

Abstract der fertigen Arbeit

Moderne Applikationen sind zunehmend von kurzen Antwortzeiten, maßgeschneiderter Hardware und dynamischer Skalierung abhängig. Edge Computing ist ein attraktives Konzept um diese Anforderungen zu erfüllen, bringt aber andere Herausforderungen mit sich. Im Gegensatz zu Cloud-Umgebungen sind bei Edge Computing Hardware und Netzwerk heterogen, was für Entwickler*Innen zusätzlichen Aufwand bedeutet. Serverless Computing, eine mögliche Lösung hierfür, ist ein Paradigma das die dahinter liegende Infrastruktur abstrahiert, wodurch Entwickler*innen mit der einhergehenden Komplexität nicht mehr umgehen müssen. Aktuelle Serverless Frameworks sind allerdings nicht für die Anforderungen von Edge Computing ausgelegt. Um dies zu verbessern wurden mehrere Prototypen entwickelt, aber in vielen Situationen, insbesondere bei netzwerkgebundenen Anfragen, was solche Anfragen sind bei denen die Netzwerktransferzeiten den größten Anteil der Antwortzeit ausmachen, ist die Leistung nach wie vor nicht ausreichend. In dieser Diplomarbeit präsentieren wir eine Methode welche die Funktionsweise von Serverless Frameworks anpasst, sodass sich die Antwortzeiten für netzwerkgebundene Anfragen deutlich verbessern. Um herauszufinden woher die schlechte Leistung hervorgerufen wird führen wir vorläufige Experimente durch. Diese zeigen, dass unpassende Platzierung und Implementierung von Load Balancern für die schlechte Leistung verantwortlich sind, da diese nicht für Edge Computing angepasst sind, und diese darüber entscheiden welchen Pfad eine Anfrage durch das Netzwerk nehmen muss. Wir schlagen daher vor diese Komponenten anzupassen, sodass sie den Anforderungen von Serverless Edge Computing besser entsprechen. Hierfür entwickeln wir eine Version von gewichtetem Round Robin Load Balancing. Weil die Leistung einzelner Geräte nicht a priori bekannt ist nutzen wir die Antwortzeit von Anfragen als Black-Box-Metrik anhand derer wir kontinuierlich die Load Balancing Gewichte anpassen. Um die Anzahl an und Position von Load Balancern zu entscheiden, schlagen wir eine von osmotischem Druck inspirierte Lösung vor, in welcher dieser dynamische Druck für die Entscheidung herangezogen wird. Der Druck selbst ist eine Funktion der Anzahl an Anfragen, sowie der Nähe zu Clients und Instanzen der angefragten Serverless Funktionen. Unsere Evaluierungen zeigen, dass unsere Lösung verglichen mit dem aktuellen Standard, nämlich zentralisiertem Round Robin Load Balancing, abhängig vom Szenario die durchschnittlichen Antwortzeiten um 30% bis 69% reduziert. Außer den Leistungsverbesserungen ermöglicht es unsere Lösung auch mit einem sich dynamisch ändernden System umzugehen wie es typischerweise bei Edge Computing vorgefunden wird, und nutzt die Systemressourcen effizienter als bestehende Methoden.