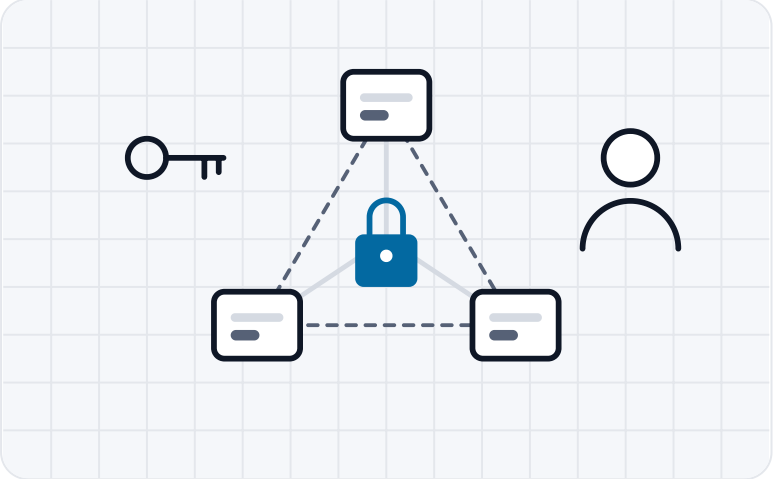
A-SIT Technology
E-Government Innovationszentrum

25. August 2023 Konzept · Privatsphäre & Krypto · Transparenz · 12 Min. Lesezeit

Zugriffskontrolle für verteilte Datenanalyse mittels Secure Multi-party Computation

Stefan More



Autorinnen und Autoren	Stefan More
Quelle	https://technology2.a-sit.at/veroeffentlichungen/zugriffskontrolle-fuer-verteilte-datenanalyse-mit-mpc/
Hinweis	Maschinell übersetzt.

Kernaussagen

- Ohne Zugriffskontrolle könnte ein kompromittierter Datenkatalog beliebige Berechnungen auslösen und Ergebnisse einsehen, ohne dass Nutzer zustimmen.
- Die Architektur erweitert MPC-Datenkataloge um Datennutzungsrichtlinien, die Käufer und zulässige Berechnungen samt Parametern einschränken.
- Jeder MPC-Knoten wertet die Richtlinien und Käufer-Nachweise aus, bevor er seine Daten-Anteile entschlüsselt und rechnet.
- Die Ergebnis-Anteile werden mit dem öffentlichen Schlüssel des Käufers verschlüsselt; nur der autorisierte Käufer kann das Ergebnis einsehen.

> - Die Durchsetzung der Richtlinien ist gewährleistet, solange mindestens ein Berechnungsknoten ehrlich ist; dem Datenkatalog muss nicht vertraut werden.

Multi-party computation (MPC) ist eine Methode aus der Kryptografie zur sicheren verteilten Durchführung von Berechnungen. In einem MPC System arbeiten mehrere Teilnehmer gemeinsam an einer Berechnung, zum Beispiel einer Datenanalyse. Jeder Teilnehmer steuert dabei einen Teil der Daten bei, wobei diese Daten auch während der Berechnung geheim bleiben. Kein Teilnehmer lernt also etwas über die Daten der anderen Teilnehmer.

Ein Spezialfall von MPC Anwendungen ist die Berechnung auf Daten aus Datenkatalogen, in welchen Benutzer die Daten verschlüsselt für ein MPC System bereitstellen. Dabei werden die Daten eines Benutzers mittels kryptografischer Verfahren (secret sharing) aufgeteilt und für einzelne Knoten des MPC Systems verschlüsselt. Ein Anwender kann dann eine Berechnung auf diesen Daten durchführen, während das MPC System nichts über die Daten lernt außer das Ergebnis der Berechnung. Dabei muss der die Daten bereitstellende Benutzer dem System vertrauen, dass nur zulässige Berechnungen auf den Daten durchgeführt werden.

Ziel dieses Projekts ist die Erweiterung eines solchen MPC Systems um ein System zur Zugriffskontrolle, um Benutzern die Kontrolle über ihre Daten zu geben. Dabei soll außerdem verhindert werden, dass ein MPC System etwas über das Ergebnis der Berechnung lernt – und nur der autorisierte Anwender dieses einsehen kann.

