig zum Pull einer neuen Regelwerksversion auffordern kann).

Der PAP ermittelt die für den anfragenden Fachdienst geltenden PolicySets und übermittelt diesen Ausschnitt des Regelwerks an den PDP. Der PDP prüft die Integrität und Authentizität jedes PolicySets durch eine Signaturprüfung mithilfe des entsprechenden öffentlichen Schlüssels und verwendet erst nach einer erfolgreichen Prüfung das neue, abgefragte Regelwerk. Jede Version des Regelwerks wird zudem mit einem Aktivierungszeitpunkt versehen und wird von dem PDP erst ab diesem Zeitpunkt verwendet.

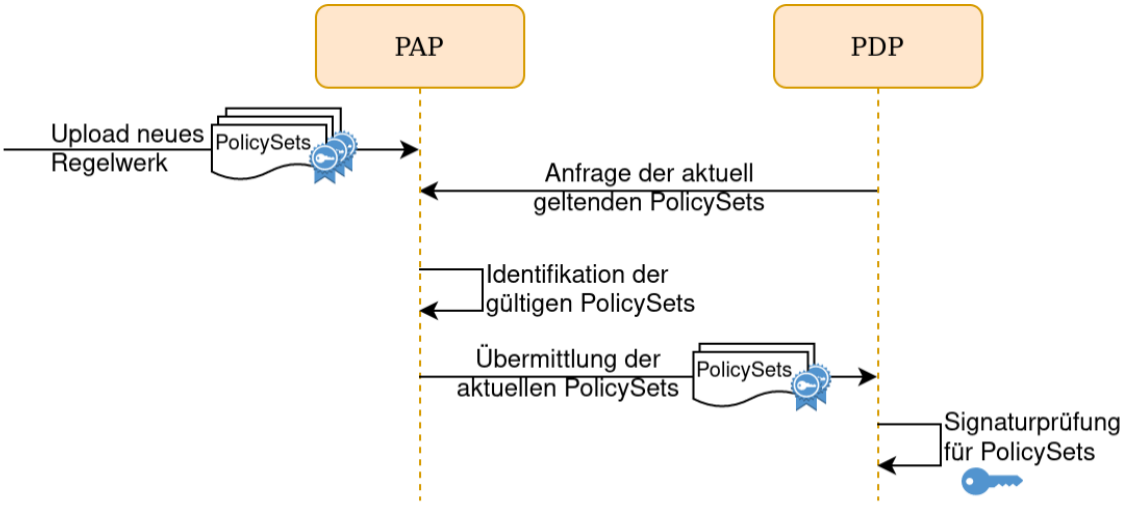

Abbildung 3.5.4-7: Datenflussdiagram für das Verteilen des Regelwerkes an die PDPs

Jedem Fachdienst werden ein oder mehrere PolicySets zugeordnet, die durch den PDP vom PAP abgefragt werden. Bei der Gestaltung des Regelwerks kann durch eine Verschachtelung von PolicySets dafür gesorgt werden, dass auch spätere Anpassungen in der Strukturierung des Regelsets möglich bleiben. So kann beispielsweise dem Fachdienst ePA grundsätzlich ein PolicySet "ePA" zugeordnet, dieses aber wiederum aus mehreren "Sub-PolicySets" (z.B. einem generischen PolicySet, einem PolicySet für alle Fachdienste, die medizinische Daten speichern, und einem PolicySet mit ePA-spezifischen Regeln) zusammengesetzt sein.

Die für das Regelwerk benötigten Soll-Attribute können je nach Regel entweder als Teil der Condition direkt in der Regel enthalten sein oder durch einen PIP für die Regelauswertung zur Verfügung gestellt werden.
Im ersten Fall, d.h. wenn die Soll-Attribute direkt in der Regel enthalten sind, erfolgt ein Update der Attribute immer als Teil eines Regelwerk-Updates und wird wie oben beschrieben vom PAP an die PDPs weitergegeben.
Wenn hingegen die Soll-Attribute durch einen PIP zur Verfügung gestellt werden, kann ein Update der gültigen Werte unabhängig von einem Update des gesamten Regelwerks direkt am jeweiligen PIP erfolgen. Die meisten Attribute werden dabei im Rahmen organisatorischer Prozesse (siehe Kapitel 5.2.2) geändert. Die durch diese Prozesse definierten Soll-Attribute sollten versioniert und signiert werden, um am PDP die

Aktualität, Integrität und Authentizität der verwendeten Informationen überprüfen zu können. Für manche Attribute kann (statt eines organisatorischen) ein automatisierter Prozess für die Festlegung der Referenzwerte durch das Monitoring über die in FA1.4 beschriebene Funktion definiert werden.

FA1.2 - Regelwerk überwachen

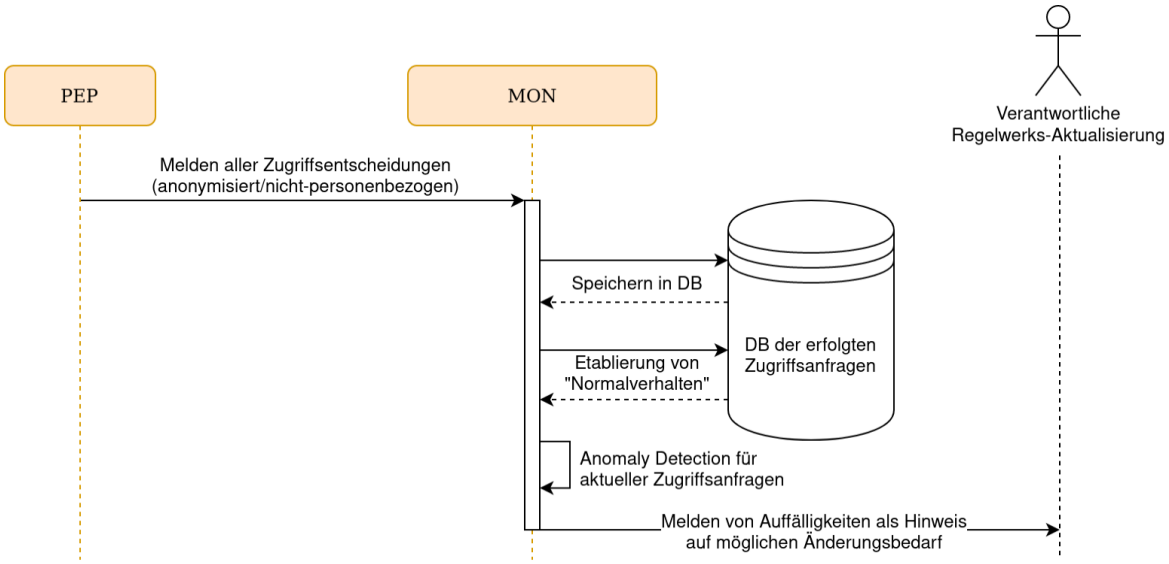

Abbildung 3.5.4-8: Datenflussdiagram für das Monitoring des Regelwerks

Um die Effektivität und Angemessenheit des Regelwerkes zu bewerten, findet ein Monitoring seiner Anwendung statt, bei dem alle Zugriffsentscheidungen in der ZTA erfasst und geloggt werden. Zu diesem Zweck werden vom PEP zu jeder Regelauswertung Informationen über die getroffenen Zugriffsentscheidungen übermittelt, d.h. welche Attribut-Ist-Werte im Rahmen der Zugriffsanfrage erhalten, welche Regelwerksversion verwendet und welche Entscheidung durch den PDP getroffen wurden. Die Informationen werden vom PEP dabei nur in anonymisierter Form, also insbesondere ohne Identifier wie UUID_Nutzer und UUID_Gerät, übertragen. Bei der Monitoring-Komponente werden die Zugriffsentscheidungen geloggt. Zudem analysiert die Monitoring-Komponente laufend die eingehenden Zugriffsentscheidungen, um signifikante Abweichungen von gelernten Erfahrungswerten zu erkennen (Anomaly Detection). Ein starkes Abweichen von etabliertem Normalverhalten, z.B. eine deutlich erhöhte Anzahl an abgelehnten Zugriffsentscheidungen wegen einer geltenden Regel, kann als Hinweis für einen Änderungsbedarf für das aktuell geltende Regelwerk dienen. Die geloggten Informationen können, wie in Funktion FA1.3 Regelwerk anpassen beschrieben, bei die Abstimmungsprozesse zu möglichen Regelwerksupdates unterstützen.

FA1.3 - Regelwerk anpassen

Um die Aktualität des Regelwerks sicherzustellen, wird das Regelwerk regelmäßig durch das zuständige Gremium begutachtet. Auslöser dieser Begutachtung kann entweder der Ablauf eines festgelegten Zeitraums seit der letzten Überprüfung sein (z.B. spätestens ein Jahr nach der letzten Überprüfung) oder ein durch das Monitoring gemeldeter möglicher Änderungsbedarf, z. B. wenn unverhältnismäßig viele Zugriffe wegen einer Regel abgewiesen werden. Die Beteiligten an Erstellung bzw. Update des Regelwerks können die gespeicherten Monitoring-Daten für die Bewertung möglicher Regeländerungen hinzuziehen. So können beispielsweise Informationen über tatsächlich verwendete Betriebssystemversionen helfen die Auswirkung einer Änderung der Soll-Werte abzuschätzen oder kaum verwendete Regelsätze identifiziert werden, die ggf. aus dem Regelwerk entfernt werden können. Änderungen werden im Rahmen der entsprechenden

Governance-Prozesse beschlossen und anschließend wird eine neue Version des Regelwerks über den oben beschriebenen Prozess, FA1.1, an PAP, PDP und ggf. PIP verteilt.

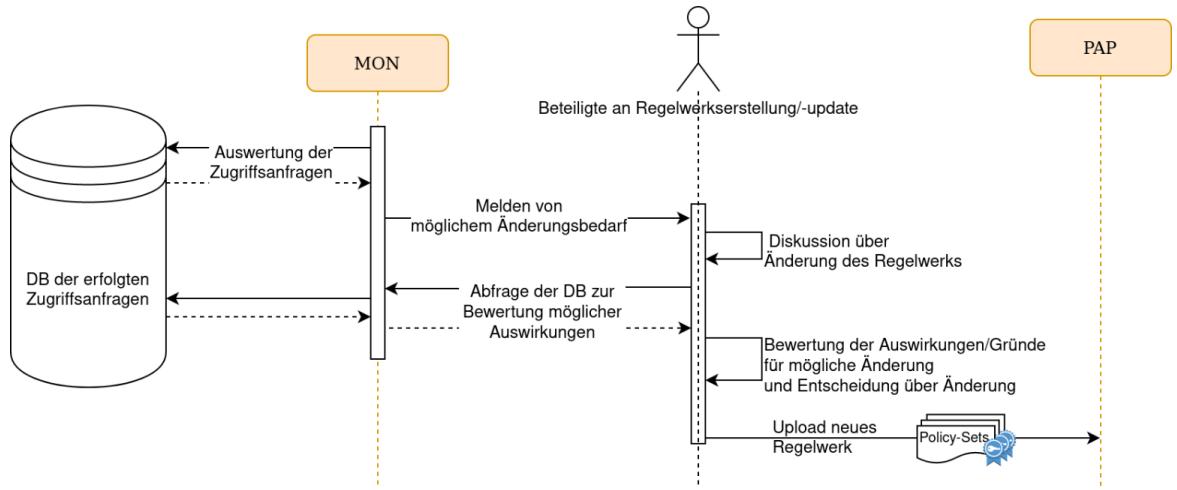

Abbildung 3.5.4-9: Datenflussdiagram für ein Update des Regelwerkes durch organisatorische Prozesse

FA1.4 - Attribute des Regelwerks anpassen

Neben den Anpassungen von Soll-Werten im Rahmen von Regelwerk-Updates bzw. durch entsprechende Governance-Prozesse gibt es auch Attribute, für die eine automatisierte Anpassung der Referenzwerte auf Basis aktueller Sicherheitsbetrachtungen sinnvoll ist. Ein Beispiel hierfür ist das Identifizieren von IP-Address-Ranges, aus denen unverhältnismäßig viele Anfragen an die Fachdienste gestellt werden.

Das Monitoring sammelt hierfür benötigte Informationen vom PEP ein und ist dafür zuständig, diese Informationen automatisiert zu bewerten und zu entscheiden, ob eine Änderung der Referenzwerte nötig ist. In diesem Fall ist das Monitoring verantwortlich für die Aktualisierung der Referenzwerte im PIP und das Loggen der getroffenen Entscheidungen und Änderungen (Change Log). Bei Bedarf (insbesondere beim Auswerten personenbezogener Daten und zur Erhöhung der Performance) findet am PEP bereits eine Vorverarbeitung (z.B. Aggregation oder Filterung) statt. Bereits vorhandene Informationen beim Monitoring (z.B. die Zugriffsentscheidungen aus FA1.3 Regelwerk anpassen) können ebenfalls in die Auswertung mit einfließen. Die Informationen für die automatisierte Anpassung von Referenzwerten werden nur über einen begrenzten Zeitraum verarbeitet und nicht längerfristig gespeichert.

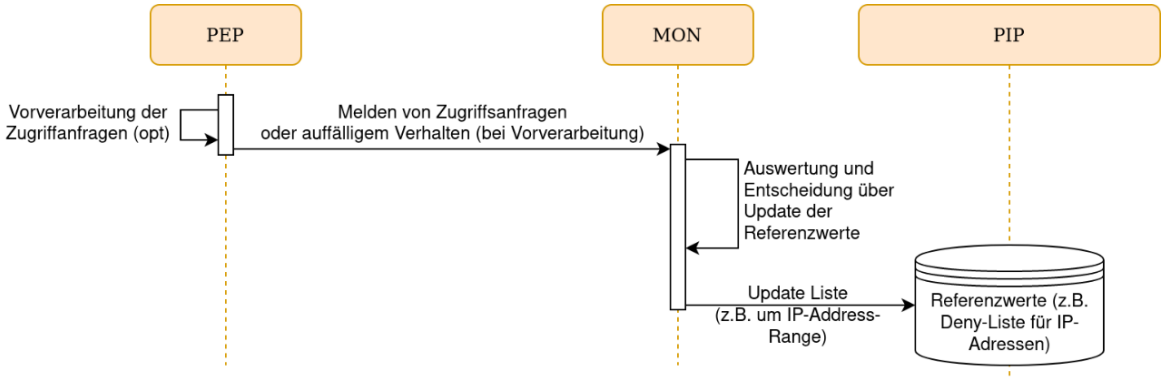

Abbildung 3.5.4-10: Datenflussdiagram für ein automatisches Update von Soll-Attributen

FA1.5 - Nutzerspezifische Regelwerkattribute anpassen

Bei der Erstellung des Regelwerks werden möglicherweise auch Attribute identifiziert, die von den Nutzungsgewohnheiten oder Präferenzen einzelner Nutzer abhängen. Beispiele hierfür sind die Definition von vertrauenswürdigen Orten und Zeiten für den Zugriff des Nutzers. Für derartige Attribute kann es schwierig sein, generische Soll-Werte zu definieren. Auch kann es die Effektivität von Regeln steigern nutzerspezifische Soll-Werte zu verwenden. Daher soll der Nutzer die Möglichkeit erhalten, über das Nutzerportal selbst Soll-Attribute zu definieren. Eine Speicherung und Abfrage der Werte im Rahmen der Regelauswertung kann z.B. beim zuständigen IDP, beim GMS oder bei einem dafür zuständigen PIP erfolgen. Die Konzeptionierung und Umsetzung der entsprechenden Funktionalität ist nicht Teil dieses Projektes.

3.5.5 Autorisierter Zugriff und Session Management

FA7 Autorisierter Zugriff

Zugriffe auf Ressourcen der ZTA müssen autorisiert und gegen ein Regelwerk geprüft werden. Die Prüfung der Anfrage erfolgt beim Policy Decision Point (PDP), welcher seine Entscheidung dem Policy Enforcement Point (PEP) zur Durchsetzung der Entscheidung mitteilt. Um die Anfragen korrekt gegen das Regelwerk prüfen zu können, müssen dem PDP alle relevanten Informationen aus der Anfrage zur Verfügung gestellt werden. Dazu erfasst der PEP alle notwendigen Informationen des aktuellen Zugriffs und übergibt sie gesammelt an den PDP. Dies schließt Informationen aus dem ID-/ oder Geräte-Token sowie der TLS/TCP/IP-Verbindung ein. Eine genauere Aufschlüsselung der Attribute und Informationsquellen wird im Kapitel 3.3- Regelwerk (Tabelle 3.3.1-1) diskutiert. Der PEP kann die Attributquellen filtern, um dem PDP nur die wirklich notwendigen Teile zugänglich zu machen. Beispielhaft können nur die relevanten Teile des ID-Tokens übertragen werden, aber der Nutzer-Identifier explizit ausgeschlossen. In einem ähnlichen Verfahren werden die Informationen ebenfalls dem Fachdienst zur Verfügung gestellt, damit dieser den Nutzer identifizieren kann. Damit kann der Fachdienst eigene Zugriffskontrollen implementieren.

In den folgenden beiden Abbildungen wird ein erfolgreicher Zugriff auf eine Ressource (Abb. 3.5.5-1), sowie die Zurückweisung einer nicht autorisierten Anfrage abgebildet (Abb. 3.5.5-2). Beide stellen beispielhaft einen durch den Nutzer angestoßenen Zugriff dar.

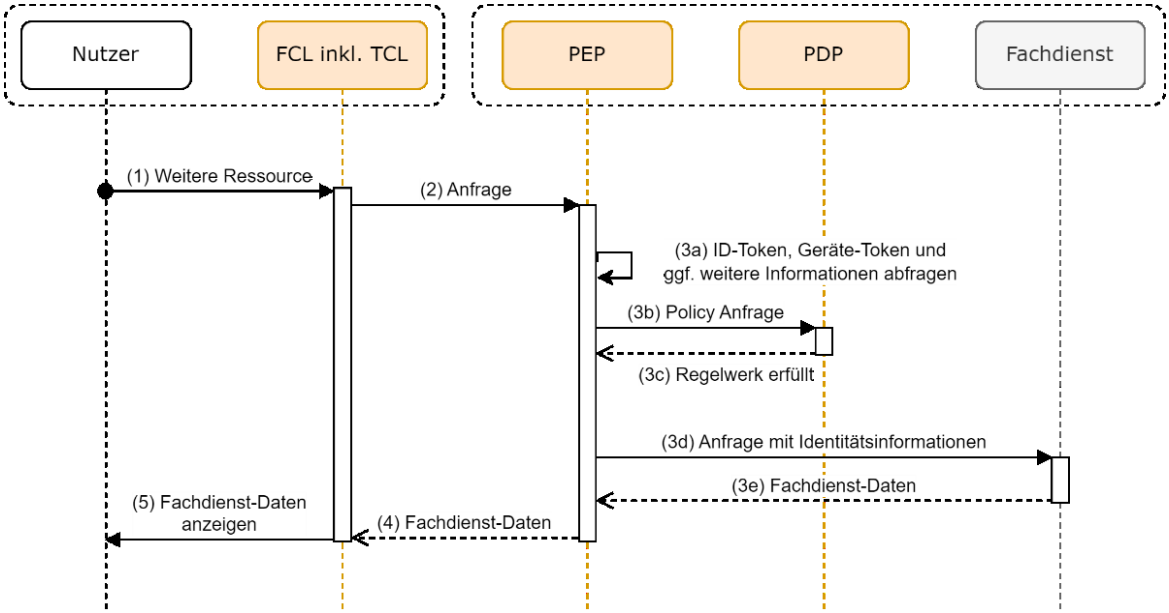

Abbildung 3.5.5-1: Erfolgreicher autorisierter Zugriff auf eine Ressource der ZTA

Der Zugriff in Abbildung 3.5.5-1 erfolgt durch Fach- und Trust-Client zum PEP (2), welcher in diesem Fall bereits alle notwendigen Informationen für den aktuellen Zugriff vorliegen hat oder selbständig abrufen kann (3a). Diese werden an den PDP zur Regelentscheidung übergeben (3b), welcher seine (in diesem Fall positive) Entscheidung dem PEP zurückgibt (3c). Anschließend reicht der PEP die Anfrage an den Fachdienst weiter (3d), zusammen mit den notwendigen Informationen, um den Nutzer identifizieren zu können. Nach einer Verarbeitung im Fachdienst, gibt der Fachdienst seine Antwort an den PEP zurück (3e), welcher sie an den Fach- und Trust-Client weiterreicht (4). Die Daten werden dann dem Nutzer angezeigt (5).

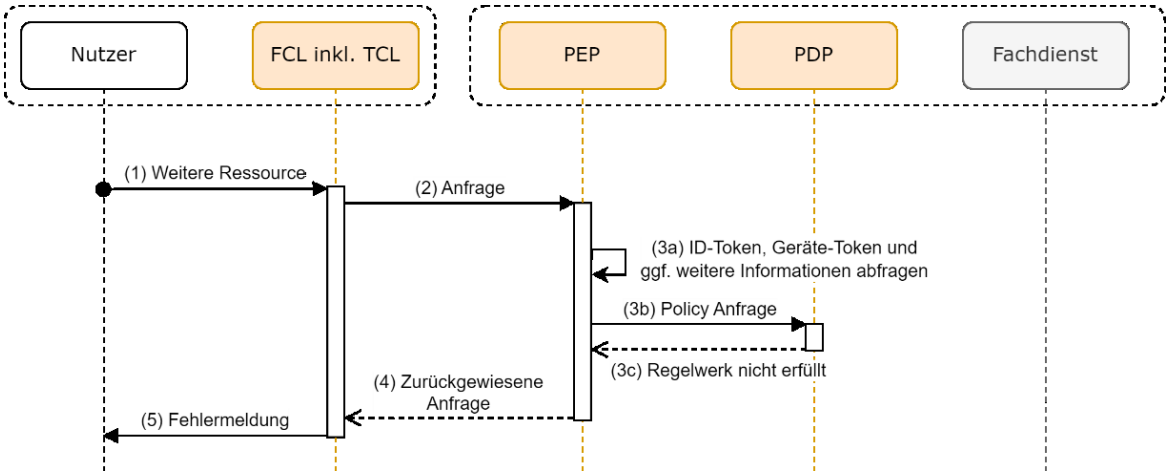

Abbildung 3.5.5-2: Zurückgewiesener Zugriff auf eine Ressource der ZTA

FA7.1 Session-Management

Die Zugriffe auf Ressourcen sollen auf Basis von Sessions erfolgen. Entsprechend der Definition in *3.1.3 Sessions,* bildet eine Session eine etablierte Kommunikationsbeziehung eines Clients mit einem Fachdienst. Mit der Session sind alle notwendigen Informationen

für den Zugriff assoziiert, um diese für zukünftige Anfragen innerhalb derselben Session wiederverwenden zu können.

Das Session-Management muss die Möglichkeit bieten, eine Session anzulegen und zu beenden, die Zugehörigkeit einer Anfrage zu einer Session zu prüfen und zugriffsrelevante Informationen mit den Sessions verknüpfen zu können (z.B. ID-Token, Geräte-Token).

Die Session wird durch einen für den Client nicht weiter interpretierbaren Wert identifiziert. Der Client speichert diesen Wert und sendet ihn bei allen nachfolgenden Anfragen mit. Der Server wiederum legt den Wert zusammen mit dazugehörigen Nachweisen (ID-Token, Geräte-Token etc.) in einer Datenbank ab.

Ein Spezialfall ist die Beendigung der Session, bei der der Server einen Backchannel-Logout gemäß der OIDC-Spezifikation ausführt, um dem IDP einen Logout zu signalisieren [OpenID-Connect_BCL]. Auch hier muss sicher OIDC-Client mittels mTLS gegenüber dem IDP authentisieren. Diese Spezifikation sieht ebenfalls vor, dass der IDP andere OIDC-Clients zu benachrichtigt. Alternativ wird der Logout bei der nächsten Verlängerung des ID-Tickets an den jeweiligen OIDC-Clients wirksam, da auch alle ID-Refresh-Token invalidiert werden. Um die aktiven Sessions für ein besseres Nutzererlebnis zu bündeln, ist es auch möglich ID-Token bei der Erstellung mit einen IDP-Spezifischen Session-Identifier zu versehen, der eine zusammenhängende Gruppe an Token identifiziert. Damit wäre beispielsweise umsetzbar, dass durch einen Logout alle Sessions auf dem gleichen Gerät geschlossen werden.

In Abbildung 3.5.5-3 sind diese drei zentralen Datenflüsse schematisch dargestellt.

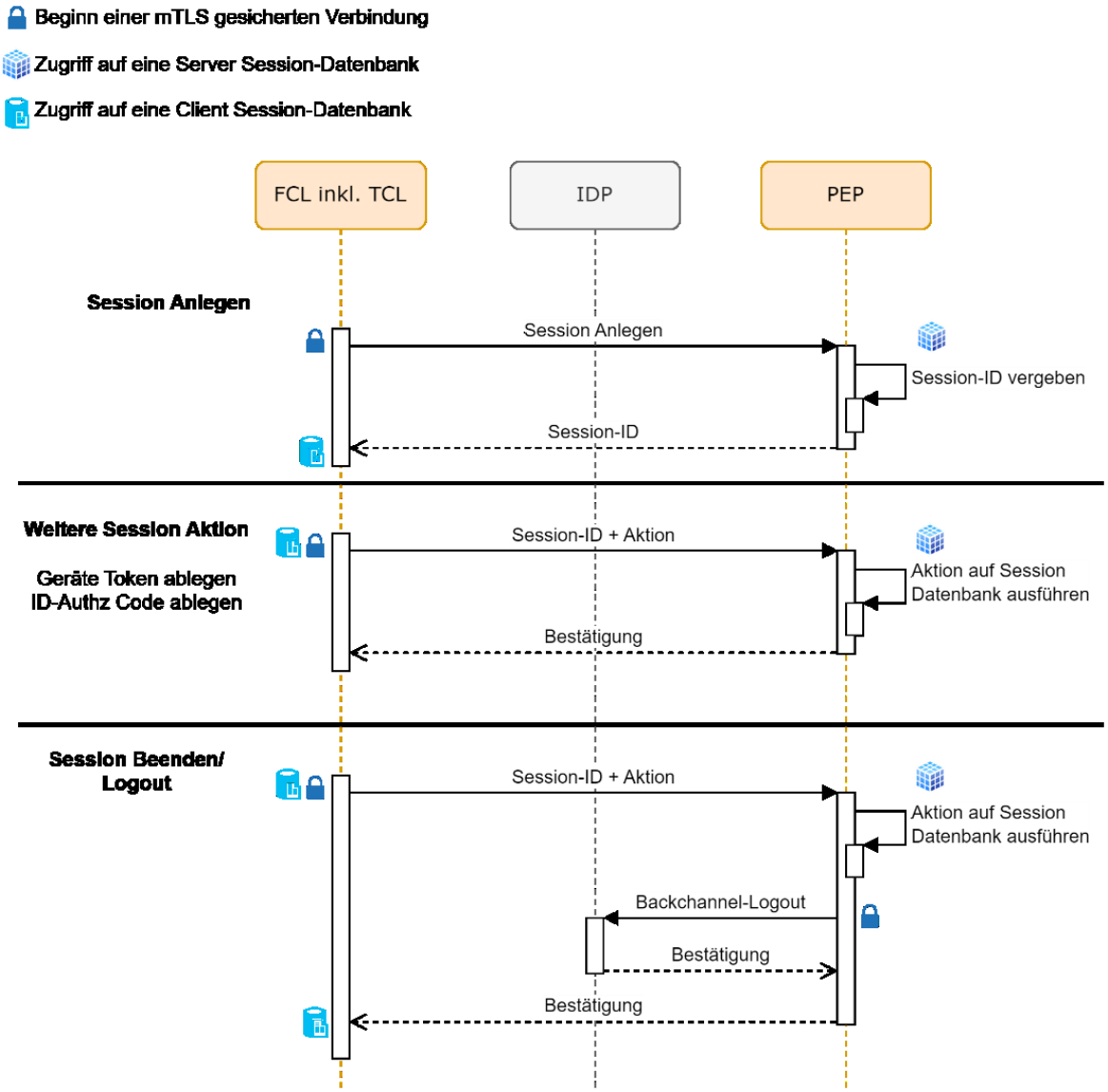

Abbildung 3.5.5-3: Kernfunktionalitäten des Sessionmanagement (Anlegen, Erweitern, Beenden)

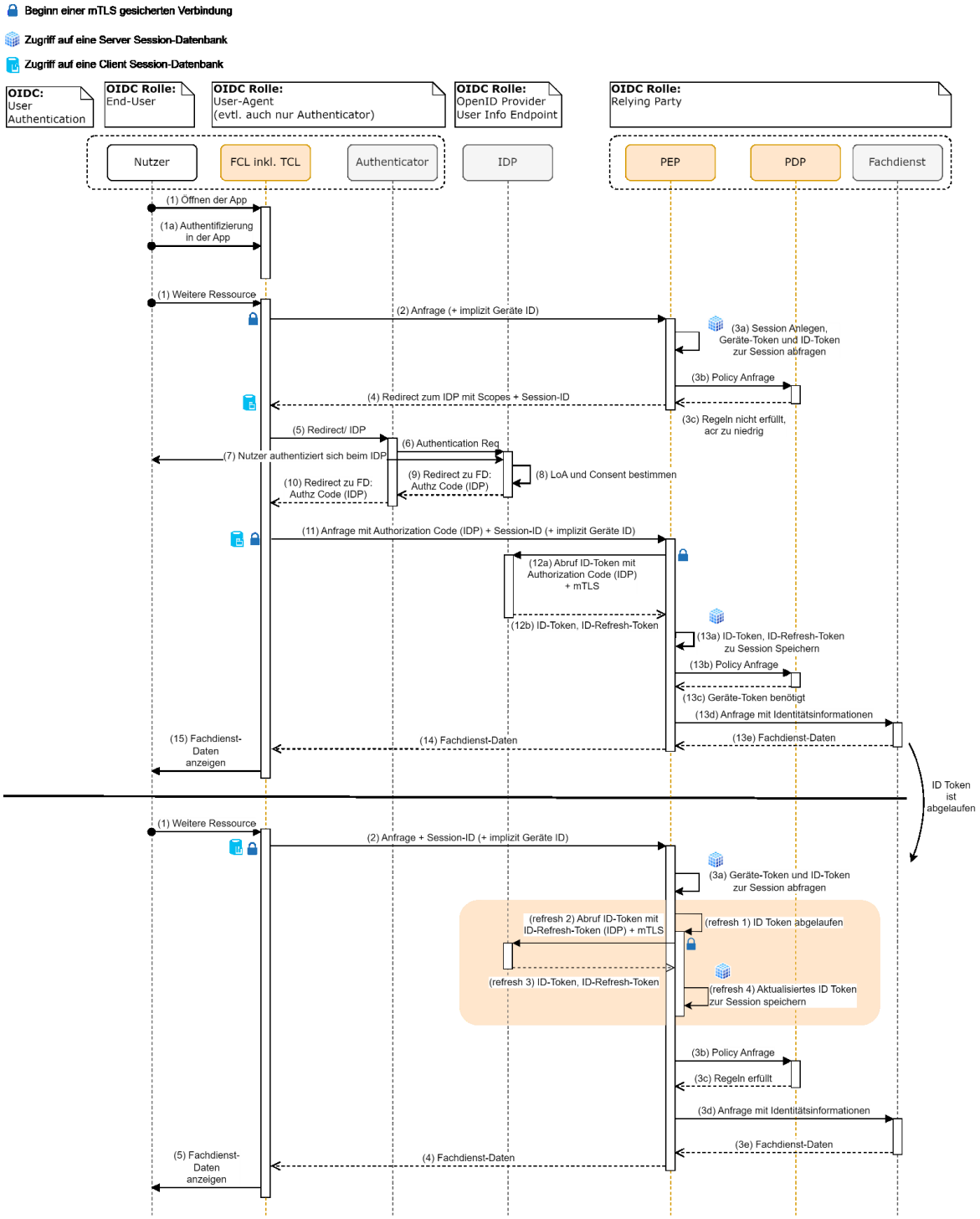

Abbildung 3.5.5-4: Initialer ID-Token Abruf, mit anschließendem Ablauf und Erneuern (Refresh)

Die Session-Aktionen können in der Implementierung mit anderen Aktionen kombiniert werden. So kann das Erstellen der Session sowohl implizit als Anfrage ohne gültige Session-ID, oder explizit an einen gesonderten Endpunkt gestaltet werden. Um DoS-Angriffen vorzubeugen ist es auch möglich die Session erst mit erfolgreichen Abrufen des ID-Token anzulegen. Gleiches gilt für die Beendigung der Session, die mit einem Abmelden am Fachdienst kombiniert werden kann.

Teil des Session-Management ist es auch zugriffsrelevante Informationen aktuell zu halten und ggf. zu aktualisieren.

In Abbildung 3.5.5-4 ist der Refresh des ID-Tokens abgebildet. Zuerst ist ein normaler Zugriff abgebildet, mit einem zusätzlichen Refresh-Token in Schritt 12b. Der Refresh des ID-Tokens ist dann im unteren Teil abgebildet (refresh 1-4). Die relevanten Teile des Diagramms sind orange hinterlegt. Für die Aktualisierung des ID-Token ist der PEP direkt zuständig. Dies erfolgt über das gespeicherte Refresh-Token, das wiederum durch den PEP beim IDP für ein neues ID-Token eingelöst werden kann (refresh 2). So ist der Fach/Trust-Client selbst nie involviert. Die Aktualisierung des Tokens kann auch proaktiv im Hintergrund und unabhängig von einer Zugriffsanfrage ausgeführt werden.

Alle notwendigen Verbindungen für den Refresh des ID-Token bestehen direkt zwischen PEP und IDP. Daher kann hier die Laufzeit des ID-Token deutlich reduziert werden, ohne zusätzliche Last bei möglicherweise ressourcenbeschränken Endgeräten zu generieren.

Die Aktualisierung des Geräte-Tokens erfolgt durch den Trust-Client, ohne dass der PEP direkt involviert ist. Daher ist diese Aktion außerhalb des Session-Management und für den PEP gleichzusetzen mit einem ersten Hinterlegen bzw. Ersetzen des Geräte-Tokens. Genauere Information dazu sind in Kapitel *3.5.2- Geräte der Nutzer authentisieren und attestierten* beschrieben.

FA7.2 StepUp-Autorisierung

Eine StepUp Autorisierung bietet die Möglichkeit innerhalb einer Session zusätzliche Berechtigungsnachweise anzufordern. Dieses Konzept beschränkt sich auf das Nachfordern von erweiterten Identitätsnachweisen. Ein Nachfordern von erweiterten Geräteinformationen innerhalb einer Session ist nicht vorgesehen, ließe sich aber ebenfalls implementieren. Das Level der geforderten Identitätsnachweise wird über das OIDC -Feld *acr_values* angefragt und durch den IDP im Feld *acr* bestätigt [OpenID-Connect].

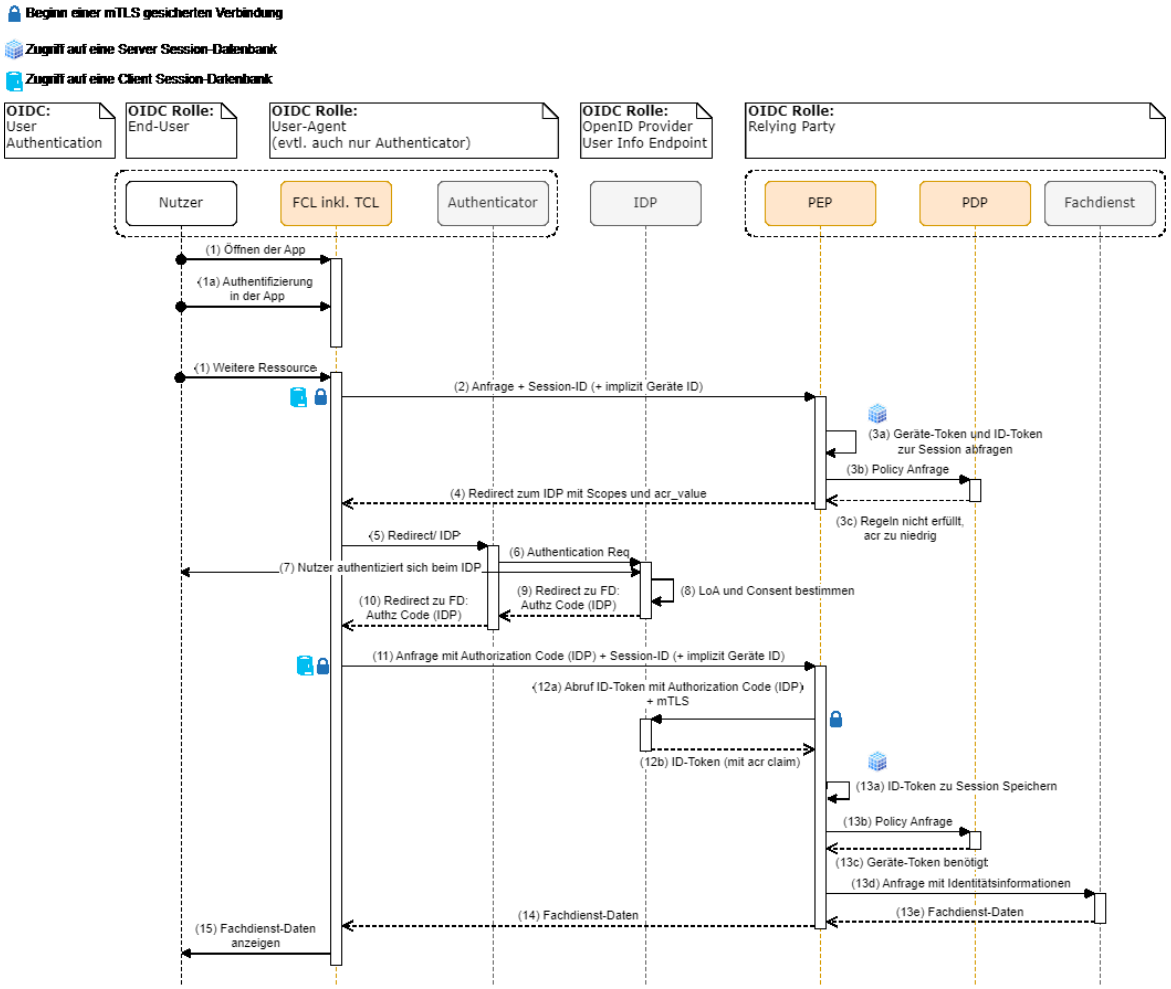

Abbildung 3.5.5-5: Step-Up Autorisierung des Nutzers

Im Rahmen der StepUp-Autorisierung soll auch einer möglichen kürzeren Lebensdauer von höher privilegierten Autorisierungen Rechnung getragen werden. So soll das höher privilegierte Token eine möglichst kurze Lebensdauer haben (NFA5, NFA6). Generell stehen zwei Möglichkeiten zur Verfügung, um dies zu implementieren.

