HOCHSCHULE
ESSLINGEN

Informatik und
Informationstechnik

IT Innovationen

Band 29
Juni 2022

HOCHSCHULE ESSLINGEN

Grußwort des Dekans

Liebe Leserinnen und Leser,

Die Zahl der Themen, die die IT-Branche derzeit besonders beschäftigt, ist enorm. Quer Beet durch die gesamte Wirtschaft ist der Bedarf an IT-Lösungen zur Realisierung der digitalen Transformation riesig. Besonders stechen aktuell die Themen Sicherheit, Data Analytics und Künstliche Intelligenz hervor.

Sicherheit von IoT-Geräten in industriellen Netzwerken spielt durch die lokale Industrie naturgemäß eine besondere Rolle in den Abschlussarbeiten unserer Absolventen und Absolventinnen. Aber auch generell die Sicherheit ganzer IT-Systeme oder Softwareanwendungen liegen im Fokus. Der Bedarf an Expertise auf diesen Gebieten lässt sich schon heute nicht bedienen und er wird zunehmend größer und für die Zukunft ein Dauerthema bleiben. Mit einem neuen Studienprogramm IT-Sicherheit wird die Hochschule Esslingen zur Befriedigung dieses Bedarfs zukünftig noch stärker beitragen können. Die Planungen dazu laufen auf Hochtouren.

Mit welcher Vielfalt unsere aktuellen Absolventinnen und Absolventen aufwarten und in welcher Komplexität sie agieren, zeigen Ihnen, lieber Leser und liebe Leserin, die in diesem Band der IT-Innovationen zusammengefassten Beiträge der Abschlussarbeiten.

Viel Spaß beim Lesen wünscht Ihnen

Prof. Jürgen Nonnast

Dekan der Fakultät Informatik und Informationstechnik

IMPRESSUM

ERSCHEINUNGSORT

73732 Esslingen am Neckar

HERAUSGEBER

Prof. Jürgen Nonnast
Dekan der Fakultät Informatik und Informationstechnik
der Hochschule Esslingen - University of Applied Sciences

REDAKTIONSANSCHRIFT

Hochschule Esslingen - University of Applied Sciences
Fakultät Informatik und Informationstechnik
Flandernstraße 101
73732 Esslingen am Neckar

Telefon	+49(0)711.397-4211
Telefax	+49(0)711.397-4214
E-Mail	it@hs-esslingen.de
Website	www.hs-esslingen.de/it

REDAKTION, LAYOUT UND DESIGN

Prof. Dr.-Ing. Reinhard Schmidt
Hochschule Esslingen - University of Applied Sciences
Fakultät Informatik und Informationstechnik
Flandernstraße 101
73732 Esslingen am Neckar

SATZ, ANZEIGEN und VERLAG

Dipl.-Inf.(FH) Rolf Gassner

Hochschule Esslingen - University of Applied Sciences
Fakultät Informatik und Informationstechnik
Flandernstraße 101
73732 Esslingen am Neckar

ERSCHEINUNGSWEISE

Einmal pro Semester, jeweils Januar und Juni

DRUCK

Pixelgurus
Werbung – Werbetechnik – Digitaldruck.
Horbstraße 8
73760 Ostfildern

AUFLAGE

500 Exemplare

ISSN 1869-6457

Validierung alternativer Technologien zur Umfeldwahrnehmung in der mobilen Reinigungsrobotik

Julian Baisch **Clemens Klöck**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Alfred Kärcher SE & Co. KG, Winnenden

Stand der Technik

Die überwiegende Mehrheit der heute gängigen Reinigungsroboter für den Privatgebrauch, sind mit 360° Lidar Sensoren zur Wahrnehmung ihres Umfelds ausgestattet. Häufig mit Kameras zur Objektbestimmung und kleinen ToF Sensoren zur Absturzerkennung kombiniert, dominiert diese Sensortechnik den Markt. Das Lidar ermöglicht es dem Roboter, seine Umgebung auf ein paar Millimeter genau zu Kartografieren, um dann mithilfe dieser Karte eine optimale Route zur Reinigung der ihm vorgegebenen Räumlichkeiten zu berechnen. Eine eventuell zusätzlich angebrachte Kamera kann während der autonomen Navigation dazu genutzt werden, Objekten dynamisch auszuweichen und/oder die Beschaffenheit des Untergrunds zu bestimmen.

Problemstellung

Nun ist diese Sensorik und deren Technologie in vielen Bereichen ausreichend, jedoch gibt es einige Szenarien, in welchen sie keine verwendbaren Ergebnisse liefert oder schlicht und ergreifend versagt. Dies ist zum Beispiel bei Glaswänden oder bodennahen Spiegel der Fall. In der Nähe einer Glaswand, erkennt das Lidar diese einfach überhaupt nicht oder zumindest nicht hinreichend genau. Im Falle eines Spiegels, welcher sich im Sichtbereich des Lidars befindet, werden dessen Laserstrahlen durch den Spiegel reflektiert und von einem anderen Hinderniss zurückgeworfen. Das Resultat ist eine Art Durchgang, welchen das Lidar, anstelle eines Spiegels, zu detektieren scheint.

