Ein Laufzeitmodell-basierter Ansatz zur Datenschutz-Prüfung von Cloud-Systemen

Eric Schmieders, Andreas Metzger, und Klaus Pohl

paluno (The Ruhr Institute for Software Technology)
Universität Duisburg-Essen, Essen, Deutschland
{eric.schmieders, andreas.metzger, klaus.pohl}@paluno.uni-due.de

Abstract: Personenbezogene Daten werden durch Datenschutzrichtlinien geschützt. Allerdings können Cloud-Systeme Komponenten zur Laufzeit migrieren und replizieren. Das kann die geographische Lage der Daten beeinflussen und damit zu einer Verletzung der Richtlinien führen. In unserer Forschung untersuchen wir den Einsatz von Laufzeitmodellen. Die vorgeschlagenen Laufzeitmodelle bilden die Architektur und die Verteilung eines Cloud-Systems ab und ermöglichen dadurch die Prüfung unterschiedlicher Verletzungs-Szenarien.

1 Einleitung

Datenschutzrichtlinien, wie zum Beispiel die Europäische Data Protection Directive, definieren geographische Grenzen, innerhalb derer personenbeziehbare Daten verarbeitet werden dürfen. Diese Richtlinien gelten insbesondere auch dann, wenn die Datenverarbeitung durch Cloud-Systeme erfolgt. Cloud-Systeme können Software-Komponenten zur Laufzeit migrieren und replizieren und somit deren geographische Lage verändern. Da die konkreten Änderungen zur Entwurfszeit unbekannt sind, ist bei Cloud-Systemen eine dynamische Prüfung der Richtlinien zur Laufzeit erforderlich. Bisherige Lösungen zur Beobachtung von Cloud-Systemen („Monitoring“) liefern nur unzureichende Informationen, um solche Datenschutz-Prüfungen durchzuführen. Sie liefern zwar Informationen über einzelne Software-Komponenten oder virtuelle Maschinen, erfassen jedoch keine Datenflüsse zwischen verteilten Komponenten.

In unserer Forschung adressieren wir diese Lücke und untersuchen den Einsatz von Laufzeitmodellen, welche ein Cloud-System sowie die Verteilung und Datenflüsse seiner Komponenten repräsentieren. Spezifische Monitoring-Sonden liefern für Datenschutzprüfungen relevante Änderungsereignisse. Diese Änderungsereignisse werden zur Aktualisierung von Laufzeitmodellen genutzt, auf Basis derer die eigentlichen Datenschutzprüfungen vorgenommen werden. Die Laufzeitmodelle bilden somit das Cloud-System ab und enthalten die zur Prüfung erforderliche Informationen.

Spezifische Forschungsfragen umfassen (*i*) die geeignete Platzierung von Monitoring-Sonden in der Cloud, (*ii*) Modelltransformationen zur Aggregation der Monitoring-Daten, sowie (*iii*) performante Prüfungen der Datenschutzrichtlinien auf Basis der Laufzeitmodelle. Dieser Beitrag geht auf Forschungsfrage (*iii*) ein und fasst die in [1] vorgestellte Datenschutzrichtlinien-Prüfung zusammen.

2 Laufzeitmodell-basierte Datenschutz-Prüfung

Die Datenschutzrichtlinien-Prüfung basiert auf den in Abb. 1 dargestellten Entitäten und Relationen des Laufzeit-Metamodells. Datenflüsse zwischen verteilten Software-Komponenten sind durch die Konzepte *VM hosts Component* und *Component executes Process* sowie *Process access Process* und *Process access Data* modellierbar.

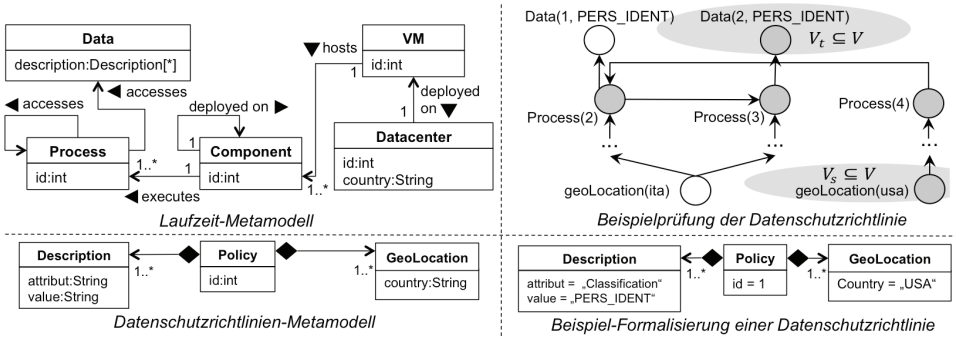

Abbildung 1: Metamodelle und Beispiel-Prüfung

Neben dem aktuellen Laufzeitmodell (einer Instanz des abgebildeten Metamodels) erhält der Prüfalgorithmus formalisierte Datenschutzrichtlinien als Eingabe. Diese werden mittels des Datenschutzrichtlinien-Metamodells aus Abb. 1 spezifiziert. Zur Prüfung des Laufzeitmodells gegen die Richtlinien führt der Algorithmus die Prüfaufgabe zurück auf das Erreichbarkeitsproblem in Graphen (auch bekannt unter *s-t-Connectivity* oder *STCON*). Der Algorithmus selektiert aus der Laufzeitmodellinstanz die Menge der in der Datenschutzrichtlinie spezifizierten Geo-Lokationen V_s (im Beispiel in Abb. 1 *USA*). Außerdem selektiert der Algorithmus die Menge der in der Richtlinie klassifizierten Daten V_t. Ist mindestens ein Element aus V_t von mindestens einem Element aus V_s erreichbar, so liegt eine Verletzung der geprüften Richtlinie vor (was im Beispiel aus Abb. 1 der Fall ist). Laut Modell-Semantik können dann zu schützende Daten (V_s) in die ausgeschlossenen Geo-Lokationen (V_t) transferiert werden.

Unsere bisherige Evaluation des Prüfalgorithmus zeigt, dass dieser Ansatz eine performante Prüfung auch von komplexen Modellen ermöglicht. Werden zur Laufzeit Datenschutzverletzungen erkannt, ermöglicht das ein schnelles Eingreifen zur Vermeidung des unrechtmäßigen Zugriffs auf personenbeziehbare Daten.

Literaturverzeichnis

[1] Schmieders, E., Metzger, A., Pohl, K., 2014. A Runtime Model Approach for Data Geo-location Checks of Cloud Services, in: Franch, X., Ghose, A.K., Lewis, G.A., Bhiri, S. (Eds.), Service-Oriented Computing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 306–320.