Die beiden Optionen sind wie folgt:

- **Option 1, zwei unabhängige ID-Token:** Die beiden Autorisierungslevel werden in unterschiedlichen ID-Token abgebildet. Beide Token haben ihre eigenen unabhängigen Scopes, Laufzeiten und Refresh-Token.
- **Option 2, ein zusammenhängendes Token:** Die höher privilegierte Autorisierung ersetzt die niedrigere Autorisierung und nimmt dessen Scopes mit auf. Nach Ablauf einer gewissen Zeit, liefert der IDP für ein Refresh-Token lediglich das ID-Token mit niedrigerem Autorisierungslevel zurück.

Option 1 bietet eine höhere Flexibilität, bedeutet allerdings auch eine höhere Komplexität in der Implementierung.

Option 2 entspricht stärker dem OIDC-Flow, und kann durch Standard-Implementierungen abgedeckt werden. Es werden aber nicht verschiedene Scopes mit unterschiedlichen Autorisierungsleveln oder Laufzeiten gleichzeitig ermöglicht.

3.5.6 Monitoring von Komponenten

In der TI ist ein externes Echtzeit-Monitoring für alle verwendeten Komponenten vorgesehen (FA8 ZTA Komponenten überwachen). Extern bedeutet hier, dass das Monitoring nicht nur durch den Betreiber des Fachdienstes oder der ZTA-Komponente erfolgt, sondern auch von außerhalb der Domäne des Fachdienstes bzw. der ZTA-Komponente. In erster Linie soll dies der gematik eine Möglichkeit geben, den Betriebs- und Sicherheit-Status der TI-Komponenten im Blick zu behalten und bei Bedarf reagieren zu können. Es ist auch denkbar, dass konsolidierte Informationen aus dem Monitoring öffentlich zur Verfügung gestellt werden, um Transparenz zu schaffen, allen Nutzern einen Einblick in den aktuellen Zustand der TI zu geben und so das Vertrauen der Nutzer in die TI zu stärken. Das Betriebsmonitoring umfasst Informationen zu Verfügbarkeit, Auslastung und Performance der Komponenten. Das Security-Monitoring bewertet die Sicherheit der TI-Komponenten z.B. auf Basis von Attestierungsinformationen über die aktuell verwendete Software und Betriebsumgebung, testet die Anfälligkeit für Angriffe und prüft die Betroffenheit der Komponenten von neu veröffentlichen CVEs.

Zur Unterstützung dieses Monitorings bieten alle Komponenten der TI (z.B. IDPs, GMSe, PEPs/PDPs/Fachdienste, ...) eine Schnittstelle an, über die eine Monitoring-Komponente in Echtzeit Informationen über den aktuellen Betriebs- und Sicherheitsstatus abrufen kann. Die Informationen werden durch das Monitoring ausgewertet und den gewünschten Nutzern/Empfängern zur Verfügung gestellt (z.B. über ein entsprechendes Dashboard). Die zugehörigen Governance-Prozesse sind in Kapitel 5.2.3 beschrieben. Informationen über den Sicherheitszustand der Komponenten und mögliche Sicherheitsvorfälle werden zudem an das SIEM/SOC weitergegeben und ggf. für andere betriebliche Zwecke genutzt.

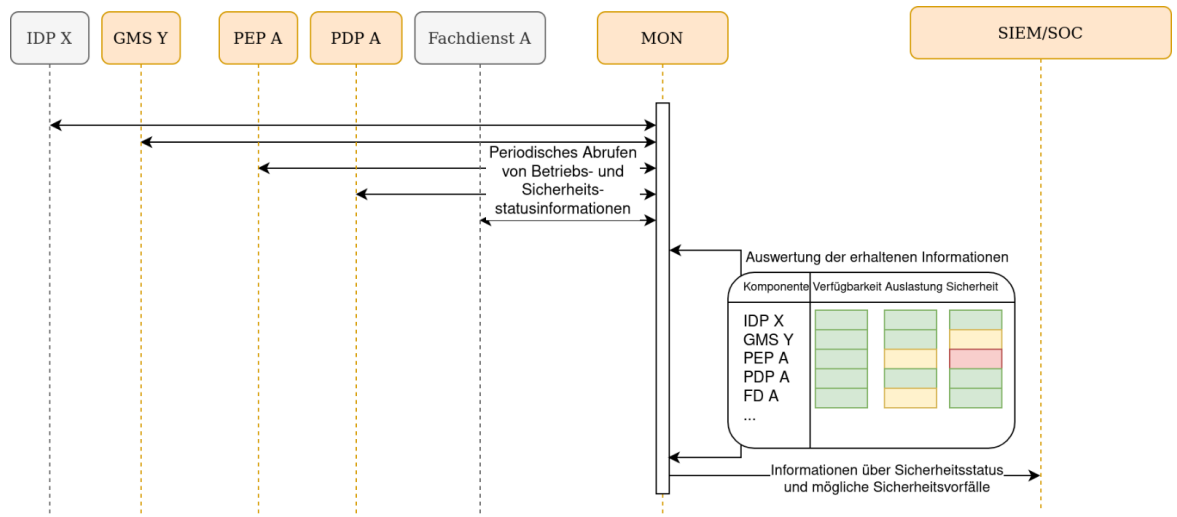

Abbildung 3.5.6-1: Datenflüsse für Betriebs- und Security-Monitoring von Komponenten

3.5.7 Zusammenfassender Datenfluss

Die bereits in Abbildung 3.5-1 in der Einleitung skizzierten Datenflüsse werden im folgenden Daten- und Kommunikationsfluss-Diagramm in Abbildung 3.5.7-1 für die zusammenfassende Darstellung von FA0 detaillierter betrachtet und erläutert. Dabei bezieht sich das Diagramm jeweils auf die bereits in den vorangegangenen Unterkapiteln eingeführten Datenflüsse. Es wird zwischen Erstzugriff und allen weiteren Zugriffen unterschieden. Der Erstzugriff beschreibt den Datenfluss, wenn bereits alle Vorbedingungen erfüllt sind, aber noch keine aktuell gültigen Nachweise abgefragt wurden

(ID-Token, Geräte-Token). Die weiteren Zugriffe beschreiben alle Zugriffe, bei denen bereits vorher alle erforderlichen Nachweise hinterlegt wurden, und auch weiterhin gültig sind.

Vorbedingungen

Der in Abbildung 3.5.7-1 dargestellte Datenfluss hat folgende Vorbedingungen:

- Der Nutzer ist bereits beim IDP registriert (Kapitel 3.5.2 - FA2.3)
- Der Nutzer ist beim Fachdienst registriert und hat entsprechende Berechtigungen, um auf den Fachdienst zuzugreifen. (Kapitel 3.5.1 - FA2.2)
- Das verwendete Gerät ist bereits beim GMS registriert. (Kapitel 3.5.2 - FA5.1)
- Die Anforderungen des Regelwerks an Nutzer, Endgerät und Umgebung sind erfüllt (Kapitel 3.5.4 FA1).

Erstzugriff

1. Der Nutzer öffnet den Fach-/Trust-Client (bspw. ePA-FdV).
   a. Eine App-Interne lokale Authentifizierung ist als zusätzliche Sicherung möglich.
2. Der Fach-/Trust-Client generiert einen PKCE Code Verifier und Code Challenge.
   a. Mit diesem Datenfluss beginnt ein OAuth2 Authorization Code Flow mit PKCE.
3. Der Fach-/Trust-Client sendet einen OAuth2 Authorization Request (mit PKCE Challenge) an den Authorization Endpoint des GMS
   a. Die Verbindung wird per mTLS gesichert, welche die Geräte-ID im Client Zertifikat beinhaltet.
   b. Durch interne Mechanismen bestimmt der Fach-/Trust-Client die notwendigen Scopes für das angefragte Geräte-Token.
4. Der Fach-/Trust-Client kooperiert mit dem GMS zur sicheren Feststellung der Geräteeigenschaften (Remote Attestation, ID etc.) Der genaue Ablauf ist plattform- und gerätespezifisch.
   a. Das GMS sendet eine initiale Anfrage mit den zu bestätigenden Daten an den Fach-/Trust-Client
   b. Der Fach-/Trust-Client startet die Attestierung unter Verwendung des Vertrauensankers des Endgerätes.
   c. Der Vertrauensanker liefert die kryptographisch gesicherten Messungen an den Fach-/Trust-Client.
   d. Der Fach-/Trust-Client reicht die Messungen des Vertrauensankers mit ggf. weiteren Anreicherungen an das GMS zurück.
5. Das GMS verarbeitet die gesammelten Daten des anfragenden Geräts
   a. Das GMS such nach assoziierten Nutzern für das Gerät. Diese Assoziation muss bei der Registrierung hinterlegt sein und kann durch die Geräte-ID gefunden werden.
   b. Die Überprüfung, dass tatsächlich der korrekte Nutzer das Gerät verwenden möchte, erfolgt erst in Schritt 17.
   c. Das GMS kodiert die Messergebnisse des Geräts in ein standardisiertes Token-Format.
6. Das GMS gibt den Authorization Code an Fach-/Trust-Client zurück.

7. Der Fach-/Trust-Client ruft das Geräte-Token mit Authorization Code und PKCE Code Verifier am Token Endpoint des GMS ab.
8. Das GMS stellt das OAuth2 Geräte-Token an den Fach-/Trust-Client aus.
   a. Das Token enthält:
      - Eine Assoziation der Geräte-ID zum Nutzer.
      - Eine Liste aller durchgeführten Messungen und deren Ergebnissen in einem standardisierten Format.
9. Der Nutzer startet die Anfrage für eine Ressource (z.B. durch Klick auf "*Blutdruckprotokoll abrufen*" aus der Favoritenliste).
10. Der Fach-/Trust-Client initiiert einen Zugriff auf die Ressource am Fachdienst.
11. Der PEP/PDP prüft den Zugriff und fragt für fehlende Nachweise nach weiteren Informationen.
    a. Der PEP prüft ob bereits eine Session zwischen Fach-/Trust-Client und PEP besteht.
       - Aktuell ist keine Session hinterlegt, daher legt der PEP serverseitig eine Session an.
    b. Der PEP lässt die Anfrage gegen das Regelwerk durch den PDP prüfen.
    c. Der PDP lehnt den Zugriff mit dem Verweis auf ein fehlendes ID-Token ab.
12. Der PEP verweist den Fach-/Trust-Client an den IDP und startet eine OIDC Authentication Sequence.
    a. Hier beginnt die OIDC Authentication Sequence nach OIDC-Standard mit dem Fachdienst als OIDC-Client. Dementsprechend ist auch der Fachdienst in der Anfrage hinterlegt.
    b. Entsprechend dem OIDC-Standard setzt der PEP die notwendigen Scopes in der Anfrage.
    c. In der Antwort ist ebenfalls die clientseitige Session-Information erhalten, mit der sich der Fach-/Trust-Client später identifizieren kann. Diese Session ist an die Geräte ID durch das mTLS-Zertifikat des Clients gebunden.
    d. Ebenfalls in der Antwort enthalten ist die Session-ID, die der Fach-/Trust-Client speichert.
13. Der Fach-/Trust-Client leitet die OIDC Authentication an den IDP weiter.
14. Der Fach-/Trust-Client erhält vom IDP einen OIDC Authorization Code.
15. Der Fach-/Trust-Client erneuert seine Anfrage auf die Ressource und übergibt damit auch den Authorization Code an den PEP.
    a. Die Anfrage enthält die Gerät-ID (implizit durch die mTLS-Verbindung), das Geräte-Token, den Authorization Code und die clientseitige Session Information.
16. Der PEP ruft mit Authorization Code und OIDC Client Secret das ID-Token und ID-Refresh-Token am IDP ab.
17. Der PEP/PDP überprüft auf Basis der Session (jetzt mit ID- und Geräte-Token) ob der Zugriff erlaubt ist und erlaubt dem Client Zugriff auf die Ressource.
    a. ID-Token und Geräte-Token werden serverseitig in Verbindung zur Session gespeichert. Dabei wird zuerst überprüft, ob die gelieferten Nachweise

entsprechend miteinander verwendet werden dürfen (mTLS Binding, Geräte-Nutzer-Bindung).

b. Der PEP lässt die Anfrage durch den PDP prüfen.

c. Der PDP meldet dem PEP zurück, dass alle Regeln erfüllt sind.

18. Der PEP gibt die Anfrage an den Fachdienst weiter.

    a. Der PEP leitet die Anfrage des Clients mit Nutzerinformationen an den Fachdienst weiter.

    b. Der PEP erhält als Antwort die angefragte Ressource des Nutzers von dem Fachdienst.

19. Der PEP gibt die erhaltenen Fachdienst-Daten zurück an den Fach-/Trust-Client.

20. Der Fach-/Trust-Client zeigt die Fachdienst-Daten dem Nutzer an.

Weitere Zugriffe

Die weiteren Schritte (wiederbeginnend bei 1) in Abbildung 3.5.7-1, beschreiben den Zugriff nach dem bereits initial erfolgten Nachweis der Berechtigungen.

1. Der Nutzer wählt eine weitere Ressource aus.
2. Der Fach-/Trust-Client ruft die Ressource am Fachdienst ab.
3. Der PEP prüft den Zugriff auf Basis der hinterlegten Nachweise und leitet die Anfrage an den Fachdienst weiter.
4. Der PEP gibt die erhaltenen Fachdienst-Daten zurück an den Fach-/Trust-Client.
5. Der Fach-/Trust-Client zeigt die Fachdienst-Daten dem Nutzer an.

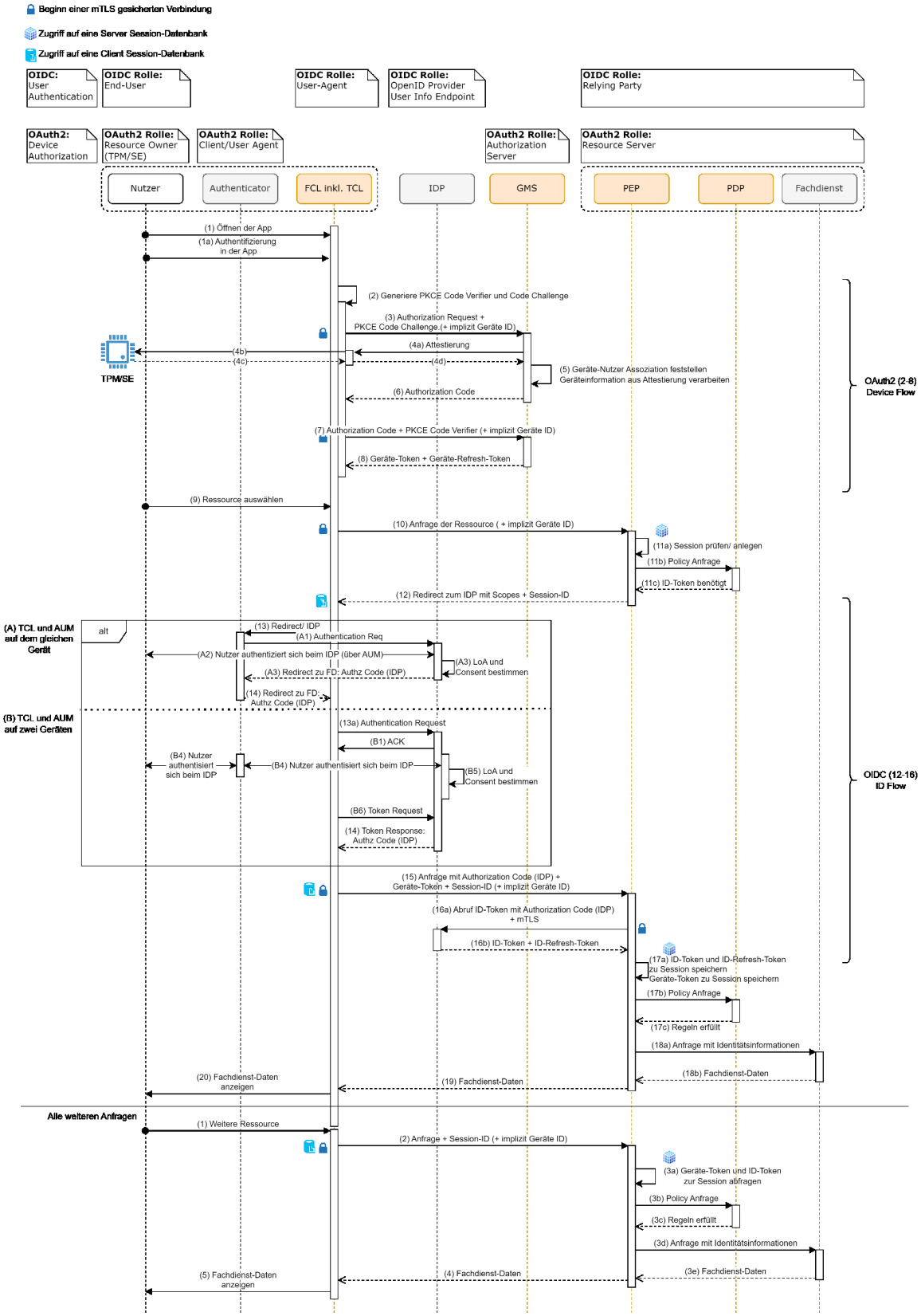

Abbildung 3.5.7-1: Detaillierte Übersicht des Daten- und Kommunikationsflusses

4 Technische Umsetzung

Ziel der technischen Umsetzung ist eine verfügbare und performante Architektur mit hoher Zukunftsfähigkeit und passend zu den hohen Datenschutz- und Sicherheitsanforderungen im Gesundheitswesen. Im Folgenden werden in 4.1 die allgemeinen Konzepte vorgestellt und wie diese in den Komponenten der ZTA prinzipiell technisch umgesetzt werden. In 4.2 wird dies für die fachdienstübergreifenden Komponenten detailliert und in 4.3 für die fachdienstnahen. Schließlich werden in 4.4 Komponenten betrachtet, die nicht selbst Bestandteil der Architektur, aber für die Funktion essenziell sind.

4.1 Komponentenübergreifende Umsetzungskonzepte

4.1.1 Verfügbarkeit und Robustheit

Eine hohe Verfügbarkeit und Robustheit werden erreicht, indem die Kopplung zwischen den Komponenten weitgehend lose ist, die Komponenten redundant ausgelegt sind und die Kommunikation weitgehend statuslos erfolgt. Damit schränkt ein temporärer Ausfall von Instanzen einzelner Komponenten nicht die Verfügbarkeit anderer Komponenten bzw. gesamter Anwendungen ein. Fachdienstübergreifend essenzielle Komponenten (4.2) werden hochverfügbar betrieben. Eine Skalierbarkeit auf viele Teilnehmer bei wechselnden Belastungen wird erreicht durch ein dynamisches Up- und Downscaling der involvierten Komponenten. Die Kommunikation erfolgt über die gut ausgebaute und robuste Infrastruktur öffentlicher Netze, wobei die Komponenten redundant über verschiedene Anbieter angebunden werden sollten. Netzwerkverbindungen und Bandbreiten in der Gesamtarchitektur sind damit nicht mehr von einzelnen Netzanbietern abhängig.

Zur Erhöhung der Robustheit und Reduktion der Abhängigkeit von spezifischen Instanzen der Komponenten bzgl. Verfügbarkeit und Skalierbarkeit ist die Kommunikation so weit wie möglich zustandslos. D.h. es ist irrelevant, zu welcher spezifischen Instanz einer redundant ausgelegten Komponente Clients sich verbinden. Bei der Kommunikation zwischen Fach-Client und Fachdienst kann es jedoch eine Session-Affinität geben. In diesem Fall muss der Fachdienst dies berücksichtigen, indem er hohe Verfügbarkeit und den Transfer von Sessions zwischen Instanzen bereitstellt.

Die Komponenten der ZTA sind, wie auch der IDP, direkt bzw. indirekt aus dem Internet erreichbar. Entsprechend ist für eine hohe Verfügbarkeit ein Schutz vor Denial-of-Service Angriffen nötig. Details dazu sind nicht im Scope des Feinkonzepts, aber es ist sowohl ein individueller Schutz der einzelnen Komponenten als auch ein komponenten-übergreifender Schutz denkbar. Dieser würde geschützte Netzbereiche schaffen, in denen die TI-Komponenten betrieben werden und über verteilte Zugänge erreichbar wären. Es wird davon ausgegangen, dass über den implementierten Mechanismus eine transparente Lastverteilung (d.h. ohne TLS-Terminierung) auf die Komponenten der ZTA vorgenommen werden kann und dass die originale Quell-IP-Adresse des Clients bzw. davon abgeleitete Informationen wie Geolokalisation für die Zugriffskontrolle innerhalb der ZTA verfügbar gemacht werden.

4.1.2 Performance und Skalierung

Eine hohe Performance wird durch kurze Kommunikationswege und optimierte Kommunikationsflüsse begünstigt. Der bzgl. Performance und Verfügbarkeit kritische Pfad

beim Zugriff auf einen Fachdienst ist der Datenaustausch über den TCL zum PEP, die dortige Zugriffskontrolle über den PDP und die Weiterleitung der Daten zum Fachdient. Dazu kommen beim Start des Fach-Clients bzw. bei der Step-Up-Authentisierung die Kommunikationen mit IDP und GMS, welche aber von Nutzerinteraktion (Start der Anwendung, Eingabe von Zugangsdaten) begleitet und dadurch weniger zeitkritisch sind. Die weitere kontinuierliche Verifikation von Nutzer und Gerät erfolgt durch Refresh der ID- und Geräte-Token frühzeitig vor Ablauf der Token im Hintergrund und ist daher unkritisch. Um die Last auf IDP und GMS zu reduzieren, sollte die Lebenszeit der Token so gewählt sein, dass die Sicherheitsanforderungen ausreichend erfüllt sind, aber keine unnötigen Refreshs erfolgen.

4.1.3 Sicherheit und Datenschutz

Robuste Sicherheit, Datenschutz und Privacy sind bereits im Design der Komponenten und deren Interaktion berücksichtigt. Gesundheitsdaten werden niemals im Klartext über vertrauensunwürdige Infrastrukturen (insb. Internet) übertragen. Zur Gewährleistung von Sicherheit und Privacy wird die Kommunikation zwischen den Komponenten mit TLS 1.3 gesichert. Damit sind sowohl Server- als auch Clientzertifikate verschlüsselt. Bei der Validierung von Server-Zertifikaten sollte OCSP-Stapling verbindlich sein, da es die Performance erhöht und die Abhängigkeit von der Verfügbarkeit von OCSP-Respondern reduziert. Die Nutzung von ECC statt RSA in Keys und Zertifikaten wird für eine bessere Performance empfohlen, da ECC sowohl weniger Rechenaufwand benötigt als auch weniger Daten übertragen muss. Das Konzept ist jedoch nicht spezifisch auf bestimmte Algorithmen festgelegt. Durch diese Krypto-Agilität ist es auch auf zukünftige Verfahren wie Post-Quantum-Kryptographie vorbereitet. Zur weiteren Erhöhung der Privacy wird insbesondere für die Kommunikation vom Fach-Client/TCL zum Fachdienst/PEP die Nutzung von ECH (Encrypted Client Hello) als sinnvoll erachtet.

4.1.4 Zukunftsfähigkeit

Zur Erreichung der gewünschten Zukunftsfähigkeit und Herstellerunabhängigkeit sollten so weit wie möglich etablierte offene Standards und Technologien zur Umsetzung genutzt werden. Wenn entsprechende Standards nicht existieren, sollten diese herstellerunabhängig spezifiziert werden.

4.2 Fachdienstübergreifende Komponenten

In diesem Kapitel werden die fachdienstübergreifenden Komponenten der ZTA beschrieben, welche in der Verantwortung der gematik liegen, aber nicht notwendigerweise von dieser auch betrieben werden. Verfügbarkeitsprobleme bei diesen Komponenten führen zu übergreifenden Störungen für alle Clients und Fachdienste. Entsprechend robust und hochverfügbar müssen diese Komponenten konzipiert, umgesetzt und betrieben werden.

4.2.1 GMS

Das GMS übernimmt die Registrierung von Geräten, die Verifikation der Gerätebindung sowie den Erhalt, die Normalisierung und ggf. partielle Konsolidierung von Informationen zum aktuellen Gerätestatus. Dabei kommuniziert es mit dem TCL auf dem zugreifenden Gerät sowie mit dem Nutzerportal zur Bereitstellung von Informationen über die Geräte eines Nutzers und zur Entfernung von Geräten eines Nutzers. Das GMS speichert einen

Fingerprint zu jedem registrierten Gerät, um die Registrierung zu überprüfen. Zur Information im Nutzerportal sowie ggf. zur Einbeziehung in verhaltensbasierte Regeln werden beschränkte Informationen zur Nutzung des Gerätes gespeichert, d.h. Zeitpunkt und Location der letzten Zugriffe auf das GMS sowie die letzten Geräte- und Attestierungsinformationen.

Die Verfügbarkeit des GMS ist essenziell für den Zugriff auf Fachdienste, da bei der Zugriffskontrolle ein gültiger Geräte-Token passend zum Client-Zertifikat des TCL erforderlich ist und der im Token transportierte Gerätestatus für die Zugriffsentscheidung im PDP benötigt wird. Entsprechend sind sehr hohe Verfügbarkeit, Performance und Robustheit dieser Komponente wichtig.

Dies wird durch einen Parallelbetrieb mehrerer GMS-Instanzen erreicht, wie in Abbildung 4.2.1-1 dargestellt. Diese Instanzen werden innerhalb von GMS-Clustern zusammengefasst, in denen ein elastisches Scaling über einen vorgeschalteten Loadbalancer auf Netz- bzw. Transportebene mit Lastmonitoring erfolgt. Dazu kommt eine Verteilung der GMS-Cluster für Betriebsredundanz über verschiedene Standorte hinweg sowie eine Anbindung über verschiedene Netzbetreiber. Clients werden auf die entsprechenden Cluster durch ein Loadbalancing innerhalb der DDoS-Protection verteilt. Damit GMS-Cluster z.B. für Wartungsarbeiten temporär abgeschaltet werden können, wird die aktuelle Verfügbarkeit in diesem Loadbalancing mitberücksichtigt. Zusätzlich dazu kann der Verbindungsaufbau im TCL robust gestaltet werden, indem beim DNS-Lookup für das GMS mehrere IP-Adressen geliefert werden und der TCL diese der Reihe nach durchprobiert, bis eine erfolgreiche Verbindung aufgebaut werden konnte. Dies verursacht im Normalfall keine zusätzliche Verzögerung, da der Verbindungsaufbau mit der ersten IP funktionieren sollte. Für den Ausnahmefall scheitert dagegen die Verbindung zur TI nicht komplett, sondern erfolgt nur verzögert.

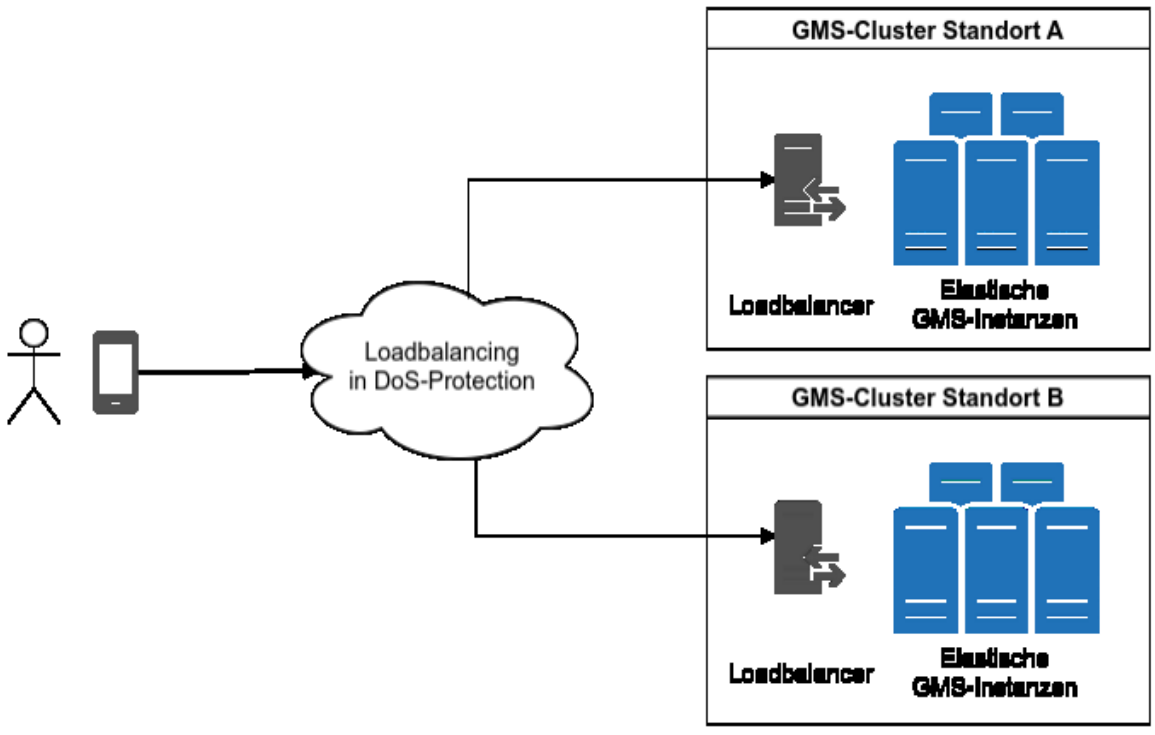

Abbildung 4.2.1-1 Skalierbarkeit des GMS

Der Parallelbetrieb der GMS-Instanzen wird unterstützt durch die Zustandslosigkeit der Anfragen, d.h. es ist egal, mit welcher GMS-Instanz ein TCL kommuniziert. Es muss im Backend des GMS jedoch eine Datenbank existieren, welche die Informationen zur Geräteregistrierung und Nutzungshistorie verwaltet. Die Anforderungen an diese Datenbank bzgl. Performance, Verfügbarkeit und niedrige Latenz bei der standortübergreifenden Synchronisation sind entsprechend hoch, die Anforderungen an Transaktionskonsistenz und Feature-Reichtum jedoch gering. Neben klassischen SQL-Datenbanken mit starken Garantien (ACID), kann hier auch über abgeschwächte Modelle wie etwa neuartige Datenbanken auf Basis von Eventual Consistency Eigenschaften nachgedacht werden, da mit diesen technologiebedingt eine besser skalierbarere Verteilung und dadurch höhere Performance und Robustheit erreicht werden kann. Um die bei der Verwendung personenbezogener Daten hohen Anforderungen an den Datenschutz zu reduzieren, kann bei der Speicherung neben der Verschlüsselung eine Pseudonymisierung des mit einem Gerät zusammengehörigen Nutzers ohne Verlust an Funktionalität vorgenommen werden.

Das GMS bekommt vom TCL Informationen zum Status des Gerätes. Umfang und Format dieser Informationen sind plattformspezifisch, d.h. unterscheiden sich zwischen verschiedenen Betriebssystemen und evtl. auch Betriebssystemversionen, installierter Software und verbauten Hardwarekomponenten. Für eine bessere Verarbeitbarkeit durch den PDP transformiert das GMS die Informationen daher in eine plattformübergreifende Form, evtl. bei teilweiser Konsolidierung von Informationen. Die konkrete Umsetzung hängt dabei davon ab, welche Informationen dem Regelwerk zur Verfügung gestellt werden sollen, d.h. eine Detaillierung kann erst nach einer detaillierten Analyse der mit dem Regelwerk abzubildenden Sicherheitsanforderungen erfolgen. Der Transport der Informationen zum PDP via PEP erfolgt über den Geräte-Token, sodass prinzipiell eine Umsetzung als separate Attribute in diesem denkbar ist und auch die zukünftige Erweiterbarkeit um weitere Informationen ermöglicht. Details zu den Attributen und deren Inhalt sind zwischen GMS und TCL abzustimmen. Interoperabilität zwischen GMS und diversen TCL-Implementationen hinsichtlich der gelieferten Daten sowie eine kontinuierliche Weiterentwicklung entsprechend Änderungen an den Geräteplattformen sind zu gewährleisten. Die Erfassung, Normalisierung und Konsolidierung der Gerätedaten im GMS erfolgt über statische Algorithmen und Parameter, die nur bei neuen GMS-Versionen aktualisiert werden. Das aus dem PAP gelieferte dynamische Regelwerk mit den dynamischen Attributen aus dem PIP gibt es nur bei den PDP, nicht im GMS.

Um eine Konsolidierung von gerätespezifischen Sicherheitseigenschaften in der heterogenen Welt der heutigen Endgeräte zu ermöglichen und aktuell zu halten, ist eine Einbeziehung von darauf spezialisierten Dienstleistern vermutlich erforderlich. Derartige Dienstleister pflegen Informationen zu einer großen Vielfalt an Endgeräten, wie z.B. die integrierten Sicherheitsfunktionen, bekannte Schwachstellen, nötige Software-Patches etc. Mit Hilfe dieser Informationen können die relevanten Sicherheitseigenschaften des Gerätes in einige wenige und geräteübergreifend vergleichbare "Scores" umgewandelt werden, die dann statt der komplexen Rohdaten im Regelwerk benutzt werden. Das Auslagern einer solchen Dienstleistung ist insbesondere deshalb interessant, da Informationen über die Sicherheit genutzter Endgeräte für eine breite Masse an Digitalisierungsvorhaben in verschiedenen Sektoren (z.B. Gesundheit, Finanzen, Dienste im Rahmen des Onlinezugangsgesetzes, Europäische digitale Identität) und regional übergreifend relevant sind. Die für einen solchen Dienst notwendigen Aufwände können sich somit entsprechend breit verteilen. Mögliche Ansätze für solche Dienste sind bereits in der Forschung sichtbar, z.B.: Android Device Security Database.

Die Standardisierung der plattformübergreifenden Sicherheit der Endgeräte von Verbrauchern steht noch am Anfang. Erste Standards (z.B. Consumer Mobile Device Protection Profile) sind aber bereits sichtbar. Mit Blick auf die kommenden, verpflichtenden Anforderungen an die Endgerätesicherheit für in Europa in Umlauf gebrachten Produkte

durch den Cyber Resilience Act (CRA) ist zu erwarten, dass solche Standards in den kommenden Jahren stark vorangetrieben werden. Der aktuelle Mangel an herstellerübergreifenden Standards für die Erfassung der Daten, für den Austausch zwischen Gerät und Gerätemanagement sowie für die Normalisierung der Gerätedaten zur Weiterverarbeitung im PDP müssen durch eine enge Zusammenarbeit zwischen den Herstellern von TCL und GMS sowie mit der Regelerstellung im PAP adressiert werden. Eine übergreifend gepflegte Open Source Implementierung für Normalisierung und Konsolidierung kann hier die Komplexität signifikant reduzieren.

4.2.2 PIP

Die PIP stellen den PDP Informationen für Zugriffsentscheidungen bereit. Sie können aus verschiedenen Quellen befüllt werden, manuell kuratiert, z.B. mit einer Liste sicherer Betriebssysteme, oder automatisch gefüllt aus dem Monitoring oder anderer Threat-Intelligence, z.B. mit auffälligen IP-Adressen. Entsprechend unterschiedlich sind die in den PIP gespeicherten Datenmengen und ihre Aktualisierungsfrequenz, was Implikationen für die Umsetzung und Anbindung an die PDP hat.