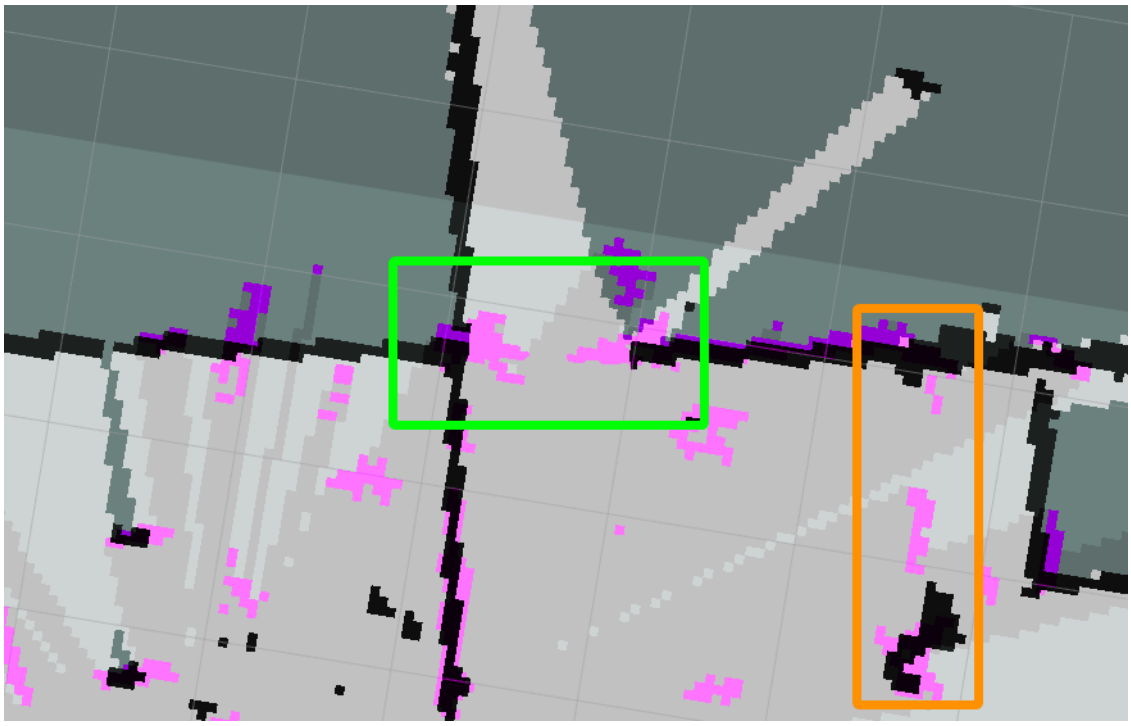

Abb. 1: Lidar bei Spiegel (Grün) und Glaswand (Orange) [1]

In Abbildung 1 zu sehen, ist eine durch den SLAM-Algorithmus Hector (siehe [3]) erstellte Karte der Testumgebung, visualisiert durch das in der Roboter Entwicklungsumgebung ROS [4] verfügbare Tool RVIZ. Die schwarzen Linien, welche den hellgrauen Raum einschließen, sind dabei alle Hindernisse die das Lidar detektiert hat. Die Grüne Box zeigt scheinbar einen Durchgang von einem Raum in den nächsten. In der Praxis handelt es sich hierbei um einen an der Wand stehenden Spiegel. Die Box in Orange markiert den Standort einer Glaswand. Diese ist für den Lidar gänzlich unsichtbar. Lediglich ein dahinter befindlicher Karton zur rechten Seite wird detektiert.

Ziel der Arbeit und Umsetzung

Diese Arbeit hat nun zum Ziel, mit anderen Sensortechnologien diese Probleme zu verbessern oder gar ganz zu verhindern. Das heißt, ein Sensor oder eine Sensorkombination, welche in der Lage ist, diese Art von Hindernisse zuverlässig erkennen zu können, zu testen und zu validieren. Darüber hinaus soll auch geprüft werden, ob sich diese etwaige Sensorik auch zur Detektion kleiner Objekte im Raum eignet, um sich auch hier nicht nur auf eine Kamera verlassen zu müssen. Dazu wurden zwei Radarsensoren verschiedener Hersteller eingekauft, und auf einer fahrbaren Testumgebung angebracht. Ein TurtleBot3 Waffle Pi der Firma ROBOTIS, bildet diese fahrbare Testumgebung. Dieser bietet viele Möglichkeiten um schnell und unkompliziert einen funktionsfähigen Prototypen zu entwickeln.