1. Einleitung

Personenbezogene Daten sind eine attraktive Quelle für Erkenntnisse in vielfältigen Bereichen von Forschung und Wirtschaft. Unsere Daten sind zwar sehr wertvoll, oft aber auch datenschutzrelevant. Vorschriften wie die DSGVO beschränken daher, welche Daten rechtmäßig veröffentlicht werden dürfen und was ein Käufer mit diesen sensiblen Daten tun darf. Personenbezogene Daten müssen geschützt werden. Dennoch lassen sich Erkenntnisse aus unseren Daten verkaufen, die unserer Privatsphäre nicht schaden. Die zentrale Herausforderung eines solchen Datenkatalogs liegt darin, einen Ausgleich zwischen wertvollen Einblicken in die Daten und den Anforderungen an den Datenschutz zu finden. Ein Lösungsansatz besteht darin, kryptografische Verfahren einzusetzen, die datenschutzfreundliche Berechnungen auf personenbezogenen Daten ermöglichen. Multi-Party Computation (MPC) ist eine mögliche Technik dafür. MPC-Berechnungen erlauben es, Statistiken zu berechnen oder Modelle des maschinellen Lernens auf personenbezogenen Daten zu trainieren, ohne im Klartext auf die Daten zuzugreifen.

Die betroffene Person kann jedoch nicht einschränken, wer Berechnungen kaufen darf oder welche Art von Berechnung auf den Daten zulässig ist.

Berechnungen auf personenbezogenen Daten kontrollieren: In diesem Bericht wird eine flexible Zugriffskontrollarchitektur für MPC-basierte Datenkataloge beschrieben, die sich auf bestehende Datenmärkte anwenden lässt. Die Architektur ermöglicht es Datenverkäufern (oder betroffenen Personen), detaillierte Richtlinien (Policies) festzulegen, die einschränken, wer ihre Daten kaufen darf. Außerdem kann

ein Verkäufer steuern, welche Berechnung ein bestimmter Käufer auf den Daten erwerben darf, und deren Parameter einschränken, um Verletzungen der Privatsphäre zu verhindern. Das Berechnungssystem setzt die Richtlinien durch, bevor eine Berechnung gestartet wird.

2. Hintergrund: Multi-Party Computation

Multi-Party Computation (MPC) bezeichnet verteilte Protokolle, die es einer Gruppe von Parteien erlauben, gemeinsam eine Funktion über ihre Eingaben zu berechnen und dabei diese Eingaben geheim zu halten [1]. Keine Partei erfährt dadurch etwas über die Eingaben der anderen Parteien, das über das hinausgeht, was sich aus der eigenen Eingabe und der Ausgabe der Funktion ableiten lässt.

In einer Vielzahl von MPC-Frameworks wurden praktisch effiziente Protokolle demonstriert. Sie ermöglichen es, MPC in realen Szenarien und Produkten einzusetzen. Besonders Programme mit Ganzzahlarithmetik lassen sich mit Techniken auf Basis von Secret Sharing sehr effizient berechnen.

- Bei der Multi-Party Computation (MPC) wird die Berechnung verteilt auf mehreren Knoten durchgeführt, und jeder Knoten erhält nur einen Teil der Daten der Benutzer.
- Die Daten werden in undurchsichtigen Anteilen (Shares) [2] zur Berechnung auf mehrere Knoten verteilt.
- Erst das abschließende Zusammensetzen aller Ausgabe-Anteile legt dem Käufer der Berechnung das Ergebnis offen.

In einer Vielzahl von MPC-Frameworks wurden praktisch effiziente Protokolle demonstriert. Sie ermöglichen es, MPC in realen Szenarien und Produkten einzusetzen. Besonders Programme mit Ganzzahlarithmetik lassen sich mit Techniken auf Basis von Secret Sharing sehr effizient berechnen.

3. Hintergrund: Datenkataloge

Um personenbezogene Datensätze besser zu nutzen, müssen verfügbare Daten effizient an relevante Abnehmer vermittelt werden. Datenkataloge (oder Marktplätze) übernehmen diese Vermittlungsaufgabe. Sie sind Online-Plattformen, die die Produzenten personenbezogener Daten mit relevanten Abnehmern zusammenbringen. Die erfassten personenbezogenen Daten sind jedoch hochsensibel, und der Gesetzgeber neigt dazu, sie umfassend zu schützen.