Die im PIP gespeicherten Informationen unterliegen einer einfachen Key-Value Struktur, wobei die Values durchaus strukturierte Informationen sein können. Beispiele sind Information zur Aktualität von Betriebssystemen (Key: Betriebssystem, Value: Liste der als aktuell akzeptierte Patch-Level) oder die Klassifikation bzw. Reputation von IP-Adressen (Key: IP-Adresse, Value: Klasse bzw. Reputation). Ein Zugriff erfolgt aus den Regeln im PDP über den entsprechenden Key. Verglichen mit der Menge an Lesezugriffen finden nur wenige Schreibzugriffe statt. Es ist in vielen Fällen nicht notwendig, dass Updates in Echtzeit propagiert werden. Dies erlaubt ein extensives Caching auf Seiten des PDP und damit performantere da lokale Zugriffe auf die Informationen während der Regelauswertung. Bei größeren Datenmengen ist eine verteilte redundante Datenhaltung möglich, bei der Informationen schrittweise und ggf. selektiv propagiert werden, um mit hoher Performance aus dem PDP erreichbar zu sein. Die verteilte Datenhaltung trägt auch zur Robustheit bei, da eine temporäre Nichtverfügbarkeit einzelner PIP-Instanzen tolerabel wird.

Die im PIP vorliegenden Daten werden als "Daten für die Zugriffssteuerung ohne Personenbezug" (DS3 entsprechend 3.4.1) eingestuft. Entsprechend muss der Zugriff auf die Daten durch die PDP gesichert sein. Typischerweise geschieht dies durch eine Authentisierung des PDP am PIP beim Abruf der Daten. Alternativ könnten die Daten im PIP selbst verschlüsselt sein, sodass nur den Schlüssel besitzende PDP die Daten entschlüsseln können. Ausnahmen beim Schutz der Daten können getroffen werden, wenn die Daten nicht sensitiv sind (z.B. öffentlich) und der Schutz einem performanten Abruf signifikant abträglich wäre.

Entsprechend den unterschiedlichen Anforderungen an PIP kann man diese auf verschiedene Weise umsetzen. Aus dem Spektrum der Möglichkeiten werden hier zwei charakteristische Fälle bzgl. Aktualisierungsfrequenz und Datenmenge der Daten aufgezeigt. Dabei wird nicht weiter beschrieben, wie eine authentisierte Aktualisierung der Daten in den PIP erfolgt. Passende und etablierte Mechanismen existieren für alle im Folgenden beschriebenen Technologien.

Seltene Aktualisierungen, wenige Daten

Diese Ausprägung ist charakteristisch für manuell kuratierte Daten wie z.B. die Liste sicherer Betriebssysteme oder die Liste unerwünschter Programme (z.B. für Jail-Breaking). Diese Informationen werden selten aktualisiert und haben eine entsprechend hohe Lebensdauer. Die geringe Datenmenge ist eine ideale Voraussetzung für ein Caching in den PDP, wodurch ein temporärer Ausfall des PIP kein wirkliches Problem darstellt und somit geringe Anforderungen an den Betrieb des PIP gestellt werden können. Eine

Umsetzung als Webserver, welche Inhalte mit passenden Cache-Control Informationen gemäß etablierter Standards ausliefern und den PDP darüber auch die Lebensdauer der Informationen und benötigte Aktualisierungszeitpunkte mitteilen, reicht hier aus.

Hohe Aktualisierungsfrequenz, viele Daten

Diese Ausprägung beschreibt z.B. den Fall von IP-basierten Reputationslisten. Die Nutzung eines solchen PIP dient der Einbeziehung der Reputation in eine komplexere Zugriffsentscheidung im PDP. Das direkte Blockieren von klar bösartigen IP-Adressen wird hingegen nicht im PDP, sondern im vorgelagerten DoS-Schutz gesehen. Der PIP für die IP-basierte Reputation wird durch das Monitoring oder andere Threat-Intelligence Quellen gefüllt, z.B. wenn von einer IP eine ungewöhnliche Anzahl von fehlgeschlagenen Zugriffsversuchen erfolgt ist. Die zeitnahe Nutzung der Informationen durch die PDP ist gewünscht, aber es muss keine Echtzeit-Propagierung von Updates sein, d.h. keine Propagierung im Bereich weniger Sekunden mit zwingend einzuhaltenden Deadlines. Eine potenzielle Umsetzung wäre eine verteilte und redundant ausgelegte Key-Value Datenbank. Aber bereits ein klassisches System wie DNS kann diese Anforderungen erfüllen: DNS-basierte Realtime-Blocklisten sind z.B. in der Domain der Spambekämpfung eine seit vielen Jahren bewährte Technologie, welche eine selektive Replikation und Caching der Daten (im lokalen DNS-Server) bietet und über etablierte Mechanismen (DNS-Lookup) abgefragt werden kann. Da eine DNS-basierte Replikation jedoch keinen Datenschutz durch Authentisierung beinhaltet, müssten die verteilten Daten entweder für die PDP verschlüsselt verteilt bzw. über ein Einweg-Hash anonymisiert werden.

Ein weiterer Anwendungsfall für diese Ausprägung stellt das sofortige Sperren von Nutzern dar. Zwar haben ID-Token eine begrenzte Laufzeit, aber es gibt ein Zeitfenster, in dem ein ID-Token für einen gesperrten Nutzer weiterhin gültig ist. In diesem Fall könnte das Token gehasht in einer entsprechenden Sperrdatenbank abgelegt und dieser PIP vom PDP bei der Zugriffsentscheidung befragt werden. Nach einiger Zeit ist der Token ungültig und kann wieder aus dem PIP entfernt werden. Analog zu IP-basieren Blocklisten könnten hier verteilte Key-Value Datenbanken oder auch (mit passendem zusätzlichem Schutz) DNS-basierte Realtime-Blocklisten verwendet werden. Auch ein geräteübergreifendes Abmelden von der TI könnte prinzipiell über einen analogen Mechanismus abgebildet werden, wird aber eher in der Eigenverantwortung des Nutzers gesehen und hier nicht weiter betrachtet.

4.2.3 PAP

Der PAP stellt die Regeln für das Regelwerk zur Verfügung, welches im PDP ausgewertet wird. Es wird davon ausgegangen, dass sich Regeln im Gegensatz zu den in den PIP gespeicherten Attributen nur selten ändern, da jede Regeländerung gründliches Design, Testen, Review und Freigabe erfordert. Da das Regelwerk sowohl verständlich als auch performant sein muss, werden die Regelsätze entsprechend klein sein. Damit reicht zum Bereitstellen der Regeln an die PDP ein Webserver mit einer entsprechenden Cache-Policy aus. Über den Cache-Control Header kann dabei granular gesteuert werden, in welchem Zeitraum ein PDP die Aktualität des Regelwerks überprüfen sollte (max-age) und wann nach fehlgeschlagenem Update das Regelwerk als ungültig anzusehen ist (max-stale). Über den ETag-Header kann der PDP die Information über seine lokale Version des Regelwerks liefern und damit nur bei einer neuen Version ein Update erhalten. Das längerfristige Caching im PDP reduziert auch die Verfügbarkeitsanforderungen an den PAP, d.h. Verfügbarkeitsprobleme selbst über mehrere Stunden beeinträchtigen die Funktionalität der Architektur nicht wesentlich. Für eine Umsetzung kann ein Webserver für statische Dateien, an dem sich die PDP per mTLS authentisieren, bereits ausreichen. Die Informationen zu den Versionen des Regelwerks, mit denen die PDP aktuell laufen, werden über das Monitoring propagiert.

Um trotz extensivem Caching ein einheitliches Inkrafttreten von Regeln über alle PDP hinweg zu ermöglichen, sollten die bereitgestellten Regelsets mit einem Zeitpunkt markiert sein, zu dem sie in Kraft treten. Wie weit in die Zukunft der Zeitpunkt festgelegt werden muss, hängt von den Cache-Policies ab: je weiter er weg ist, desto länger kann man cachen und desto unabhängiger wird das System von der Verfügbarkeit des PAP.

Der Prozess der Erstellung und Prüfung von Regeln ist primär eine Frage der Governance. Technisch unterstützt werden kann die Governance z. B. mittels Autorisierung für das Schreiben auf den PAP oder mittels Signaturen für Regelsets. Sowohl Autorisierung als auch Signaturen sind etablierte Technologien, auf die hier nicht weiter eingegangen wird. Welche technische Unterstützung es bei der Erstellung und Prüfung von Regeln gibt, hängt von der spezifischen Regel-Sprache und Regel-Engine ab und wird hier ebenfalls nicht weiter betrachtet.

Der Prozess für Emergency-Updates ist ebenfalls primär eine Frage der Governance. Wie dieser Prozess aussieht, ob eine Out-of-Band Benachrichtigung der PDP über ein zu aktualisierendes Regelwerk nötig ist und wie sie ggf. implementiert sein könnte, wird in diesem Feinkonzept nicht betrachtet.

4.2.4 Monitoring

Im betrieblichen Monitoring werden Informationen über den aktuellen Status (Verfügbarkeit, Funktionsfähigkeit, Performance) der Komponenten gesammelt. Dazu müssen die Komponenten zum einen die nötigen Loginformationen bereitstellen als auch über APIs Betriebsinformationen wie z.B. zur Auslastung zur Verfügung stellen. Die Auswertung dieser Informationen wird als Teil eines TI-übergreifenden Monitoring nicht speziell für die ZTA gesehen und hier nicht weiter betrachtet.

Zum Monitoring gehört auch das Sammeln von Informationen, welche direkt die Funktion der ZTA beeinflussen. Das betrifft sicherheitsrelevante Events z.B. über fehlgeschlagene Logins, DoS-Angriffsversuche u.ä. Diese Informationen werden über etablierte Protokolle an Analyse- und ggf. SIEM-Systeme übergeben. Dort werden sie aggregiert, analysiert und konsolidiert und die relevanten Ableitungen dann in die PIP eingespeist, um darauf aufbauend in den PDP z.B. auffällige IP-Adressen vom Zugriff auszunehmen bzw. für betroffene Accounts eine höhere Authentisierung zu erfordern. Die Verfügbarkeitsanforderungen für das Monitoring sind moderat: zwar sollten möglichst alle Logdaten erfasst werden, aber leichte Verzögerungen in der Analyse haben keine wirklich kritischen Auswirkungen und kurzzeitige Lücken in den Daten sind auch tolerabel. Ebenso ist die Speicherung der Daten vergleichsweise unkritisch, da personenbezogene Informationen entweder nicht bzw. nur in pseudonymisierter Form im Monitoring erfasst werden sollten.

Ebenfalls Bestandteil des Sicherheitsmonitoring ist die Überwachung der Komponenten der ZTA auf konkrete Angriffsversuche basierend auf fehlerhaften Anfragen oder bekannten Angriffsmustern. Eine kontinuierliche Analyse basierend auf Informationen zu neuen Sicherheitslücken sollte ebenfalls stattfinden. Die aktive Überprüfung der Komponenten auf derartige Sicherheitslücken sollte jedoch nicht in Produktivumgebungen, sondern nur in Testumgebungen stattfinden. Es ist dabei sicherzustellen, dass alle in der Praxis verwendete Versionen und Konfigurationen von diesen Tests abgedeckt werden.

Und schließlich wird die Ausführung des Regelwerks überwacht. Dabei werden Informationen über die abgearbeiteten Regeln sowie die in Entscheidungen einbezogenen Geräteinformationen gesammelt, jedoch keine personenbezogenen Daten. Die Informationen werden im MON aggregiert und ausgewertet, um sowohl unbenutzte Regeln zu ermitteln als auch die Auswirkung von Änderungen des Regelwerks (z.B. Verschärfung, welche Betriebssysteme zugelassen sind) vorab zu evaluieren. Aufgrund der Menge der Zugriffe ist eine Vorab-Aggregierung der Daten vor der Weitergabe an das

Monitoring empfehlenswert. Die Verfügbarkeitsanforderungen für dieses Monitoring sind ebenfalls moderat: eine Verzögerung bei der Datensammlung und Auswertung ist unkritisch und kurzzeitige Lücken in der Datenerfassung sind tolerabel.

4.2.5 Nutzerportal

Über das Nutzerportal kann der Anwender der TI (Versicherte, Leistungserbringer) sich betreffende Sicherheitseinstellungen einsehen und konfigurieren. Dabei ist das Nutzerportal nur die Oberfläche, welche über APIs mit anderen Komponenten der TI interagiert, jedoch selbst keine eigene Business-Logik und Datenspeicherung implementiert. Das Nutzerportal ermöglicht dem Nutzer das Einsehen der registrierten Geräte und ihrer Nutzungshistorie sowie die Möglichkeit zum Entfernen von Geräten und damit zum Sperren für die Nutzung. Dafür kommuniziert es mit dem GMS, welches über APIs die nötige Funktionalität bereitstellt. Das Zusammenspiel mit weiteren Komponenten z.B. zur Konfiguration nutzerspezifischer Sicherheitseinstellungen ist denkbar.

Die Verfügbarkeit des Nutzerportals ist wichtig, damit der Nutzer aktuelle Sicherheitseinstellungen wie z.B. die registrierten Geräte einsehen bzw. ändern kann. Der Zugriff auf die Fachdienste ist mit bereits registrierten Geräten jedoch auch bei temporärer Nichtverfügbarkeit des Nutzerportals gewährleistet. Die Umsetzung des Nutzerportals könnte als Applikation für Endgeräte erfolgen, welche die APIs der einzelnen ZTA-Komponenten direkt anspricht. Falls eine eigenständige App realisiert wird, muss diese allerdings auf allen relevanten Plattformen implementiert werden. Eine Integration der entsprechenden Funktionalität in etablierte Applikationen der Krankenkassen kann hier sowohl die Aufwände senken als auch nutzerfreundlicher sein. Eine alternative Möglichkeit wäre der Zugriff über ein Web-Frontend mit einem zentralen hochverfügbaren Applikationsserver. Wie in 4.3.2 tiefer beschrieben geht jedoch ein browser-basierter Zugriff mit größeren Herausforderungen bzgl. der Authentisierung von Nutzer und ggf. Gerät sowie bei der Absicherung gegen web-basierte Angriffe einher.

4.3 Fachdienstnahe Komponenten

Die in diesem Kapitel beschriebenen fachdienstnahen Komponenten liegen nicht in der Betriebsverantwortung der gematik, sondern beim Anbieter des Fachdienstes bzw. des Fach-Clients auf dem Endgerät. Nicht-systemische Verfügbarkeitsprobleme an diesen Komponenten betreffen nicht die komplette TI, sondern nur die Verfügbarkeit von Fachdiensten bzw. die Möglichkeit für einzelne Clients die Dienste zu nutzen.

4.3.1 am Fachdienst: PEP, PDP, fachdienstspezifischer PIP

Am PEP wird die TLS-Verbindung vom Fach-Client/TCL terminiert, welche nutzerspezifische Meta-Daten der Kommunikation und medizinische Daten enthält. Damit muss der PEP unter den gleichen Sicherheitsmaßnahmen wie der Fachdienst selbst laufen. Das kann ein Betrieb in der sicheren Ablaufumgebung des Fachdienstes sein, oder in einer mit dem Fachdienst verbundenen eigenen sicheren Ablaufumgebung, die die gleichen hohen Sicherheitsanforderungen erfüllt. Analoge Sicherheits-betrachtungen gelten für den PDP, da dieser Entscheidungen auf der Basis von Zugriffs-URLs fällt, welche bereits medizinisch relevante und/oder personenbezogene Indikatoren enthalten können. Dasselbe gilt für einen potenziellen fachdienstspezifischen PIP, welcher vom PDP bei der Ausführung von Regeln befragt werden kann.

Letztlich unterliegen damit Verfügbarkeit, Performance und Sicherheit dieser Komponenten der gleichen betrieblichen Verantwortung wie der Fachdienst selbst und

müssen von dem jeweiligen Betreiber entsprechend umgesetzt werden. Der Betrieb innerhalb von elastischen Clustern mit dynamischem Up- und Downscaling kann hier die entsprechende Performance und Skalierbarkeit schaffen, insb. wenn kombiniert mit zustandslosen Anfragen (die nicht an eine spezifische PEP-Instanz gebunden sind) und einem Cluster-übergreifenden Session-Management sowohl für PEP als auch den Fachdienst selbst.

Die genaue Umsetzung dieser Komponenten ist abhängig von der Ablaufumgebung. Um Interoperabilität und Betrieb für den Fachdienst zu vereinfachen, wird empfohlen, geprüfte und für den produktiven Einsatz ausgereifte Implementierungen von PEP und PDP für die typischen Ablaufumgebungen als Open Source verfügbar zu machen und zu pflegen.

PEP

Der PEP ist ein TLS-terminierender HTTP Reverse Proxy, der erst nach erfolgreicher Zugriffskontrolle durch den PDP die Kommunikation zwischen Fach-Client/TCL und Fachdienst ermöglicht. Im Rahmen der Authentisierung bekommt der PEP ein ID-Token vom IDP und ein Geräte-Token vom GMS. Der PEP verifiziert im Rahmen des Session-Managements bei jeder Anfrage die aktuelle Gültigkeit der Token und das sie zueinander sowie zum über mTLS übermittelten Geräte-Zertifikat des TCL passen. Die hierzu nötige Funktionalität ist einfach und kann z.B. als Plugin in etablierten Reverse-Proxys wie Envoy implementiert werden. Der PEP refresht in regelmäßigen Abständen den ID-Token über den vom IDP zusammen mit dem ID-Token gelieferten Refresh-Token. Ein derartiger Refresh kann zur Verbesserung des Nutzererlebnisses bzgl. Performance im Hintergrund erfolgen, solange der ID-Token noch gültig ist. Ein Refresh im Hintergrund wird durch Anfragen eines Fach-Client ausgelöst, findet also nicht unabhängig von einem verbundenen Fach-Client statt.

ID-Token und Geräte-Token bzw. die daraus extrahierten Informationen sowie die Quell-IP des TCL als Umgebungsinformation werden dem PDP als Grundlage der Zugriffsentscheidung übergeben. Nur nach erfolgreicher Entscheidung wird der Request an den Fachdienst weitergereicht. Die Kommunikation zwischen PEP und PDP ist nicht standardisiert. Für maximale Flexibilität und Entwicklungsgeschwindigkeit bei minimalen Interoperabilitätsproblemen ist es daher sinnvoll, dass PDP und PEP vom gleichen Hersteller kommen bzw. ggf. in Kombination als Open Source produziert werden. Es wäre auch möglich, dass PDP und PEP als logische Komponenten innerhalb des gleichen Prozesses laufen, um den Kommunikationsoverhead gering zu halten. Eine enge Abstimmung und geringer Kommunikationsoverhead von PEP und PDP bilden auch eine gute Grundlage für Performanceoptimierungen bei der Überprüfung des Regelwerks, welche beim PDP weiter beschrieben werden.

PDP

Im PDP wird die Zugriffsentscheidung getroffen basierend auf dem Regelwerk für den Fachdienst, den vom PEP gelieferten Informationen zum aktuellen Request sowie den übergreifenden Kontextinformationen aus den PIP. Für das Regelwerk holt sich der PDP periodisch die für den Fachdienst relevanten Regelsets vom PAP und aktiviert diese nach Überprüfung der Signaturen zum je Regelset gesetzten Aktivierungszeitpunkt. Bzgl. der mit PAP und PIP ausgetauschten Daten und dafür benutzten Protokolle siehe die Beschreibung von PIP und PAP und Kapitel 4.2.

Zentrales Element im PDP ist die Regel-Engine. Diese muss für die Ausführung an die Request-spezifischen (vom PEP) und Request-übergreifenden (PIP) Informationsquellen angebunden werden. Wie das genau passiert und ob eine Konsolidierung der Informationsquellen im PDP nötig ist, hängt von der gewählten Regel-Engine und ihren Möglichkeiten sowie Integrations-APIs ab. Wie in 3.5.4 beschrieben gibt es hier verschiedene am Markt befindliche Umsetzungen, wie z.B. auf XAMCL aufbauende Lösungen sowie der Open Policy Agent (mit eigener Policy-Sprache Rego). Diese

Umsetzungen sollten auf eine Eignung für eine Nutzung in der Architektur tiefer evaluiert werden. Neben der Performance bei der Ausführung von Regeln spielen dabei auch die für eine maximale Zukunftsfähigkeit angestrebten Eigenschaften eine wichtige Rolle: Leistungsfähigkeit und Flexibilität für die Umsetzung des Regelwerks und dessen zukünftige Anforderungen, Integrierbarkeit in PDP sowohl bzgl. Komplexität und Eignung für verschiedene Ablaufumgebungen, ausreichende Herstellerunabhängigkeit und eine absehbar langfristige Unterstützung sowohl von Industrie als auch von Entwicklern. Da Flexibilität und Komplexität oft einhergehen und sowohl Sicherheit als auch Wartungsaufwände negativ beeinflussen, sollte bei dieser Evaluation nur die tatsächlich in der Praxis benötigte Leistungsfähigkeit und Flexibilität für das Regelwerk als Kriterium genutzt werden. Insbesondere wenn eine Umsetzung der PDP als Open Source Lösung angestrebt wird, sollte neben der Nutzung von bestehenden Umsetzungen auch die Möglichkeit aufgenommen werden, eine auf den konkreten Anwendungsfall der TI bestens zugeschnittene Regel-Engine selbst zu implementieren (ggf. basierend auf einer etablierten Regelsprache) und unter voller Kontrolle der gematik zu pflegen und weiterzuentwickeln. Als eine Möglichkeit sehen wir hierbei (aber evtl. auch im Kontext existierender Regel-Engines) die Umsetzung in einer Web Assembly (WASM)-Umgebung an, d.h. zum einen die Integration der Regelausführung mit den PIP, zum anderen die Ausführung der Regel-Engine als WASM bzw. die Kompilierung von Regeln nach WASM. Dies bietet eine hohe Flexibilität und Plattformunabhängigkeit bei der Umsetzung. Die Nutzung von WASM als Ablaufumgebung ist auch aus Sicherheitssicht attraktiv, da sie eine Sandbox für mächtige aber gleichzeitig kontrollierte und limitierte Codeausführung bildet.

Zur Optimierung der Performance ist es angebracht, dass nicht bei jedem Zugriff erneut das komplexe Regelwerk im PDP ausgeführt werden muss. Anfragen, die im gleichen "Realm" wie vorhergehende Anfragen liegen und damit den gleichen Pfad im Regelwerk durchlaufen und zu der gleichen Entscheidung kommen würden, können in den meisten Fällen durch eine gecachte Entscheidung für diesen Realm beantwortet werden. Was genau einen solchen Realm und damit den beim Caching benutzten Key ausmacht ist jedoch abhängig von dem konkreten Regelwerk. Je nach dessen Regeln kann es sich um Eigenschaften der Anfrage, wie Teilen des Pfades, Eigenschaften des Gerätes oder auch Eigenschaften der Umgebung wie IP-Adresse der Clients bzw. Geolokation handeln. Dazu kann die Entscheidung von veränderlichen Informationen außerhalb des Realms abhängig sein, wie der aktuellen Zeit oder der aktuellen Reputation der Client-IP-Adresse. Entsprechend maßgeschneidert müssen die Zusammensetzung des Realms und die Zeitdauer des Cachings der Entscheidung definiert werden. Eine möglichst automatische Herleitung derartiger Informationen aus dem Regelwerk sollte bei der Evaluation der Kandidaten für ein Regelwerk mit betrachtet werden.

Fachdienstspezifischer PIP

Es ist möglich, dass ein Fachdienst eigene PIP betreibt, um lokal erfasste Informationen in Entscheidungen des Regelwerks einfließen zu lassen. Im Gegensatz zu übergreifenden PIP unterliegen diese fachdienstspezifischen PIP nur den Verfügbarkeitsanforderungen des Fachdienstes selbst, analog zu den PEP und PDP. Eine Authentisierung des PDP am fachdienstspezifischen PIP muss ebenfalls nicht über ZTA-globale Mechanismen erfolgen, sondern kann auf eine fachdienstspezifische Weise umgesetzt werden. Ansonsten können diese fachdienstspezifischen PIP prinzipiell auf analoge Weise umgesetzt werden wie fachdienstübergreifende.

4.3.2 am Fach-Client: TCL

Über den Trust-Client (TCL) kommuniziert der Fach-Client mit dem Fachdienst. Der TCL managt dabei die Registrierung, Authentisierung und Attestierung am GMS, die nachfolgende authentisierte Kommunikation mit dem Fachdienst sowie den Refresh des Geräte-Token.

Für den Schutz des zum Gerätezertifikat gehörenden privaten Keys benutzt der TCL einen plattformspezifisch geschützten Keystore. Der private Key wird in diesem geschützten Keystore erzeugt und kann diesen auch nicht verlassen, d.h. eine Nutzung des Gerätezertifikats durch andere Geräte ist nicht möglich. Unter Nutzung des geschützten Zertifikats baut der TCL eine mTLS-Verbindung zum GMS auf und überträgt, nach der initialen einmaligen Registrierung des Geräte-Zertifikats, bei Bedarf die aktuellen Attestierungsinformationen und ggf. weitere Informationen über das Gerät an den GMS. Für die Attestierung kann eine Kommunikation mit Diensten der Plattformanbieter (Google, Apple, ...) notwendig sein, weshalb die dortigen Rate-Limitierungen zu beachten sind.

Ebenfalls unter Nutzung des geschützten Geräte-Zertifikats erfolgt die mTLS-Kommunikation zwischen TCL und PEP/Fachdienst. Da das Geräte-Zertifikat im vom GMS ausgestellten Geräte-Token als Fingerprint angegeben ist und ein Transfer des Zertifikats auf andere Geräte daher nicht möglich ist, kann der PEP sicher sein, dass das Token tatsächlich für das aktuell zugreifende Gerät ausgestellt wurde.

Durch Kommunikation mit dem GMS aktualisiert der TCL das Geräte-Token vor Ablauf im Hintergrund. Das ermöglicht eine kontinuierliche Überwachung des Gerätestatus ohne negative Auswirkungen auf die Usability, wie z.B. zusätzliche Wartezeiten.

Da die mTLS-Verbindung zum PEP am TCL initiiert wird, müssen Fach-Client und TCL sicher miteinander verbunden sein. Im einfachsten Fall wird dies durch eine Integration des TCL als SDK in den Fach-Client erreicht. Die Kommunikation zwischen diesen Komponenten erfolgt dann innerhalb des gleichen Prozesses. Ein standardisiertes API gibt es dafür nicht, d.h. es ist zu spezifizieren, um TCL als SDK plattformübergreifend und herstellerunabhängig mit verschiedenen Fach-Clients zusammenspielen zu lassen. Neben der Integration als SDK sind auch Umsetzungen als eigenständige Applikation oder eigenständiges Gerät (Infrastruktur-TCL) denkbar. Diese können auf dem SDK basieren, brauchen aber noch eine sichere Kommunikation zwischen Fach-Client und TCL und im Falle von Infrastruktur-TCL auch eine Integration in lokale ISMS und IAM. Im Rahmen des Feinkonzepts wird sich auf die Integration als SDK in eine plattform-native Anwendung beschränkt.

Einen Spezialfall stellen Fach-Clients dar, welche als Webapplikationen realisiert sind. Hierbei kann es sich um Browser-Frontends handeln, welche mit einem Webapplikations-Server als Fach-Client-Backend kommunizieren, in dem die Business-Logik und der TCL implementiert ist. Wie in 3.2 für Client-Server-Infrastrukturen beschrieben, könnte das Fach-Client-Backend eine Identifikation des angeschlossenen Clients über den TCL an die TI weiterleiten, um Transparenz herzustellen. Eine Unterscheidung einzelner Browser finden hierbei nicht statt.

Moderne Browser bieten jedoch zunehmende komplexere Ablaufumgebungen, bei denen es möglich ist, den kompletten Fach-Client inkl. Business-Logik im Browser zu implementieren. Hier kommuniziert der Browser nicht mit einem Applikationsserver als Fach-Client-Backend, sondern direkt mit dem Fachdienst in der TI. Bei derartigen "Rich" Applikationen muss die Logik des TCL innerhalb des Frontends im Browser realisiert, oder ein eigenständiger TCL-Prozess als HTTP-Proxy vom Browser integriert werden. Die direkte Integration in das Frontend der Webapplikation ist dabei in ihren Möglichkeiten eingeschränkt, kann aber je nach Anforderungen des Dienstes ausreichen. Hier können z.B. Informationen zur Browsersicherheit über das Web Authentication API übermittelt werden. Ist ein lokaler Sicherheits-Agent installiert, so können Informationen von diesem über lokale Webzugriffe oder Browsererweiterungen eingebunden werden. Unabhängig von der konkreten Umsetzung sollte beim Einsatz von Webapplikationen berücksichtigt werden, dass es für in einem universellen Browser laufende Applikationen im Vergleich zu eigenständigen Applikationen mehr sicherheitskritische Fallstricke wie CSRF und XSS gibt, die eine sichere Ausführungsumgebung gefährden können. Im Feinkonzept wird sich daher

auf Standalone-Applikationen beschränkt, welche ebenfalls mit Webtechnologien wie WebView realisiert sein können. Wir erwarten für die Zukunft weitere Verbesserungen bzgl. der Sicherheit und Integrationsfähigkeit in Zero Trust Szenarien, die universelle Browser als flexible und überall verfügbare Ablaufumgebung auch für sensitive Anwendungen attraktiver machen.

Der TCL muss plattformspezifische Methoden nutzen, um den privaten Schlüssel des Gerätezertifikats vor Missbrauch zu schützen und die als API angebotenen kryptografischen Methoden in einer sicheren Umgebung auszuführen. Welche Methoden dies im Einzelnen sind und welche Garantien durch diese geboten werden, hängt von der jeweiligen Plattform ab. Neuere Systeme bieten hierzu typischerweise ein hardware-geschütztes Key-Management und gesicherte Ausführungsumgebungen an, bei älteren Systemen ist dies aber evtl. nicht oder nur in geringerem Umfang verfügbar. Die vom TCL erreichbare Sicherheit wird im Rahmen der Geräteattestierung an das GMS übermittelt und dort direkt oder als Teil von konsolidierten Informationen in den Geräte-Token eingebettet. Das ermöglicht es, die Sicherheit der Plattform des Fach-Clients bzgl. kryptografische Operationen in die Zugriffsentscheidungen des PDP einfließen zu lassen.

Neben dieser Information werden im Rahmen der Attestierung auch weitere Geräteinformationen vom TCL an das GMS übermittelt. Dazu gehören plattformspezifische kryptografisch gesicherte Attestierungen über die Eigenschaften und Integrität der Plattform (Secure Boot, Jail-Breaking, ...) aber auch nicht zwingend kryptografisch geschützte Informationen über Aktualität des Antivirus, laufende Programme etc. Das bedeutet, dass der TCL dem GMS Informationen unterschiedlicher Zuverlässigkeit übermittelt und dass das GMS diesen Informationen nur beschränkt vertrauen kann - ein manipulierter TCL könnte hier andere Informationen übermitteln. Eine entsprechende Manipulation kann nicht ausgeschlossen werden, da das Endgerät unter der Kontrolle des Benutzers und nicht unter der Kontrolle des GMS bzw. Fachdienstes steht. Zwar gibt es z.B. auf mobilen Geräten zuverlässige plattformspezifische Attestierungsmechanismen, die kryptografisch abgesichert Auskunft über die Integrität des Systems geben. Diese Integritätsbetrachtungen sind jedoch eventuell nur ein Teil der Geräte-Informationen, die man in eine Entscheidung einbeziehen will. Die Zuverlässigkeit der erfassten Informationen muss daher beim Design der Zugriffsregeln bekannt sein und mit betrachtet werden.

Die Erhebung der Attestierungsinformationen ist plattformspezifisch. Unterschiedliche Plattformen stellen unterschiedliche Informationen auf verschiedene Art zur Verfügung. Grundlegend für das Erheben von Informationen zum Endgerät ist zunächst die Integrität der Plattform und des Trust-Client selbst, also die Sicherheit, dass der Trust-Client und die Endgeräteplattform selbst nicht manipuliert wurden.

Auf stationären Geräten sind TPM vorherrschend und von allen aktuellen Betriebssystemen unterstützt. Darüber lassen sich gesicherte Informationen insbesondere zum Boot Vorgang gewinnen. Während des Bootens schreiben das UEFI sowie der Bootloader Daten in das TPM Event Log, welches später ausgelesen werden kann. Es enthält die Information, ob Secure Boot verwendet wurde und es protokolliert die Inhalte der Secure Boot Schlüsseldatenbank. Damit kann sichergestellt werden, dass lediglich bekannte Secure Boot Signaturschlüssel von vertrauenswürdigen Herstellern installiert sind und die am Boot Vorgang beteiligten Komponenten sowie der Betriebssystemkern integer sind. Die Log Informationen können mithilfe eines im TPM vorher installierten Schlüssels kryptographisch abgesichert werden. Dieser Schlüssel wird durch ein standardisiertes kryptographisches Verfahren an ein vom TPM-Hersteller eingebrachtes Zertifikat bzw. Schlüsselpaar gebunden.

Exemplarisch für mobile Geräte können Android Geräte betrachtet werden. Als Plattformhersteller bietet Google die sogenannte Google Play Integrity API an, welche Informationen ("Verdicts") über die Integrität des Geräts, sowie der Anwendung

bereitstellt. Das exakte Verfahren zum Prüfen dieser Informationen ist nicht öffentlich verfügbar. Das Resultat enthält lediglich zusammengefasste Informationen, wie etwa "MEETS_DEVICE_INTEGRITY", "MEETS_BASIC_INTEGRITY", "MEETS_STRONG_INTEGRITY" oder "PLAY_RECOGNIZED". Zusätzlich kann festgestellt werden, ob etwa Technologien wie Android Verified Boot verwendet werden. Apple liefert auf seinen Geräten (iOS/iPadOS/MacOS) mit dem DeviceCheck Framework ähnliche Funktionalität.

In einem zweiten Schritt können über den Trust-Client selbst je nach Plattform weitere Informationen (Geräteklasse, Betriebssystem, Patch-Level, usw.) über das Endgerät gesammelt und zur Verfügung gestellt werden. Die Vertrauenswürdigkeit dieser Informationen ist dann von dem zuvor bestimmten Vertrauen in die Integrität der Plattform und des Trust-Client abhängig. Für einige aber nicht alle Plattformen ist es sogar möglich, auch solche detaillierten Informationen über die PKI des Plattformherstellers abzusichern (z.B. Certificate extension data schema im Rahmen der Android Key Attestation).

Es ist nicht standardisiert, wie oder welche Attestierungsinformationen zwischen einem TCL und einem GMS übertragen werden, oder wie diese in Regelwerken verwendet werden. Jede Zero-Trust-Lösung hat hier ihre herstellerspezifische Umsetzung. Entsprechend muss es eine enge Zusammenarbeit zwischen den Herstellern von TCL und GMS sowie den Erstellern des Regelwerks geben, damit diese Teile sowohl gut aufeinander abgestimmt sind als auch zukunftsfähig weiterentwickelt werden können. Herstellerübergreifend genutzte Open Source Implementierungen von TCL können hier die Interoperabilität vereinfachen.

4.4 Abhängigkeiten von weiteren Komponenten

Die im Folgenden beschrieben Dienste liegen außerhalb der Zero-Trust-Architektur. Ihre Verfügbarkeit ist jedoch essenziell für die Funktionsfähigkeit der ZTA, sodass sie hier kurz betrachtet werden.

4.4.1 IDP-Infrastruktur und Authenticator

Das Zusammenspiel zwischen Authenticator am/im Fach-Client und der IDP-Infrastruktur dient zur initialen Authentisierung des Nutzers (Ausstellen des ID-Token) sowie zur regelmäßigen Verlängerung ihrer Gültigkeit (Refresh des ID-Token). Der Refresh des ID-Tokens wird vom PEP mittels Refresh-Token bei Aktivität des Fach-Clients des Benutzers durchgeführt.

4.4.2 Fachdienst und Fach-Client

Der Fach-Client kommuniziert mit dem Fachdienst und der ZTA über den Trust-Client. Entsprechende Robustheitsbetrachtungen befinden sich bereits in Kapitel 4.3.2. Der Fachdienst selbst unterliegt hohen Verfügbarkeits- und Skalierbarkeitsanforderungen, welche in der Spezifikation bzw. Zulassung des Fachdienstes betrachtet werden und hier nicht weiter betrachtet sind.

4.4.3 Federation Master/Main (FEM)

Der Federation Master verwaltet die Authentisierungsmerkmale (Zertifikate, öffentliche Schlüssel) für alle in der IDP-Föderation beteiligten Teilnehmer. Seine Verfügbarkeit ist für

die ZTA zur Überprüfung von ID-Token undGeräte-Token wichtig. Da sich die Authentisierungsmerkmale nicht oft ändern, sollten sie von den sie nutzenden Diensten gecacht werden, um die Verfügbarkeitsanforderungen an den FEM zu senken.

4.4.4 Zeitsynchronisierung, PKI

Eine zuverlässige Zeitsynchronisation ist wichtig für die Überprüfung der Gültigkeit von Zertifikaten und Tokens, wobei hierfür eine Granularität im Minutenbereich ausreicht. Es ist nicht zwingend nötig, dafür eigene Zeitserver bereit zu halten. Die im Internet bereits verfügbaren Server sind ausreichend und erfüllen die Anforderungen sowohl bzgl. der Redundanz als auch der Präzision.

