Datenzentrische Softwarearchitekturen zur Analyse von Vertraulichkeit[1]

Stephan Seifermann[2], Robert Heinrich[2], Ralf Reussner[2]

Keywords: Softwarearchitektur; Datenflussanalyse; Vertraulichkeit

1 Übersicht

Die Definition und Umsetzung von Sicherheitsanforderungen ist für alle Arten von Anwendungen essentiell. Insbesondere die Wahrung von Vertraulichkeit ist von hoher Bedeutung, da ansonsten Reputationsverlust oder hohe Strafen drohen. Mit Inkrafttreten der europäischen Datenschutzgrundverordnung sind die rechtlichen Rahmenbedingungen zur Wahrung der Vertraulichkeit sogar noch einmal verschärft worden.

Die Einhaltung von Anforderungen zu überprüfen ist nicht trivial und gerade in komplexen Softwaresystemen nicht mehr ohne unterstützende Analysewerkzeuge möglich. Im Allgemeinen gilt, dass verletzte Anforderungen so früh wie möglich identifiziert werden müssen, um teuere Nachbesserungen in späteren Entwicklungsphasen zu vermeiden.

In frühen Entwicklungsstadien können Entwurfszeitanalysen genutzt werden. Dabei ist zwischen kontrollfluss- und datenflussorientierten Analysen zu unterscheiden. Die als Bedingungen an Daten formulierten Anforderungen können in datenflussorientierten Analysen genutzt werden, um in der gleichen Terminologie Analyseziele zu formulieren. Bestehende Ansätze integrieren sich jedoch nicht gut in existierende architekturelle Beschreibungssprachen (ADLs), sind eingeschränkt in ihrer Ausdrucksmächtigkeit oder erfordern detaillierte Beschreibungen auf dem Niveau von Quelltext.

In unserem Ansatz [SHR19] haben wir daher Datenflussmodellierung in die ADL Palladio [Re16] integriert. Über Propagation von Datencharakteristiken und Abgleich dieser Charakteristiken mit denen von Verarbeitungsschritten können Analysen formuliert werden. Mehr dazu findet sich in Abschnitt 2. Die Nutzung von Geschäftsprozessen [PSH18] und grobgranularen Datenflussdiagrammen [SWE19] zur Erzeugung und Parametrisierung unserer Modelle wurde initial untersucht. Eine Evaluierung mittels zweier Fallstudien zeigte, dass eine gute Präzision mittels der Analyse erreicht werden kann.

[1] Diese Arbeit wurde gefördert durch das deutsche Bundesministerium für Bildung und Forschung unter dem Förderkennzeichen 01IS17106A (Trust 4.0).

[2] Karlsruher Institut für Technologie (KIT), IPD Reussner, Am Fasenengarten 5, 76131 Karlsruhe, Deutschland, firstname.lastname@kit.edu

doi:10.18420/SE2020_11

2 Ansatz

Der Ansatz unterteilt sich in einen Modellierungs- und einen Analyseteil.

Modellierungsansatz Der Modellierungansatz erweitert die Palladio-ADL um Annotationen an Rechenknoten und bestehenden kontrollflussorientierten Verarbeitungsoperationen. Rechenknoten können mit Charakteristiken, wie der geographischen Lokation, annotiert werden. Kontrollflussorientierte Verarbeitungsoperationen können mit einer Menge von Datenverarbeitungsoperationen annotiert werden. Datenverarbeitungsoperationen berechnen Charakterstiken von ausgehenden Daten basierend auf den Charakteristiken von eingehenden Daten, sowie der Semantik der Operation. Einige Datenverarbeitungsoperationen sind vordefiniert, es ist jedoch auch möglich, eigene Operationen zu definieren. Der Verknüpfung von Operationen ist durch die Datenabhängigkeiten und Aufrufe aus dem kontrollflussorientierten Palladio-Modell gegeben. Charakteristiken können grundsätzlich frei definiert werden, sodass der Einsatz in verschiedenen Anwendungsdomänen möglich ist.

Analyseansatz Der Ansatz zur Analyse basiert auf der Propagation von Datencharakterstiken und einem Vergleich dieser Charakteristiken mit Vorgaben oder Charakterstiken von Ressourcen. Die Propagation von Charakterstiken erfolgt durch Auswertung der für die Datenverarbeitungsoperation definierten Berechnungsvorschrift und der Anwendung der berechneten Charakteristiken auf die Ausgabedaten. Dabei ist entscheidend, welche Charakteristiken die eingehenden Daten haben. Da es grundsätzlich mehrere mögliche Sequenzen von Datenverarbeitungsoperationen und initialen Datencharakterstiken geben kann, wird während der Analyse jeden mögliche Sequenz berücksichtigt. Eine konkrete Fragestellung für eine Analyse ist durch Formulierung einer Anforderung gegeben. Dabei werden berechnete Datencharakteristiken mit einer fixen Charakteristik oder einer Ressourcencharakterstik verglichen. Das Ergebnis der Analyse ist dann eine Verletzung der Anforderung inklusive der dazu führenden Sequenz von Verarbeitungsoperationen.

Literatur

[PSH18] Pilipchuk, R.; Seifermann, S.; Heinrich, R.: Aligning Business Process Access Control Policies with Enterprise Architecture. In: Proceedings of the Central European Cybersecurity Conference 2018. CECC'18, ACM, 2018.

[Re16] Reussner, R. H.; et al.: Modeling and Simulating Software Architectures – The Palladio Approach. MIT Press, 2016.

[SHR19] Seifermann, S.; Heinrich, R.; Reussner, R. H.: Data-Driven Software Architecture for Analyzing Confidentiality. In: IEEE International Conference on Software Architecture, ICSA 2019. IEEE, S. 1–10, 2019.

[SWE19] Seifermann, S.; Werle, D.; Ebada, M.: Mapping Data Flow Models to the Palladio Component Model. Softwaretechnik-Trends/, 2019.