Private Datenkataloge versuchen, diese Probleme mithilfe moderner datenschutzfördernder Technologien (Privacy-Enhancing Technologies) abzumildern. Diese Technologien ermöglichen Berechnungen auf personenbezogenen Daten, ohne die Daten selbst offenzulegen. Dafür kann MPC eingesetzt werden, um die Privatsphäre der Benutzer zu wahren. Die Daten werden in undurchsichtigen Anteilen zur Berechnung auf mehrere Knoten verteilt. Erst das abschließende Zusammensetzen aller Ausgabe-Anteile legt dem Käufer der Berechnung das Ergebnis offen.

4. Konzept der verteilten Zugriffskontrolle

Ein MPC-System, das datenschutzfreundliche Berechnungen auf den Daten eines Datenkatalogs durchführt, bildet die Grundlage dieser Arbeit. In diesem Kapitel werden zunächst die Architektur und die Datenschutzannahmen eines solchen Systems beschrieben. Außerdem wird gezeigt, wie dieser Entwurf Benutzer daran hindert, Berechnungen auf ihren Daten zu kontrollieren. Im nächsten Abschnitt wird die Lösung für dieses Problem beschrieben: Das verteilte Berechnungssystem wird um Zugriffskontrollmechanismen erweitert.

4.1 Ausgangslage: Architektur

Das betrachtete Berechnungssystem besteht im Allgemeinen aus den folgenden Komponenten:

Betroffene Person (oder Verkäufer): Der Akteur, der Daten erzeugt und anderen Instanzen anbieten möchte. Zur Speicherung dieser Daten nutzt der Datenverkäufer einen öffentlichen Cloud-Speicher. Manche Modelle unterteilen den Datenverkäufer weiter in einzelne Rollen, nämlich Datenproduzent/-generator, betroffene Person und Datenanbieter.

Datenkäufer: Der Akteur, der Berechnungen auf den Daten mehrerer Datenverkäufer kaufen möchte. Er wählt ein oder mehrere Datenprodukte aus einem Katalog aus und entscheidet, welche Berechnungen ausgeführt werden. Der Datenkäufer wird manchmal auch als Datenkonsument bezeichnet.

Datenkatalog und Vermittlungssystem: Die Online-Plattform, die als Vermittler Datenverkäufer mit relevanten Datenkäufern verbindet und den Datenhandel ermöglicht. Der Katalog stellt eine Liste von Datenprodukten bereit, zu denen ein Datenverkäufer seine Datensätze hinzufügen kann. Außerdem hilft der Katalog dem Datenkäufer, passende Datenprodukte zu finden, und verkauft die Nutzung der Daten auf seiner Recheninfrastruktur. Weitere Aufgaben des Katalogs, etwa die Zahlungsabwicklung, liegen außerhalb des Umfangs dieses Berichts.

In diesem Bericht wird ein datenschutzfreundliches Berechnungssystem verwendet, um die vom Datenkäufer angeforderte Berechnung auszuführen. Die Anzahl der an dieser Berechnung beteiligten **Berechnungsknoten** N hängt vom kryptografischen Verfahren ab, das der Katalog einsetzt. Bei der Multi-Party Computation (MPC) wird die Berechnung verteilt auf mehreren Knoten (N - 1) durchgeführt, und jeder Knoten erhält nur einen Teil der Daten des Benutzers. Alternative Techniken, die in diesem Bericht nicht betrachtet werden, sind Functional Encryption (FE) und Fully Homomorphic Encryption (FHE), die auf einem einzelnen Knoten (N = 1) durchgeführt werden.

Konzeptionell umfasst die Vermittlung und Berechnung von Daten die folgenden Schritte:

1. Der Datenverkäufer wählt Daten aus, die er verkaufen möchte, und wendet darauf eine kryptografische Funktion an (z. B. Secret Sharing [2]). Anschließend verschlüsselt er jeden „Daten-Anteil" für einen der Knoten des Berechnungssystems.