Für die Authentisierung von Fachdiensten und Komponenten der ZTA sowie der IDP-Infrastruktur sollte auf eine eigene Private Key Infrastruktur zurückgegriffen werden, so dass bei der Überprüfung von Dienst-Zertifikaten nur der eigenen CA vertraut werden muss. Da diese PKI für die IDPs bereits existieren muss, wird sie für die ZTA als gegeben vorausgesetzt. Für die Kommunikation zwischen Diensten der TI darf nur diese PKI für als vertrauenswürdig angesehen werden. Es wird empfohlen, so weit wie möglich OCSP-Stapling zu benutzen, um sowohl die Abhängigkeit von als auch die Last der OCSP-Responder zu reduzieren (mit Auswirkung auf Verfügbarkeits- und Skalierungsanforderungen) und um eine höhere Performance beim TLS-Verbindungsaufbau zu ermöglichen.

4.4.5 DNS

Für die öffentlichen Schnittstellen der ZTA ist ein eigener, vom öffentlichen DNS entkoppelter Namensraum nicht erforderlich. Eine Integration in den öffentlichen Namensraum kann sogar Implementationen vereinfachen, da Clients nicht je nach Zieldomain mit verschiedenen DNS-Providern reden müssen. DNS bietet bereits per Design eine hohe Robustheit und Verfügbarkeit, sodass hier keine zusätzlichen Maßnahmen getroffen werden müssen. Für eine verbesserte Privacy ist dabei DoH (DNS über HTTPS) bzw. DoT (DNS über TLS) empfehlenswert, welches von vielen Betriebssystemen bereits nativ unterstützt wird. Es wäre möglich DoH/DoT-Implementierungen mit passenden Zugangspunkten spezifisch für Fachdienst/TCL einzusetzen. Dies könnte jedoch zu Problemen in restriktiven Umgebungen (Firmen, Captive-Portals, ...) führen, sodass empfohlen wird, für das DNS-Handling lieber auf die vom Betriebssystem anwendungsübergreifend bereitgestellten Mechanismen zur Namensauflösung zurückzugreifen.

Die Nutzung von DNSSec ist als zusätzliche Sicherungsschicht möglich. Sie ist jedoch nicht zwingend erforderlich, da sämtliche Kommunikation zu bzw. zwischen den Komponenten der ZTA per TLS erfolgt und die TLS-Zertifikatsüberprüfung ein MITM durch DNS-Spoofing bereits zuverlässig verhindert. Von daher stellen DNS-Forwarder und DNS-Stacks, welche DNSSec nicht unterstützen (diverse Router, ältere Betriebssysteme) kein zusätzliches Sicherheitsrisiko dar.

5 Governance und Betrieb

Gesamtverantwortlich für den zuverlässigen Betrieb der TI und damit zukünftig der ZTA ist die gematik. Damit sie dieser Verantwortung für die ZTA gerecht werden kann, delegiert die gematik die Verantwortung an die Anbieter, Betreiber und Hersteller der einzelnen Komponenten und Dienste der Zero-Trust-Architektur. Die gematik trifft weiterhin die Regelungen zur Funktionalität, Interoperabilität, Sicherheit und zur Betriebsführung der TI in Spezifikationen und setzt diese in Zulassungsverfahren bzw. in direkten Beauftragungsverhältnissen und in ihrer Rolle als koordinierende Instanz der Betriebsführung durch.

Die entscheidenden Neuerungen, die die Einführung der ZTA für die Governance der TI mit sich bringen wird, sind:

- Steuerung von Zugriffen auf der Grundlage des dynamischen Regelwerks der Zero-Trust-Architektur:
  Mit der Anwendung des Regelwerks werden die wesentlichen Anforderungen an erlaubte Zugriffe auf Dienste der TI zur Laufzeit durchgesetzt. Das Regelwerk wird unter der Federführung der gematik definiert, kontinuierlich weiterentwickelt und damit an aktuelle Bedrohungslagen und technologische Entwicklungen angepasst.
- Übergreifendes Monitoring aller Komponenten und Dienste auf nutzungsrelevante Informationen:
  Das Monitoring von Komponenten und Diensten wird so weit ausgeweitet, dass die erfassten Informationen (beispielsweise zum Verhalten von Nutzern oder zum Einsatz konkreter Technologien und ihrer Versionsstände) zur Anpassung des dynamischen Regelwerks herangezogen werden können.

Die Gestaltung des Betriebs und der relevanten Betriebsprozesse leitet sich aus der Architektur und den involvierten Akteuren ab. Die Governance beschreibt deren Verantwortlichkeiten und Interaktionen miteinander. Im Folgenden werden das Betreibermodell (Kapitel 5.1) und die Betriebsprozesse (Kapitel 5.2) beschrieben.

5.1 Betreibermodell

Nicht in der öffentlichen Version enthalten.

5.2 Betriebsprozesse

In den nachfolgenden Kapiteln werden die für den Betrieb der ZTA relevanten Prozesse und Rollen näher betrachtet. Es wird auf On- und Offboarding von Diensten (Kapitel 5.2.1), die Administration des Regelwerks (Kapitel 5.2.2), das Sicherheits- und Betriebsmonitoring (Kapitel 5.2.3), das übergreifende IT-Servicemanagement (Kapitel 5.2.4), Supportprozesse (Kapitel 5.2.5) sowie auf die administrativen Prozesse der Nutzer (Kapitel 5.2.6) eingegangen. Damit wird ein Zielbild gezeigt bzw. eine Empfehlung zur Gestaltung der Prozesse abgegeben. Es werden die Akteure beschrieben, die über die

im Kapitel 2.1- Akteure genannten hinaus eine Rolle spielen, z.B. Stakeholder wie BSI und BfDI.

5.2.1 On- und Off-Boarding von Diensten

Das On- und Offboarding von Diensten soll sowohl für Hersteller, Anbieter und Prüfinstanzen als auch für die gematik selbst automatisierter, transparenter und schneller möglich sein. Aufwändige, zum Teil organisatorisch ausgeführte Prozesse sollen weitgehend vereinfacht werden. Die Beschaffung und Inbetriebnahme von "Sicheren zentralen Zugangspunkten" (SZZP) für den Anschluss an die zentrale TI, einschließlich aller Serviceinteraktionen zwischen dem designierten Fachdienstanbieter, dem Anbieter der zentralen TI-Plattform und der gematik entfällt in der ZTA.

Die gematik sollte einenüber das Internet erreichbaren Dienst "Selfserviceportal für Anbieter und Hersteller" (im Folgenden als "Anbieterportal" bezeichnet) bieten, der es Dienstanbietern ermöglicht, sich als Hersteller bzw. Teilnehmer der TI zu registrieren und z. B. einen Zulassungsantrag zu stellen. Das Anbieterportal sollte jedem Anbieter einen Überblick zum aktuellen Status, Sicherheitsstand und Dauer der Zulassung seiner Dienste geben und es ermöglichen, die weiteren Schritte des On- bzw. Offboardings zu steuern. Im Folgenden wird skizziert, wie das Portal beim On- und Offboarding genutzt werden kann. Die Überwachung des Betriebsstatus wird in Kapitel 5.2.3.1 beschrieben. Separat ausgeführt wird das Selfservice-Portal für Nutzer der TI in Kapitel 5.2.6 .

Voraussetzung für die Nutzung des Anbieterportals ist eine Identifikation der genannten Akteure sowie die Erteilung einer Zugangsberechtigung zu den nachfolgend aufgeführten Diensten des Selfserviceportals.

5.2.1.1 Onboarding

Die folgenden logischen Schritte wird ein Hersteller bzw. Anbieter über das Anbieterportal ausführen bzw. auslösen können, um den Prozess der Zulassung eines Produkts bzw. des Anschlusses eines Dienstes an die TI selbständig zu steuern:

- Erfassung der Absicht eines Herstellers/Anbieters zur Bereitstellung von Produkten und Diensten. Das umfasst die Erfassung der Basisdaten des Herstellers bzw. Anbieters, z. B.:
  - Daten zum in die TI einzubringenden Produkt bzw. Dienst: Bezeichnung, Zielversion, Zeitplanung für Go-Live, etc.
  - Kontaktdaten
  - Servicezeiten sofern nicht vorgegeben
- Anzeige des Sicherheitsstatus für die eigenen Produkte basierend auf Security-Scans, Audits und Pentest-Ergebnissen von der gematik
- Abrufen aller Informationen und Unterlagen, die als Grundlage der Zulassung benötigt werden. Das umfasst Informationen zum Konfigurationsmanagement, die aktuell geltenden Spezifikationsversionen, die Kompatibilitätsmatrix und Zeitleisten zur geplanten Fortschreibung sowie ggf. entsprechenden Migrationspläne. Ein solcher, konkret auf den laufenden Zulassungsvorgang bezogener Service, würde die derzeitige Veröffentlichung der Spezifikationen im Fachportal ergänzen.
- Abrufen von Referenzimplementierungen und Testsuiten für eigenverantwortliche Tests (EVT ggf. im Selfservice), die durch die gematik (oder die Community) bereitgestellt werden

- Bereitstellung von Testinstallationen der Komponenten der Hersteller in einer für die gematik Testumgebung geeigneten Form, z.B. als Container und ohne weitere Abhängigkeiten
- Verwaltung der vom Antragsteller berufenen Prüfer und Gutachter für Registrierung, Prüfung, Zertifizierung und Bekanntmachung
- Einreichung von für das Onboarding notwendiger Unterlagen durch Anbieter und Prüfer, z.B.
  - Test-/Prüfberichte
  - Gutachten
  - Zertifikat
  - Quellcode-Referenzen (sofern Open Source und nicht bereits im Gutachten enthalten)
- Aufsetzen automatisierter Tests für Sicherheitsanalyse und Betriebsmonitoring des Fachdienstes. Das betrifft Tests für die Zulassung selbst aber auch für ein kontinuierliches Monitoring des Betriebs und regelmäßige Attestierung des Sicherheitsstatus. Diese Tests können von der gematik bereitgestellt werden oder auch von Anbietern spezifisch für ihren Fachdienst erstellt werden.
- Steuerung des Stagings der Produkte in den unterschiedlichen Umgebungen (Referenzumgebung, Testumgebung, Produktivumgebung), einschließlich Ausstellung der zur Kommunikation und Teilnahme an der Föderation erforderlichen Zertifikate und Registrierung im Federation-Master.
- Schaffung von Transparenz und Koordination mit anderen Herstellern bzw. Anbietern für Interoperabilitäts-Tests durch Verfügbarmachung von Kontaktdaten, Testkalender, Monitoringinformationen und Statusüberblick
- Abrechnung der Zulassungsgebühren

5.2.1.2 Offboarding

Nicht nur das Onboarding, sondern auch das Offboarding von Diensten soll im Anbieterportal durchgeführt werden. Das Offboarding kann vom Hersteller / Anbieter ausgelöst werden, wenn er seinen Dienst zurückziehen möchte. Aber auch die gematik kann ein Offboarding bewirken, wenn sie z.B. die Zulassungsvoraussetzungen als nicht mehr gegeben sieht (bspw. End of Life). Diese Prozesse sollen weitestgehend automatisiert ablaufen. Dazu zählen:

- Sperren von Diensten (z.B. zeitweise Zugriffe ausschließen)
- Entfernen von Diensten
- Federation Master Eintrag löschen
- Zertifikate widerrufen

Die Ausführung dieser Prozesse darf allerdings nur mit entsprechender Autorisierung der Anbieter und unter Einhaltung der in der Governance festgelegten Entscheidungsinstanzen erfolgen. Das ist zum Beispiel ein 4-Augen-Prinzip zwischen gematik und Anbieter beim Entfernen von Diensten oder 4-Augen-Prinzip innerhalb der gematik / SOC beim Sperren von Diensten bspw. aufgrund eines Notfalls.

5.2.2 Administration Regelwerk

Im folgenden Kapitel wird näher auf die Rollen und Verantwortlichkeiten bei der Erstellung und Pflege des Regelwerks eingegangen. Das Regelwerk muss erarbeitet, verteilt, getestet, simuliert und auf die gewünschte Wirkung hin validiert werden. Die Administration des Regelwerks sollte in enger Abstimmung mit dem Sicherheitskonzept und dem Informationssicherheitsmanagement der TI erfolgen und im Rahmen organisatorischer Prozesse unter Integration aller relevanten Stakeholder (z.B. BSI, BfDI, Hersteller, Anbieter, Betreiber). Dabei muss jeweils die Ressource, für die der Zugriff geregelt werden soll, zusammen mit den beteiligten Akteuren betrachtet werden. Es ist immer eine Fallbetrachtung durchzuführen und dementsprechend sind Regeln nur in der Verantwortung und Moderation der gematik und unter Einbeziehung der relevanten Stakeholder, wie z.B. BSI, BfDI, Hersteller, Anbieter, Betreiber änderbar.

Die mit der Administration des Regelwerks einhergehenden Prozesse werden in Abbildung 5.2-1 zusammengefasst und nachfolgend im Einzelnen beschrieben.

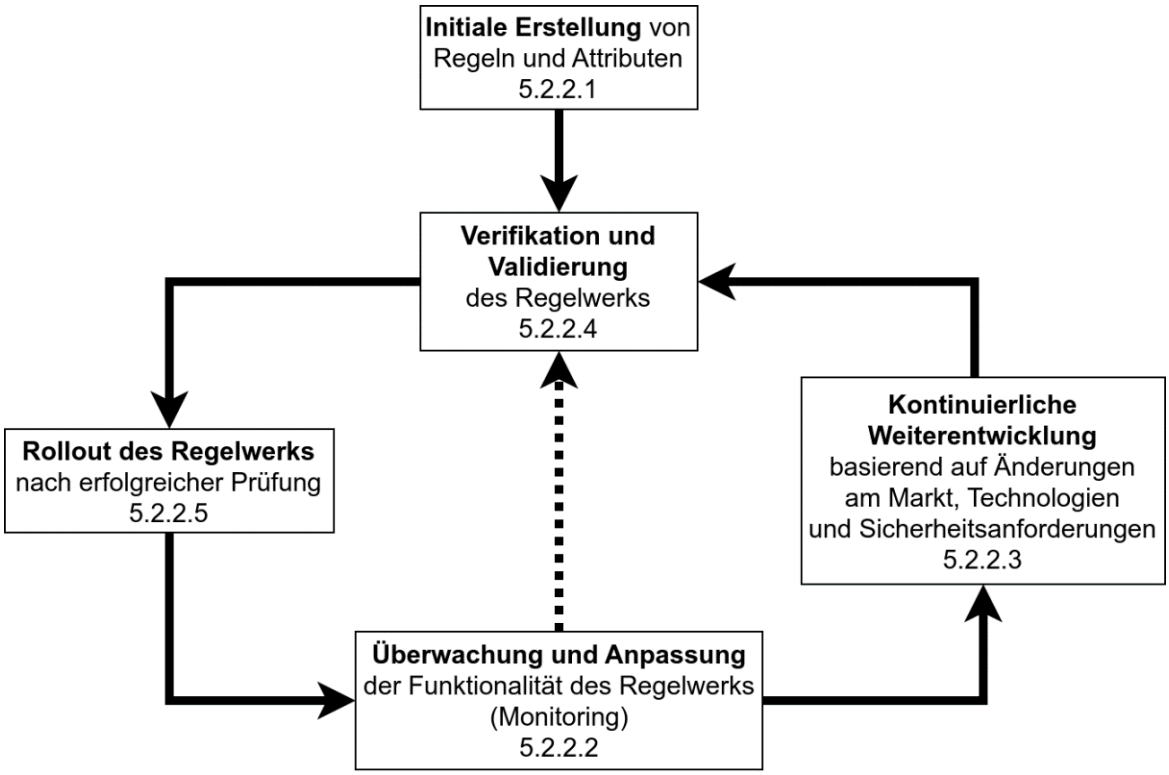

Abbildung 5.2-1: Übersicht über die Prozesse zur Administration des Regelwerks

5.2.2.1 Erstellung eines Regelwerks

In Kapitel 3.3.2 ist ausgeführt, wie sich der Prozess der Erstellung des dynamischen Regelwerks im Rahmen eines Risikoanalyseprozesses gestaltet, nämlich zweistufig, erst übergreifend und dann fachdienstspezifisch. Der Prozess muss in der Verantwortung der gematik ausgeführt werden und die Vorgaben aus dem Informationssicherheitskonzept der TI berücksichtigen. Die gematik sollte das BfDI, BSI und Komponentenhersteller, sowie Anbieter und Betreiber von Fachdiensten und IDP einbeziehen. Insbesondere in der übergreifenden Analyse von Bedrohungen und bei der Festlegung der Attribute für die

ABAC sowie bei der Definition von Regeln fließen die Expertisen der zuständigen Aufsichtsbehörden aber auch der herstellenden bzw. betreibenden Industrie mit ein.

Sollten fachdienstspezifische Regeln benötigt werden, muss für deren Erstellung fachliche Expertise zu den vom Dienst angebotenen Schnittstellen, ihrem Bedrohungsmodell und Schutzbedarf hinzugezogen werden. Dazu können fachliche Expertengruppen einbezogen werden, beispielsweise Nutzervertreter im Rahmen von Nutzerstudien, Fachverbände und spezialisierte Expertenkreise der zuständigen Regulierungsbehörden. Der Prozess läuft ebenfalls unter der Moderation und in der Verantwortung der gematik.

5.2.2.2 Überwachung und Anpassung der Funktionalität des Regelwerks

Das Überwachen und Anpassen des Regelwerkes unterliegt den Informationssicherheitsmanagement-prozessen sowie den operativen Managementprozessen der TI. Eine Anpassung kann zum einen aus Erkenntnissen aus dem regelmäßigen Review von Artefakten dieser Prozesse hervorgehen, z.B. basierend auf Artefakten der dem Regelwerk zugrundeliegenden Risikoanalyse. Zum anderen kann sie auch anlassbezogen auf Grund von Ereignissen außerhalb der regelmäßigen Reviews erfolgen, z.B. ausgehend von Auffälligkeiten im Monitoring oder vom Informationssicherheitsvorfallmanagement.

In Kapitel 3.1.4- Dynamisches Regelwerk wurde bereits auf die dynamische Regelauswertung, Aktualisierung von Referenzwerten und Anpassungen des Regelwerkes eingegangen. Die dynamische Regelauswertung erfolgt im Rahmen der Zugriffsentscheidung im PDP. Die für eine Zugriffsentscheidung herangezogenen Informationen sollten ins Monitoring einfließen (s. auch Kapitel 3.5.4) und bilden dort die Basis für die Überwachung und Analyse des Regelwerks in der Verantwortung der gematik. Darauf aufbauend kann eine anlassbezogene Aktualisierung von Referenzwerten und die Änderung, Löschung bzw. Erstellung von Regeln vorgenommen werden. Während die Aktualisierung der Referenzwerte in der Verantwortung der gematik liegt, müssen die Erstellung, Änderung und Löschung von Regeln mit den zur Erstellung des Regelwerkes notwendigen Expertengruppen gemeinsam durchgeführt werden.

5.2.2.3 Weiterentwicklung des Regelwerks

Die Weiterentwicklung des Regelwerkes sollte ebenfalls einem regelmäßigen Managementprozess unterliegen, der den Status des Regelwerks (z.B. in Form definierter KPIs) und die Anfragen relevanter Stakeholder (z.B. Fachdienste Hersteller, IDP, BSI und BfDI) berücksichtigt. Dabei lässt sich unterscheiden in einen strategischen Lifecycle- und Change-Management-Prozess und einen anlassbezogenen Ad-hoc-Prozess bspw. aufgrund einer aktuellen Bedrohungslage.

- **Regelmäßige Aktualisierung im Rahmen des Lifecycle-Managements**: Im Zuge des Lifecycle-Managements ist eine kontinuierliche, langfristige (Risiko-)Analyse des Regelwerks und der Regelwerksattribute notwendig. Dies ermöglicht einerseits eine fokussierte Optimierung der Regeln auf Wirksamkeit und Performance und stärkt andererseits die proaktive Adressierung von Bedrohungen, indem z.B. alte und unsichere Softwareversionen ausgesperrt werden. Die Überarbeitung läuft entsprechend einem vordefinierten organisatorischen Prozess in regelmäßigen Abständen, z.B. jährlich, quartalsweise, monatlich.
- **Anlassbezogene Aktualisierung**: Wenn aus dem Security-Monitoring (s. Abbildung 5.2-2) Auffälligkeiten bekannt werden, die eine dringende Änderung des Regelwerks verlangen (z.B. Change mit unerwarteten Nebenwirkungen, Notfall oder Informationen über neue Sicherheitslücken), so müssen Anpassungen kurzfristig und außerhalb des Lifecyclemanagements durch die gematik erfolgen können. Sollten die normalen Update-Zyklen für das Regelwerk aus den PDP für

derartige Situationen nicht ausreichen, muss hierzu ein expliziter Notifizierungsmechanismus etabliert werden.

Bei der Weiterentwicklung des Regelwerks werden Aufbau, Akzeptanzund Wirksamkeit von Regelwerk und Zugriffsprofilen regelmäßig im Kontext aktueller Entwicklungen von Technologien und Bedrohungslage betrachtet und wenn nötig angepasst. Auslöser für Änderungen sind beispielsweise:

- die Einführung neuer Anwendungen bzw. neuer Releases mit neuen bzw. geänderten Ressourcen
- Aufnahme neuer Dienste / Anbieter
- Bekanntwerden von Risiken / Schwachstellen
- Geänderte technologische Rahmenbedingungen, wie z. B. neue Releases von Betriebssystemen oder neue Möglichkeiten der Detektion von Umgebungsparametern

Die Überarbeitung erfolgt dabei über einem Change Request Prozess: In einem Change Request muss das Ziel der Änderung beschrieben werden. Anhand dessen wird eine fachliche Analyse durchgeführt, um herauszufinden welche Komponenten beteiligt sind und wie weitreichend die Änderung ist. Auf dieser Basis werden Expertengruppen einbezogen, um gemeinsam die neuen Regeln bzw. die Regelanpassungen, wie bereits in Kapitel 3.5.4 beschrieben, auszuformulieren. Der Gesamtprozess bis hin zur Koordination der Simulation, Test und Rollout liegt in der Verantwortung der gematik.

Änderungen am Regelwerk können auf zwei Ebenen vorgenommen werden: einerseits am Regelwerk selbst, andererseits an den Referenzwerten, die bei der Regelevaluation herangezogen werden (Listen zulässiger Betriebssystemversionen etc.). Abhängig von Kriterien wie Auslöser, Dringlichkeit, Kritikalität, Auswirkung und Reichweite der Änderung ist im Einzelfall festzulegen, wer in den Freigabeprozess einzubeziehen ist. So ist beispielsweise denkbar, dass die gematik ohne umfassende Abstimmung mit Gremien, einzelne, obsolet gewordene und im Monitoring nicht mehr wahrgenommene Client-Gerätetypen aus Allow-Listen streichen darf, wohingegen die Einführung der Bindung bestimmter Ressourcenzugriffe an Rollen mit den Betroffenen abgestimmt werden muss. Ein Kriterienkatalog für die Entscheidung bei der Wahl des Freigabeverfahrens ist auf der Grundlage der Erfahrungen aus der initialen Regelerstellung abzuleiten. Art und Umfang der Änderungen bestimmen auch, wie und an wen die Änderung über die unmittelbar involvierten Parteien hinaus zu kommunizieren ist.

Abbildung 5.2.2.3-1 fasst abschließend noch einmal zusammen, über welche Wege das Regelwerk oder Teile davon erstellt und aktualisiert werden können.

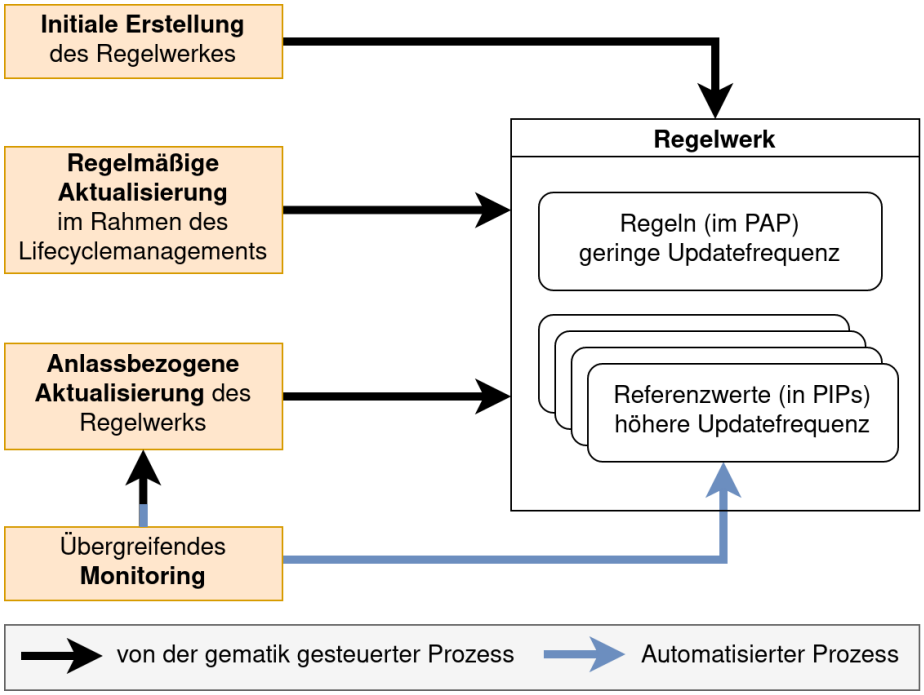

Abbildung 5.2.2.3-1: Update-Prozesse für das Regelwerk

5.2.2.4 Verifikation und Validierung des Regelwerks

Sämtliche Regeln und Referenzwerte durchlaufen einen Prozess der technischen und fachlichen Prüfung hinsichtlich ihrer Wirksamkeit und ihrer Zielstellungen sowie einen Prozess der formalen Freigabe. Dazu werden die Regeln und Referenzwerte in einer Referenzumgebung der TI ausgerollt und getestet. Die Ergebnisse der Prüfungen werden von der gematik und relevanten Stakeholdern freigegeben. Erst dann erfolgt ein Rollout der Regeln in die Produktivumgebung.

5.2.2.5 Verteilung des Regelwerks

Für die Überführung von Regeln und Referenzwerten in den produktiven Betrieb werden voraussichtlich unterschiedliche Verfahrensmodi benötigt - abhängig von Art, Umfang und damit Auswirkung der Änderung. Welcher der Verfahrensansätze zu wählen ist, muss im Einzelfall von der gematik entschieden werden.

Die gematik sollte in die Lage sein, Änderungen an Regeln und insbesondere Referenzwerten **kurzfristig für alle** Fachdienste (und deren PDP) **durchzusetzen**. Betreiber der Fachdienste würden dabei lediglich informiert. Eine explizite Mitwirkung der Betreiber für das Wirksamwerden der Änderung ist nicht erforderlich. Damit kann beispielsweise in Notfällen Bedrohungsszenarien flächendeckend begegnet werden, um z.B. Clientgeräte mit kritischen Sicherheitslücken temporär vom Zugriff auf die Fachdienste auszuschließen.

Alternativ muss die Möglichkeit bestehen, Änderungen am Regelwerk und an Referenzwerten **selektiv bzw. schrittweise und unter Einbeziehung der Fachdienstbetreiber** (in die Planung und die Freigabeentscheidung) produktiv zu setzen. Das kann insbesondere dann sinnvoll sein, wenn geplante Änderungen weitreichende Auswirkungen auf die Zugriffsentscheidungen haben, wie z.B. eine spürbare Verschlechterung der Usability durch gehäufte Authentisierungsaufforderungen, Notwendigkeit der Aktualisierung der Anwendungen oder gar den Ausschluss größerer Nutzerkreise. Derartige Änderungen müssen entsprechend vorbereitet und gegenüber den Betroffenen kommunikativ begleitet werden.

Ein abgestufter Rollout von Regeln bzw. Referenzwerten an bestimmte PDPs / Fachdienste, oder auch selektiv für bestimmte Nutzer und Nutzergruppen, kann genutzt werden, um die Wirksamkeit der Änderungen in der Produktivumgebung in kleinem, definiertem Rahmen (z. B. mit "Friendly Users") vor einem Flächenrollout zu erproben.

5.2.3 Sicherheits- und Betriebsmonitoring

In Kapitel 3.5.6 wird das Monitoring bereits funktional beschrieben. Es sammelt und aggregiert Informationen zum laufenden Betrieb und zur aktuellen Sicherheit der ZTA. Dieses übergreifende Monitoring ermöglicht der gematik die Überwachung aller Komponenten und das Reporting gegenüber den Gesellschaftern, Nutzern und der Öffentlichkeit. Es stellt die Basis für die betriebliche Governance und die Durchsetzung der vereinbarten bzw. spezifizierten SLAs dar.

Sowohl die Schnittstellenspezifikation als auch die Entwicklung und der Betrieb des Monitorings liegen in der Verantwortung der gematik. Die Bedienung der Schnittstellen wird mit der Zulassung durchgesetzt. Damit ist gewährleistet, dass der zugelassene Dienst im Betrieb die spezifizierten Informationen entweder zum Abruf oder aktiv bereitstellt.

Die Informationen werden im Monitoring gesammelt und aggregiert und anschließend für Analysen und Entscheidungen vom SOC-Team herangezogen. Das Monitoring wird, wie in Abbildung 5.2-2 "Monitoring Input für Lifecycle-, Change- und Emergency-Management" dargestellt, in drei Varianten unterschieden: Betriebs-, Security- und Regelwerksmonitoring. Die bereitgestellten Informationen aus diesen drei Varianten des Monitorings sind Grundlage für die langfristige Bewertung der Situation und bspw. für die Optimierung des Regelwerks (Betriebsmonitoring und Regelwerksmonitoring im Rahmen des Lifecycle-Managements), für kurzfristige Reaktionen auf Ausfälle (Betriebsmonitoring) oder für das Auslösen einer Reaktion aufgrund einer aktuellen Bedrohungslage (Security-Monitoring im Rahmen des Emergency Managements). Auf die einzelnen Varianten wird in den nachfolgenden Kapiteln näher eingegangen.

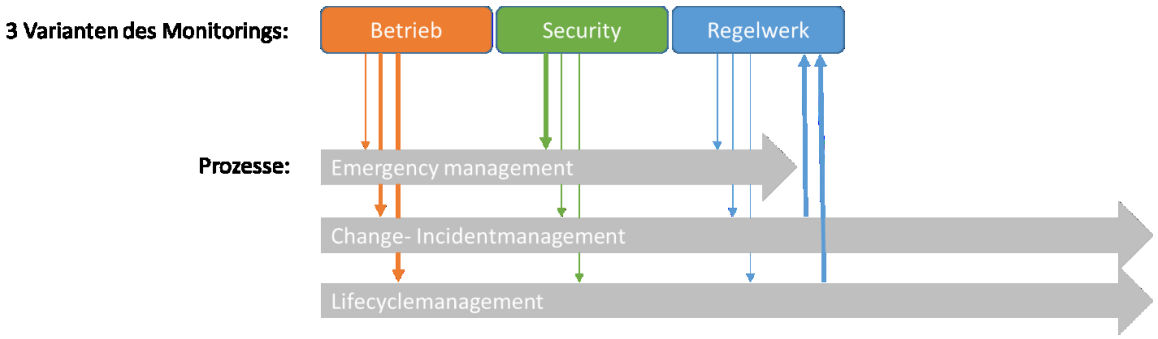

Abbildung 5.2-2: Monitoring Input für Lifecycle-, Change- und Emergency-Management

5.2.3.1 Betriebsmonitoring

Das Betriebsmonitoring liefert aggregierte Informationen zum aktuellen Betriebsstatus aller ZTA-Komponenten und Dienste und ggf. Alarmmeldungen an die gematik. Darüber hinaus kann die gematik Dienste auch aktiv abfragen, um ihren Status festzustellen. Nutzern könnte in etwas reduzierter Detaillierung der aktuelle Betriebsstatus transparent im Nutzerportal dargestellt werden. Anbieter und Betreiber sollten zusätzlich zu ihrem eigenen Monitoring individuelle dienstbezogene Sichten erhalten, die Laufzeitinformationen und Abhängigkeiten ihrer Dienste von Diensten anderer Anbieter oder Betreiber darstellen. Die gematik stellt den Anbietern, über ihre eigenen Dienste hinaus, eine gesamtheitliche Sicht im Anbieterportal bereit. So könnten Abhängigkeiten

zwischen Diensten bzw. übergreifende Probleme besser identifiziert werden. Mit dieser systemweiten Sicht lassen sich Laufzeitprobleme sowie Ausfälle erkennen und ggf. nachsteuern. Das Nachsteuern erfolgt entweder automatisiert, teilautomatisiert oder manuell unter Einhaltung eines für diesen Fall (z.B. Erweiterung der Anbieter für einen Dienst, Ausfall einer ZTA-Komponente) definierten Prozesses (s. Kapitel 5.2.1- On- und Off-Boarding von Diensten). Alle automatisierten Anpassungen am System müssen zur Nachvollziehbarkeit protokolliert werden. Darüber hinaus ermöglicht das Betriebsmonitoring

- die Anzeige von Betriebsstatus und Verfügbarkeit auf Anwendungsebene der einzelnen Komponenten / Dienste / Nodes / Schnittstellen
- die Überwachung der Performance, der Auslastung und des Security Status (Auswertung und Aggregation auf Basis von sicherheitsrelevanten Parametern oder bekannten und offenen Schwachstellen aus Scans oder CVE-Einträgen)
- die Anzeige des Status und die Überwachung der Verteilung, Nutzung und Aktualität von Regelwerken (Was läuft wo und auf welchem Stand?)

um z.B. die folgenden Reaktionsmöglichkeiten abzuleiten

- Skalierung
- Lastplanung, Überlastbehandlung
- Priorisierung von Diensten (bewusste Steuerung der Zugriffe bspw. aus Testgründen eines neuen Regelwerks oder um Auffälligkeiten entgegenwirken zu können).

5.2.3.2 Security-Monitoring

Das Security-Monitoring umfasst das anonymisierte Monitoring der Zugriffsmuster von Nutzern und der Zugriffsentscheidungen (Kapitel 3.5.4) sowie das Monitoring des Sicherheitszustandes aller TI-Komponenten. Ziel ist es, durch das Monitoring ungewöhnliches Nutzerverhalten oder ungewöhnliche Nutzungen von Fachdiensten erkennen und frühzeitig auf Angriffe bzw. Missbrauch reagieren zu können.

Beispiele eines ungewöhnlichen Verhaltens sind:

- ungewöhnlich gehäufte Zugriffsversuche mit den gleichen Eigenschaften (z.B. IP-Adresse, Nutzer)
- viele fehlgeschlagene Zugriffsversuche aus demselben Grund
- Angriffsversuche z.B. durch Manipulation von Authentisierungs-Token

Es sind Schwellwerte zu bestimmen, ab denen Nutzerverhalten als bedrohlich einzustufen ist und Gegenmaßnahmen eingeleitet werden müssen. Diese Schwellwerte können manuell definiert werden oder, aus bisherigem Verhalten automatisch oder semi-automatisch erlernt werden (Anomalieerkennung). Gegenmaßnahmen können durch manuelle oder automatisierte Updates von Referenzwerten im Regelwerk bewirkt werden (siehe Kapitel 5.2.2).

Durch das Monitoring sicherheitsrelevanter Parameter der TI-Komponenten werden z.B. mögliche Schwachstellen erkannt und angezeigt oder auf möglicherweise bereits erfolgte Angriffe bzw. manipulierte Komponenten hingewiesen. Auf dieser Basis kann mit entsprechenden Maßnahmen (bspw. Aktualisierung der Software) reagiert werden, um weiteren Angriffen vorzubeugen.

Beispiele für mögliche Schwachstellen einer TI-Komponente sind:

- Sicherheitslücken in Software-Komponenten, die aufgrund von ungewöhnlichen Zugriffen entdeckt werden.

- Sicherheitslücken in Software-Komponenten, die z.B. durch Abgleich mit aktuellen Sicherheitsmeldungen gefunden werden
- Veraltete Software bei den Fachdiensten (z.B. Scan ergibt zu niedrige Software-Versionen)
- Sicherheitsvorfälle in der Betriebsumgebung

Möglicher Input für die Bewertung sind Informationen aus den Zulassungen, aus der Attestierung der Fachdienste oder aus den Monitoring-Schnittstellen der Komponenten. Über das SIEM werden Monitoringdaten aus den ZTA-Komponenten gesammelt, aggregiert, analysiert und visualisiert. Darauf aufbauend identifiziert das SOC-Team sicherheitsrelevante Ereignisse und reagiert auf Bedrohungslagen. Dabei wird jedes Mal ein organisatorischer Prozess ausgelöst: Ein Gremium aus Experten bewertet das gemeldete ungewöhnliche Verhalten sofort bzw. je nach Schwere oder Ausmaß im Rahmen eines mehrstufigen Verfahrens. Das kann insbesondere bei höherer Komplexität der Fall sein, bspw., wenn viele Dienste in unterschiedlichen Detailtiefen betroffen sind. Es entscheidet über weitere nötige Maßnahmen, z.B. betroffene Betreiber oder Hersteller informieren, betroffene Nutzer informieren, Abschottungen von Diensten oder Ausschluss von Nutzern oder Geräten über IDP/GMS sein.