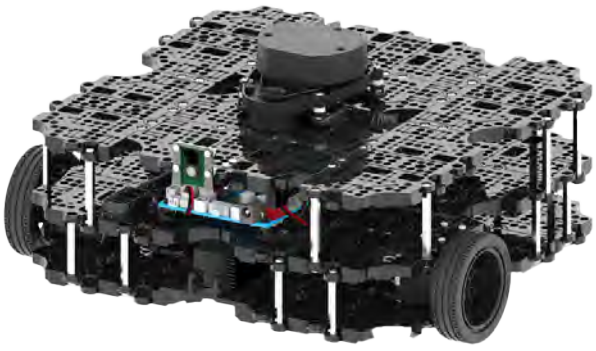

Abb. 2: ROBOTIS TurtleBot3 Waffle Pi [1]

Erste Daten wurden bereits gesammelt und aufbereitet. Abbildung 1 zeigt erste Radardaten als magentafarbene Flächen. Große Teile des Spiegels und etwa die Mittelsektion der Glaswand werden durch den Radarsensor erkannt und in der Karte markiert. Dazu wird die mithilfe des Hector-SLAM [3] berechnete Position im Raum verwendet, um eine weitere Karte der Radar-Punktwolke mittels Octomap (siehe hierzu [2]) zu generieren. Man sieht auch die durch das Lidar erkannten Hindernisse, welche zuverlässig erkannt und kartografiert werden. Zudem fallen dem Betrachter weitere magentafarbene Flächen in der Karte auf. Diese sind allesamt Hindernisse, welche physisch nicht hoch genug sind um durch das Lidar erkannt zu werden. Der Umriss dieser Objekte stimmt jedoch nicht exakt mit den durch den Radarsensor erkannten Flächen überein. Das liegt zum einen daran, dass der Radarsensor, welcher für diesen Versuch im Einsatz war, eine Entfernungsauflösung von ungefähr 47 mm aufwies. Dieser Wert ist von der Konfiguration des Sensors abhängig, welche wiederum von vielen weiteren Einzelfaktoren abhängig ist, und kann effektiv nicht weniger als 39 mm betragen. Zu dieser relativ groben Entfernungsauflösung kommt zudem eine Azimuth-Winkel-Auflösung von 30 Grad. Diese Eigenschaften des Radarsensor in Kombination mit der Einbaulage am Roboter und die Bewegungen desselben, bestimmen auch dessen Fähigkeit, wie genau Objekte detektierbar sind. So wurde die Glasscheibe im Versuch besonders gut detektiert, wenn in einem Winkel von nahezu 90 Grad zwischen Glasscheibe und Radarsensor gefahren wurde. In diesem Fall kann der Radarsensor die Reflektionen der Glasscheibe sehr gut von Reflektionen anderer Objekte unterscheiden. Noch ist also die Lidartechnologie bei der Genauigkeit der Messung weiterhin im Vorteil. Ein Radarsensor hat jedoch in vielen anderen Punkten Vorteile. So ist dieser zum Beispiel deutlich weniger Anfällig für Ausfälle durch bewegte und strapazierte Bauteile. Außerdem sieht der Radarsensor auch bei widrigen Witterungsverhältnissen wie Regen oder Dampf noch zuverlässig Hindernisse. Des Weiteren gibt es den Unterschied im Preis. Der auf dem TurtleBot3 angebrachte Lidarsensor, ist etwa 4 mal teurer im Vergleich zu dem verwendeten Radarsensor. Das bedeutet auch einen enormen Kostenunterschied in einem späteren Produkt.

Ausblick

Im weiteren Verlauf der Arbeit gilt es nun, die Möglichkeiten des Radarsensors immer weiter auszutesten und diesen mit einer Stereokamera zu ergänzen. Da der verwendete Radarsensor alleine nur sehr wenige Datenpunkte pro Messung bereitstellt, kann durch diese Kombination die Informationsmenge deutlich erhöht werden. Somit hat dann ein SLAM-Algorithmus wieder genügend Daten zur Verfügung und es kann eventuell in Zukunft auf einen Lidarsensor zur Positionsbestimmung und Kartografierung des Raumes verzichtet werden.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. {OctoMap}: An Efficient Probabilistic {3D} Mapping Framework Based on Octrees. *Autonomous Robots*, 2013.

[3] Stefan Kohlbrecher, Johannes Meyer, Oskar von Stryk, and Uwe Klingauf. A Flexible and Scalable SLAM System with Full 3D Motion Estimation. In *Proc. IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE, 2011.

[4] Jason M. O'Kane. A Gentle Introduction to ROS. https://www.cse.sc.edu/~jokane/agitr/agitr-letter.pdf, 2014.

Co-Scheduling of Kubernetes Pods in a SLURM HPC Cluster

Simon Bauer **Rainer Keller**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Department of Computer Science and Engineering, Esslingen

Introduction

Modern science is greatly dependent on high-performance computing (HPC) in order to solve computational expensive calculations and complex simulations. Often, these HPC systems are designed as a compute cluster consisting of many individual nodes that are connected through fast networks. A software scheduler is deployed to the HPC cluster in order to combine the computation power of all the individual compute nodes and make it available as one HPC system. The scheduler is responsible for monitoring the individual systems, managing a queue of pending and running tasks, allocating exclusive or non-exclusive access to resources, mainly CPU, GPU and memory, and collecting accounting information to ensure the users stays within their allocated budget. A common scheduler that is used in many of the top 500 supercomputers is the free and open-source SLURM Workload Manager [12] [14]. Features of SLURM include high-availability, high-performance and high-configurability with over 100 optional plugins that can be used to extend the functionality.

