

Plattform zur Bereitstellung und standortunabhängigen Ausführung von Analyseverfahren in der kollaborativen Forschung

Maximilian Beier^{1,2}, Dagmar Krefting¹, Thomas Penzel²

¹Hochschule für Technik und Wirtschaft Berlin, ²Charité – Universitätsmedizin Berlin

Ziel

Ziel ist die Untersuchung von Möglichkeiten, Datenverarbeitungsverfahren lokal im Web-Browser ausführen zu können. Das ermöglicht beispielsweise die standortunabhängige, reproduzierbare Auswertung von Forschungsdaten in der kooperativen Forschung, ohne dass die Daten selber übertragen werden müssen. Diese Art der Verarbeitung ist vorteilhaft bezüglich datenschutzrechtlicher Bestimmungen, bei divergierender Hard- und Software und in Arbeitsumfeldern mit beschränkten Benutzerrechten.

Ausführungsverfahren

Aus Performance- und Datenschutzgründen ist eine externe Ausführung von Software nicht vorgesehen und es wird sich nur auf die lokale Ausführung konzentriert. Hierbei kann zwischen verschiedenen Verfahren unterschieden werden, vor allem ob die Anwendung im originalen Quellcode, als Kompilat oder in einem Zwischenformat vorliegt.

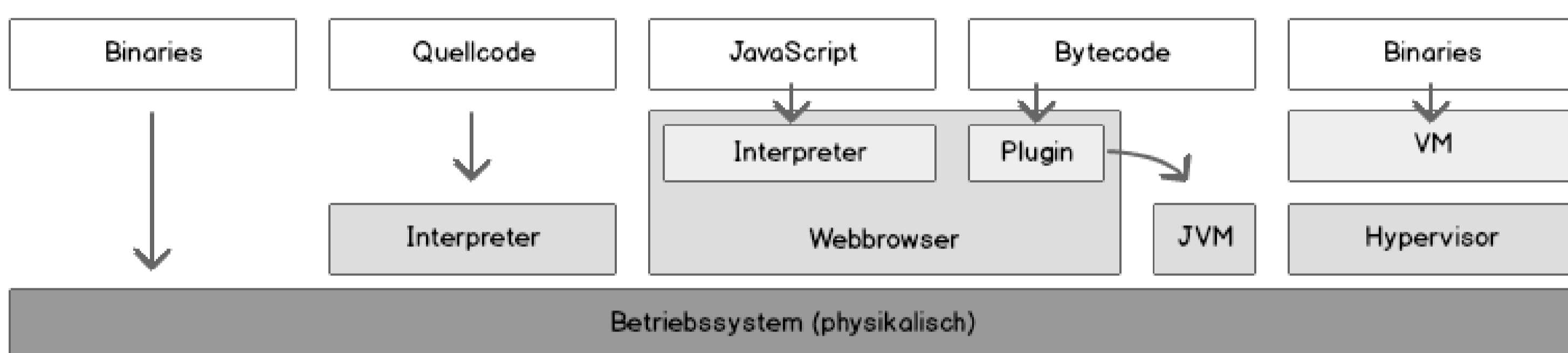


Abb 1: Verschiedene Arten von Software und ihre Laufzeitumgebungen.

JavaScript und Emulation

JavaScript ist eine schwach typisierte, interpretierte Scriptsprache. Letzteres ermöglicht eine plattformunabhängige Ausführung. Als quasi Programmiersprache des Webs ist sie in allen modernen Browsern verfügbar. Allerdings sind in der Ausführung Grenzen gesetzt, wodurch sie in ihren Möglichkeiten systemnahen Umgebungen gegenüber benachteiligt ist.