5.2.3.3 Monitoring des Regelwerks

In Kapitel 3.5.4 (zu FA1.2) ist dargestellt, wie die Effektivität und Angemessenheit des Regelwerkes überwacht werden können. Dazu ist es nötig, die im PEP/PDP anfallenden Informationen über Zugriffsentscheidungen im ersten Schritt automatisiert zu aggregieren, darauf aufbauend Auffälligkeiten und Trends zu analysieren, um schließlich Änderungsbedarf festzustellen und Änderungsprozesse anzustoßen. Diese Tätigkeiten werden von einem zu etablierenden Analyseteam bzw. Entscheidungsgremium durchgeführt. Das Monitoring des Regelwerks ist unterstützend für die Weiterentwicklung des Regelwerks (s. Kapitel 5.2.2.3) im Rahmen der Lifecycle- und Changemanagement Prozesse sowie anlassbezogenen Änderungen des Regelwerkes. Dabei werden zum einen aktuelle Bedrohungslagen wie z.B. neu entdeckte Sicherheitslücken einbezogen, die ggf. eine kurzfristige anlassbezogene Modifikation des Regelwerks nötig machen. Zum anderen werden im Rahmen des Lifecyclemanagements technologische Entwicklungen z.B. bei den Nutzerendgeräten oder Veränderungen der Markt- und Bedrohungslage betrachtet und davon ausgehend mittel- bzw. längerfristige Änderungsprozesse angestoßen.

5.2.4 Support

Jeder Hersteller und Anbieter muss für seinen Dienst bzw. sein Produkt Support leisten. Es besteht ein Vertragsverhältnis zwischen den jeweiligen Parteien, bspw. zwischen Nutzer und Fach-Client-Anbieter durch Bezug der Software. In diesem ist geregelt, wer in welchem Umfang Support leistet. Ziel ist es, dass der Nutzer genau weiß, wen er im Supportfall kontaktieren kann und dort auch Hilfe bekommt. Das kann am besten vom Anbieter des Frontend geleistet werden. Die zentralen ZTA-Dienste müssen von der gematik supportet bzw. der Support organisiert werden. Zur Erreichbarkeit des Supports sollte eine Hotline und Helpdesk eingerichtet werden. Über eine Statusübersicht zu den ZTA-Diensten kann der Hersteller/Anbieter/Betreiber sich selbst informieren, ob der zentrale ZTA-Dienst online ist, bevor ggf. ein Supportvorfall eröffnet wird.

Es ist davon auszugehen, dass Nutzer sich mit Support-Bedarf immer zunächst an den Anbieter des Fach-Clients wenden. Die Möglichkeit der Kontaktaufnahme wird beispielsweise direkt in einer App angeboten. Der Anbieter selbst pflegt in der Regel den Kontakt zu den Nutzern bzw. delegiert dies an von ihm beauftragte Dienstleister. Nutzer wenden sich mit allen Supportbedarfen an den Anbieter des Fach-Clients.

Im Folgenden werden zwei Ansätze für den Support vorgeschlagen, bei denen eine zentrale Instanz (gematik bzw. von ihr beauftragt) entweder eine nur bei Bedarf koordinierende Rolle im Hintergrund einnimmt oder eine deutlich führendere Rolle. Welche dieser Ansätze realisierbar und gewünscht ist hängt auch davon ab, welche zukünftige Rolle die gematik aus politischer Sicht einnimmt und wie diese Rolle mit Ressourcen und Durchsetzungskraft gegenüber den anderen Playern in der TI hinterlegt wird.

Ansatz 1: Supportnetzwerk

Supportbedarf zu allen in eine Anwendung eingebundene Komponenten und Diensten wird vom Nutzer zunächst beim Anbieter der Anwendung angemeldet und muss von diesem einer Klärung zugeführt werden. Das bedeutet beispielsweise (wie mit Abbildung 5.2-2 illustriert), dass der Anbieter einer App "A", der den User Help Desk für die entsprechende Anwendung aus Sicht des Nutzers in Gänze anbietet, auch Support bieten muss, wenn sich Probleme ergeben, mit:

- einem eingebundenen Fachdienst eines dritten Fachdienstanbieters "C",
- dem zugelieferten Trust-Client (z. B. Open Source hergestellt im Auftrag der gematik),
- den Diensten der ZTA-Plattform
- dem von Nutzer verwendeten IDP, der ggf. nutzerindividuell abhängig von des Versicherten Krankenversicherung ausgeprägt ist

Der UHD des Anbieters des Fach-Clients muss dazu Servicebeziehungen zu diesen Parteien pflegen und zumindest so viel Lösungskompetenz zu den einzelnen Komponenten und Diensten aufbauen, um den Supportbedarf bei der Koordination der Lösungsfindung korrekt adressieren zu können. Service-Level und Reaktionszeiten der einzelnen Beteiligten können über die Anbieterzulassung reguliert werden.

Die gematik sollte als zentrale Eskalationsinstanz ansprechbar sein. Nur sie ist in Zweifelsfällen in der Lage, auf Basis der Anbieterzulassungsverträge Mitwirkungspflichten durchzusetzen. Die gematik sollte, z. B. als Teil des Anbieterportals, mit den aggregierten Ergebnissen des Monitorings, jederzeit ein Lagebild anbieten, auf dessen Basis der UHD selbstständig eine Erstanalyse von Problemlagen vornehmen kann.

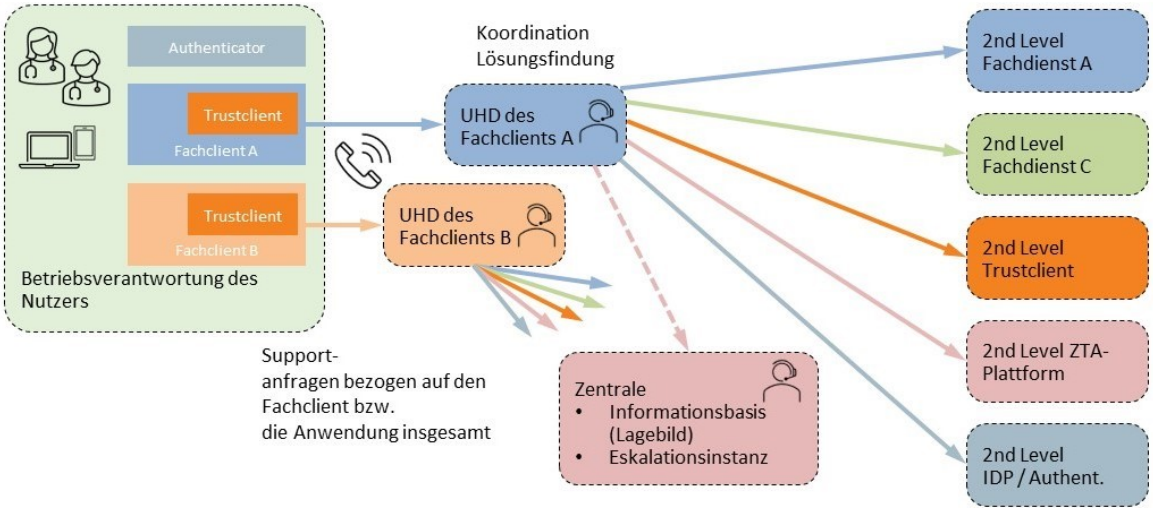

Abbildung 5.2-2: Organisation Nutzersupport im Vernetzungsmodell

Ansatz 2: zentrale Supportinstanz

Der in Abbildung 5.2-2 vorgestellte Ansatz 1 der Bildung von Supportketten stellt insbesondere kleine Anbieter von Fach-Clients vor die Herausforderung der individuellen Organisation der Supportbeziehungen mit vielfältigen Anbietern und Herstellern. Der folgende Alternativvorschlag für die Organisation der Supportkette vereinfacht diese Beziehungen.

Ist der Hersteller mit seiner eigenen Organisation nicht in der Lage den Supportbedarf zu befriedigen, weil externe eingebundene Dienste bzw. Komponenten ursächlich sind, wendet sich der Hersteller an eine zentrale Supportinstanz, die die Koordination der Lösungsfindung bzw. Lösung gesamtheitlich übernimmt. Mit dieser zentralen Instanz ist es für Hersteller von Fach-Clients nicht erforderlich, Servicebeziehungen mit der Vielzahl von Anbietern (Fachdienste, IDP, ZTA-Plattform, Trust-Client) zu unterhalten, die in der Kette der Dienste und Komponenten der Gesamtanwendung eingebunden sein können. Dies senkt insbesondere für kleinere Anbieter von Anwendungen die Schwelle für den Einstieg in die TI und fördert damit u. U. das Innovationspotential der TI als Ökosystem. Hersteller können durch das Zusammenführen der Angebote unterschiedlicher Fachdienste über den Fach-Client innovative und integrierte Anwendungserfahrungen für den Nutzer schaffen. Die zentrale Support-Instanz fungiert dabei primär als Enabler- bzw. Eskalationsinstanz, die es insbesondere auch kleinen und neuen Markteilnehmern ermöglicht, gegenüber ihren Kunden Ende-zu-Ende-Lösungsverantwortung zu übernehmen, ohne umfassende Vertrags- und Servicebeziehungen mit dritten Diensteanbietern eingehen zu müssen.

Wie mit dem "Shortcut" in Abbildung 5.2-3 angedeutet, muss nicht jede Supportanfrage durch die zentrale Support-Instanz verarbeitet werden. Um hier den Effekt eines "Flaschenhalses" oder "Durchlauferhitzers" im Prozess zu vermeiden, sollte die Funktion tatsächlich nur eingebunden werden, wenn Hersteller bzw. Anbieter, Supportanfragen nicht aus eigener Kraft, mit vorhandenen Servicebeziehungen lösen können.

Idealerweise wird die Rolle der zentralen koordinierenden Service-Instanz von der gematik oder in deren unmittelbarem Auftrag ausgefüllt, da diese mit allen Herstellern und Anbietern Service- bzw. Vertragsbeziehungen unterhält - entweder aus der Zulassung oder der direkten Beauftragung. Außerdem besitzt die gematik mit den Daten des Betriebsmonitorings den besten Einblick in den Zustand der einzelnen Dienste und den Überblick über den Gesamtzustand der TI.

Den Kontaktweg und damit auch die Servicequalität gegenüber dem Nutzer können die Hersteller der Apps gemäß ihren Bedarfen und Fähigkeiten individuell gestalten (die Telefonsymbolik in der Abbildung "5.2-2: Organisation Nutzersupport" dient der Anschaulichkeit und ist beispielhaft). Für die Entgegennahme und Weiterverarbeitung bzw. Weitergabe der Supportanfragen durch den zentralen Support (also für die Verwaltung der Vorgänge bzw. Tickets) empfiehlt sich der Einsatz eines Ticketsystems. Das Selfserviceportal des Nutzers kann neben Telefonhotlines eine Schnittstelle bieten, über die die Nutzer an dieses System angeschlossen sind.

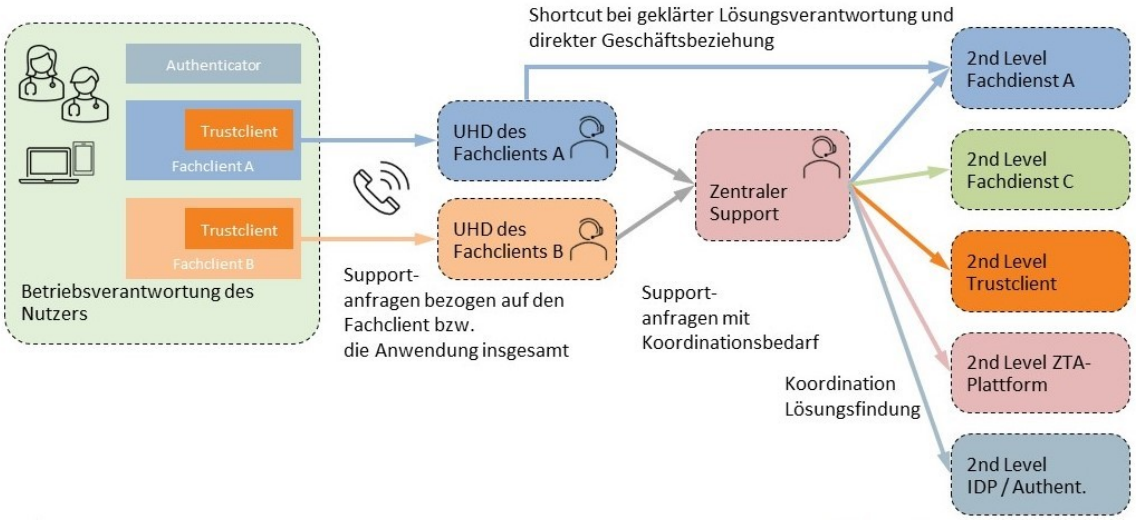

Abbildung 5.2-3: Organisation mit zentralisiertem Nutzersupport

5.2.5 Übergreifendes IT-Servicemanagement (ITSM)

Die Prozesse innerhalb derer die Dienste und Komponenten der TI 2.0 im laufenden Betrieb gemanagt werden, orientieren sich an dem Best-Practice-Leitfaden ITIL und haben Aufgabe wie:

- Änderungen einzubringen
- Incidents und Probleme zu Lösungen zu führen.
- Service-Management zu betreiben
- Konfigurationsmanagement
- etc.

Die Prozesse sollen hier nicht vertieft werden. Die Abbildung "ITSM" zeigt lediglich exemplarisch und stark vereinfacht, anhand des Incident-Managements, das Zusammenspiel unterschiedlicher Hersteller und Anbieter, für die übergreifende Durchführung der ITSM-Prozesse.

Kern des Vorschlags ist, dass alle innerhalb der TI zu verarbeitenden Service-Vorfälle, die ein Anbieter oder Hersteller nicht lokal verarbeiten kann, an eine zentrale TI-Serviceplattform übergeben werden. Vorgänge werden designierten Adressen möglichst automatisiert zugeordnet, z. B. anhand fester Prozesstemplates mit a priori definierten Verantwortlichen. Fälle, in denen dies nicht möglich ist, müssen von einer Eskalationsinstanz koordiniert werden. Idealerweise wird diese Rolle von der gematik ausgefüllt. Aufgrund der Tatsache, dass die gematik für den Großteil der Dienste der ZTA verantwortlich ist, wird sie ohnehin als direkt verantwortliche Instanz in den Großteil der übergreifend wirksamen ITSM-Vorgänge eingebunden sein. Bei diesen Diensten handelt es sich um diejenigen, von denen Fachdienstbetreiber im Wesentlichen abhängig sind.

Es kann als wesentlicher Vorteil der ZTA gewertet werden, dass Abhängigkeiten zwischen Diensten von Marktteilnehmern weitgehend aufgelöst sind. Übergreifend wirkende Service-Prozesse wie z. B. Incidents, werden in der Regel zwischen dem betroffenen Fachdienst und den von der gematik verantworteten PAP, PIP, Monitoring und den IDP zu koordinieren sein. Es sollte angestrebt werden, dass dritte Organisationen in unmittelbare Vertragsverhältnisse mit der gematik eintreten, in deren Rahmen klare Servicebeziehungen mit geeigneten Mitteln der Durchsetzung (z. B. Pönale) vereinbart werden. Insbesondere für Dienste wie die IDP, von denen Anwendungen

elementar abhängig sind, ist eine engere und direktere Mitwirkung als im bisherigen Anbietervertragsverhältnis erforderlich. Zukünftige Gesundheitsanwendungen der TI werden nicht mehr pauschal über eigene Servicestrukturen der Anwender (DVO etc.) verfügen oder längere (anbieterübergreifende) Störungen über Papier-Ersatzverfahren kompensieren können. Allein deshalb ist eine effektive Koordination zwischen allen Beteiligten erforderlich.

Die Vorgangsverarbeitung sollte dabei über ein zentrales ITSM-System von der gematik gesteuert werden: das heißt beispielsweise, dass die lokal beim Anbieter erfassten Service-Tickets über geeignete Schnittstellen in das zentrale ITSM überführt werden müssen. Dort werden diese Tickets bestandsführend bearbeitet, sobald der übergreifende Charakter bestätigt ist. Betroffene bzw. an der Bearbeitung beteiligte Anbieter müssen das zentrale ITSM-System nutzen bzw. sich geeignet anschließen, um Informationen und Steuerungsflüsse (Statusübergänge, Aufträge etc.) wahrzunehmen und in ihre interne Arbeitsorganisation zu übernehmen.

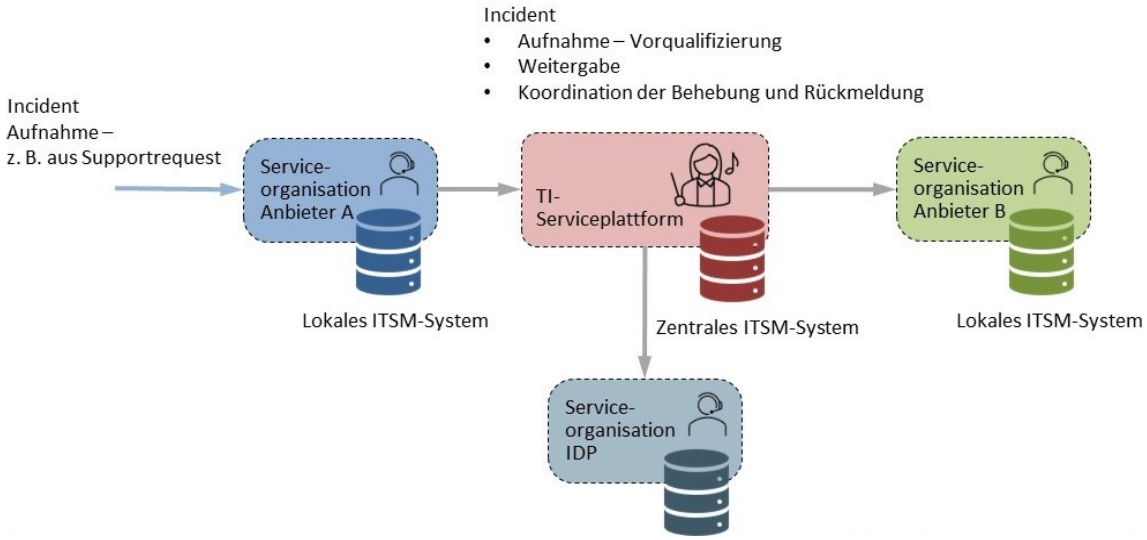

Abbildung 5.2-4: Organisation ITSM

Mit dem Ansatz wird eine Vereinheitlichung der Leistungs- und Serviceerbringung über das ITSM-System erreicht. Er führt zu kurzen Kommunikationswegen ohne Medienbrüche. Übergreifende Vorfälle werden ebenfalls in demselben System abgebildet. Über automatische Zuweisungsmechanismen wird sichergestellt, dass Informationen zu eingegebenen Service-Tickets oder übergreifenden Incident-, Problem- und Change-Tickets an die entsprechenden Verantwortlichen weitergegeben werden. Dies führt zu einer stets aktuellen Informationslage für den aktuellen oder späteren Bearbeiter.

Die Prozesse des Supports und des ITSM werden ihrerseits kontinuierlich überwacht und anhand typischer Kennzahlen (Durchlaufzeiten, etc.) gesteuert und optimiert.

5.2.6 Administrative Prozesse der Nutzer

Die Nutzer der ZTA bekommen die Möglichkeit ihre Zugriffsmöglichkeiten selbst zu verwalten. Nutzer müssen sich authentisieren, ihre Endgeräte einrichten sowie Standorte organisatorisch und technisch attestierbar machen. Die Nutzer benötigen in der ZTA gültige Identitäten und Authentisierungsmittel zum Zugang zu ihren Identitäten.

Als Frontend für die nachfolgend beispielhaft aufgelisteten administrativen Prozesse der Nutzer wird ein Nutzerportal zur Verfügung gestellt. Im Kontext der ZTA-Komponenten werden besonders Funktionen zum Gerätemanagement benötigt, insbesondere die

Registrierung neuer Geräte, das Verwaltung der registrierten Geräte und das Entfernen oder Sperren nicht mehr verwendeter Geräte.

Weitere Prozesse, wie die Verwaltung von Identitätsdaten, Einsichtnahme in vergangene Anmeldungen sowie das Management von Authentisierungsmitteln sind Funktionalitäten des IDP. Ob diese Prozesse in das Portal integriert werden können, ist zu prüfen.

Aus Nutzersicht ist ein zentraler Anlaufpunkt mit folgenden Funktionen wünschenswert:

- Einsicht in die letzten Zugriffe im Sinne der Transparenz für den Nutzer
  - Anzeige der Zugriffe auf Dienste und ggf. von Änderungen/Autorisierungen durch den Nutzer
  - Anzeige der Zugriffe auf das Nutzerportal und ggf. der Änderungshistorie
- Konfiguration der Vertreterregelung
- Identitätsdatenmanagement
  - Änderung von Personendaten (Name, Anschrift etc.)
- Verwaltung der Zugangsdaten zur Authentisierung an Diensten
  - Benutzername / Passwort
  - zusätzliche bzw. alternative Authentisierungsfaktoren wie YubiKey, FiDo Token oder Biometrie (Gesicht, Fingerabdruck)
- Anpassung von nutzerspezifischen Regelwerkattributen
  - Geolocation-Steuerung, bspw. kein Zugriff auf die ePA aus dem Ausland u.a. vgl. "FA1.5 Nutzerspezifische Regelwerkattribute anpassen"
  - zeitbasierte Beschränkungen
- individuelle Festlegung des Sicherheitsniveaus pro Dienst, da das Sicherheitsempfinden sehr subjektiv ist, beispielsweise
  - ePA hoch
  - e-Rezept normal
  - Fitness- und Ernährungsapp niedrig
- Einwilligungsmanagement - Einwilligung erteilen / ändern / löschen
  - Behandlungsverträge
  - Nutzung SSO
  - niederschwellige Verfahren (Biometrie) zur Erhöhung der Usability
  - Datenweitergabe und -nutzung z.B. Telemetriedaten oder Gesundheitsdaten zu Forschungszwecken

Das Portal muss weiterentwickelt werden können, wenn sich weitere Bedarfe nach Administrationsmöglichkeiten durch den Endnutzer ergeben.

6 Evaluation

In Kapitel 2.2 sind Anwendungsfälle benannt, welche hinsichtlich der Konzeption der Architektur im Vordergrund stehen und aus denen die funktionalen Anforderungen an die Architektur abgeleitet wurden. Im folgenden Kapitel 6.1 wird nun dargelegt, wie diese Anwendungsfälle im Kontext von Gut- und Fehlerfällen mit der in den vorangehenden Kapiteln beschriebenen Architektur funktionieren. Im Anschluss wird in Kapitel 6.2 evaluiert, inwiefern die beschriebene Architektur die in Kapitel 2.4 dokumentierten, nicht-funktionalen Anforderungen erfüllt.

6.1 Anwendungsfälle

In diesem Kapitel werden Beispielanwendungen beschrieben, die die Abläufe der Zero-Trust-Mechanismen deutlich machen. Dafür werden zuerst Gutfälle beschrieben, welche die Abläufe der Anwendungsfälle mit einem positiven Verlauf darstellen. Im Anschluss werden Fehlerfälle beschrieben. Mit diesen Fehlerfällen werden relevante Use Cases dargestellt, bei denen z.B. ein Zugriff auf die TI nicht erfolgreich realisiert werden kann.

Abweichend zu den in der Leistungsbeschreibung und im Kapitel 2.2- Anwendungsfälle aufgeführten 7 Use Cases gehen wir nach Abstimmung mit der gematik in diesem Kapitel auf 4 Gutfälle und Fehlerfälle aus 5 Fehlerklassen ein. Damit werden die Abläufe in der Zero-Trust-Architektur deutlicher dargestellt und die Lesbarkeit verbessert.

Die Use Cases wurden dabei wie folgt zu 4 Gutfällen zusammengefasst:

- Gutfall für mobiles Gerät: Zugriff eines Leistungserbringers auf die ePA eines Versicherten adressiert UC2 und UC4, da sich lesender und schreibender Zugriff im Regelfall lediglich durch die anzuwendenden Regeln aus dem Regelwerks unterscheiden.
- Gutfall für stationäres Gerät: Zugriff eines Leistungserbringers auf die ePA eines Versicherten im Praxisumfeld adressiert UC1 und UC3.
- Gutfall: Zugriff des Versicherten über ein mobiles Endgerät auf seine ePA beschreibt UC6 und zeigt, dass sich die Use Cases UC5 und UC6 von den entsprechenden Gutfällen für LEs nur dadurch unterscheiden, dass ggf. andere Authentisierungsmittel und Regeln im Regelwerk verwendet werden.
- Gutfall: Ein Fachdienst kommuniziert mit einem anderen Fachdienst der TI adressiert UC7.

Jeder dieser Gutfälle besteht aus den in Abbildung 6.1-1 dargestellten Unter-Use-Cases.

Als Fehlerfälle wurden basierend auf den in der Abbildung dargestellten 5 Fehlerfall-Klassen 7 Beispiele ausgewählt und näher beschrieben.

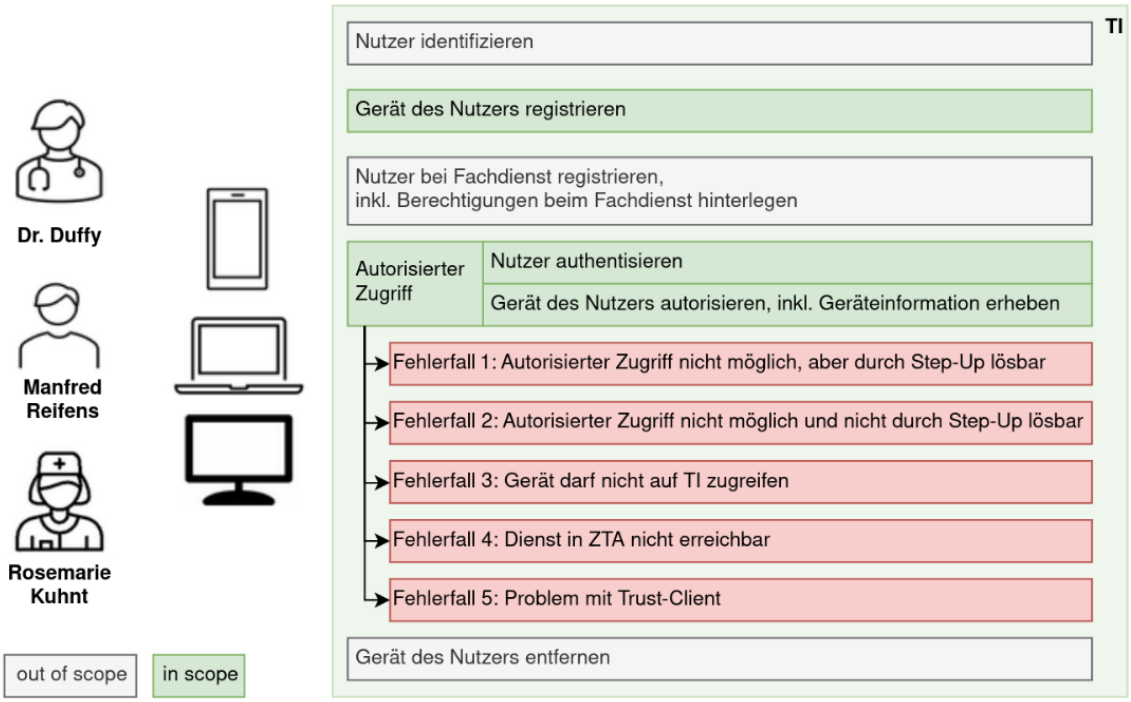

Abbildung 6.1-1 Übersicht über die Use Cases und Fehlerfälle

In den nachfolgenden Anwendungsfällen kommen 3 Personen mit unterschiedlichen Rollen vor. Abbildung 6.1-2 stellt eine Übersicht über diese Personen dar.

Name	Dr. Erik Duffy	Manfred Reifens	Rosemarie Kuhnt
Alter	44 Jahre	56 Jahre	29 Jahre
Rolle	Arzt	Patient	Krankenpflegerin

Abbildung 6.1-2 Übersicht über die in den Anwendungsfällen verwendeten Personen

6.1.1 Gutfall für mobiles Gerät: Zugriff eines Leistungserbringers auf die ePA eines Versicherten

Annahmen

- Primäridentität erhalten: ***Dr. Duffy*** (Nachweis elektronische Funktion des Personalausweises)
- Rolle bezogen: Arzt (HBA mit entsprechender Profession OID)
- Patient: ***Manfred Reifens***
- Fachdienst: EPA
- Nötige Geräte
  - Mobiles Endgerät mit WLAN-Access oder mit mobilen Daten ins Internet und NFC-Schnittstelle für Interaktion mit HBA (Handy: ***Duffys iPhone***)
- Nötige Software
  - ePA-App mit Arzt Modul: APP ePA 2.0 (Hersteller: Industrie)
  - Authenticator-Modul (AUM): KVNO Ident (Hersteller: KV)
  - Trust-Client (TCL): TI connect Version 2.044 (Hersteller: gematik)
- Nötige Zugangsmittel
  - HBA + PIN

1. Use Case: Nutzer identifizieren

Die Identifizierung der Personen mit mindestens einer Rolle beim IDP wurde erfolgreich durchgeführt, um die Authentisierung des Nutzers gemäß Funktion FA2.1 Nutzer authentifizieren zu ermöglichen.

1. Registrierung des Nutzers und seiner Rolle(n) (z.B. Arzt, Versicherter, medizinische Einrichtung,) bei einem IDP der Föderation der TI inkl. Identifizierung
2. Verfügbarkeit des zugehörenden Authentisierungsmittels (Authenticator auf Mobiltelefon, ggf. eGK, HBA, SMC-B), siehe FA2.3 eID-Lifecycle IDP

2. Use Case: Gerät des Nutzers registrieren

Im Folgenden wird beschrieben, wie ein mobiles Gerät gemäß Funktion 5.1 "Gerät registrieren" beim GMS für die Verwendung in der ZTA registriert werden kann. Die Schritte für eine Geräteregistrierung sind:

1. ***Dr. Duffy*** installiert und öffnet die ePA-App auf seinem Gerät ***Duffys iPhone.***
2. ***Dr. Duffy*** bekommt einen Hinweis: „Wir prüfen nun, ob Sie berechtigt sind auf die ePA zuzugreifen. Sind Sie Leistungserbringer oder Patient? Bitte wählen Sie!"
3. ***Dr. Duffy*** bekommt mögliche Nutzeridentifikationsmittel angezeigt und wählt "HBA" aus.
4. ***Dr. Duffy*** bekommt in der ePA-App folgende Nachricht angezeigt: "Bitte HBA an das Gerät halten, um Ihre Identität zu bestätigen." und hält seine HBA an das Handy (bis der Vorgang abgeschlossen ist).
5. ***Dr. Duffy*** bekommt in der ePA-App folgende Nachricht angezeigt: "Bitte PIN für HBA eingeben." und tippt diese ein.
6. ***Dr. Duffy*** bekommt in der ePA-App die Info angezeigt: "Berechtigungsprüfung erfolgreich."

3. Use Case: Nutzer beim Fachdienst registrieren

Die Registrierung des Nutzers beim Fachdienst ist gemäß Funktion FA2.2 Nutzer bei Fachdienst registrieren erfolgt, d.h. es gibt eine existierende ePA für den entsprechenden Versicherten Manfred Reifens.

4. Use Case: Autorisierter Zugriff (lesend)

1. Der Leistungserbringer öffnet seine App ePA 2.0. Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden.
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck*: Entspricht ***Duffys iPhone*** den vorgegebenen Bedingungen zum lesenden Zugriff auf die ePA
   b. Vertrauenswürdiges Betriebssystem ✅
   c. Sichere Ausführungsumgebung ✅
   d. System Sicherheit (Gerätesicherheit) ✅
   e. *Hintergrundcheck*: Entsprechen die ***Duffys iPhone*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
3. GEO-Location ✅
4. Systemzeit passt zur TI-Zeit ✅
   a. *Hintergrundcheck*: Hat Dr. Duffy eine aktive Rolle "Allgemeinmedizin"?
5. Wurde die Identität innerhalb der letzten 10 Stunden mit einem Authenticator bestätigt und sein HBA vorgelegt? ✅
6. Anzeige der Daten von ***Manfred Reifens***

5. Use Case: Gerät des Nutzers entfernen

Der Nutzer verwaltet seine App-Geräte-Zuordnung und kann Einträge selbständig löschen.

6.1.2 Gutfall für stationäres Gerät: Zugriff eines Leistungserbringers auf die ePA eines Versicherten im Praxisumfeld

Annahmen

- Primäridentität erhalten: ***Dr. Duffy***
- Rollen-Attribut erhalten: Arzt (HBA mit entsprechender Profession OID)
- Patient: ***Manfred Reifens***
- Fachdienst/Ressource: Medikationsplan in der ePA
- Nötige Geräte
  - Stationärer PC mit LAN-Bindung (PC: ***Behandlung 1***)
  - Mobiles Endgerät mit WLAN-Access oder mit mobilen Daten ins Internet und NFC-Schnittstelle für Interaktion mit HBA (Handy: ***Duffys iPhone***)

- Nötige Software
  - ePA-Anwendung im PVS (Fach-Client mit ePA-Modul) auf ***Behandlung 1***: Primärsystem Roxit (Hersteller: Industrie)
  - Authenticator-Modul (AUM) mit Karten-Lese-Funktion auf Duffys ***iPhone***: KVNO Ident (Hersteller: KV)
  - Trust-Client auf ***Behandlung 1***: TI connect Version 2.144 (Hersteller: gematik)
- Nötige Zugangsmittel
  - HBA + PIN

1. Use Case: Nutzer identifizieren

Die Identifizierung der Personen mit mindestens einer Rolle beim IDP wurde erfolgreich durchgeführt, um die Authentisierung des Nutzers gemäß Funktion FA2.1 Nutzer authentifizieren zu ermöglichen.

1. Registrierung des Nutzers und seiner Rolle(n) (z.B. Arzt, Versicherter, medizinische Einrichtung, ...) bei einem IDP der Föderation der TI inkl. Identifizierung
2. Verfügbarkeit des zugehörenden Authentisierungsmittels (Authenticator auf Mobiltelefon, ggf. eGK, HBA, SMC-B), siehe FA2.3 eID-Lifecycle IDP

2. Use Case: Geräte der Nutzer registrieren

Die Funktionalität orientiert sich an dem bestehenden Mechanismus für die Verwendung eines elektronischen Personalausweises an einem Gerät ohne Kartenleser (Fernzugriff)(Kopplung der AusweisApp2 mit einem Smartphone — AusweisApp2 1.14.2 Dokumentation (bund.de): https://www.ausweisapp.bund.de/ausweisapp2/handbuch/1.14/de/Windows/settings-pairing-with-remote-reader.html). Sollte das stationäre Gerät selbst über einen Kartenleser verfügen, entfällt der Bedarf für ein zweites Gerät und der Ablauf ist vergleichbar mit dem Use Case "Geräte der Nutzer registrieren" beim "Gutfall für mobiles Gerät". Die Schritte für die Geräteregistrierung sind:

1. ***Dr. Duffy*** installiert und öffnet die ePA-Anwendung auf seinem Gerät ***Behandlung 1.***
2. ***Dr. Duffy*** klickt in der ePA-Anwendung auf ***Behandlung 1*** auf und bekommt einen Hinweis: "Wir prüfen nun, ob Sie berechtigt sind auf die ePA zuzugreifen. Sind Sie Leistungserbringer oder Patient? Bitte wählen Sie!"
3. ***Dr. Duffy*** bekommt mögliche Nutzeridentifikationsmittel angezeigt und wählt "HBA" aus.
4. ***Dr. Duffy*** bekommt in der ePA-Anwendung auf ***Behandlung 1*** folgende Nachricht angezeigt: "Bitte verbinden Sie Ihren HBA mit dem Gerät, um Ihre Identität zu bestätigen."
5. ***Dr. Duffy*** wählt in der ePA-Anwendung die Option "Kein Kartenterminal/NFC vorhanden. Zweitgerät für Kommunikation mit HBA nutzen." und bekommt einen QR angezeigt.
6. ***Dr. Duffy*** öffnet auf ***Duffys iPhone Handy*** die Authenticator-App mit Karten-Lese-Funktion für den HBA und scannt den angezeigten QR-Code ab.
7. ***Dr. Duffy*** bekommt in der ePA-Anwendung auf ***Behandlung 1*** folgende Nachricht angezeigt: "Gerät verbunden. Bitte PIN für HBA eingeben." und tippt diese ein.
8. ***Dr. Duffy*** bekommt in der Authenticator-App auf ***Duffys iPhone*** folgende Nachricht angezeigt: "Bitte halten Sie Ihre HBA ans Gerät, um Ihr neues Gerät zu bestätigen." und hält seine HBA an das Handy.

9. ***Dr. Duffy*** bekommt in der ePA-Anwendung auf ***Behandlung 1*** die Info angezeigt: "Berechtigungsprüfung erfolgreich."

3. Use Case: Nutzer beim Fachdienst registrieren

Die Registrierung des Nutzers beim Fachdienst ist gemäß Funktion FA2.2 Nutzer bei Fachdienst registrieren erfolgt, d.h. es gibt eine existierende ePA für den entsprechenden Versicherten Manfred Reifens.