While SLURM is a great fit for a lot of scientific purposes, there exist numerous alternative that target similar audiences. One of them is Kubernetes, release in 2014, which recently gained a lot of attraction. Kubernetes is an open-source container orchestration system that is now being used or evaluated by 96% of organizations with approximately 5.6 million developers using it worldwide [3]. This huge popularity is caused by the fact that Kubernetes provides a lot of important feature that makes managing applications at scale a lot easier. This includes service discovery, load balancing, storage orchestration, automated rollouts and rollbacks, self-healing, and secret and configuration management [1].

Unlike SLURM, Kubernetes is focusing on deploying and managing production applications rather than executing computational expensive scientific workloads. This, however, does not mean that there is not a substantial amount of overlap between both technologies. One big overlapping topic is the need for containerization to enable reproducible outcomes and ease the deployment and management overhead. While using containers is a requirement in Kubernetes, with SLURM clusters it is an optional technology that is only recently gaining popularity. Numerous papers have been published in the last couple of years to investigate how containers can be used effectively in a SLURM HPC environment [2] [11] [13]. In addition, a handful of papers examined how both these technologies, Kubernetes and SLURM, can work together [5] [6] [8] [9] [16]. The main driver for this research are deep learning applications that neither work well with only a SLURM cluster, as it lacks microservice support and does not support the complex environment requirements, nor a Kubernetes cluster, as it lacks proper resource allocation and isolation support that are required by these kinds of applications [15].

Related Work

As discussed above, the integration of Kubernetes with HPC clusters is an ongoing research topic. Several papers have been published that introduce different approaches on how to incorporate both technologies to combine the advantages. V. Pisaruk and S. Yakovtseva developed the WLM-operator [6], which got extended by Zhou et al. in their Torque-operator [15]. In addition, Lublinsky et al. proposed a Kubernetes Bridge operator which extends the concept of the before mentioned operators with a generic approach that allows controlling multiple different external systems [8]. López-Huguet et al. developed a similar approach called the hpc-connector [9]. These proposed solutions are focusing on one direction of integration. Specifically, they target the use-case that all workloads, no matter if they should be executed on a Kubernetes worker or on a SLURM node, are scheduled through the Kubernetes API. To accomplish this, custom resource definitions (CRDs) are created in Kubernetes

to represent a SLURM job , as well as virtual nodes that are created for each existing SLURM compute partition. Therefore, the user can use the default Kubernetes deployment mechanisms and tooling to interact with both, Kubernetes resources as well as SLURM jobs. If the user starts a SLURM job via the Kubernetes API, a dummy pod is created in Kubernetes that tracks the status of the SLURM job, including the status, the stdout and stderr stream, and the output files.
Piras et al. proposed another approach (ge-k8s) in which the Kubernetes cluster is running on the HPC system (Oracle Grid Engine). Using the Grid Engine batch workload manager, the user is able to transiently add or remove nodes. The solution is using the Grid Engine Parallel Environment (PE) which allows cluster administrators to configure setup and teardown scripts that are respectively executed before and after the job execution. The scripts are responsible for bootstrapping and decommissioning the transient Kubernetes nodes [10].

Motivation

The above mentioned solutions can be divided into two groups: dual-level schedulers (WLM-operator [6], Torque-operator [15], Bridge Operator [8] and hpc-connector [9]) and homogenous job management (ge-k8s [10]).
While the dual-level schedulers require minimal modifications to an existing HPC cluster, they pose two major issues:

- Workloads that run in Kubernetes bypass the SLURM accounting system and therefore user budget limits are not enforced.
- Two clusters, SLURM and Kubernetes, need to exists and have sufficient resources available in order to run the requested workload. This generally results in lower resource utilization compared to only one cluster and therefore results in higher cost.

In contrast, the ge-k8s approach by Piras et al. solves those issues by properly integrating with the accounting system of the HPC cluster and reusing the same nodes [10]. It however has other limitations that need to be addressed:

- No auto-scaling of the Kubernetes cluster is possible as no integration from Kubernetes to HPC is provided, only vice versa.
- User-impersonation is not possible, meaning the Kubernetes cluster is not able to request resource on behalf of a user.

Proposed Solution

To overcome the above mentioned limitations, a new solution to integrate SLURM and Kubernetes is proposed that extends the concept of ge-k8s [10]. By extending the concept with a Kubernetes operator, a deeper integration of the two systems is possible that results in the support for further use-cases. Major advantages of the proposed solution include:

Single resource pool

With the existing dual-level scheduling integration approaches, the SLURM and Kubernetes cluster are running on completely different hardware. Some of the existing approaches even go as far as to offload the Kubernetes cluster to a public cloud provider. While this approach might work well for certain HPC clusters, it has several limitations and downsides that prevent it from being used widely. One major drawback is that the computing power of the HPC cluster cannot be shared with Kubernetes. A bursty Kubernetes workload will require the Kubernetes cluster to scale in order to meet the demands of the submitted job. At the same time, the SLUM cluster might be underutilized and have some spare compute power available that it could allocate for the Kubernetes workload. With the proposed solution, this limitation will be overcome and allow the clusters to be seen as one big resource pool. Depending on the current usage, the overall computing power can be divided up between the SLURM and Kubernetes cluster as needed.