JavaScript wird in diesem Projekt benutzt, um die Komponenten eines Computers nachzubilden, beispielsweise CPU, RAM oder Festplatte. Darauf läuft Linux als Betriebssystem, in dem wiederum vorkompilierte Anwendungen, sogenannte Binaries, ausgeführt werden können. Die Ausführungsgeschwindigkeit in diesem emulierten System liegt derzeit bei circa 0,25% verglichen mit dem gleichen Programm im Host-System.

Besonderheiten in der Schlafforschung

In der schlafmedizinischen Diagnostik spielt die Biosignalanalyse eine herausragende Rolle. Dafür gibt es allerdings kaum Standard-Software, die von allen benutzt wird, sondern viele kleine, spezialisierte Programme. Bei multizentrischen Forschungsprojekten, wie sie aufgrund der Struktur der Schlaflabore und Patientenkollektive üblich sind, ist eine gemeinsame Nutzung dieser Programme dringend erforderlich. Allerdings ist die Installation von Software aufgrund von Sicherheitsrichtlinien in medizinischen Einrichtungen häufig schwierig. Wegen datenschutzrechtlicher Aspekte ist auch der Transport der Daten aus den Laboren heraus, beispielsweise zu einem zentralen Auswertungsdienst, problematisch; zudem sind die Daten signifikant größer als die Auswertungsverfahren, weswegen eine lokale Auswertung vorteilhaft ist.

Plattform

Während des Projektes wird eine Plattform entwickelt, auf der Forscherinnen und Forscher eigene Analyseverfahren hochladen und deren Signaturen hinterlegen können. Für jene wird dann automatisch ein Image einer virtuellen Maschine angelegt.

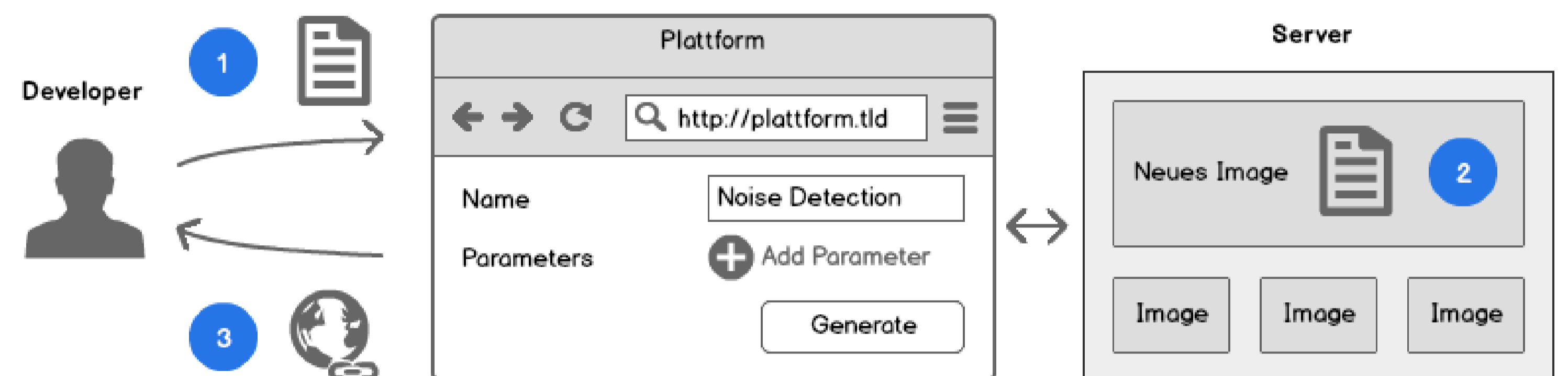


Abb 2: Use-Case: Benutzung einer Analyse-VM. ForscherInnen wählen die passende Software in der Plattform aus (1) und öffnen eigene Biosignaldaten im Browser (2), womit das Analyseprogramm in der lokalen VM ausgeführt wird (3). Anschließend werden die Ergebnisse dargestellt (4).