4. Use Case: Autorisierter Zugriff (erst lesend, dann schreibend)

1. Der Leistungserbringer öffnet seine Primärsystem Roxit.
   Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen, ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden.
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" (lesender Zugriff) wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Hat Dr. Duffy eine aktive Rolle "Arzt"?
      - Wurde die Identität inklusive Rollen-Attribut innerhalb der letzten 24 Stunden bestätigt? ✅
   b. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Ist das Gerät Behandlung 1 auf Dr. Duffy registriert? ✅
   c. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entspricht der Rechner ***Behandlung 1*** den vorgegebenen Bedingungen zum lesenden Zugriff auf die ePA
      - Vertrauenswürdiges Betriebssystem ✅
      - Sichere Ausführungsumgebung ✅
      - System Sicherheit ✅
   d. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entsprechen die ***Behandlung 1*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
      - GEO-Location: ✅
      - Systemzeit passt zur TI-Zeit: ✅
3. *Hintergrundcheck (durch Fachdienst)*: Hat Dr. Duffy die Berechtigung, auf die ePA von Manfred Reifens zuzugreifen? ✅
4. Anzeige der Daten von ***Manfred Reifens***
5. Beim Klick "Neue Datei hochladen" (schreibenden Zugriff) wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Hat Dr. Duffy eine aktive Rolle "Arzt"?
      - Wurde die Identität inklusive Rollen-Attribut innerhalb der letzten 24 Stunden bestätigt? ✅
      - Wurde eine starke Authentifizierung genutzt? ✅
   b. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Ist das Gerät Behandlung 1 auf Dr. Duffy registriert? ✅

c. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entspricht der Rechner ***Behandlung 1*** den vorgegebenen Bedingungen zum lesenden Zugriff auf die ePA

- Vertrauenswürdiges Betriebssystem ✅
- Sichere Ausführungsumgebung ✅
- System Sicherheit ✅

d. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entsprechen die ***Behandlung 1*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?

- GEO-Location: ✅
- Systemzeit passt zur TI-Zeit: ✅

6. *Hintergrundcheck (durch Fachdienst)*: Hat Dr. Duffy die Berechtigung, Daten in der ePA von Manfred Reifens zu ergänzen? ✅
7. Hinzufügen der neuen Datei zur ePA von ***Manfred Reifens***

5. Use Case: Geräte der Nutzer entfernen

Der Nutzer verwaltet seine App-Geräte-Zuordnung und kann Einträge selbständig löschen.

6.1.3 Gutfall: Zugriff des Versicherten über ein mobiles Endgerät auf seine ePA

Annahmen für diesen Use Case

- Primäridentität erhalten: ***Manfred Reifens*** (Nachweis elektronische Funktion des Personalausweises)
- Rolle bezogen: Patient der CyberMed Versicherung (eGK)
- Fachdienst: ePA
- Nötige Geräte
  - Android Handy mit WLAN-Access ins Internet (Handy: ***Manfreds Galaxy S24***)
- Nötige Software
  - ePA-App aus dem Playstore: CyberMed EPA (Hersteller: Industrie)
  - Authenticator-Modul (AUM) aus dem Playstore: CyberMed Ident (Hersteller: Industrie)
  - Trust-Client (TCL) aus dem Playstore: TI connect Version 2.244 (Hersteller: gematik)
- Nötige Zugangsmittel
  - eGK + PIN

1. Use Case: Nutzer identifizieren

Die Identifizierung der Personen mit mindestens einer Rolle beim IDP wurde erfolgreich durchgeführt, um die Authentisierung des Nutzers gemäß Funktion FA2.1 Nutzer authentifizieren zu ermöglichen.

1. Registrierung des Nutzers und seiner Rolle(n) (z.B. Arzt, Versicherter, medizinische Einrichtung, ...) bei einem IDP der Föderation der TI inkl. Identifizierung

2. Verfügbarkeit des zugehörenden Authentisierungsmittels (Authenticator auf Mobiltelefon, ggf. eGK, HBA, SMC-B), siehe FA2.3 eID-Lifecycle IDP

2. Use Case: Gerät des Nutzers registrieren

Im Folgenden wird beschrieben, wie ein mobiles Gerät gemäß Funktion 5.1 "Gerät registrieren" beim GMS für die Verwendung in der ZTA registriert werden kann. Die Schritte für Geräteregistrierung sind:

1. ***Manfred Reifens*** installiert und öffnet die App CyberMed EPA auf seinem Gerät ***Manfreds Galaxy S24***
2. ***Manfred Reifens*** bekommt einen Hinweis: "Wir prüfen nun, ob Sie berechtigt sind auf die ePA zuzugreifen. Sind Sie Leistungserbringer oder Patient? Bitte wählen Sie!"
3. ***Manfred Reifens*** bekommt mögliche Nutzeridentifikationsmittel angezeigt und wählt "eGK" aus
4. ***Manfred Reifens*** bekommt in der ePA-App folgende Nachricht angezeigt: "Bitte eGK an das Gerät halten, um Ihre Identität zu bestätigen." und hält seine eGK an das Handy (bis der Vorgang abgeschlossen ist).
5. ***Manfred Reifens*** bekommt in der ePA-App folgende Nachricht angezeigt: "Bitte PIN für eGK eingeben." und tippt diese ein.
6. ***Manfred Reifens*** bekommt in der ePA-App die Info angezeigt: "Berechtigungsprüfung erfolgreich. Sie können Ihre eGK entfernen. Viel Freude mit Ihrer ePA."

3. Use Case: Nutzer beim Fachdienst registrieren

Die Registrierung des Nutzers beim Fachdienst ist gemäß Funktion FA2.2 Nutzer bei Fachdienst registrieren erfolgt, d.h. es gibt eine existierende ePA für den entsprechenden Versicherten Manfred Reifens.

4. Use Case: Autorisierter Zugriff (lesend)

1. Der Patient öffnet seine APP CyberMed EPA. Er möchte in seiner Akte prüfen, ob sein Arzt die neuen Medikationen hinterlegt hat. Der Login erfolgt mit seinem von IDP erhaltenen "All 4 one" Login
2. Beim Klick auf Login "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck*: Entspricht das Handy ***Manfreds Galaxy S24*** den vorgegebenen Bedingungen zum Zugriff auf die ePA?
3. NFC aktiviert ✅
4. Betriebssystem ✅
5. System Konsistenzcheck ✅
6. System Sicherheit ✅
   a. *Hintergrundcheck*: Entsprechen die ***Manfreds Galaxy S24*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
7. GEO-Location: ✅
8. Systemzeit passt zur TI-Zeit: ✅
   a. *Hintergrundcheck*: Hat ***Manfred Reifens*** eine aktive Rolle "Patient der CyberMed"? ✅

9. Wurde die Identität innerhalb der letzten 24 Stunden mit einem Authenticator bestätigt? ✅
10. Anzeige der Daten von ***Manfred Reifens***

5. Use Case: Geräte der Nutzer entfernen

Der Nutzer verwaltet seine App-Geräte-Zuordnung und kann Einträge selbständig löschen.

6.1.4 Gutfall: Ein Fachdienst kommuniziert mit einem anderen Fachdienst der TI

Betrachtet wird hier ein Abruf von Notfalldaten eines Versicherten durch den Deutscher National Contact Point eHealth (DNCP) in der ePA. Voraussetzung hierfür ist eine eingerichtete Vertreterbeziehung zwischen DNCP und des Versicherten wie in Kapitel 3.5.1 FA2.4 beschrieben.

Use Case: Freigabe für Datenweitergabe durch Nutzer

1. ***Manfred Reifens*** loggt sich beim Nutzerportal ein (mit seinem ID-Token).
2. ***Manfred Reifens*** wählt "Datenfreigabe zwischen Fachdiensten konfigurieren".
3. ***Manfred Reifens*** wählt für "Freigabe des DNCP auf Notfalldaten in der ePA" die Option "ja" aus.

Danach erfolgt der Zugriff des Fachdienstes "Deutscher National Contact Point eHealth" auf die ePA wie im Folgenden beschrieben.

Annahmen für diesen Use Case

Der Fachdienst agiert eigenständig als technischer Nutzer.

- Primäridentität erhalten: Deutscher National Contact Point eHealth "***DNCP"*** (angelegt über Governance-Prozesse für Registrierung von Fachdienst als Nutzer)
- Rolle bezogen: Fachdienst
- Ressource: Notfalldaten in der ePA
- Nötige Geräte
  - Server mit Internetzugriff (NCA-Server)
- Nötige Software (zusätzlich zur Software des DNCP selbst)
  - Authenticator-Modul (AUM): FachdienstIdent (Hersteller: Industrie)
  - Trust-Client (TCL): TI connect Version 2.244 (Hersteller: gematik)
- Nötiges Zugangsmittel:
  - Privater Schlüssel zu registriertem Zertifikat (für ***National Contact Point eHealth)***

1. Use Case: Nutzer identifizieren

Die Identifizierung des Fachdienstes mit der Rolle "Fachdienst" beim IDP wurde erfolgreich durchgeführt, um die Authentisierung des Nutzers gemäß Funktion FA2.1 Nutzer authentifizieren zu ermöglichen.

1. Registrierung des Nutzers und seiner Rolle als "Fachdienst" bei einem IDP der Föderation der TI inkl. Identifizierung

2. Verfügbarkeit des zugehörenden Authentisierungsmittels (z.B. privater Schlüssel des Fachdienstes), siehe FA2.3 eID-Lifecycle IDP

2. Use Case: Gerät des Nutzers registrieren

Im Folgenden wird beschrieben, wie der ***NCA-Server*** gemäß Funktion 5.1 "Gerät registrieren" beim GMS für die Verwendung in der ZTA registriert werden kann.

- Schritte für Geräteregistrierung:
  a. ***DNCP*** triggert bei seinem Trust-Client das Senden folgender Nachricht an GMS: "Gerät registrieren"
  b. Auf die erhaltene Challenge vom GMS, erstellt der ***DNCP*** eine mit dem privaten Schlüssel des DNCP signierte Response (z.B. QES).
  c. Mithilfe des Trust-Clients sendet der ***DNCP*** die Response zurück an den GMS.

3. Use Case: Nutzer beim Fachdienst registrieren

Die Registrierung des Nutzers beim Fachdienst ist gemäß Funktion FA2.2 Nutzer bei Fachdienst registrieren erfolgt, d.h. es gibt eine existierende ePA für den entsprechenden Versicherten Manfred Reifens.

4. Use Case: Autorisierter Zugriff (lesend)

1. ***DNCP*** triggert über den Trust-Client einen Zugriff auf die Notfalldaten von Manfred Reifens.
   a. *Hintergrundcheck*: Entspricht **der NCA-Server des DNCP** den vorgegebenen Bedingungen zum Zugriff auf die ePA?
2. Betriebssystem ✅
3. System Konsistenzcheck ✅
4. System Sicherheit ✅
   a. *Hintergrundcheck*: Entsprechen die **dem NCA-Server des DNCP** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
5. GEO-Location: ✅
6. Systemzeit passt zur TI-Zeit: ✅
   a. *Hintergrundcheck*: Hat sich der DNCP als Fachdienst korrekt authentisiert? ✅
7. *Hintergrundcheck (durch Fachdienst)*: Hat DNCP die Berechtigung, auf die Daten in der ePA von Manfred Reifens zuzugreifen (entsprechende Regelung beim Vertreterdienst hinterlegt)? ✅
8. Übertragen der Daten von ***Manfred Reifens***

5. Use Case: Geräte der Nutzer entfernen

Der DNCP verwaltet seine App-Geräte-Zuordnung und kann Einträge selbständig löschen.

6.1.5 Fehlerfall 1: Autorisierter Zugriff nicht möglich, aber durch Step-Up lösbar

Ein Nutzer hat einen Aspekt für den autorisierten Zugriff nicht erfüllt. Es wird eine Step-Up-Authentisierung durchgeführt und der Zugriff ist dann für den Nutzer möglich. Als Beispiel wird folgender Fall dargestellt:

Zugriff eines Leistungserbringers auf die ePA eines Versicherten erst nach Step-Up-Authentisierung möglich.

Der Leistungserbringer öffnet seine Primärsystem Roxit. Er möchte in der Akte von Patient Manfred Reifens prüfen, ob alle Medikationen in die Akte übernommen und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden.

1. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Hat Dr. Duffy eine aktive Rolle "Arzt"?
      - Wurde die Identität inklusive Rollen-Attribut innerhalb der letzten 24 Stunden bestätigt? ✅
   b. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Ist das Gerät Behandlung 1 auf Dr. Duffy registriert? ✅
   c. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entspricht der Rechner ***Behandlung 1*** den vorgegebenen Bedingungen zum lesenden Zugriff auf die ePA
      - Vertrauenswürdiges Betriebssystem ✅
      - Sichere Ausführungsumgebung ✅
      - System Sicherheit ✅
   d. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entsprechen die ***Behandlung 1*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
      - GEO-Location: ✅
      - Systemzeit passt zur TI-Zeit: ✅
   e. *Hintergrundcheck (durch Fachdienst)*: Hat Dr. Duffy die Berechtigung, auf die ePA von Manfred Reifens zuzugreifen? ✅
2. Anzeige der Daten von ***Manfred Reifens***
3. Beim Klick "Neue Datei hochladen" (schreibenden Zugriff) wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Hat Dr. Duffy eine aktive Rolle "Arzt"?
      - Wurde die Identität inklusive Rollen-Attribut innerhalb der letzten 24 Stunden bestätigt? ✅
      - Wurde eine starke Authentifizierung genutzt? ❌
   b. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Ist das Gerät Behandlung 1 auf Dr. Duffy registriert? ✅
   c. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entspricht der Rechner ***Behandlung 1*** den vorgegebenen Bedingungen zum lesenden Zugriff auf die ePA
      - Vertrauenswürdiges Betriebssystem ✅
      - Sichere Ausführungsumgebung ✅

- System Sicherheit ✅

d. *Hintergrundcheck (durch PDP gemäß Regelwerk)*: Entsprechen die ***Behandlung 1*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?

- GEO-Location: ✅
- Systemzeit passt zur TI-Zeit: ✅

4. Abfrage zur Bestätigung der Rolle: "Bitte nutzen Sie Ihren Authenticator, um Ihre Identität nachzuweisen! Für den Zugriff auf die Akte müssen wir Ihren HBA prüfen".
5. Das Authenticator Modul wird auf ***Duffys iPhone*** angezeigt und verlangt das Vorhalten des HBA: "Bitte bestätigen Sie, dass Sie als ***Dr. Duffy*** auf die ePA von ***Manfred Reifens*** zugreifen möchten"
6. Bestätigung der Authenticator Abfrage
7. *Hintergrundcheck (durch Fachdienst)*: Hat Dr. Duffy die Berechtigung, Daten in der ePA von Manfred Reifens zu ändern? ✅
8. Hinzufügen der neuen Datei zur ePA von ***Manfred Reifens***

6.1.6 Fehlerfall 2: Autorisierter Zugriff nicht möglich und nicht durch Step-Up lösbar

Im Folgenden werden drei Beispiele aufgeführt, unter welchen Umständen ein Zugriff durch das Regelwerk abgelehnt werden kann und was das für den Nutzer bedeutet:

Beispiel 1: Zugriff eines Leistungserbringers auf die ePA eines Versicherten nicht möglich, weil kein vertrauenswürdiges Betriebssystem benutzt wird.

Ein Nutzer hat einen Aspekt für den Zugriff nicht erfüllt und dieser kann auch durch eine Step-Up-Authentisierung nicht ermöglicht werden. In diesem Beispiel ist das Betriebssystem, mit dem auf die TI zugegriffen wird, nicht vertrauenswürdig. Dem Nutzer wird kein Zugriff auf die TI ermöglicht und eine entsprechende Fehlermeldung mit Lösungsmöglichkeiten angezeigt.

Use Case: Autorisierter Zugriff

1. Der Leistungserbringer öffnet seine App ePA 2.0 auf dem mobilen Endgerät ***Duffys iPhone.*** Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden.
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst

   a. *Hintergrundcheck*: Entspricht der Rechner ***Duffys iPhone*** den vorgegebenen Bedingungen zum Zugriff auf die ePA

3. Vertrauenswürdiges Betriebssystem ❌
4. Sichere Ausführungsumgebung ✅
5. System Sicherheit (Gerätesicherheit) ✅

   a. *Hintergrundcheck*: Entsprechen die ***Duffys iPhone*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?

6. GEO-Location ✅
7. Systemzeit passt zur TI-Zeit ✅
   a. *Hintergrundcheck*: Hat Dr. Duffy eine aktive Rolle "Arzt"? ✅
8. ***Duffys iPhone*** zeigt eine Fehlermeldung an: "Sie nutzen kein vertrauenswürdiges Betriebssystem und können somit auf die Akte von Manfred Reifens nicht zuzugreifen. Bitte aktualisieren Sie ihr Betriebssystem oder nutzen Sie ein Gerät mit vertrauenswürdigem Betriebssystem."

Beispiel 2: Schreibender Zugriff eines Leistungserbringers auf die ePA eines Versicherten nicht möglich, weil die Berechtigung dafür fehlt.

Ein Nutzer hat einen Aspekt für den schreibenden Zugriff nicht erfüllt und dieser kann auch durch eine Step-Up-Authentisierung nicht ermöglicht werden. In diesem Beispiel versucht eine Krankenpflegerin schreibend auf die ePA eines Versicherten zuzugreifen. Nach § 352 SGB V ist ein schreibender Zugriff aktuell nicht erlaubt. Der Krankenpflegerin wird kein schreibender Zugriff auf die TI ermöglicht und eine entsprechende Fehlermeldung mit Lösungsmöglichkeiten angezeigt.

Annahmen

- Primäridentität erhalten: ***Rosemarie Kuhnt***
- Rollen-Attribut erhalten: Krankenpfleger (HBA mit entsprechender Profession OID)
- Patient: ***Manfred Reifens***
- Fachdienst: ePA
- Nötige Geräte
  - Mobiles Endgerät mit WLAN-Access oder mit mobilen Daten ins Internet und NFC-Schnittstelle für Interaktion mit HBA (Handy: ***Rosemarie Kuhnts iPhone***)
- Nötige Software
  - ePA-App: APP ePA 2.0 (Hersteller: Industrie)
  - Authenticator-Modul (AUM): KVNO Ident (Hersteller: KV)
  - Trust-Client (TCL): TI connect Version 2.044 (Hersteller: gematik)
- Nötige Zugangsmittel
  - HBA + PIN

Use Case: Autorisierter Zugriff

1. Die Leistungserbringerin Rosemarie Kuhnt öffnet ihre App ePA 2.0. Sie möchte in der Akte von Patient ***Manfred Reifens*** ein neues Medikament in die Akte eintragen
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck*: Entspricht ***Rosemarie Kuhnts iPhone*** den vorgegebenen Bedingungen zum schreibenden Zugriff auf die ePA
3. Betriebssystem ✅
4. System Konsistenzcheck ✅
5. System Sicherheit ✅

a. *Hintergrundcheck*: Entsprechen die ***Rosemarie Kuhnts iPhone*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?

6. GEO-Location ✅
7. Systemzeit passt zur TI-Zeit ✅
   a. *Hintergrundcheck*: Hat Rosemarie Kuhnt eine aktive Rolle "Arzt, Zahnarzt, Physiotherapeut oder Apotheker" und damit die Berechtigung in die ePA von **Manfred Reifens** zu schreiben? ❌
8. ***Rosemarie Kuhnts iPhone*** zeigt eine Fehlermeldung an: "Der schreibende Zugriff auf den Medikationsplan ist Krankenpflegern nicht gestattet. Sie besitzen somit aktuell keine Berechtigung für den schreibenden Zugriff."

Beispiel 3: Zugriff eines Leistungserbringers auf die ePA eines Versicherten nicht möglich, weil auf dem mobilen Endgerät keine Gerätesicherheit eingerichtet ist.

Ein Nutzer hat einen Aspekt für den Zugriff nicht erfüllt und dieser kann auch durch eine Step-Up-Authentisierung nicht ermöglicht werden. In diesem Fehlerfall ist das Betriebssystem, mit dem auf die TI zugegriffen wird, nicht vertrauenswürdig. Dem Nutzer wird kein Zugriff auf die TI ermöglicht und eine entsprechende Fehlermeldung mit Lösungsmöglichkeiten angezeigt.

1. Der Leistungserbringer öffnet seine App ePA 2.0. Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck*: Entspricht der Rechner ***Duffys iPhone*** den vorgegebenen Bedingungen zum Zugriff auf die ePA
      i. Vertrauenswürdiges Betriebssystem ✅
      ii. Sichere Ausführungsumgebung ✅
      iii. System Sicherheit (Gerätesicherheit) ❌
   b. *Hintergrundcheck*: Entsprechen die ***Duffys iPhone*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
      i. GEO-Location ✅
      ii. Systemzeit passt zur TI-Zeit ✅
   c. *Hintergrundcheck*: Hat Dr. Duffy eine aktive Rolle "Arzt"? ✅
3. ***Duffys Handy*** zeigt eine Fehlermeldung an: "Sie haben auf Ihrem Endgerät keine Gerätesicherheit aktiviert. Bitte ändern Sie Ihre Gerätesicherheit, indem Sie den Zugriff auf Ihr Gerät durch eine PIN, ein Muster oder Biometrie absichern. Anschließend starten Sie bitte den Zugriff auf die ePA von Manfred Reifens erneut."

6.1.7 Fehlerfall 3: Gerät darf nicht auf TI zugreifen

Der Fehlerfall beschreibt, dass ein Gerät nicht auf die TI zugreifen kann, weil es kein Geräte-Token erhält. Das folgende Beispiel wird nachfolgend beschrieben:

Zugriff eines Leistungserbringers auf die ePA eines Versicherten nicht möglich, weil das Gerät nicht registriert ist.

Das Gerät, mit dem auf die TI zugegriffen wird, ist nicht registriert und somit ist eine Attestierung nicht möglich ist. Dem Nutzer wird kein Zugriff auf die TI ermöglicht. Er erhält eine Fehlermeldung in der App mit Lösungshinweisen.

1. Der Leistungserbringer öffnet seine App ePA 2.0. Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck*: Entspricht der Rechner ***Duffys iPhone*** den vorgegebenen Bedingungen zum Zugriff auf die ePA
      - Gerät registriert ❌
3. ***Duffys iPhone*** zeigt eine Fehlermeldung an: "Sie haben keine Berechtigung auf die Akte von Manfred Reifens zuzugreifen, weil ihr Gerät nicht registriert ist. Bitte folgen Sie den Schritten für die Registrierung von Geräten oder klicken Sie *hier*". Mit Klick auf *hier* wird ***Dr. Duffy*** direkt zur Geräteregistrierung weitergeleitet.

6.1.8 Fehlerfall 4: Dienst in ZTA nicht erreichbar

Ist eine ZTA-Komponente oder ein Fachdienst in der Zero-Trust Architektur nicht verfügbar, so können manche oder alle Zugriffsanfragen nicht erfolgreich bearbeitet werden. Der Nutzer kann den somit nicht auf die ePA zugreifen. Dem Leistungserbringer wird auf dem Endgerät eine Fehlermeldung mit Supporthinweisen angezeigt, an welcher Komponente es hängt und an wen er sich wenden kann. Folgendes Beispiel wird hierzu im Folgenden ausgeführt:

Zugriff eines Leistungserbringers auf die ePA eines Versicherten nicht möglich, weil der Fachdienst nicht verfügbar ist.

1. Der Leistungserbringer öffnet seine App ePA 2.0. Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" wird folgendes automatisch ausgelöst
   a. *Hintergrundcheck*: Entspricht der Rechner ***Duffys iPhone*** den vorgegebenen Bedingungen zum lesenden Zugriff auf die ePA
      i. Vertrauenswürdiges Betriebssystem ✅
      ii. Sichere Ausführungsumgebung ✅
      iii. System Sicherheit (Gerätesicherheit) ✅
   b. *Hintergrundcheck*: Entsprechen die ***Duffys iPhone*** zugeordneten Umgebungswerte den Anforderungen der ePA und der TI ZT?
      i. GEO-Location ✅
      ii. Systemzeit passt zur TI-Zeit ✅
   c. *Hintergrundcheck*: Hat Dr. Duffy eine aktive Rolle "Arzt"?

i. Wurde die Identität innerhalb der letzten 24 Stunden mit einem Authenticator bestätigt und seinen HBA vorgelegt? ✅

3. ***Duffys iPhone*** zeigt eine Fehlermeldung an: "Leider steht der ePA-Dienst aktuell nicht zur Verfügung. Bitte wenden Sie sich an den Anbieter."

6.1.9 Fehlerfall 5: Problem mit TCL

Es existiert ein Fehler beim Trust-Client und die App des Nutzers stürzt ab. Hier wird der autorisierte Zugriff gleich zu Beginn des Prozesses abgebrochen, weil sich die App nicht starten lässt. Hier muss sich der Kunde über den App-Anbieter Hilfe suchen. Wir beschreiben dazu folgendes Beispiel:

Zugriff eines Leistungserbringers auf die ePA eines Versicherten nicht möglich, weil ein Fehler beim Trust-Client aufgetreten ist.

1. Der Leistungserbringer öffnet seine App ePA 2.0. Er möchte in der Akte von Patient ***Manfred Reifens*** prüfen ob alle Medikationen in die Akte übernommen wurden und ob weitere Medikamente von anderen Leistungserbringern hinzugefügt wurden.
2. Beim Klick "Anzeige der Akteninhalte von Manfred Reifens" reagiert die App nicht.
3. ***Duffys iPhone*** öffnet die App ePA 2.0 nicht. Dr. Duffy muss sich an den Hersteller der App wenden. Die Informationen bzgl. dem Support wird mit der App zur Verfügung gestellt werden (z.B. im App-Store, auf entsprechenden Websites, ...).

6.2 Nichtfunktionale Anforderungen

Im Folgenden wird je nicht-funktionaler Anforderung aus Kapitel 2.4 dargestellt, inwiefern die in diesem Konzept beschriebene Architektur die entsprechende Anforderung erfüllt.

6.2.1 NFA1 - Sichere Kommunikation

Das Ziel von NFA1 besteht darin sicherzustellen, dass die gesamte Kommunikation innerhalb der ZTA durchweg gesichert ist, unabhängig von Standort und Netzwerkzugehörigkeit.

Gemäß Paradigma P2 "Sichere Kommunikation" erfolgt die Übertragung aller Daten zwischen Komponenten der ZTA grundsätzlich gemäß dem Stand der Technik verschlüsselt und gegenseitig authentisiert. Dies gilt, auch wenn Komponenten der ZTA innerhalb organisationsspezifischer Netzwerke kommunizieren. Dabei werden die gültigen Standards, wie zum Beispiel BSI TR-02102 [BSI_TR02102], herangezogen. Dies führt zu einem Schutz aller Daten "in Transit".

Im Folgenden wird gesondert die TLS gesicherte Kommunikation des Nutzers zum Zugriff auf eine Ressource der ZTA betrachtet. Es wird angenommen, dass mittels entsprechender Maßnahmen sichere Implementierungen von TLS für Clients und Server verfügbar sind.

Im Rahmen des mTLS-Verbindungsaufbaus müssen sich beide Teilnehmer gegenseitig authentifizieren. Dies geschieht durch Zertifikate und Signaturen innerhalb des TLS-Handshakes. Der Trust-Anchor der Zertifikate kann in Nutzer und Infrastruktur unterteilt werden.

Die Nutzer-Zertifikate identifizieren ein bestimmtes hardwaregebundenes Schlüsselpaar, dessen öffentlicher Schlüssel unter anderem im Geräte-Token als

Authentifizierungsmerkmal hinterlegt ist. Dieser wird durch den GMS legitimiert. Dazu prüft das GMS bei der Geräte-Registrierung die Hardwarebindung, die er in den Geräte-Token bescheinigt.

Bei Infrastruktur-Zertifikaten ist eine zentrale PKI vorgesehen, die die Authentizität bei Ausstellung des Zertifikats prüft.

6.2.2 NFA2 - Data (Datenschutz)

Das Ziel von NFA2 besteht darin den Schutz der Daten gemäß den in Kapitel 3.4 eingeführten Datenklassen und ihren Schutzbedarfen sicherzustellen. Während das Konzept der ZTA selbst den Schutz der Daten vor unberechtigtem Zugriff durch einen Nutzer zum Ziel hat, liegt der Fokus dieser Anforderung darauf auch den Schutz vor unberechtigtem Zugriff durch Akteure der Infrastruktur zu gewährleisten.

Gemäß NFA1 "Sichere Kommunikation" ist der Schutz von Daten bei der Übertragung zwischen Komponenten der ZTA gewährleistet. Je nach Datenklasse werden Daten darüber hinaus auf einzelnen Komponenten der ZTA verarbeitet und ggf. gespeichert. In Tabelle 6.2.2-1 sind diese Informationen zusammen mit vorgesehenen technischen und organisatorischen Maßnahmen zum Schutz dieser Daten "in Use" und "at Rest" dargestellt.

DS1

Die Datenklasse DS1 gehört gemäß DSGVO [EU_DSGVO] zu den besonderen Kategorien personenbezogener Daten, für welche auf Grund ihrer Sensibilität ein besonders hoher Schutzbedarf anzunehmen ist. Für einen Teil der Komponenten der ZTA ist der Zugriff auf diese Daten deshalb bereits durch das Design der Architektur selbst ausgeschlossen. Für Komponenten, welche Daten dieser Datenklasse zwangsläufig verarbeiten oder speichern (FCL, TCL, Fachdienst, PEP, PDP, IDP), sieht das Konzept folgende Maßnahmen vor, um einen unberechtigten Zugriff auf die Daten zu verhindern.

Fachdienst

Eine Verarbeitung von Daten der Kategorie DS1 erfolgt zentral durch den Fachdienst für alle Nutzer des Fachdienstes zum Zweck der Erbringung der Leistung des Fachdienstes. Um die für Betrieb und Herstellung des Fachdienstes relevanten Akteure von einem Zugriff auf Daten der Kategorie DS1 auszuschließen, erfolgt die Verarbeitung dieser Daten innerhalb einer Vertrauenswürdigen Ausführungsumgebung. Das persistente Speichern von Ergebnissen der Datenverarbeitung erfolgt aus der VAU heraus verschlüsselt. Die Umsetzung dieser Anforderungen wird im Rahmen der Anbieter- und Produktzulassung des Fachdienstes überprüft.

Fach-Client (FCL)/Trust-Client (TCL)

Daten der Kategorie DS1 werden durch Fach- und Trust-Client dezentral in ihrer Nutzerumgebung verarbeitet bzw. gespeichert. Wie der Fachdienst verarbeitet der Fach-Client personenbezogene medizinischen Daten zum Zweck der Erbringung der Leistung des Fachdienstes. Der Trust-Client verschlüsselt und entschlüsselt Daten für den Fach-Client und hat somit grundsätzlich ebenso Zugriff auf personenbezogene medizinische Daten. Je nach Nutzer werden von Fach- und Trust-Client jedoch nur Daten verarbeitet, welche entweder dem Nutzer selbst gehören (z.B. Versicherter) oder für deren Verarbeitung sich der Nutzer zuvor legitimiert hat (z.B. Leistungserbringer im Rahmen einer Patienten-Arzt-Beziehung) und die Verantwortung trägt.

Es muss verhindert werden, dass durch den Hersteller von Fach- und Trust-Client z.B. im Rahmen des Applikations-Monitorings (Telemetrie, Unterstützung von Debugging) Daten abgeschöpft werden, die eine Profilbildung ermöglichen. Dies wird für den Trust-Client durch entsprechende Anforderungen im Rahmen einer Produktzulassung sichergestellt.

Für den Trust-Client ist darüber hinaus die Realisierung der Kernfunktionalität als Open Source Implementierung empfohlen. Für den Fach-Client ist ein Zulassungsverfahren nicht vorgesehen. Wesentliche Anforderungen an die Informationssicherheit aus der ZTA heraus werden durch den Trust-Client oder den Fachdienst gekapselt. Darüberhinausgehende Anforderungen an die Informationssicherheit des Fach-Client (z.B. das verschlüsselte Speichern der Daten der Kategorie DS1) und den Datenschutz (z.B. das DSGVO-konforme Applikationsmonitoring) werden durch direkte Wirkung gesetzlicher Anforderungen (z.B. DSGVO) auf den Anbieter des Fach-Client bzw. auf spezielle Nutzergruppen (z.B. Leistungserbringer) durchgesetzt.

PEP, PDP, IDP

PEP, PDP und IDP verarbeiten ebenfalls Daten der Kategorie DS1. Diese umfassen (neben Fachdaten beim PEP) auch Metadaten wie Zugriffsanfragen, aus denen bestimmte Indikationen oder die Erbringung bestimmter Gesundheitsleistungen hervorgehen können. Um die für Betrieb und Herstellung der Komponenten relevanten Akteure von einem Zugriff auf diese Daten auszuschließen, erfolgt ihre Verarbeitung innerhalb einer Vertrauenswürdigen Ausführungsumgebung. Die Umsetzung dieser Anforderungen wird im Rahmen der Anbieter- und Produktzulassung der Komponenten überprüft.

Es soll mit weitgehend technischen Mitteln ein Zugriff seitens der Akteure aus dem Umfeld der Infrastruktur der Komponenten verhindert werden. Anbieter und Betreiber werden daher durch eine Vertrauenswürdige Ausführungsumgebung (VAU) vom Zugriff auf die Daten ausgeschlossen. Sie schützt die Daten "in Use". Eine persistente Speicherung der Daten erfolgt nur in verschlüsselter Form. Die korrekte Umsetzung dieser Anforderungen wird im Rahmen der Anbieter- und Produktzulassung der Komponenten überprüft.

DS2

Daten der Datenklasse DS2 müssen gemäß DSGVO [EU_DSGVO] insbesondere vertraulich behandelt werden. Für Komponenten, welche Daten dieser Datenklasse speichern, wird deshalb gefordert, dass das Speichern verschlüsselt erfolgt. Darüber hinaus unterliegt die Verarbeitung der Daten dem jeweiligen Datenschutzkonzept des Infrastrukturbetreibers. Die Umsetzung dieser Anforderungen wird im Rahmen der Anbieter- und Produktzulassung bzw. durch Abnahme der Komponenten durch die gematik überprüft.

DS3

Daten der Datenklasse DS3 müssen insbesondere hinsichtlich ihrer Integrität geschützt werden, um eine Manipulation der Zugriffsteuerung zu verhindern. Für Komponenten, welche Daten dieser Datenklasse speichern, wird deshalb gefordert, dass das Speichern integritätsgesichert erfolgt. Darüber hinaus unterliegt die Verarbeitung der Daten dem jeweiligen Sicherheitskonzept des Infrastruktur-Betreibers. Die Umsetzung dieser Anforderungen wird im Rahmen der Anbieter- und Produktzulassung bzw. durch Abnahme der Komponenten durch die gematik überprüft.

DS4

Daten der Datenklasse DS4 müssen insbesondere hinsichtlich ihrer Integrität geschützt werden, um eine Manipulation des Monitorings und damit ggf. der Zugriffsteuerung zu verhindern. Für Komponenten, welche Daten dieser Datenklasse speichern, wird deshalb gefordert, dass das Speichern integritätsgesichert erfolgt. Darüber hinaus unterliegt die Verarbeitung der Daten dem jeweiligen Sicherheitskonzept des Infrastruktur-Betreibers. Die Umsetzung dieser Anforderungen wird im Rahmen der Anbieter- und Produktzulassung bzw. durch Abnahme der Komponenten durch die gematik überprüft.

DS5

Daten der Datenklasse DS5 bezeichnet öffentlich zugängliche Daten ohne erhöhten Schutzbedarf. Die Verarbeitung der Daten obliegt dem Sicherheitskonzept der betreffenden Komponenten.