Unified Accounting

With both, the SLURM and Kubernetes cluster being tightly coupled, it allows both services to communicate and therefore properly report accounting information and enforce budget limits. This is a vital topic as Kubernetes itself does not have any accounting or budgeting functionality, which is very much needed when a cluster is shared between hundreds of people who only get a small share of the overall computing power. To achieve this, a Kubernetes operator is created that communicates with *slurmdbd* to report usage information and query budget limits. It then enforces these limits by killing Kubernetes workloads that are exceeding the users budget.

User Impersonation

Certain use-cases, like running a multi-user JupyterHub [7] instance, requires the ability to impersonate users and request resources on behalf of the user from the SLURM workload manager. This is possible by having a deep bi-directional connection between SLURM and Kubernetes.

Approach

To develop and test the proposed integration approach, a test setup has been created, which can be summa-

rized as follows:

- **Hardware**: Mainboard: Supermicro X9SCi-LN4F, CPU: Intel(R) Xeon(R) CPU E31270 (4 Cores, 8 Threads), Memory: 4x 8GB DDR3 ECC UDIMM 1333MHz
- **Hypervisor**: Proxmox Virtual Environment 6.4
- **Virtual Machines**: 4 VMs with 2 Cores, 4 GB Memory, Ubuntu 20.04
- **Automation**: Custom Ansible roles have been created to automate the complete provisioning process.

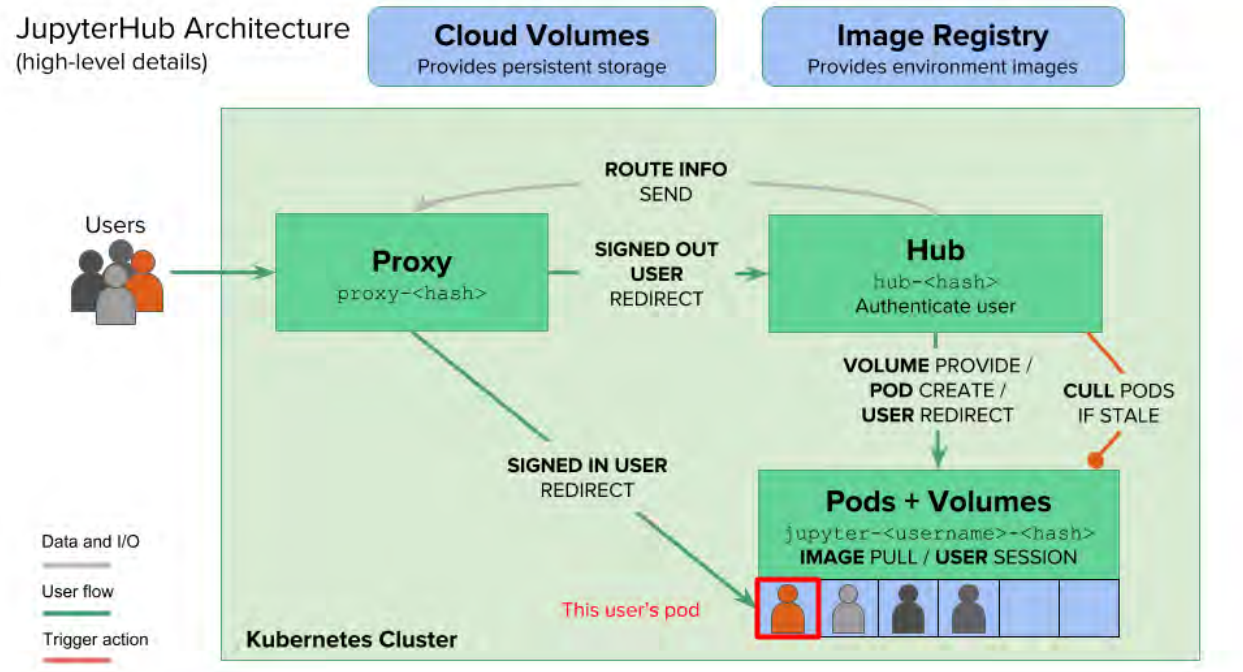

Fig. 1: The JupyterHub Architecture [4]

The implementation goal is to run a JupyterHub [7] service on Kubernetes which is tightly integrated with SLURM and correctly integrates with the SLURM accounting service. The high-level architecture of JupyterHub is highlighted in Figure 1 and shows how the allocation of user resources work. When a user visits JupyterHub and logs in with his credentials, a dedicated volume and pod is created on behalf of the user. By default, JupyterHub is handling its own resource guarantees and limits regarding CPU, GPU and memory. This process will be intercepted in order to read the resource information out of the SLURM accounting service and enforce those limits in Kubernetes.

References and figures

[1] The Kubernetes Authors. What is Kubernetes? https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/, 04 2022.

[2] Abdulrahman Azab. Enabling Docker Containers for High-Performance and Many-Task Computing. *2017 IEEE International Conference 2017*, pages 279–285, 2017.