Auf der Plattform hinterlegte Images können von anderen Forschenden über eine spezielle URL geladen werden, woraufhin die virtuelle Maschine gestartet und das darin verpackte Programm ausgeführt wird. Die Seite bietet zur Signatur passende Eingabemöglichkeiten, worüber BenutzerInnen die Anwendung mit eigenen Daten, beispielsweise lokal gespeicherte Schlafaufzeichnungen, ausführen können. Die Ergebnisse der Auswertung werden letztlich aus der VM heraus in die Oberfläche der Website überführt.

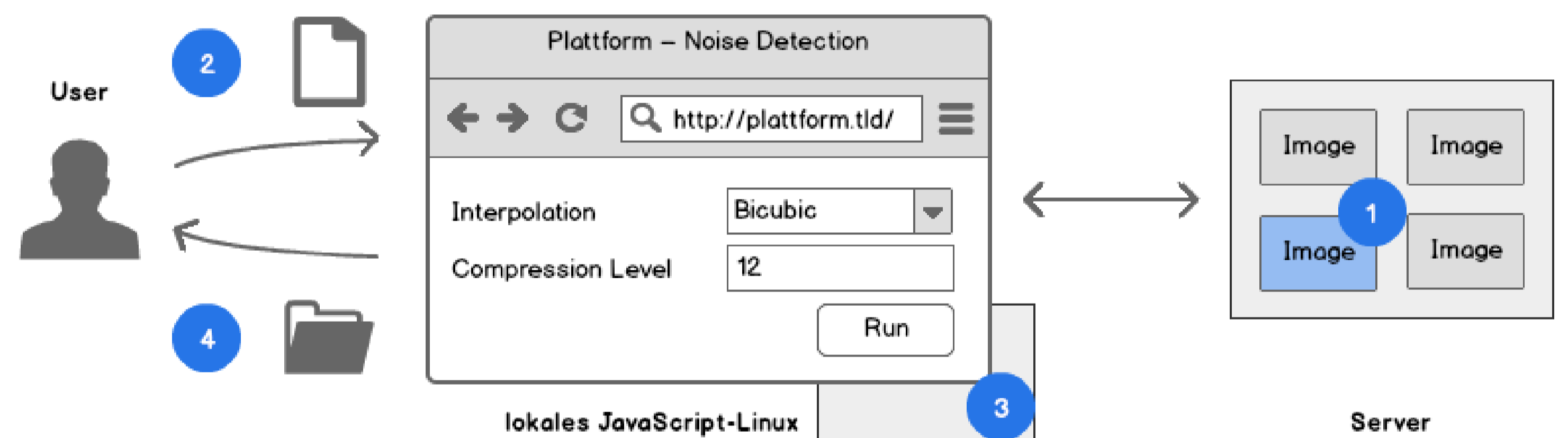


Abb 3: Use-Case: SoftwareentwicklerIn. Analysesoftware wird samt Metadaten auf die Plattform geladen (1), woraus eine VM und eine passende Projektseite erzeugt werden (2), deren URL zurück gegeben wird (3).

Ergebnisse

Es existiert ein virtueller Computer samt Kernkomponenten wie CPU, RAM und einem Dateisystem, in dem sich verschiedene Betriebssysteme starten lassen, u.a. Linux. In diesem ist es möglich, nach x86 kompilierte Binaries im Browser ausführen zu lassen. Das System kann über WebWorker in einem separaten Thread laufen, wodurch die Oberfläche nicht blockiert wird. Der Performanceverlust liegt derzeit circa bei Faktor 400, abhängig von der Art der Anwendung.

Ausblick

Das derzeit größte Problem ist die Performance der virtuellen Maschine. Mit WebAssembly steht ein neues Binär-Format für performance-kritische Anwendungen im Web bereit, das als Gemeinschaftsprojekt mehrerer großer Browserhersteller entwickelt wird. Emulatoren sind explizites Designziel, weswegen hiervon deutliche Verbesserungen zu erwarten sind.