2. Der Datenverkäufer lädt alle Daten-Anteile in das Datenkatalogsystem hoch.

3. Der Datenverkäufer fügt den hochgeladenen Daten Metadaten hinzu, um im Datenkatalogsystem ein Angebot (d. h. ein Datenprodukt) zu erstellen.

4. Der Datenkäufer durchsucht den Datenkatalog, wählt eine Reihe von Datenprodukten aus, die ihn interessieren, und legt eine Berechnung fest, die auf den Daten durchgeführt werden soll (z. B. durch Auswahl aus einer Menge von Standardberechnungen oder durch Angabe einer Formel).

5. Der Datenkatalog sendet die Anfrage und die Daten der ausgewählten Datenprodukte an alle Knoten des Berechnungssystems. Jeder Knoten erhält nur den Anteil der Datenprodukte, der für sein Schlüsselmaterial verschlüsselt wurde, da er nur diesen Anteil öffnen kann. Letztlich entschlüsselt jeder Knoten die für ihn bestimmten Anteile und erhält somit nur einen Anteil der Daten des Datenprodukts.

6. Die Knoten des Berechnungssystems arbeiten über ein MPC-Protokoll zusammen, um die beauftragte Berechnungsfunktion auszuführen. Nach Abschluss der Berechnung sendet jeder Knoten seinen Anteil des Ergebnisses an den Datenkatalog.

7. Der Datenkatalog fügt alle Anteile der Berechnung zum Endergebnis zusammen und sendet das Ergebnis an den Datenkäufer.

4.2 Ausgangslage: Datenschutzannahmen

Durch den Einsatz von kryptografischem Secret Sharing auf den Datenprodukten erhält jeder Knoten nur einen Anteil der Daten. Da dieser Anteil zusätzlich für den jeweiligen Knoten verschlüsselt wird, haben andere Knoten keinen Zugriff auf den Anteil eines bestimmten Knotens. Außerdem hat das Katalogsystem selbst keinen Zugriff auf Daten und sieht nur die Metadaten (d. h. das Produktangebot im Katalog).

Der beschriebene Ansatz hat jedoch mehrere Nachteile.

Da das MPC-Berechnungssystem Berechnungsanfragen vom Katalog erhält und ausführt, könnte ein böswilliges (oder kompromittiertes) Katalogsystem beliebige Berechnungsanfragen auf den Daten auslösen, ohne dass die Benutzer zustimmen oder überhaupt von der Anfrage erfahren. Dies könnte zum Beispiel dazu genutzt werden, die verschlüsselten Daten im Klartext wiederherzustellen, indem eine Identitätsfunktion auf den Daten angefordert wird.

Ein weiteres Problem besteht darin, dass selbst das Ergebnis einer legitimen Berechnung, die ein legitimer Datenkäufer veranlasst hat, an den Datenkatalog zurückgesendet wird. Der Datenkatalog erfährt damit die Berechnungsergebnisse. Der Katalog könnte sie beispielsweise speichern und Ergebnisse verschiedener Berechnungen kombinieren, um mehr über die Daten zu erfahren, oder die Ergebnisse an andere Instanzen verkaufen.

Außerdem verlässt sich der beschriebene Ansatz darauf, dass der Datenkatalog durchsetzt, welche Berechnungsfunktionen oder sonstigen Berechnungsparameter auf ein Datenprodukt angewendet werden dürfen. Ein Datenverkäufer muss sich also darauf verlassen, dass der Katalog seine Datenschutzerwartungen und -anforderungen durchsetzt. Das ist problematisch, wenn der Marktplatz böswillig handelt oder wenn die im Katalog gespeicherten Daten abfließen und böswillige Akteure sie nutzen, um direkt Berechnungen auszulösen.

4.3 Datenschutz- und Sicherheitsziele

Im vorigen Abschnitt wurden die Nachteile und Risiken MPC-basierter Berechnungssysteme erörtert. Ziel dieses Berichts ist es daher, diese Probleme abzumildern. Insbesondere sollen Benutzer eigene Regeln festlegen können, wer Berechnungen auf ihren Daten durchführen darf und welche Art von Berechnungen zulässig ist. Zusätzlich wird verhindert, dass ein neugieriger Katalog die Ergebnisse einer Berechnung erfährt.