[3] Cloud Native Computing Foundation. CNCF Sees Record Kubernetes and Container Adoption in 2021 Cloud Native Survey. https://www.cncf.io/announcements/2022/02/10/cncf-sees-record-kubernetes-and-container-adoption-in-2021-cloud-native-survey/, 2021.

[4] Project Jupyter Contributors. The JupyterHub Architecture. https://zero-to-jupyterhub.readthedocs.io/en/latest/administrator/architecture.html, 2022.

[5] Rim Doukha, Sidi Ahmed Mahmoudi, Mostapha Zbakh, and Pierre Manneback. Deployment of Containerized Deep Learning Applications in the Cloud. *2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech)*, pages 1–6, 2020.

[6] Sylabs Inc. WLM-operator. https://github.com/sylabs/wlm-operator, 2019.

[7] Project Jupyter. Project Jupyter | JupyterHub. https://jupyter.org/hub, 2022.

[8] Boris Lublinsky, Elise Jennings, and Viktória Spišaková. *A Kubernetes 'Bridge' operator between cloud and external resources*. Springer, 2022.

[9] Sergio López-Huguet, J. Damià Segrelles, Marek Kasztelnik, Marian Bubak, and Ignacio Blanquer. Seamlessly Managing HPC Workloads Through Kubernetes. In *High performance computing*, pages 310–320. Springer, 2020.

[10] Marco Enrico Piras, Luca Pireddu, Marco Moro, and Gianluigi Zanetti. Container Orchestration on HPC Clusters. In *High Performance Computing*, pages 25–35. Springer International Publishing, 2019.

[11] Andrey Sheka, Alexander Bersenev, and Victor Samun. Containerization in Scientific Calculations. *International Multi-Conference on Engineering, Computer and Information Sciences (2019 SIBIRCON)*, pages 793–798, 2019.

[12] Erich Strohmaier, Jack Dongarra, Horst Simon, and Martin Meuer. TOP500 LIST - NOVEMBER 2021. https://www.top500.org/lists/top500/list/2021/11/, 11 2021.

[13] Peter Z. Vaillancourt, J. Eric Coulter, Richard Knepper, and Brandon Barker. Self-Scaling Clusters and Reproducible Containers to Enable Scientific Computing. *2020 IEEE High Performance Extreme 2020*, pages 1–8, 2020.

[14] SLURM Workload Manager. SLURM Overview. https://slurm.schedmd.com/overview.html, 2021.

[15] Naweiluo Zhou, Yiannis Georgiou, Marcin Pospieszny, Li Zhong, Huan Zhou, Christoph Niethammer, Branislav Pejak, Oskar Marko, and Dennis Hoppe. Container orchestration on HPC systems through Kubernetes. *Journal of Cloud Computing*, 10, 2021.

[16] Naweiluo Zhou, Yiannis Georgiou, Li Zhong, Huan Zhou, and Marcin Pospieszny. Container Orchestration on HPC Systems. *Proceedings, 2020 IEEE 13th International Conference on Cloud Computing*, pages 34–36, 2020.

Untersuchung und Verbesserung der Usability der Fakultätsseite der Fakultät Informationstechnik der Hochschule Esslingen mittels Eye-Tracking

Michael Beyer **Reinhard Schmidt**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung:

Durch den digitalen Wandel haben heutzutage auch Hochschulen vermehrt eine starke Onlinepräsenz. Die meisten Studieninteressierten informieren sich inzwischen online über Studiengänge und Hochschulen. Auch die Hochschule Esslingen bietet daher eine allgemeine Website sowie eine separate Website für die Fakultäten. Im Zuge der Arbeit wird hierbei nur die Fakultätsseite der Fakultät Informationstechnik betrachtet und nicht die allgemeine Seite der Hochschule. Aufgrund des kontinuierlichen Fortschritts sollte die Seite auch stets gut nutzbar für die Zielgruppe sein. Da ein großer Anteil der Besucher der Seite aus Studieninteressierten bzw. deren Eltern besteht, liegt es im Interesse, dieser Gruppe die Nutzung der Seite möglichst einfach und interaktiv zu gestalten. Im ersten Schritt wurde daher der aktuelle Stand der Fakultätsseite in der Prototyping-Software Adobe XD nachgebaut. Aufgrund der zahlreichen Verlinkungen und der zu einem gewissen Grad beschränkten Möglichkeiten, die Adobe XD bietet, ist der Prototyp relativ aufwändig bzw. komplex gestaltet (vgl. Abb. 1).