Datenklasse / Komponente	DS1	DS2	DS3	DS4	Technische Maßnahme	Organisatorische Maßnahme
FCL	V/S***	V/S**	--	--	Speicherung erfolgt verschlüsselt (DS1/DS2)	direkte Wirkung gesetzlicher Anforderungen z.B. DSGVO [EU_DSGVO]
TCL	V	V/S**	--	--	Speicherung erfolgt verschlüsselt (DS2)	Produktzulassung, Open Source Implementierung empfohlen
AUM	--	V/S**	--	--	Speicherung erfolgt verschlüsselt (DS2)	Produktzulassung, Anbieterzulassung (IDP)
Nutzerportal	--	V	--	V	--	Abnahme gematik
Fachdienst	V/S***	V/S**	--	V	Vertrauenswürdige Ausführungsumgebung (DS1), Speicherung erfolgt verschlüsselt (DS1/DS2)	Produktzulassung, Anbieterzulassung
PEP	V	V/S**	V	V	Vertrauenswürdige Ausführungsumgebung (DS1)	Produktzulassung, Anbieterzulassung
PDP	V	V	V	V	Vertrauenswürdige Ausführungsumgebung (DS1)	Produktzulassung, Anbieterzulassung
IDP	V	V/S**	--	V	Vertrauenswürdige Ausführungsumgebung (DS1), Speicherung erfolgt verschlüsselt (DS1/DS2)	Produktzulassung, Anbieterzulassung
FEM	--	--	V/S*	V	Speicherung erfolgt integritätsgesichert (DS3)	Abnahme gematik

Datenklasse / Komponente	DS1	DS2	DS3	DS4	Technische Maßnahme	Organisatorische Maßnahme
PIP	--	(V/S**)	V/S*	V	Speicherung erfolgt verschlüsselt (DS2) und integritätsgesichert (DS3)	Abnahme gematik
GMS	--	V/S**	V/S*	V	Speicherung erfolgt verschlüsselt (DS2) und integritätsgesichert (DS3)	Abnahme gematik
PAP	--	--	V/S*	V	Speicherung erfolgt integritätsgesichert (DS3)	Abnahme gematik
MON	--	(V/S**)	V/S*	V/S*	Speicherung erfolgt verschlüsselt (DS2) und Integritätsgesichert (DS3/DS4)	Abnahme gematik

Tabelle 6.2.2-1: Übersicht der Komponenten, welche Daten der Kategorie DS1-4 speichern oder verarbeiten und der grundlegenden Maßnahmen zum Schutz der Daten (V = verarbeitet, S* = speichert integritätsgeschützt, S = speichert vertraulich/verschlüsselt)**

6.2.3 NFA3 - Privacy/Datenschutz (Profilbildung)

Ziel der Anforderung NFA3 ist es, eine unberechtigte Profilbildung hinsichtlich der Nutzung von Gesundheitsdiensten, insbesondere durch die für den Betrieb der ZTA-Infrastruktur relevanten Akteure (Anbieter, Betreiber, Hersteller) durch das Design der Architektur auszuschließen bzw. durch wirksame Maßnahmen zu verhindern. Eine Profilbildung ist grundsätzlich nur dann möglich, wenn eine Komponente, für die ein Akteur verantwortlich ist, personenbezogene medizinische Daten (DS1), insbesondere in Form von Metadaten über Zugriffe auf Ressourcen, verarbeitet. Durch das Design der Architektur gem. Kapitel 3.2 bzw. Kapitel 3.4 ist die Verarbeitung dieser Datenkategorie bereits auf 6 Komponenten beschränkt. Eine Übersicht der Komponenten, welche Daten der Kategorie DS1 verarbeiten oder speichern, sowie die jeweiligen Maßnahmen, welche einem Zugriff durch Akteure in der Infrastruktur entgegenwirken und eine Profilbildung verhindern, ist in Tabelle 6.2.2-1 dargestellt und bereits in Kapitel 6.2.2 beschrieben.

6.2.4 NFA4 - Data (just-in-time)

Ziel der Anforderung NFA4 ist es, den Zugriff auf Daten möglichst auf den benötigten Zeitraum (just-in-time-Access) in Abwägung gegen Usability- und Performanceanforderungen zu limitieren. Für die Evaluierung wird der Begriff Daten als

Ressourcen im Sinne der ZTA interpretiert. Diese sind in Kapitel 3.1.1- Ressourcen genauer definiert.

Zentraler Bestandteil der Architektur ist die zeitlich limitierte Autorisierung eines Nutzers zum Zugriff auf eine Ressource. Diese zeitliche Limitierung wird durch Laufzeiten des ID-Tokens sowie des Geräte-Tokens implementiert. Beide Laufzeiten sind frei wählbar, und können auch während des laufenden Betriebs angepasst werden. Bei der Ausstellung der Token können die zur Laufzeit zur Verfügung stehenden Informationen herangezogen werden. Dies können unter anderem die im Token hinterlegten Scopes sein und beim ID-Token zusätzlich die Client-Anwendung (OIDC-ClientID). Die aktuelle Gültigkeit wird bei jedem Zugriff auf eine Ressource geprüft. Dies kann in mehreren Abstufungen implementiert werden. Jedes Token besitzt eine begrenzte Laufzeit. Zusätzlich kann über ein Refresh-Token die Laufzeit des Tokens verlängert werden, ohne entsprechende Nachweise neu erbringen zu müssen. Dies bietet die Möglichkeit, zeitnah auf Ereignisse zu reagieren, die den ursprünglichen Nachweis kompromittiert haben könnten. Die Zahl der Refresh-Token sowie deren Laufzeit kann ebenfalls frei gewählt und dynamisch angepasst werden. Nach Ablauf der Laufzeit müssen entsprechend aktuelle Nachweise geliefert werden, um erneut Zugriff auf Ressourcen zu erhalten. So kann die Zugriffsberechtigung zeitlich feingranular gesteuert werden.

Die Abwägung zwischen den Zielen der Usability (Häufigkeit des Nachweises durch den Nutzer), der Performance (Anzahl an Token) und der Minimierung der Gültigkeit der Nachweise und damit des erlaubten Zugriffzeitraums (just-in-time) muss je Ressource getroffen werden.

6.2.5 NFA5 - Data (just-enough)

Ziel der Anforderung NFA5 ist es, den Zugriff auf Daten möglichst nur mit den nötigen Privilegien (just-enough-Access) auszustatten, in Abwägung gegen Usability- und Performanceanforderungen.

Für die Evaluierung wird der Begriff Daten als Ressourcen im Sinne der ZTA interpretiert. Diese sind in Kapitel 3.1.1- Ressourcen genauer definiert. Die feinste Granularität für Ressourcen ist dabei eine einzelne URL, die keine personenidentifizierenden Daten enthält.

Die Zugriffe auf eine Ressource werden durch das Regelwerk geprüft. Berechtigungen sollten als Allow-Liste ausgelegt werden, d.h., Zugriffe werden abgelehnt, sofern keine entsprechende Regel den Zugriff erlaubt. Grundlage dafür sind die direkt beim Zugriff erhobenen Daten und die in ID- und Geräte-Token hinterlegte Nachweise. Die in den Token hinterlegten Zugriffberechtigungen (Scopes) und deren Granularität sind frei wählbar. Damit kann sichergestellt werden, dass ein Token nur die minimal notwendigen Zugriffsberechtigungen enthält. Dies ist bis zur Granularität einer einzelnen Ressource möglich. Es können aber auch mehrere Ressourcen zusammengefasst werden, wenn dies in Abwägung mit Anforderungen an die Usability sinnvoll erscheint.

Durch die Step-Up Autorisierung ist es möglich, mit einer frei wählbaren zeitlichen Granularität (vgl. NFA 5) zusätzliche Nachweise nachzufordern. Zur Erhöhung der Nutzbarkeit und Reduzierung der benötigten Step-Up Autorisierungen können auch mehr als die für den ersten Zugriff minimal benötigten Nachweise und Zugriffsberechtigungen bescheinigt werden, um bei zukünftigen Zugriffen keine neuen Nachweise anfordern zu müssen.

Die Abwägung zwischen den Zielen der Usability (Nachweise durch den Nutzer), der Performance (Anzahl an Tickets) und der Minimierung der Berechtigungen (Scope) der Nachweise und damit der zugreifbaren Ressourcen (just-enough) muss je Ressource getroffen werden.

6.2.6 NFA6 - Keine Allmacht (Zugriff auf medizinische Daten)

Das Ziel von NFA6 besteht darin, die in der TI2.0 verarbeiteten medizinischen Daten des Versicherten, d.h. DS1 gemäß Kapitel 3.4, vor unberechtigtem Zugriff durch Akteure der TI zu schützen. Insbesondere darf keiner der Infrastruktur-Akteure, d.h. Betreiber, Anbieter oder Hersteller einer TI-Komponente allein alle nötigen Mittel besitzen, um sich unberechtigt Zugriff auf diese Daten zu verschaffen (Allmacht).

Gemäß Kapitel 3.4 wird der Zugriff auf Daten der Klasse DS1 den technischen Komponenten TCL, FCL, Fachdienst, PEP, PDP und IDP ermöglicht. Der Schutz der Daten, insbesondere das Verhindern eines Zugriffs auf die verarbeiteten Daten durch Infrastruktur-Akteure, wird hierbei durch technische und organisatorische Maßnahmen sichergestellt und ist in NFA2 "Data (Datenschutz)" beschrieben.

Hersteller, Anbieter oder Betreiber von TI-Komponenten gemäß Kapitel 3.5 könnten sich Zugriff auf bei einem Fachdienst A gespeicherte medizinische Daten eines Versicherten verschaffen, falls sie

1. einen Authorization Code für das Abrufen eines gültiges ID-Token eines Zugriffsberechtigten für den Fachdienst sowie
2. ein gültiges Geräte-Token mit der entsprechenden UUID_Nutzer

vorlegen könnten.

Die nachfolgende Tabelle 6.2.6-1 stellt einen Überblick dar, welche Infrastruktur-Komponenten Zugriff auf Authorization Code oder Token erhalten oder die Prüfung der Token beeinflussen könnten:

Komponente	(Auth Code für) ID-Token mit UUID_Nutzer für Fachdienst A	Geräte-Token mit UUID_Nutzer und UUID_Gerät
Nutzerportal	-	(x)
PEP von Fachdienst A	x	x
PDP von Fachdienst A	-	-
Fachdienst B inklusive PEP und PDP	-	x
IDP	x	-
FEM	(x)	(x)
PIP	-	-
GMS	-	x
PAP	(x)	(x)
MON	-	-

Tabelle 6.2.6-1: Übersicht, welche Komponenten Zugriff auf benötigte ID- oder Geräte-Token haben oder deren Prüfung beeinflussen könnte

Nutzerportal

Das Nutzerportal dient gemäß diesem Konzept als Zugriffsportal auf Schnittstellen von fachdienstübergreifende Komponenten in der ZTA. Es hat somit keinen Zugriff auf Auth Code oder ID-Token für Fachdienste.

Im vorliegenden Feinkonzept ist noch keine Funktion für das Nutzerportal definiert, die eine Gerätebindung erfordert. Sollte eine zukünftige Funktion eine Gerätebindung benötigen, würde das Nutzerportal im Rahmen dieser Funktion ein Geräte-Token verarbeiten. Ohne die Kenntnis des zugehörigen privaten Schlüssels könnte das Nutzerportal dieses Geräte-Token aber nicht verwenden und hätte zudem auch kein ID-Token, um auf Ressourcen bei einem Fachdienst zuzugreifen.

PEP von Fachdienst A

Als Endpunkt der TLS-Verbindung zum Fachdienst erhält der PEP die ID- und Geräte-Token für den Zugriff auf die angefragten Ressourcen beim Fachdienst. Die Daten innerhalb von Fachdienst A, zu dem der PEP gehört, sind durch technische und organisatorische Maßnahmen (VAU, Produkt- und Anbieterzulassung) geschützt.

PDP von Fachdienst A

Der PDP erhält vom PEP nur die benötigten Ist-Attribute, nicht aber Authorization Code, ID-Token, oder Geräte-Token und kann deshalb selbst nicht auf die Ressourcen beim Fachdienst zugreifen.

Fachdienst B inkl. PEP und PDP

Fachdienst B sowie dessen PEP erhalten im Rahmen einer Anfrage ein Geräte-Token sowie einen Authorization Code, über den das ID-Token vom IDP abgerufen werden kann und damit auch das ID-Token. Der Authorization Code ist spezifisch für den Fachdienst B und kann deshalb nicht für den Abruf von ID-Token mit Access Scope Fachdienst A verwendet werden. Das abgerufene ID-Token hat einen Access Scope Fachdienst B und daher nicht für einen Zugriff auf Ressourcen von Fachdienst B verwendet werden. Das Geräte-Token ist ohne den zugehörigen privaten Schlüssel auf dem Endgerät des Nutzers ebenfalls nicht nutzbar. Demnach kann ein Fachdienst also nur über die vorgesehenen Zugriffswege gemäß Use Case 7 (Kapitel 6.1) als autorisierter Nutzer auf Daten eines anderen Fachdienstes zugreifen.

IDP

Der IDP stellt die Authorization Codes aus und liefert basierend darauf die signierten ID-Token für die Fachdienste. Der IDP könnte also (einen Authorization Code für) ein gültiges ID-Token für einen beliebigen Akteur ausstellen. Da auf Seiten des Fachdienstes jedoch nicht nur das ID-Token, sondern auch ein dazu passendes Geräte-Token für den Zugriff auf die Nutzer-Daten nötig ist und der IDP nicht über das registrierte Gerät verfügt, kann der IDP allein nicht auf die Daten beim Fachdienst zugreifen. Auch bei Nutzung eines Pairwise-Identifiers als Pseudonym für einen Fachdienst, erfolgt das Mapping auf ein Geräte-Token über die UUID_Nutzer und kann - da der Pairwise-Identifier von dieser UUID_Nutzer abgeleitet wird - durch den IDP nicht manipuliert werden. Eine Registrierung eines neuen Gerätes für den Nutzer durch einen Angreifer beim IDP wird dadurch verhindert, dass für diese Registrierung gerade kein IDP-Token als Identitätsnachweis verwendet wird, sondern ein Authentisierungsmittel mit souveräner ID.

FEM

Der Federation Master verwaltet die öffentlichen Schlüssel für IDP und GMS, die für die Prüfung der ID- und Geräte-Token verwendet werden, ebenso wie die öffentlichen Schlüssel der Fachdienste. Ein kompromittierter FEM könnte Schlüsselmaterial austauschen oder ergänzen, so kompromittierte Komponenten in die ZTA einbringen und sich in die Lage versetzen, auf beliebige Daten der TI zuzugreifen. Um eine

Allmachtstellung des FEM zu verhindern, sollte die Integrität und Authentizität der öffentlichen Schlüssel z.B. durch Einbettung in Zertifikate von einer oder mehreren außerhalb des FEM betriebenen CAs sichergestellt werden. Alternativ kann das Risiko eines Allmacht-Missbrauchs durch das Vorsehen entsprechend sicherer und durch Gewaltenteilung gekennzeichneter Prozesse für das Ausstellen von Schlüsseln reduziert werden. Ein solches Vorgehen sollte durch ein Monitoring ergänzt werden, welches den Einsatz von irregulär eingebrachtem Schlüsselmaterial erkennt.

PIP

PIPs stellen nur Referenzwerte zur Verfügung, haben aber keinen Zugriff auf Token und Ist-Werte im Rahmen einer Zugriffsanfrage.

GMS

Ein GMS stellt Geräte-Token aus und bestätigt dabei die Bindung einer UUID_Nutzer an eine UUID_Gerät. Damit könnte ein GMS ein gültiges Geräte-Token ausstellen, das das Gerät eines Angreifers mit einer beliebigen Nutzer-ID verknüpft. Da ein Angreifer jedoch beim zuständigen IDP kein ID-Token für den entsprechenden Nutzer erhält, ist ein Zugriff auf die Daten des Versicherten beim Fachdienst nicht möglich.

PAP

Der PAP hat keinen direkten Zugriff auf ID- oder Geräte-Token, stellt jedoch das Regelwerk zur Verfügung, welches die Anforderungen an ID- und Geräte-Token für den Zugriff auf Ressourcen festlegt (z.B. Anforderungen an die Vertrauenswürdigkeit). Es wäre somit möglich, durch Manipulation des Regelwerks die Anforderungen für den Zugriff auf das zulässige Minimum zu reduzieren und so beispielsweise die Gerätebindung zu umgehen. Dennoch müsste ein Angreifer ein von einem zugelassenen IDP signiertes ID-Token für den gewünschten Fachdienst vorlegen.

Um dem Risiko entgegenzuwirken, dass die Zugriffsanforderungen wie beschrieben abgesenkt werden, sollte der PAP so umgesetzt werden, dass kein Akteur allein die hinterlegten Regeln aktualisieren kann. Die Integritätssicherung des Regelwerkes durch Einsatz eines Signaturverfahrens muss deshalb unter Einbeziehung mehrerer unabhängiger Akteure umgesetzt werden.

MON

MON erhält von den Komponenten der TI keine Daten mit explizitem Nutzerbezug, d.h. insbesondere weder ID-/Geräte-Token noch die darin enthaltenen UUIDs.

Mehrere Rollen bei einem Akteur

Im Rahmen der TI kann ein Akteur auch mehrere Rollen übernehmen und z.B. gleichzeitig mehrere Komponenten betreiben/anbieten. Durch das Zusammentreffen mehrerer Rollen bei einem Akteur kann unter Umständen ein erhöhtes Risiko entstehen. Insbesondere ein Zusammenfallen von GMS und IDP bei einem Akteur ermöglicht diesem Akteur potenziell einen systematischen Zugriff auf die Daten von Versicherten und sollte verhindert werden. Dies schließt jedoch nicht aus, dass die entsprechenden Komponenten innerhalb einer Rechenzentrumsumgebung betrieben werden, wenn diese Umgebung entsprechend starke Mechanismen zur Trennung von Mandanten bereitstellt oder wenn die Dienste innerhalb einer VAU ausgeführt werden.

6.2.7 NFA7 - Identitäten

Ziel der Anforderung NFA7 ist es sicherzustellen, dass sich die bereits durch die gematik in der Spezifikation befindliche Identity Provider Föderation der TI [gem_IDP_Federation] in die ZTA integrieren lässt. Das Konzept der ZTA berücksichtigt gem. Kapitel 3.5.1 explizit diese Integration über das Protokoll OpenID Connect.

6.2.8 NFA8 - Performanz

Das Ziel von NFA8 besteht darin, dass die Anwendungen der Telematikinfrastruktur von allen Akteuren jederzeit performant benutzt werden können, d.h. die ZTA-Architektur nicht zu störenden Performanceproblemen führt. Konkret muss die Lösung für mehr als 80 Millionen Versicherte und über 200.000 Leistungserbringer praktikabel, d.h. ohne störende Antwort-, Lauf- oder Reaktionszeiten, nutzbar sein.

Die Architektur hat keine performance-relevanten Bottlenecks und kann durch Erhöhung der Anzahl der Komponenteninstanzen linear mit der Anzahl der Teilnehmer skalieren. Auch ein temporärer Ausfall einzelner Komponenteninstanzen führt zu keinen spürbaren Verfügbarkeitsproblemen.

Die beschriebene Architektur bietet optimale Kommunikationswege zwischen Fach-Client und Fachdienst, welche eine kurze Latenz in der Kommunikation ermöglichen. Im typischen Fall einer bereits authentisierten Session auf Anwendungsebene und einer bereits bestehenden TLS-Verbindung zwischen TCL und PEP, wird die Zugriffgeschwindigkeit durch die Latenz der Kommunikation zwischen TCL und PEP, der Kommunikation zwischen PEP und PDP, der für die Regelausführung im PDP benötigten Zeit sowie der Kommunikation zwischen PEP und Fachdienst bestimmt. Der längste Kommunikationsweg ist dabei zwischen TCL und PEP, für den bei einer direkten Verbindung eine Roundtrip-Zeit von 30ms angenommen wird. Diese kann sich durch den vorgeschalteten DDoS-Schutz noch leicht erhöhen, was hier aber mangels bestehenden Konzepts für einen solchen Schutz nicht berücksichtigt wird. Für die Kommunikation zwischen PEP und Fachdienst werden aufgrund deren örtlicher Nähe maximal 15ms für den Roundtrip angenommen. Die Kommunikation zwischen PEP und PDP ist sehr eng (evtl. innerhalb des gleichen Prozesses) und wird als insignifikant betrachtet. Allerdings müssen im PEP die Anfrage analysiert und die Informationen zu Gerät und Nutzeridentität aus der Session ermittelt werden. Dafür werden 50ms veranschlagt. Für die Ausführung des Regelwerks werden 100ms angenommen, wobei dieser Wert sehr grob geschätzt ist und von der Komplexität des Regelsatzes, der Performance der Regel-Engine, der für den PDP und die Datenbanken reservierten Ressourcen des Rechenzentrums sowie von ggf. nötigen Lookups in externen PIP abhängt. Die Optimierung durch Caching von Entscheidungen, wie in 4.2 beschrieben, wird als kritisch für eine hohe Performance gesehen. Die Abarbeitung der Anfrage im Fachdienst wird hier zeitlich nicht eingerechnet. Zusammen genommen bedeutet das für die Beantwortung einer Anfrage (ohne fachliche Bearbeitungszeit) ca. 195ms, wobei der Overhead durch die ZTA selbst 165ms ist. Kann wie in 4.2 vorgeschlagen auf eine gecachte Entscheidung des PDP zugegriffen werden, reduziert sich der Overhead auf 65ms.

Besteht noch keine gültige Authentisierung, so müssen ein ID-Token vom IDP und Geräte-Token vom GMS erlangt werden. Hierbei sind, wie in Kapitel 3.5 dargestellt, eine Vielzahl von Roundtrips nötig: Authentisierung am IDP, Abrufen des ID-Token vom IDP-Token-Endpunkt, Attestierung am GMS, Abrufen des Geräte-Tokens des GMS-Token-Endpunktes. Dazu kommt ggf. noch eine Kommunikation mit Diensten der Plattformanbieter der Endgeräte zur Attestierung der Geräte sowie die Zeiten bei der Bearbeitung der Anfragen in den entsprechenden Endpunkten. Grob geschätzt ist man damit im Bereich von 1 bis 2 Sekunden. Allerdings findet die initiale Authentisierung nur selten statt und ist i.A. mit Nutzerinteraktionen wie dem Start der Anwendung und der Bereitstellung von Zugangsinformationen verbunden. Der Refresh der Authentisierungsinformationen hingegen findet ohne Unterbrechungen für den Nutzer im Hintergrund statt.

6.2.9 NFA9 - Skalierbarkeit (kurzfristig)

Das Ziel von NFA9 ist, dass die Lösung skalierbar ist und auch bei kurzfristigen Lastschwankungen performant bleibt.

Dies wird ermöglicht durch Nutzung etablierter Technologien zur dynamischen Skalierung in Verbindung mit einem komponenteninternen Monitoring der Auslastung. Damit kann zeitnah auf steigende Last reagiert werden, indem die Anzahl der Komponenteninstanzen innerhalb der Cluster erhöht wird. Dies kann sowohl reaktiv bei messbarer Laststeigerung als auch bereits proaktiv erfolgen, um zu erwartende Spitzenzeiten bei der Nutzung abzufangen.

6.2.10 NFA10 - Skalierbarkeit (langfristig)

Das Ziel von NFA10 ist eine zukunftsfähige Architektur, mit der man auch langfristig in der Lage ist, dem absehbaren Wachstum von Anwender- und Leistungserbringergruppen, einer großen Anzahl neuer Digital-Health-Anwendungen, Diensten, größeren Datenmengen und weiteren Einsatzszenarien gerecht zu werden.

Dies wird erreicht, indem das Design der Architektur keine performance-relevanten Bottlenecks aufweist. Das ermöglicht eine lineare Skalierung mit der Anzahl der Teilnehmer, Fachdienste, Anfragenmenge und Bandbreitenbedarf durch eine Erhöhung der Anzahl der Komponenteninstanzen - ohne dass es zu negativen Auswirkungen auf die Performanz kommt (siehe auch 6.2.8 NFA 8). Die Erhöhung der Anzahl der Komponenten kann dabei sowohl durch Ausbau bestehender Cluster vorgenommen werden als auch durch den Aufbau neuer Standorte von Clustern. Dabei können Erweiterungen problembezogen erfolgen: bei zunehmender Anzahl der Teilnehmer im System müssen die fachübergreifenden Komponenten ausgebaut werden, bei einer stärkeren Nutzung eines spezifischen Fachdienstes primär dieser Fachdienst inkl. PEP und PDP selbst, aber nicht gleichzeitig auch andere Fachdienste oder in größerem Umfange übergreifende Komponenten.

6.2.11 NFA11 - Auswirkungen auf Leistungserbringerprozesse (Erstanwendungen)

Ziel der Anforderung ist es, die Abläufe beim Leistungserbringer möglichst reibungsarm zu gestalten und auf ein Minimum zu reduzieren.

Zur Teilnahme an der TI 2.0 müssen die Teilnehmer identifiziert sein. Dabei muss die Identifizierung durch eine vertrauenswürdige Instanz erfolgen. Hierbei ist es unerheblich, ob das Vertrauen aus einem föderierten IDM-Prozess oder einer lokalen IAM-Lösung abgeleitet wird. Jeder Identität sind 1 bis n Rollen zuzuweisen. Identitäten ohne Rollen-Attribut (z.B. Hebamme, Apotheker, Fachangestellte oder Patient bzw. Versicherter) sind in der TI nicht vorgesehen. Die Rollen werden von bestehenden Credentials (wie HBA, eGK oder SMC-B) oder über neue Registrierungsprozesse für digitale Identitäten abgeleitet. Die bestehenden Credentials können auch als weiterer Faktor verwendet werden.

Das Onboarding der Geräte zum Zugriff auf die TI 2.0 sollte sich möglichst eng am IDP-Enrollment orientieren. Bei der initialen Bestätigung der Identität kann ein Gerät registriert werden. Anschließend sollte eine Kombination aus Login am IDP und Zugang zum GMS auf einem bereits registrierten Gerät dem Nutzer die Registrierung weiterer Geräte ermöglichen. Für den Leistungserbringer stellt sich der Vorgang des Enrollments bei IDP und GMS möglichst als ein zusammenhängender Schritt dar.

6.2.12 NFA12 - Auswirkungen auf Leistungserbringerprozesse (Regelnutzung)

Ziel der Anforderung ist es, die Abläufe beim Leistungserbringer möglichst reibungsarm zu gestalten und auf ein Minimum zu reduzieren.

Der Zugriff auf die TI 2.0 wird durch Regeln gesteuert. Diese Regeln greifen auf definierte Attribute zurück und stellen diese in Abhängigkeit. Die Ausgestaltung des Regelwerks kann sich unmittelbar auf die Leistungserbringer auswirken. Es obliegt der die Regeln definierenden Stelle, ob z.B. ein Windows XP Rechner für den schreibenden Zugriff auf eine ePA genutzt werden darf oder nicht. Eine vereinfachte Regel, die den Zugriff gestattet, sähe z.B. so aus: Der Zugriff auf die Aktion "ePA schreiben" ist für Nutzer mit derRolle Arzt erlaubt, wenn sie ein vertrauenswürdiges Betriebssystem nutzen und der Zugriff aus Deutschland erfolgt.

Die Werte der Attribute (unterstrichten dargestellt) werden innerhalb des Governance Prozesses auf der Grundlage sicherheitstechnischer und organisatorischer Bedingungen belegt. Erfüllt ein Anwender eine Anforderung nicht, so kann dies Implikationen haben (hier: nötiges Update des Betriebssystems).

6.2.13 NFA13 - Auswirkungen auf Leistungserbringerprozesse (Geräteregistrierung)

Ziel der Anforderung ist es, die Auswirkungen auf Prozesse der Leistungserbringer auf ein Minimum zu reduzieren.

Die Registrierung von Geräten variiert je nach Umgebungsart des Benutzers:

Gemanagte Umgebung (z.B. medizinische Versorgungszentren, Kliniken, Hebammenzentrum)

Gehören LE einer Institution mit einer gemanagten IT-Umgebung an, so werden die Geräte über dieses Management registriert und für den Zugriff auf die Ressourcen der Organisation eingerichtet. Die Registrierung erfolgt durch den Administrator der Organisation und hat damit keine direkten Auswirkungen auf die LE. Der Administrator ordnet Benutzer Accounts und Geräte zentral zu, im besten Fall über eine Management Oberfläche. Die Verantwortung für Sicherheit, Aktualisierung und Verwaltung der Geräte liegt bei der Institution.

Ungemanagte Umgebung (z.B. Praxis, mobiler LE)

Damit Geräte in der TI 2.0 genutzt werden können, müssen sie zunächst mit einem legitimen TI Identifikationsmerkmal verbunden werden. So ein Merkmal kann z.B. ein HBA oder eine SMC-B sein. Analog verhält es sich mit digitalen Ableitungen der Karten. Es ist unerheblich, ob eine physikalische Karte oder eine elektronische Identität zum Einsatz kommt. Das Identifikationsmerkmal wird genutzt, um Geräte der LE mit einer Gesundheits-App zu verknüpfen. Dieser Vorgang erfolgt einmalig bei der Installation der App auf dem jeweiligen Device. Nach der Zuordnung von App, Gerät und Identifikationsmerkmal kann die Verwendung der App mit dem IDP-Login des LE erfolgen. Neben der Zuordnung des TI-Identifikationsmerkmals zu Gerät und App ergeben sich keine weiteren Hürden für den LE.

6.2.14 NFA14 - Usability

Die Usability spielt für die Zero-Trust-Architektur eine entscheidende Rolle. Auch wenn die wenigsten Aspekte der Architektur für den Nutzer direkt sichtbar sind, hat diese einen

großen Einfluss auf die Nutzerfreundlichkeit der TI-Anwendungen. Zugangshürden und die Unterbrechung von Arbeitsabläufen müssen, sowohl beim Leistungserbringer als auch beim Patienten vermieden werden. Transparente und einfache Prozesse sowie eine hohe Stabilität fördern die Nutzung und den Mehrwert der Anwendungen.

Um zu darzustellen, was nach heutigem Stand der Technik und Digitalisierung der Gesellschaft geleistet werden kann, dient die folgende Tabelle.

Prozess	Nutzer ToDo	Frequenz	Auslöser	Erwarteter Aufwand Nutzer
Identifizierung beim IDP und Erhalt einer Rolle	Identifizierung mit einem Identitätsmittel auf hohem Vertrauensniveau (z.B. Online-Ausweisfunktion mit dem Personalausweis, Identifikation mit eGK, HBA, usw. über NFC am Smartphone, Vor-Ort-Identifikation mit dem Personalausweis eGK, HBA, usw. Erfolgt die Identifizierung mit einem Identitätsmittel der TI kann die Rolle in der TI übernommen werden. Erfolgt die Identifizierung z.B. mit dem Personalausweis, muss die Rolle in der TI durch eine berechtigte Stelle der TI bestätigt werden.	Abhängigkeit von der technischen Umsetzung des Identitätsmittels des IDP. Im Idealfall 1x bei Registrierung beim IDP	Identifizierung einer natürlichen Person oder funktionalen Entität. Erhalt der Rolle Arzt, Patient, Hebamme, Notfallsanitäter usw.	gering bis hoch (abhängig von den Verfahren zur Identifizierung und Authentisierung)
Bezug von IDP-Authentisierungsmitteln zur Bestätigung der Identität	Nach der Identifizierung beim IDP erhält der Benutzer **einen TI-Account**. Zum Zugriff auf diesen Account können unterschiedliche Faktoren hinterlegt werden. Für den Fall des Verlusts der Login-Daten bzw. Credentials zum TI-Account sollte Restore-Weg ausgewählt werden.	Abhängigkeit von der technischen Umsetzung des Identitätsmittels des IDP. Im Idealfall 1x bei Registrierung beim IDP	Registrierung beim IDP zur Teilnahme an der TI	gering bis hoch (abhängig von den Verfahren zur Ausstellung des Authentisierungsmittels)
Identifizierung für das Registrieren einer Endgeräte-Fach-Client Kombination	Zur Legitimation eines Gerätes zur Ausführung von TI-Applikationen muss ein valides Identifizierungs- bzw. Authentisierungsmittel vorgelegt werden.	1x je Installation eines Fach-Client/Trust-Client auf einem Gerät	Erstmalige Verwendung eines Fach-Client/Trust-Client auf einem Gerät	gering bis hoch (abhängig von den Verfahren zur Identifizierung und Authentisierung)

Prozess	Nutzer ToDo	Frequenz	Auslöser	Erwarteter Aufwand Nutzer
Authentisierung des Nutzers zur Verwendung von TI-Applikationen	Abhängig von der Implementierung des IDP und dem ausgestellten Authentisierungsmittel erfolgt die Authentisierung des Nutzers durch Nachweis verschiedener Faktoren aus dem Bereich Wissen z.B. PIN, Besitz z.B. privater Schlüssel in Chipkarte oder mobilem Endgerät oder Inhärenz z.B. biometrisches Merkmal wie Fingerabdruck oder Gesichtserkennung	Bei jedem Aufbau einer Session zum Zugriff auf Ressourcen der TI ggf. optimiert durch Implementierung von SSO im IDP	Zugriff auf eine Ressource der TI	gering
Einrichten einer Vertreterbeziehung für den Zugriff auf Ressourcen der TI	Damit ein Vertreter eines Nutzers an seiner Stelle auf Ressourcen der TI zugreifen darf, muss der Nutzer zunächst eine Vertreterbeziehung für den Vertreter hinterlegen. Dies erfolgt nach Identifizierung oder Authentisierung des Nutzers beim Dienst für Vertreterbeziehungen.	1 x für jede Vertreterbeziehung	Ein Nutzer möchte einem anderen Nutzer (ggf. auch einem Dienst) Zugriff auf seine Ressourcen gewähren	gering bis hoch (abhängig von den Verfahren zur Identif.. und Authentisierung)
Offline-Bereitstellung von TI-Daten für Patienten mit Einschränkung oder andere berechtigte Stellen	Patient: Bittet den Arzt während dem Besuch um eine Offline-Kopie der TI-Inhalte. LE: Identifikation des Patienten durch Sichtbestätigung und anschließende Bereitstellung der vom Patienten gewünschten Inhalte als Papierausdruck. Es wird ein Vermerk, dass der Ausdruck erstellt wurde, in der Akte des Patienten hinterlegt.	1x auf Wunsch des Patienten	Wunsch des Patienten bzw. anderer berechtigter Stelle wg. fehlender technischer Ausstattung oder sonstiger Einschränkung	gering

Tabelle 6.2.14-1: Usability

Die Nutzerfreundlichkeit hängt zum großen Teil von einfachen, aber sicheren Identifizierungs- und Authentisierungslösungen ab. Dies betrifft insbesondere die Erstidentifizierung bei einem IDP der Föderation der TI, die Authentisierung im Rahmen der Registrierung von Endgeräten sowie das Identifizieren im Rahmen der Konfiguration von Vertretungsbeziehungen. Der Zugriff auf Dienste der TI über die ZTA selbst erzeugt keine hohen Aufwände für den Nutzer.

Regelmäßig wiederkehrende Aufwände für den Nutzer müssen auf ein absolutes Minimum reduziert werden. Selten oder einmalig vorkommende Prozesse, müssen in jedem Fall durch einfach verständliche Softwareassistenten oder Anleitungen gestützt werden. Medienbrüche oder Prozessunterbrechungen sind dringend zu vermeiden.

Die Nutzerakzeptanz ist umgekehrt proportional zur Komplexität der für eine Teilnahme an der TI nötigen Schritte. Um die Usability für den Anwender möglichst optimal zu gestalten, sollten Usability- und Nutzer-Test zum Teil von Zulassungen und Abnahmen werden.

6.2.15 NFA15 - Standards

Das Ziel der Anforderung NFA15 besteht insbesondere darin, so weit wie möglich international anerkannte und in ihrem Kern wissenschaftlich geprüfte Standards zu nutzen, um so an den Stand der Technik anzuknüpfen und davon ausgehend die ZTA zukunftsfähig und souverän mit dem Stand der Technik weiterentwickeln zu können.

Die in diesem Konzept vorgeschlagene Architektur basiert, gem. Kapitel 3.2, auf den grundlegenden Strategien einer ZTA, wie in NIST SP 800-207 [NIST-ZeroTrust] erläutert. Sie nutzt dafür die gem. ISO/IEC 29146 [ISO29146] standardisierten und im Rahmen von Zero-Trust etablierten logischen Komponenten. Zur Kommunikation zwischen den Komponenten werden etablierte Standards wie TLS und HTTP benutzt. Die Identifizierung bzw. Authentifizierung von Nutzern und Geräten erfolgt über OpenID Connect [OpenID-Connect] und OAuth2 [RFC6749_OAuth2]. Hinsichtlich der Attestierung von Endgeräten setzt das Konzept auf den bei den Nutzern verfügbaren Stand der Technik. Da es hierbei keine herstellerübergreifenden Standards gibt, muss allerdings die je Endgeräteklasse vorliegende proprietäre Implementierung der Endgerätehersteller berücksichtigt werden (z.B. Google, Apple, Microsoft). Bzgl. des Regelwerks wird eine herstellerunabhängige und idealerweise standardisierte Lösung bevorzugt, allerdings sind entsprechende Lösungen ggf. nicht mit der gewünschten Zukunftsfähigkeit verfügbar. Für die konkrete Umsetzung von PEP und PDP gibt es keine Standards. Für maximale Zukunftsfähigkeit wird daher die Erstellung und Pflege einer herstellerunabhängigen produktionstauglichen Open Source Implementierung empfohlen.

6.2.16 NFA16 - Verfügbarkeit

Ziel der Anforderung NFA16 ist es, den Nutzern störungsfrei Zugang zu allen angefragten Diensten oder Leistungen zu gewährleisten. Dazu ist es notwendig, dass sowohl jede einzelne ZTA-Komponente als auch die Gesamtheit der zur Ausführung der Anfrage notwendigen Komponenten verfügbar sind. Um diese Verfügbarkeit zu gewährleisten, werden die ZTA-Komponenten robust gegen Lasten und ausfallsicher betrieben (vgl. NFA8 - Skalierbarkeit). Die Anbieter werden zu einer definierten Verfügbarkeit vertraglich verpflichtet.