Dadurch entfällt die Notwendigkeit, dem Datenkatalog zu vertrauen. Das Vertrauen wird stattdessen auf das verteilte MPC-Berechnungssystem verlagert. Dabei gelten dieselben Vertrauensannahmen, die bereits für die Korrektheit der MPC-Berechnungen selbst gelten: Verhält sich mindestens einer der Berechnungsknoten ehrlich, ist die Korrektheit des Berechnungsergebnisses garantiert. Dieselbe Annahme gilt auch für die hier vorgestellte Zugriffskontrollarchitektur.

4.4 Zugriffskontrolle ergänzen

Um die genannten Ziele zu erreichen, wird eine Architektur für eine Erweiterung MPC-basierter Datenkataloge eingeführt. Diese Erweiterung bietet einen flexiblen Zugriffskontrollmechanismus auf Basis eines Richtliniensystems (Policy-Systems). Im resultierenden Katalogsystem können Datenverkäufer ausdrucksstarke Richtlinien festlegen, um die Nutzung ihrer Daten zu kontrollieren. Diese Richtlinien werden den im Katalog angebotenen Datenprodukten beigefügt. Kauft ein Käufer eine Berechnung auf Datenprodukten, wird er aufgefordert, eine Reihe von Nachweisen (Credentials) vorzulegen, die seine Identität und weitere Attribute bestätigen. Das Berechnungssystem des Katalogs prüft dann, ob die Nachweise die Richtlinie für die gewählten Daten erfüllen. Zusätzlich prüft das System anhand der Richtlinie, ob der (nun authentifizierte) Käufer berechtigt ist, die angeforderte Berechnung auszuführen. Nur dann fährt das System fort und führt die Berechnung aus.

Außerdem werden die Nachweise des Käufers verwendet, um das Ergebnis einer Berechnung zu verschlüsseln. Dadurch ist sichergestellt, dass nur der legitime Käufer auf das Ergebnis zugreifen kann.

Der Entwurf konzentriert sich auf private Datenkataloge, die eine Berechnung auf Benutzerdaten ohne Beteiligung des Benutzers erlauben (nicht-interaktiv). Systeme, in denen der Benutzer an der Berechnung auf seinen Daten teilnimmt (was ein solches Richtliniensystem nicht erfordert), werden daher nicht betrachtet.

Zu diesem Zweck wird der Architektur des Berechnungssystems ein Policy Interpreter hinzugefügt. Der Katalog nutzt diese Softwarekomponente, um zu entscheiden, ob ein bestimmter Käufer berechtigt ist, (eine Berechnung auf) bestimmten Datensätzen zu erwerben. Als Eingabe erhält der Interpreter eine vom Verkäufer für seine Daten festgelegte Datennutzungsrichtlinie sowie eine Reihe von Nachweisen des Käufers, zusammen mit Metadaten zur angeforderten Berechnung.

Die resultierende Architektur ist in Abbildung 1 dargestellt.

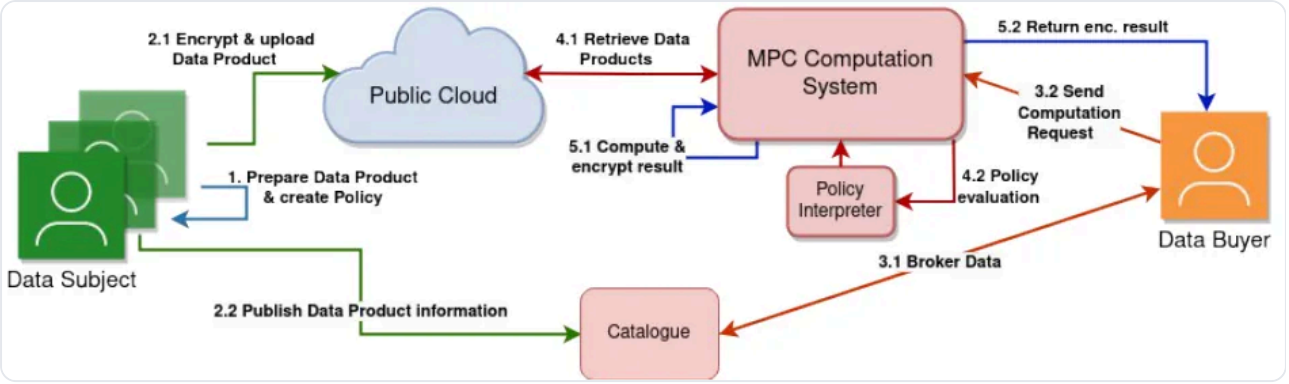


Abbildung 1: Architektur und Datenfluss eines MPC-Berechnungssystems mit richtlinienbasierter Zugriffskontrolle

Die Architektur erweitert den oben in Abschnitt 4.1 beschriebenen Ablauf. Dem Datenkatalog und dem Berechnungssystem werden insbesondere die folgenden Schritte hinzugefügt.