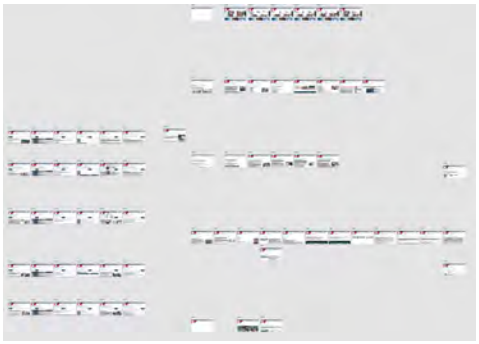

Abb. 1: Gesamtansicht des Adobe XD Prototyps [1]

Dieser Prototyp wird dann in einer Testinstanz auf seine Usability getestet. Der Begriff Usability hat viele verschiedene Definitionen und Bedeutungen. In diesem Zusammenhang wird Usability hier als einfache und schnelle Bedienbarkeit der Seite definiert [6]. Die Usability einer Anwendung bzw. Seite kann mithilfe eines sogenannten Usability Tests getestet werden. Ein großer Vorteil an Usability Tests ist, dass bereits eine sehr kleine Gruppe an Testpersonen (5-7 Testpersonen) den Großteil der Probleme und Schwierigkeiten aufdeckt. Usability Tests sind daher sehr gut geeignet, um Tests mit relativ geringen zeitlichen, sowohl als Personalaufwand durchzuführen (siehe Abb. 2).

Abb. 2: Relation Testnutzer/gefundene Probleme [4]

Ziel der Arbeit:

Im Umfang der Bachelorarbeit wird neben den herkömmlichen Mitteln wie generellen Fragen und Aufgaben an die Testpersonen, das Eye-Tracking Gerät „Tobii Pro Fusion“ verwendet. Es ermöglicht nachzuverfolgen, auf welchen Bereich der Seite die Testperson schaut. Diese Informationen können in der Auswertung der Daten hilfreich sein, um den Gedankengang der Testperson bei der Durchführung des Tests nachvollziehen zu können. Hinzu kommt, dass hiermit außerdem herausgefunden werden kann, in welchem Bereich der Seite die Testperson eine Funktion instinktiv erwartet. Die von Tobii mitgelieferte Software „Tobii Pro Lab“ bietet zudem noch weitere

Optionen zur Datenauswertung. Beispielsweise kann die Augenbewegung des Nutzers in einer Heatmap dargestellt werden. Diese Heatmap visualisiert Bereiche des Bildschirms in unterschiedlichen Farben, je nachdem wie viel Zeit der Nutzer den Bereich des Bildschirms betrachtet. Sie liefert daher eine datenbasierte Grundlage, die für Änderungen bzw. Umplatzierungen von Elementen helfen kann. Zum Zeitpunkt des Verfassens dieses Artikels wurden die Tests noch nicht durchgeführt, wodurch nicht final gesagt werden kann, ob diese Daten im Falle der Fakultätsseite aussagekräftig genug sind, um verwendet können zu werden oder ob die Testgruppe zu klein für eine eindeutige Auswertung ist. Ziel der Arbeit ist daher, mithilfe der Usability Tests, Probleme und Verbesserungen an der am Status Quo der Seite zu finden und Lösungen für diese zu implementieren. Im nachfolgenden, zweiten Testlauf soll dann überprüft werden, ob die Änderungen zu merkbaren Verbesserungen in der Usability der Seite geführt hat.

Adobe XD:

Adobe XD ist das Prototyping Tool, welches von Adobe zur Verfügung gestellt wird. Für die Arbeit wurde die Pro Version genutzt, welche kleinere Vorteile gegenüber der kostenlosen Version bietet. Für die meisten Anwendungen ist die kostenlose Version ausreichend, allerdings bietet die Pro Version beispielsweise deutlich mehr Schriftarten, welche hilfreich sind, die Fakultätsseite möglichst detailgetreu nachzubauen. Für die Arbeit wurde dieses Tool gewählt, da es eine sehr interaktive Bedienung zur Erstellung von Prototypen bietet. Zusätzlich bietet es sehr viele nützliche Plug-ins, welche beispielsweise eine Extrahierung eines Prototyps als HTML-Datei oder eine Vielzahl an vorgefertigten Icons, ermöglichen [5].

Eye-Tracking mittels Toobi:

Eye Tracking ist prinzipiell ein simpler Prozess. Zu Beginn wird eine Leiste am unteren Ende des Bildschirms befestigt. Diese Hardware sendet dann infrarotnahes Licht aus. Dieses Licht wird von den Augen reflektiert. Die Reflektionen werden dann wieder von der Augensteuerung aufgenommen. Durch vorherige Kalibrierung und nachträgliche Berechnungen, sowie Filter kann die Software dann berechnen, wohin der Nutzer sieht. Dafür muss die Augensteuerung die Pupillen des Nutzers finden. Durch die Position der Pupillen des Nutzers kann berechnet werden, wohin der Nutzer schaut [3] (vgl. Abb. 3). Der Vorteil an der Hardware von Tobii ist, dass der Nutzer weder Hardware wie beispielsweise eine Brille tragen noch sein Kopf fixiert sein muss, wie bei anderen Verfahren. Ein großer Vorteil bei Usability Tests mit einem Eye Tracking Gerät liegt darin, dass die Augenbewegungen auf dem Bildschirm aufgezeichnet werden. Daher können sie später jederzeit erneut angeschaut und analysiert werden. Bei normalen Tests stehen hierfür nur Ton- oder Videoaufnahmen zur Verfügung, welche ohne den Kontext des Bildschirms allerdings oft nicht den vollen Kontext bieten [2].