6.2.17 NFA17 - Betreibbarkeit

Das Ziel der Anforderung NFA17 besteht darin, das ZTA-Komponenten von allen Anbietern sicher, effizient und mit etablierten Verfahren in vergleichbarer Qualität bereitgestellt werden. Die ZTA-Komponenten werden so entwickelt, dass sie möglichst unabhängig lauffähig sind. Das bedeutet, dass sie definierte Schnittstellen nach außen haben. Im Falle einer Integration (z.B. in einen Fach-Client) werden diese nach außen gelegt. Ziel ist es, die Komponenten auf RZ-Plattformen oder in der Cloud sicher und effizient zu betreiben. Die ZTA-Komponenten werden möglichst hardwareunabhängig entwickelt. Open Source Implementierungen, insbesondere für Komponenten, wie z.B. PEP und PDP können so auf einer Vielfalt von Plattformen betrieben werden.

7 Zusammenfassung und Ausblick

Die mit diesem Feinkonzept vorgeschlagene Architektur, sowie die Umsetzungsmöglichkeiten und Governanceaspekte haben das Potential, sowohl die heutigen als auch die zukünftigen Anforderungen bei der Digitalisierung des Gesundheitswesens bedienen zu können und den Anwendern einen nutzerfreundlichen, verfügbaren, skalierbaren, sicheren und privacy-freundlichen Zugang zu bestehenden und zukünftigen Fachdiensten zu geben. Die dargestellte Umsetzung des Zero Trust Paradigmas ermöglicht eine hohe Granularität und Dynamik bei der Zugriffskontrolle und wächst mit den zukünftigen technologischen Weiterentwicklungen sowohl auf der Seite der Zugangsgeräte als auch der Fachdienste mit. Insbesondere ist abzusehen, dass die technologische Entwicklung der Zugangsgeräte in Zukunft bessere Usability bei gleichzeitig höherer Sicherheit ermöglichen werden und dass zukünftige Cloudumgebungen den Betreibern von Fachdiensten mehr Sicherheitsgarantien bieten können.

Ausgehend von einer Definition der funktionalen und nicht-funktionalen Anforderungen als Rahmen für die Architektur in Kapitel 2, befasst sich der Schwerpunkt des Dokuments mit der detaillierten Beschreibung der Architektur in Kapitel 3, d.h. mit den beteiligten Komponenten, ihren Aufgaben, dem Zusammenspiel mit einem dynamischen Regelwerk für eine granulare und adaptive Zugriffskontrolle, sowie mit der konzeptionellen Betrachtung von Datenschutz- und Privacy-Eigenschaften. Dabei wurde großer Wert daraufgelegt, dass die Architektur weitestgehend mit bereits etablierten und standardisierten bzw. anderweitig herstellerunabhängigen Technologien umsetzbar ist. Das ermöglicht aus technologischer Sicht sowohl eine zeitnahe Umsetzung als auch eine langfristig souveräne Weiterentwicklung.

Die tiefere Beschreibung entsprechender Umsetzungsmöglichkeiten wurde in Kapitel 4 vorgenommen. Der Fokus wurde dabei auf hohe Verfügbarkeit, Performance und Skalierbarkeit gelegt, aber Entwicklungsgeschwindigkeit und Interoperabilität der Komponenten sowie die Zukunftsfähigkeit spezifischer Umsetzungen wurden in die Betrachtungen einbezogen. Essenzielle Prozesse beim Betrieb der Architektur wurden in Kapitel 5 beschrieben und es wurde herausgearbeitet, wie die Prozesse für die essenziellen Stakeholder, Leistungserbringer, Versicherte und Dienstanbieter, im Vergleich zum aktuellen Stand vereinfacht werden. Auf den politischen Rahmen wurde nicht tiefer eingegangen, obwohl die wettbewerblichen und rechtlichen Rahmenbedingungen für Umsetzung und Betrieb der TI-Komponenten, Endgeräte und Fachdienste durchaus einen signifikanten Einfluss auf die Komplexität und Zuverlässigkeit des Betriebs haben können. Bei der abschließenden Evaluation wurde in Kapitel 6.1 die Umsetzbarkeit und die Usability konkreter Anwendungsfälle in der TI aus Sicht der Anwender beschrieben. Kapitel 6.2 rundet die Betrachtungen der vorhergehenden Kapitel ab durch einen zusammenfassenden Abgleich der initial aufgestellten funktionalen und nicht-funktionalen Anforderungen mit den beschriebenen Umsetzungen in Architektur, Technik und Governance.

Ausgehend vom eingereichten Grobkonzept und in enger Zusammenarbeit mit der gematik wurden mit diesem Feinkonzept die grundlegenden Details für eine zukunftsweisende Umgestaltung der TI ausgearbeitet und mit dem PoC die technologische Machbarkeit grundlegender Teile gezeigt. Für die Vertiefung und Realisierung des Konzepts ist die Einbeziehung weiterer Stakeholder nötig. Welche Tiefe im Feinkonzept adressiert wurde und welche Folgearbeiten darauf aufbauend nötig werden, wird im Weiteren beschrieben.

7.1 Regelwerk

Im Feinkonzept wurde ein flexibles und mächtiges Regelwerk konzipiert, um einer Vielfalt an aktuellen und zukünftigen Anforderungen gerecht werden zu können. Flexibilität und Mächtigkeit können allerdings negative Auswirkungen auf Komplexität und Performance haben. Daher ist, ausgehend von den im Konzept aufgezeigten Möglichkeiten, eine Konkretisierung der Anforderungen an das Regelwerk nötig, um einen passenden Trade-Off zwischen Komplexität und Performanceeinschränkungen auf der einen und Flexibilität und Mächtigkeit auf der anderen Seite zu finden. Bei der Konkretisierung sollten als Stakeholder zum einen die Industrie mit ihren Erfahrungen bei Umsetzung und Betrieb einbezogen werden, als auch BSI und BfDI für die Betrachtung von Sicherheit und Datenschutz.

Ausgehend von dieser Konkretisierung der Anforderungen sollte eine Evaluation technischer Umsetzungsmöglichkeiten erfolgen, die sowohl existierende Regel-Engines und Regel-Sprachen einbezieht als auch die Möglichkeit einer auf die Anforderungen der TI zugeschnittenen eigenen Sprache. Neben den technischen Fähigkeiten spielen bei der Evaluation auch die Komplexität und Wartbarkeit sowie die langfristige Verfügbarkeit, Kontrolle und Offenheit für Weiterentwicklungen eine essenzielle Rolle.

Die konkretisierten Anforderungen an die Regeln haben Auswirkungen auf die auf dem Endgerät zu erfassenden Informationen (TCL, GMS) und damit auf die Komplexität der Umsetzung. Ggf. werfen die benötigten Informationen auch Datenschutz- und Privacy-Fragen auf und haben damit Einfluss auf den Schutzbedarf der erfassten Daten. An den konkreten Anforderungen der Regeln hängen auch die Art, Menge und Updatehäufigkeit der über PIP bereitzustellenden Informationen. Diese Details können maßgeblichen Einfluss auf die Performance der Regelausführung haben, z.B. wenn nicht mehr alle Informationen lokal im PDP gecacht werden können, sondern bei jeder Regelevaluation eine Abfrage von PIP erfolgen muss.

7.2 Identifikation der Nutzer am GMS

Um eine Allmacht des IDP beim Zugriff auf Nutzerdaten zu vermeiden, soll ein Zugriff auf Fachdienste nur von registrierten Geräten möglich sein. Nicht geklärt wurde dabei bisher die initiale Identifikation der Nutzer bei der Registrierung von Geräten - eine direkte Kopplung mit der Authentisierung am IDP ist zur Vermeidung einer Allmacht-Stellung explizit unerwünscht. Die Herausforderung hier ist ein Vorgehen zu entwickeln, welches sowohl ausreichend sicher als auch ausreichend nutzerfreundlich und akzeptabel aus der Perspektive der Nutzer ist. Neben der Evaluation verschiedener technischer Möglichkeiten durch Zusammenarbeit mit der Industrie sind für die Klärung eines Kompromisses zwischen Sicherheit und Usability auch Abstimmungen mit dem BSI und ggf. auf politischer Ebene nötig.

7.3 Integration mit IAM und ISMS in Krankenhäusern und größeren Praxen

Es wird als nicht realistisch angesehen, in größeren, komplexeren und entsprechend dynamischen Infrastrukturen von Leistungserbringern sämtliche Nutzer und Geräte der Institution innerhalb der TI zu registrieren und aktuell zu halten. Der im Konzept vorgeschlagene Ansatz ist ein übergreifender Trust-Client, welcher in das bestehende ISMS und IAM integriert wird und auf diese Weise die Zero-Trust Prinzipien auf die Geräte und Nutzer innerhalb der Institutionen ausdehnen kann. Ziel ist hier, die Kontrolle über

die internen Nutzer und Geräte bei der Institution zu belassen, gleichzeitig aber ausreichend Transparenz und Sicherheit beim Zugriff auf die TI zu ermöglichen. Eine Vertiefung des Konzeptes in dieser Richtung sollte in Zusammenarbeit mit Stakeholdern aus Industrie, Krankenhäusern und auch BSI und BfDI erfolgen.

7.4 Zusammenspiel mit bereits bestehenden Komponenten

Im Rahmen der vorgeschlagenen Architektur werden bestimmte Komponenten als extern gegeben vorausgesetzt: die föderierte IDP-Infrastruktur, der Federation Master, die PKI, DNS und DDoS-Schutz. Zwar wird hier auf Standard-Protokolle aufgesetzt, aber Details der Kommunikation, z.B. die ausgetauschten Attribute, müssen spezifiziert und ggf. auch bisherige Spezifikationen der Komponenten angepasst werden.

7.5 Übergreifende Analyse bzgl. Sicherheit und Datenschutz

In der vorgeschlagenen Architektur wird soweit möglich auf etablierte Protokolle wie OIDC und OAuth aufgesetzt, auch um bereits vorhandene Sicherheitsbetrachtungen nutzen zu können. Diese Protokolle sind jedoch nicht trivial und in der Vergangenheit gab es durchaus Lücken sowohl bei der konkreten Umsetzung als auch beim konzeptionellen Einsatz. Es können sich Probleme beim Zusammenspiel der Komponenten ergeben und evtl. auch durch Aspekte der Governance. Im Rahmen der Spezifikation der einzelnen Komponenten sollte daher eine übergreifende Sicherheits- und Datenschutzanalyse unter Einbeziehung des BSI und des BfDI erfolgen. Ergebnis dieser Sicherheitsanalyse sollte auch die Festlegung der Gültigkeitszeiten für die im Datenfluss involvierten ID- und Geräte-Token, ggf. auch für die Session-Informationen sein. Bei der Wahl der Gültigkeitszeiten ist die Balance zwischen Sicherheit und Performance zu beachten: Kürzere Zeiten erhöhen die Sicherheit führen aber zu einer höheren Last in der TI.

7.6 Interoperabilität im Betrieb auf technischer und organisatorischer Ebene

In der vorgeschlagenen Architektur wird so weit wie möglich auf Standardprotokolle aufgesetzt, die aber passend parametrisiert werden müssen. An einigen Stellen, wie z.B. beim Zusammenspiel zwischen Trust- und Fach-Client oder bei der Weitergabe von Geräteinformationen vom Trust-Client an das GMS, gibt es jedoch keine passenden Standards. Entsprechend komplex kann sich die Interoperabilität auf technischer Ebene gestalten. Detaillierte technische Spezifikationen kombiniert mit Interoperabilitätstests können eine hohe Interoperabilität zwischen Komponenten verschiedener Hersteller befördern, gehen jedoch mit einer deutlich reduzierten Agilität bei der Weiterentwicklung einher. Hier sollte zusammen mit Stakeholdern aus Industrie und Politik evaluiert werden, für welche Teile der Umsetzung Wettbewerb zwischen Herstellern förderlich oder eher hinderlich ist und wo herstellerübergreifende Umsetzungen (z.B. als Open Source) eine bessere Wahl darstellen.

Analoge Betrachtungen sind zum Support und zu anderen Betriebsaspekten anzustellen. Während aus Nutzersicht ein zentraler Support mit starker Durchsetzungskraft gegenüber den in der TI beteiligten Playern attraktiv ist, könnte dies die beteiligten Player jedoch in der Entwicklung tragfähiger Geschäftsmodelle beeinträchtigen. Diese Fragen müssen auf politischer Ebene geklärt werden.

7.7 Migration

Im Rahmen des Feinkonzepts wurde eine Zukunftsvision für eine Zero-Trust-basierte TI für ein digitalisiertes Gesundheitssystem entwickelt. Um den Umstieg von der derzeitigen TI 1.0 in diese visionäre Architektur zu vollziehen, braucht es ein realistisches Migrationskonzept, welches sowohl Investitionssicherheit als auch Zukunftsfähigkeit vereinigt. Dieses Migrationskonzept wird im Anschluss an das Feinkonzept realisiert.

8 Abkürzungsverzeichnis

Abkürzung	Ausgeschriebener Begriff
ABAC	Attribute-based Access Control
AUM	Authenticator-Module
CA	Certificate Authority
DiGA	Digitale Gesundheitsanwendungen
eGK	elektronische Gesundheitskarte
ePA	elektronische Patientenakte
eRezept	Elektronisches Rezept
FCL	Fach-Client
FEM	Federation Master
FHIR	Fast Healthcare Interoperability Resources
FHIR	Fast Healthcare Interoperability Resources
GMS	Geräte Management Service
HBA	Heilberufsausweis
IAM	Identity and Access Management
IDP	Identity Provider
ITSM	IT-Service-Management
LE	Leistungserbringer
LoA	Level of Assurance
MON	übergreifendes Monitoring
(m)TLS	(mutual) Transport Layer Security
NCPeH	National Contact Point eHealth

Abkürzung	Ausgeschriebener Begriff
OIDC	OpenID Connect
OPA	Open Policy Agent
PAP	Policy Administration Point
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PIP	Policy Information Point
SE	Secure Element
SIEM	Security Information and Event Management
SMC-B	elektronischer Praxis- oder Institutionsausweis
SOC	Security Operations Center
SZZP	Sicherer Zentraler Zugangspunkt
TCL	Trust-Client
TEE	Trusted Execution Environment
TI	Telematik-Infrastruktur
TIM	Telematikinfrastruktur-Messenger
TPM	Trusted Platform Module
UHD	User Help Desk
VAU	Vertrauenswürdige Ausführungsumgebung
WANDA	weitere Anwendungen für den Datenaustausch in der Telematikinfrastruktur
XACML	eXtensible Access Control Markup Language
ZTA	Zero Trust Architektur

9 Literaturverzeichnis

[BSI_TR02102]	BSI TR-02102. Kryptographische Verfahren: Empfehlungen und Schlüssellängen, BSI, https://www.bsi.bund.de/dok/e6615148
[EU_DSGVO]	Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG, Amtsblatt der Europäischen Union L 119/1, 4.5.2016, https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679
[gem_Leistung]	2022-0054: Zero Trust Architektur für die Telematikinfrastruktur. Funktionale Leistungsbeschreibung (Anlage-02), gematik, 08.04.2022
[gem_Glossar]	Glossar der Telematikinfrastruktur, gematik, V5.2.0, 20.01.2022, https://fachportal.gematik.de/fileadmin/Fachportal/Glossar/gemGlossar_V5.2.0.pdf
[gem_IDP_Federation]	Feature: IDP Föderation, gematik, V 1.0.0 (öffentlicher Entwurf), 09.12.2022
[gem_Spec_IDP_Sek]	Spezifikation Sektoraler Identity Provider, gematik, V2.0.0 (öffentlicher Entwurf), 30.09.2022
[gem_Whitepaper_TI2.0]	Whitepaper: TI 2.0 - Arena für digitale Medizin, gematik, 21.01.2021, https://www.gematik.de/media/gematik/Medien/Newsroom/Publikationen/Informations-materialien/gematik_Whitepaper_Arena_digitale_Medizin_TI_2.0_Web.pdf
[ISO29146]	ISO/IEC 29146:2016. Information technology - Security techniques - A framework for access management, ISO/IEC, Juni 2016, https://www.iso.org/standard/45169.html
[NIST_ZeroTrust]	NIST Special Publication 800-207. Zero Trust Architecture, NIST, August 2020 https://csrc.nist.gov/publications/detail/sp/800-207/final

[NIST_ABAC-Comparison]	NIST Special Publication 800-178. A Comparison of Attribute Based Access Control (ABAC) Standards for Data Service Applications, NIST, Oktober 2016, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-178.pdf
[NIST_PolicyVerification]	NIST Special Publication 800-192. Verification and Test Methods for Access Control Policies/Models, NIST, Juni 2017, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-192.pdf
[OPA_Rego]	Policy Language, Open Policy Agent, v0.48.0, https://www.openpolicyagent.org/docs/latest/policy-language/
[OpenID-Connect]	OpenID Connect Core 1.0 incorporating errata set 1, OpenID Foundation (OIDF), 8.11.2014, https://openid.net/specs/openid-connect-core-1_0.html
[OpenID-Connect_CIBA]	OpenID Connect Client Initiated Backchannel Authentication Flow, OpenID Foundation (OIDF), 01.09.2021, OpenID Connect Client-Initiated Backchannel Authentication Flow - Core 1.0
[OpenID-Connect_BCL]	OpenID Connect Back-Channel LogoutOpenID Foundation (OIDF) ,12.09.2022, OpenID Connect Back-Channel Logout 1.0
[OASIS_XACML]	eXtensible Access Control Markup Language (XACML) Version 3.0 Plus Errata 01, OASIS, V3.0, 12.07.2017, http://docs.oasis-open.org/xacml/3.0/errata01/os/xacml-3.0-core-spec-errata01-os-complete.html
[RFC5280-Certs]	RFC5280. Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, Internet Engineering Task Force (IETF), Mai 2008, https://www.ietf.org/rfc/rfc5280.txt
[RFC6749_OAuth2]	RFC 6749. The OAuth 2.0 Authorization Framework, Internet Engineering Task Force (IETF), Oktober 2012, https://www.ietf.org/rfc/rfc6749.txt
[RFC7636_OAuth2_PKCE]	RFC 7636. Proof Key for Code Exchange by OAuth Public Clients, Internet Engineering Task Force (IETF), September 2015, https://www.ietf.org/rfc/rfc7636.txt

[RFC8693_OAuth2_Token_Exchange]	RFC 8693, OAuth 2.0 Token Exchange, Internet Engineering Task Force (IETF), January 2020, https://www.ietf.org/rfc/rfc8693.txt
[RFC8705_OAuth2_mTLS]	RFC 8705, OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens, Internet Engineering Task Force (IETF), Februar 2020, https://www.ietf.org/rfc/rfc8705.txt
[RFC9126_OAuth2_PAR]	RFC 9126, OAuth 2.0 Pushed Authorization Requests, Internet Engineering Task Force (IETF), September 2021, https://www.ietf.org/rfc/rfc9126.txt
[RFC8446-TLS]	RFC 8446. The Transport Layer Security (TLS) Protocol Version 1.3, Internet Engineering Task Force (IETF), August 2018, https://www.ietf.org/rfc/rfc8446.txt

10 Anhang

10.1 Anhang 1: "Beschränkung der Geräte-Identität"

Das Konzept sieht, wie in Kapitel 3.5.2 beschrieben, eine gerätegebundene Identität UUID_Gerät vor, die für alle Anwendungen auf dem gleichen Gerät verwendet werden kann. Die technische Umsetzung als ein einziges geräteweit gültiges Schlüsselpaar für das Gerät kann jedoch durch gerätespezifische Gegebenheiten erschwert oder verhindert werden. Deshalb können bei Bedarf stattdessen auch gerätegebundener App-Zertifikate mit eigenen Schlüsselpaaren für die einzelnen Fach-/Trust-Clients verwendet werden. Im Folgenden wird eine grobe Übersicht über die am stärksten verbreiteten Geräteklassen und der aktuellen Einschätzung zur Umsetzbarkeit eines einzelnen geräteweit gültigen gerätegebunden Schlüsselpaars (UUID_Gerät) gegeben. Eine abschließende Aussage zur Umsetzbarkeit auf geforderten Endgeräten erfordert eine tiefergehende technische Analyse.

iOS/iPadOS/MacOS

Die Geräte und das darauf aufsetzende Betriebssystem samt Bibliotheken bietet unterschiedliche Möglichkeiten einen kryptographischen Schlüssel sicher in Hardware zu hinterlegen. Erste Untersuchungen deuten darauf hin, dass es unter iOS nicht möglich ist, einen in Hardware generierten und hinterlegten Schlüssel mit allen Apps zu teilen. Mögliche Ansätze, wie eine geräteweite Identität dennoch umgesetzt werden könnte, wären unter anderem CryptoKit, um Schlüssel in der *SecureEnclave* zu generieren, und AppGroups, um das Schlüsselmaterial mit anderen Anwendungen des gleichen Entwicklers zu teilen. Ein weiterer Ansatz wäre möglicherweise auch der systemweite Schlüsselbund, lokale VPN-Anwendungen (Network Extension) oder die integrierte FIDO2-Schnittstelle, welche aber nicht frei von Anwendungen verwendet werden können oder andere Nachteile haben. MacOS stellt prinzipiell die gleichen Schnittstellen bereit, allerdings ist hier durch die Desktop-Umgebung die Trennung zwischen Anwendungen weniger strikt. Ohne eine genaue technische Untersuchung lassen sich aber keine genauen Angaben machen.

Android

Die Gerätelandschaft unter Android Geräten ist sehr heterogen. Es bestehen zwar hardwareabstrahierende Layer zur Schlüsselverwaltung, einige Funktionen sind aber von bestimmter Hardware abhängig. Unter Android kann über die KeyStore API ähnlich zu iOS Schlüssel in einer gesicherten Umgebung (TrustZone, Strongbox) erzeugt und verwendet werden. Schlüssel allgemein können über den Android KeyStore Provider entweder anwendungsbezogen oder über die KeyChain systemweit adressiert werden. Der anwendungsbezogene Android KeyStore Provider scheint alle Funktionen der KeyStore API verwenden zu können, während die KeyChain diese API nur im Hintergrund verwendet und nicht die gesamte Funktionalität ermöglicht. Aber auch hier sind tiefergehende technische Analysen notwendig, um genauere Angaben machen zu können.

Windows

Unter Windows wird seit Windows 10 die *Cryptography API: Next Generation* (CNG) bereitgestellt. Diese kann Schlüssel in Hardware erzeugen und verwenden. Ohne weitere Schutzmechanismen stehen diese Schlüssel dann auch allen Anwendungen auf dem System zur Verfügung. Für eine Hardwarebindung muss auch entsprechende Hardware im PC verbaut sein. Seit 2016 ist das TPM Teil der Windows Minimum Hardware Requirements.

Linux

Auf Linux ist bisher keine flächendeckende übergreifende Schnittstelle verfügbar, die den Hardwarezugriff abstrahiert. Hier ist aktuell die TPM-Spezifikation maßgeblich, die durch Proxy-Anwendungen, etwa als PKCS11, abstrahiert werden kann. Auch hier stehen ohne weitere Schutzmaßnahmen alle Schlüssel allen Anwendungen zur Verfügung.

10.2 Anhang 2: "Verwendung gerätegebundener App-Zertifikate mit mTLS"

Disclaimer

Im Kontext der Betrachtungen aus Anhang 1 können im Folgenden die Begriffe Geräte-Zertifikat und Geräte-UUID stets auch für eines von mehreren geräte-gebundenen applikationsspezifischen Zertifikaten stehen.

Ziel

Das Gerät eines Nutzers muss in der ZTA für diesen Nutzer registriert sein, bevor der Nutzer damit auf einen Fachdienst zugreifen kann. Daraus ergeben sich folgende Teilziele:

- Das Gerät muss eine eindeutige Identität (UUID_Gerät) haben, welche beim Geräte Management Service (GMS) mit der Identität des Nutzers (UUID_Nutzer) assoziiert wird
- Nur ein am GMS registriertes Gerät darf vom GMS für den Zugriff auf einen Fachdienst authentisiert werden, d.h. ein Geräte-Token bekommen
- Der Geräte-Token darf (insb. gegenüber dem PEP/Fachdienst) nur von dem Gerät benutzt werde, für welches er ausgestellt wurde

Im Folgenden werden 4 Optionen dargestellt und verglichen:

	Option 1: Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens	**Option 2: GMS-signierte Zertifikate ohne Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens**	**Option 3: GMS-signierte Zertifikate inklusive Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens**	**Option 4: Nur server-side TLS ohne Certificate-Bound Access Tokens**
Relevante Standards	RFC 8705 on "OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens" mit Verwendung von Section 2.2. "Self-Signed Certificate Mutual-TLS Method"	RFC 8705 on "OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens" mit Verwendung von Section 2.1. "PKI Mutual-TLS Method"	RFC 8705 on "OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens" mit Verwendung von Section 2.1. "PKI Mutual-TLS Method"	https://www.rfc-editor.org/rfc/rfc8446(einseitig, nur für Server-Authentifizierung) and OAuth Access Tokens gemäß Section 1.4 in RFC6749 (ohne Bindung ans Client-Zertifikat)
UUID_Gerät	Hash des selbst-signierten Zertifikats	Hash des GMS-signierten Zertifikats	Hash des GMS-signierten Zertifikats	Hash des öffentlichen Schlüssels

Ablauf Geräte-Registrierung	• TCL generiert Schlüsselpaar • TCL erstellt **selbst-signiertes** Zertifikat für öffentlichen Schlüssel • TCL registriert sich beim GMS mit dem selbst-signierten Zertifikat • Der Nutzer authentisiert sich im Rahmen des Registrierungsprozesses als UUID_Nutzer • Der GMS speichert die Zuordnung UUID_Nutzer ↔ UUID_Gerät (Zertifikatshash) ab	• TCL generiert Schlüsselpaar • TCL registriert sich beim GMS mit dem öffentlichen Schlüssel • Der Nutzer authentisiert sich im Rahmen des Registrierungsprozesses als UUID_Nutzer • Der **GMS stellt für den öffentlichen Schlüssel des Gerätes ein Zertifikat aus**, das nur den öffentlichen Schlüssel enthält und damit einen Nachweis, dass dieses Gerät beim GMS registriert wurde • Der GMS speichert die Zuordnung UUID_Nutzer ↔ UUID_Gerät (Zertifikatshash) ab	• TCL generiert Schlüsselpaar • TCL registriert sich beim GMS mit dem öffentlichen Schlüssel • Der Nutzer authentisiert sich im Rahmen des Registrierungsprozesses als UUID_Nutzer • Der **GMS stellt für den öffentlichen Schlüssel des Gerätes ein Zertifikat aus,** das die Zuordnung UUID_Nutzer ↔ UUID_Gerät enthält • Der GMS speichert die Zuordnung UUID_Nutzer ↔ UUID_Gerät (Zertifikatshash) ab (für Revokationszwecke und spätere Gerätemanagement-Aktivitäten)	• TCL generiert Schlüsselpaar • TCL registriert sich beim GMS mit dem öffentlichen Schlüssel • Der Nutzer authentisiert sich im Rahmen des Registrierungsprozesses als UUID_Nutzer • Der GMS speichert die Zuordnung UUID_Nutzer ↔ UUID_Gerät (Hash öffentlicher Schlüssel) ab

Gerät authentisieren beim Fachdienst	• TCL führt mTLS-Handshake mit dem GMS durch (beinhaltet Nachweis über Besitz des privaten Schlüssels durch das Gerät) • TCL fragt Geräte-Token an • GMS stellt Geräte-Token aus (für das verwendete Client-Zertifikat) • TCL führt mTLS-Handshake mit dem PEP durch (beinhaltet Nachweis über Besitz des privaten Schlüssels durch das Gerät) • TCL übermittelt das Geräte-Token • PEP prüft, dass das Client-Zertifikat zu dem Hash im Geräte-Token passt	• TCL führt mTLS-Handshake mit dem GMS durch (beinhaltet Nachweis über Besitz des privaten Schlüssels durch das Gerät) • TCL fragt Geräte-Token an • GMS stellt Geräte-Token aus (für das verwendete Client-Zertifikat) • TCL führt mTLS-Handshake mit dem PEP durch (beinhaltet Nachweis über Besitz des privaten Schlüssels durch das Gerät **und** Prüfung, dass das Client-Zertifikat durch einen GMS ausgestellt wurde) • TCL übermittelt das Geräte-Token • PEP prüft, dass das Client-Zertifikat zu dem Hash im Geräte-Token passt	• TCL führt mTLS-Handshake mit dem GMS durch (beinhaltet Nachweis über Besitz des privaten Schlüssels durch das Gerät) • TCL fragt Geräte-Token an • GMS stellt Geräte-Token aus (für das verwendete Client-Zertifikat) • TCL führt mTLS-Handshake mit dem PEP durch (beinhaltet Nachweis über Besitz des privaten Schlüssels durch das Gerät **und** Prüfung, dass das Client-Zertifikat durch einen GMS ausgestellt wurde) • TCL übermittelt das Geräte-Token • PEP prüft, dass das Client-Zertifikat zu dem Hash im Geräte-Token passt	• TCL führt TLS-Handshake (nur server-seitige Zertifikatsprüfung) mit dem GMS durch • TCL fragt Geräte-Token an • GMS prüft, dass TCL im Besitz des entsprechenden privaten Schlüssels ist (z.B. Challenge-Response Verfahren) • GMS stellt Geräte-Token aus • TCL führt TLS-Handshake (nur server-seitige Zertifikatsprüfung) mit dem PEP durch • TCL übermittelt das Geräte-Token • PEP prüft, dass TCL im Besitz des entsprechenden privaten Schlüssels ist (z.B. Challenge-Response Verfahren)

	Option 1: Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens	Option 2: GMS-signierte Zertifikate ohne Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 3: GMS-signierte Zertifikate inklusive Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 4: Nur server-side TLS ohne Certificate-Bound Access Tokens
Ziel bei Verwendung des Zertifikats	• Authentifizierung des Gerätes über privatem Schlüssel (durch Challenge-Response im TLS-Handshake)	• Authentifizierung des Gerätes über privatem Schlüssel (durch Challenge-Response im TLS-Handshake) • Im PEP beim Verbindungsaufbau Verbindungsanfragen von nicht-registrierten Geräten rausfiltern	• Authentifizierung des Gerätes über privatem Schlüssel (durch Challenge-Response im TLS-Handshake) • Im PEP beim Verbindungsaufbau Verbindungsanfragen von nicht-registrierten Geräten rausfiltern • Zuordnung UUID_Nutzer ↔ UUID_Gerät unabhängig von einer zentralen Datenbank zwischen GMS-Instanzen (und evtl. auch woanders) nachweisbar machen	-- (kein Zertifikat für Nutzerendgerät)

	Option 1: Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens	Option 2: GMS-signierte Zertifikate ohne Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 3: GMS-signierte Zertifikate inklusive Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 4: Nur server-side TLS ohne Certificate-Bound Access Tokens
Laufzeit des Zertifikats	langfristige Gültigkeit des Zertifikats möglich (Änderungen erfordern Update der Gerät-Nutzer-Assoziation am GMS)	langfristige Gültigkeit des Zertifikats möglich (Änderungen erfordern Update der Gerät-Nutzer-Assoziation am GMS)	Zwei Optionen: 1. Langfristig gültiges Zertifikat (Änderungen erfordern Update der Gerät-Nutzer-Assoziation am GMS), GMS bietet Mechanismus für Revokations-Prüfung (OCSP), um De-Registrierung von Geräten zu ermöglichen 2. Kurzfristig gültiges Zertifikat, sodass Revokation nicht nötig ist -> dafür regelmäßige Erneuerung nötig (mit Update der Gerät-Nutzer-Assoziationen)	-- (kein Zertifikat für Nutzerendgerät)

	Option 1: Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens	**Option 2: GMS-signierte Zertifikate ohne Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens**	**Option 3: GMS-signierte Zertifikate inklusive Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens**	**Option 4: Nur server-side TLS ohne Certificate-Bound Access Tokens**
Inhalt Zertifikat	Öffentlicher Schlüssel für Nutzerendgerät	Öffentlicher Schlüssel für Nutzerendgerät	Öffentlicher Schlüssel für Nutzerendgerät **Zuordnung zu UUID_Nutzer (dadurch mit UUID_Gerät aus Zertifikat die Nutzer-Geräte-Assoziation gespeichert)**	-- (kein Zertifikat für Nutzerendgerät)
Inhalt Geräte-Token	Zuordnung UUID_Nutzer ↔ UUID_Gerät + Attestierungsinformationen zum Gerät	Zuordnung UUID_Nutzer ↔ UUID_Gerät + Attestierungsinformationen zum Gerät	bei Option 1 zur Laufzeit: Revokationsstatus für das verwendete Client-Zertifikat (OCSP-Response) + Attestierungsinformationen zum Gerät	Zuordnung UUID_Nutzer ↔ UUID_Gerät + Attestierungsinformationen zum Gerät

	Option 1: Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens	Option 2: GMS-signierte Zertifikate ohne Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 3: GMS-signierte Zertifikate inklusive Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 4: Nur server-side TLS ohne Certificate-Bound Access Tokens
Vorteile:	• Zertifikat wird durch TCL selbst erzeugt, keine PKI nötig • Kein Prüfen von Zertifikatskette und/oder Revokation für Client-Zertifikat nötig • Bindung der mTLS-Verbindung an das authentifizierte Endgerät (durch Certificate-Bound Access Tokens)	• Frühzeitige Erkennung nicht registrierter Geräte am PEP (beim mTLS-Verbindungsaufbau) • Revokation nicht zwingend nötig, da Zertifikat lediglich für "Vor-Filterung" der Zugriffsanfragen beim PEP verwendet wird • Bindung der mTLS-Verbindung an das authentifizierte Endgerät (durch Certificate-Bound Access Tokens)	• Frühzeitige Erkennung nicht registrierter Geräte am PEP (beim mTLS-Verbindungsaufbau) • Nachweis über erfolgte Zuordnung UUID_Nutzer und UUID_Gerät kann auch ohne (Echtzeit-)Zugriff auf Datenbank registrierter Geräte erfolgen, ggf. auch über mehrere GMS-Anbieter hinweg • Bindung der mTLS-Verbindung an das authentifizierte Endgerät (durch Certificate-Bound Access Tokens)	• Kein Zertifikat für Nutzerendgerät nötig (aber weiterhin Schlüsselpaar)

	Option 1: Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens	Option 2: GMS-signierte Zertifikate ohne Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 3: GMS-signierte Zertifikate inklusive Assoziation Gerät-Nutzer für mTLS mit Certificate-Bound Access Tokens	Option 4: Nur server-side TLS ohne Certificate-Bound Access Tokens
Nachteile:	• GMS muss für jede Zugriffs-Anfrage Datenbank registrierter Geräte (mit aktuell gültiger Geräte-Nutzer-Assoziationen) abfragen • Erkennen von Anfragen nicht-registrierter Geräte am PEP erst bei Prüfung des Access-Tokens	• Eigene PKI für Gerätezertifikate nötig (mit GMS als (intermediate) CA) • GMS muss für jede Zugriffs-Anfrage Datenbank registrierter Geräte (mit aktuell gültigen Geräte-Nutzer-Assoziationen) abfragen	• Eigene PKI für Gerätezertifikate nötig (mit GMS als (intermediate) CA) • bei Option 1: GMS muss für jede Zugriffs-Anfrage den aktuellen Revocation-Status prüfen (entweder durch Abfrage der Datenbank oder durch Information von Gerät - OCSP-Stapling) • bei Option 2: Regelmäßiges Renewal der Client-Zertifikate nötig	• Eigener Mechanismus für Nachweis über Besitz des privaten Schlüssels aus dem Schlüsselpaar nötig (z.B. Challenge-Response Verfahren), Prüfung findet auf Anwendungsebene statt • Bindung des Geräte-Tokens an den Kommunikationskanal fehlt, ggf. zusätzliche Mechanismen zum Schutz vor Session-Hijacking auf Anwendungsebene nötig

Analyse der Optionen der Verwendung gerätegebundener App-Zertifikate mit mTLS

Fazit

- Im Rahmen des Projektes wurde über mögliche Sicherheitsprobleme durch Parsing von ASN.1 diskutiert. Die Verwendung einer geeigneten (sicheren und praxiserprobten) Implementierung des (mutual) TLS-Handshakes ist für die Sicherheit der TI jedoch in jedem Falle nötig. Durch die Verwendung etablierter Implementierungen/Libraries sollten Parsing-Probleme o.ä. aber vermieden werden können. mTLS ist auch eine etablierte Technologie in vielen Umgebungen (z.B. Cloud-Mesh). Parsen von Zertifikaten ist ebenfalls etabliert, z.B. im Kontext der Verifikation von Server-Zertifikaten oder bei anderen Kommunikationsprotokollen wie IPSec/IKEv2 etc.
- Die Verwendung von einseitigem TLS (Option 4) ist nicht sinnvoll, da für die Authentifizierung des Nutzerendgerätes und Bindung des Kommunikationskanals an diese Authentifizierung dann auf der Anwendungsschicht eigene Alternativen zu diesen Standard-Mechanismen geschaffen werden müssen.
- Der Aufwand für eine eigene Geräte-PKI (ggf. mit Revokationsmechanismen) - egal ob mit Option 2 oder 3 - ist unverhältnismäßig in Bezug auf den eher geringen Mehrwert, den die Erkennung nicht-registrierter Geräte am PEP bereits im mTLS-Handshake bringt.
- Deshalb haben wir uns im Feinkonzept für die Verwendung von Option 1: "Selbst-signierte Zertifikate für mTLS mit Certificate-Bound Access Tokens" entschieden und diese in Kapitel3.5.2 näher beschrieben.