Die betroffene Person berechnet wie zuvor kryptografische Anteile und verschlüsselt sie für jeden MPC-Knoten. Zusätzlich zu den verschlüsselten Anteilen hängt die betroffene Person ihre Datennutzungsrichtlinie an das Datenprodukt an und lädt beides in eine öffentliche Cloud (oder direkt in den Katalog) hoch. In dieser Richtlinie legt die betroffene Person (als Datenverkäufer) Regeln für ihre Daten fest.

Einerseits kann der Datenverkäufer festlegen, wer Berechnungen auf seinen Daten durchführen darf. Er kann den Kreis der Datenkäufer auf eine bestimmte Instanz oder auf eine Art von Instanz beschränken. Die Richtlinie kann beispielsweise Regeln zur Identität (zum Nachweis) des Datenkäufers enthalten oder dazu, welches Trust Scheme (z. B. welche Menge von Ausstellern) der Verkäufer für vertrauenswürdig hält.

Andererseits legt der Datenverkäufer Regeln für die Berechnung selbst fest. Er kann zum Beispiel einschränken, welche Funktion auf seinen Daten ausgeführt werden darf und welche Parameter für diese Funktion zulässig sind. Zusätzlich kann er festlegen, dass mindestens eine bestimmte Anzahl weiterer Produkte Teil der Berechnung sein muss. Das kann eine Wiederherstellung der Daten verhindern, etwa wenn ein Datenkäufer eine Berechnung auf einem einzelnen Produkt durchführt.

Nachdem der Datenkäufer einige Datenprodukte ausgewählt hat, sendet der Datenkatalog eine Berechnungsanfrage an die MPC-Berechnungsknoten. Zusätzlich zur Berechnungsanfrage übermittelt der Datenkäufer dem System eine Reihe von Identitätsnachweisen. Die Einzelheiten des Berechnungsablaufs sind in Abbildung 2 dargestellt.

Jeder Berechnungsknoten des MPC-Systems erhält die für ihn verschlüsselten Anteile und die Nutzungsrichtlinien aller beteiligten Datenprodukte. Bevor er Berechnungen durchführt, wertet der Berechnungsknoten alle beteiligten Richtlinien aus. Dabei prüft der Knoten, ob der Käufer die Berechnung durchführen darf (durch Verifizieren der Nachweise des Käufers) und ob die Berechnungsanfrage den spezifischen Einschränkungen für die Berechnung entspricht. Nur wenn alle Richtlinien erfüllt sind, entschlüsselt der Knoten seine Daten-Anteile, erlaubt die Berechnung und startet den MPC-Dienst. Nachdem alle Knoten festgestellt haben, dass die Richtlinie erfüllt ist, und den MPC-Dienst gestartet haben, läuft das MPC-Protokoll wie im Ausgangsfall weiter.

Sobald der MPC-Prozess abgeschlossen ist, besitzt jeder Knoten einen Anteil des Berechnungsergebnisses. Jeder Knoten verschlüsselt diesen Anteil dann mit dem öffentlichen Schlüssel des Käufers (aus den Nachweisen entnommen) und sendet das verschlüsselte Ergebnis an den Katalog zurück. Dadurch ist sichergestellt, dass nur der (zuvor autorisierte) Käufer auf das Ergebnis zugreifen kann.

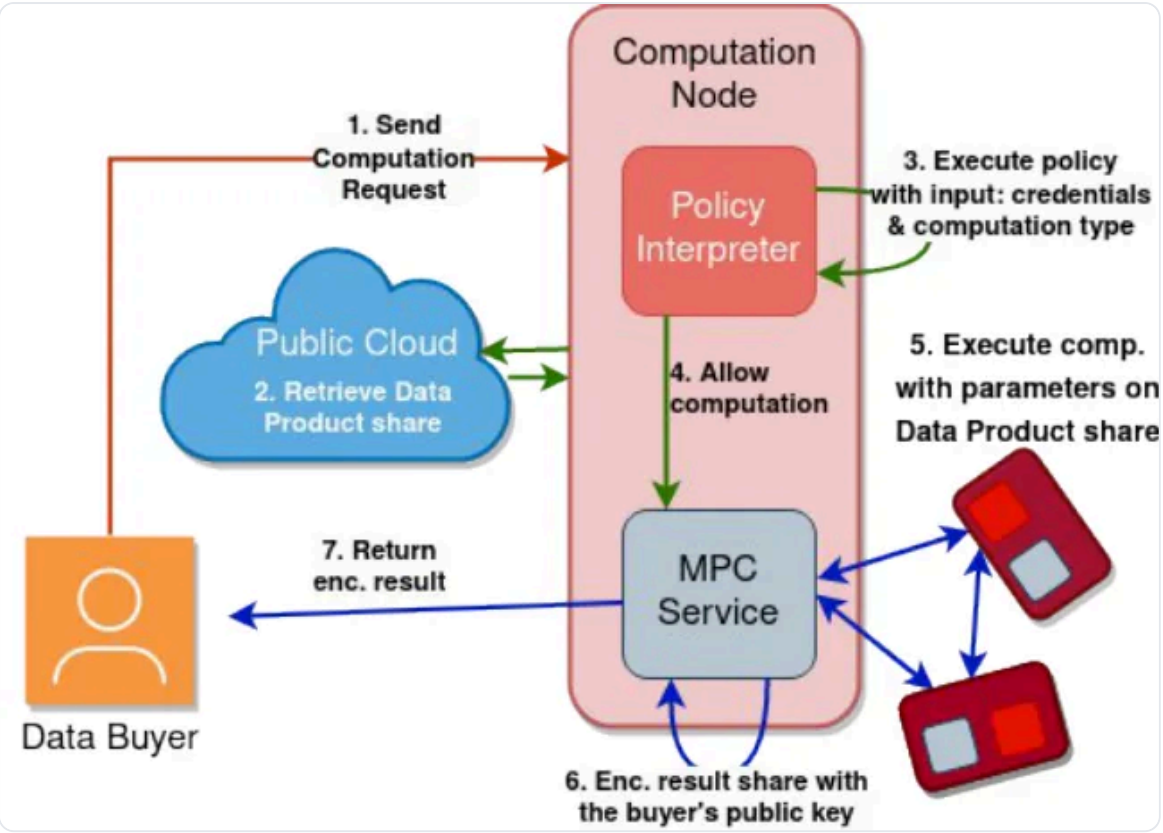


Abbildung 2: Architektur und Datenfluss eines Berechnungsknotens

5. Schlussfolgerungen

Bei der Multi-Party Computation (MPC) wird die Berechnung verteilt auf mehreren Knoten durchgeführt, und jeder Knoten erhält nur einen Teil der Daten der Benutzer. Die Daten werden in undurchsichtigen Anteilen zur Berechnung auf mehrere Knoten verteilt. Erst das abschließende Zusammensetzen aller Ausgabe-Anteile legt dem Käufer der Berechnung das Ergebnis offen.

In diesem Bericht wird eine Strategie beschrieben, wie sich ein MPC-basierter Datenkatalog um Zugriffskontrollfunktionen erweitern lässt. Insbesondere können Benutzer ihren Daten Datennutzungsrichtlinien beifügen. Diese Datennutzungsrichtlinien werden dann gemeinsam von den Berechnungsknoten durchgesetzt. Dadurch ist sichergestellt, dass keine Berechnung auf Daten möglich ist, solange mindestens ein Berechnungsknoten die Richtlinie ehrlich durchsetzt.

Literaturverzeichnis

[1] Andrew Chi-Chih Yao. Protocols for Secure Computations. In 23rd FOCS. IEEE Computer Society Press, 1982.

[2] Adi Shamir. How to Share a Secret. In Communications of the Association for Computing Machinery, 1979.