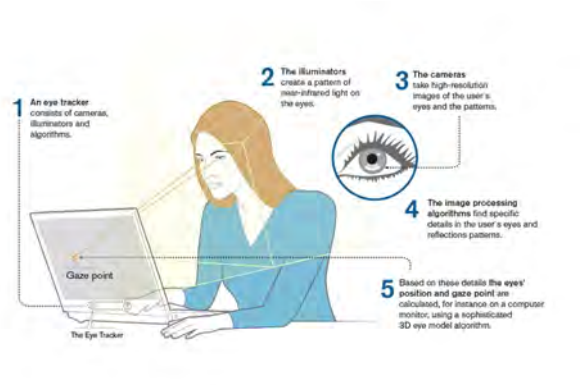

Abb. 3: Wie Eye-Tracking funktioniert [2]

Ausblick:

Die Lösungen und Verbesserungsvorschläge, die im Umfang der Usability Tests erlangt werden können später nicht nur auf die Seite der Fakultät Informationstechnik angewandt werden. Da sämtliche Fakultätsseiten in Design und Aufbau sehr ähnlich sind können viele Änderungen bei Bedarf auch auf diese Seiten übernommen werden. Des Weiteren wird sich auch zeigen, wie sinnvoll und praktikabel die Anwendung von Eye Tracking in diesem Zusammenhang ist bzw. ob die Ergebnisse des Eye Trackings aussagekräftig sind und zielorientiert genutzt werden können.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Maria Gordon. Wie funktioniert Eyetracking? https://de.tobiidynavox.com/pages/funktioniert-eyetracking, 2022.

[3] Kenneth Holmqvist, Marcus Nyström, et al. *Eye Tracking A comprehensive guide to methods and measures*. Oxford University Press, 1 edition, 2011.

[4] Jakob Nielsen. *Designing Web Usability*. Markt + Technik-Verlag, 2 edition, 2001.

[5] Matt Rae. What is Adobe XD and what is it used for? https://www.adobe.com/de/products/xd/learn/get-started/what-is-adobe-xd-used-for.html, 2022.

[6] Stephan Thesmann. *Interface Design Usability, User Experience und Accessibility im Web gestalten*. Springer Vieweg Verlag, 2 edition, 2016.

Konzeptionierung und prototypische Implementierung eines App-Stores zur Verteilung von Micro-Frontend Anwendungen

Maxim Bickel **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz Tech Motion GmbH, Leinfelden-Echterdingen

Einführung

Microservices sind aus der heutigen IT nicht mehr wegzudenken, da sich diese durch zahlreiche Vorteile auszeichnen. Beispielsweise wird durch die Skalierbarkeit und die Unabhängigkeit der einzelnen Services die Belastbarkeit des Gesamtsystems erhöht, wodurch eine robuste Anwendung mit geringerem Ausfallrisiko erstellt werden kann. Micro Services können dynamisch hoch- und runtergefahren werden und sind zudem leicht austauschbar. Zusätzlich ist es möglich, jeden Service mit einer unterschiedlichen Technologie umzusetzen, um die bestmögliche Performance des Gesamtsystems zu ermöglichen. [2] Das Konzept der Microservices wird aktuell hauptsächlich im Backend einer Anwendung verwendet, weswegen die meisten zugehörige Frontends meist noch eine monolithische Struktur aufweisen. Abhilfe sollen Micro-Frontends schaffen, welche den Gedanken bzw. die Vorteile des Microservice auf das Frontend erweitern sollen.

Grundlagen

Micro-Frontends stellen laut Geers [1] keine konkrete Technologie dar, sondern sollen eher als alternativer organisatorischer und architektonischer Ansatz verstanden werden. Die Abbildung 1 zeigt dabei einen Überblick, wie ein solcher Ansatz umgesetzt werden kann und welche Bestandteile nötig bzw. voneinander abhängig sind.
Grundlegend können Micro-Frontends vollständige Seiten oder auch nur eine einzelne Komponente, wie beispielsweise ein Kommentarfeld oder eine Suchleiste, einer Webanwendung darstellen. Diese Seiten bzw. Fragmente können dann in eine Host-Applikation eingebunden werden. Diese fungiert als eine Art Grundlage, welche alle Bestandteile in einer Anwendung miteinander verbindet und dafür sorgt, dass die Bestandteile eingebunden werden können. Dabei übernimmt die Host-Applikation zudem Aufgaben wie Routing und die Kommunikation zwischen den einzelnen Micro-Frontends.
Ein Vorteil bei der Verwendung von Micro-Frontends ist, dass die zuständigen Teams unabhängig voneinander sind und eigenständige Entscheidungen, in Bezug auf die Art und Weise der Umsetzung des Moduls, treffen können. Ist beispielsweise ein Team für die Erstellung des Warenkorbs zuständig und ein anderes für die Auflistung der einzelnen Produkte in einem Online-Shop, so können beide Teams unterschiedliche Frameworks verwenden, beispielsweise React.js und Angular.js. Eine Absprache bzw. generelle Kommunikation zwischen den einzelnen Teams ist nicht zwingend notwendig.

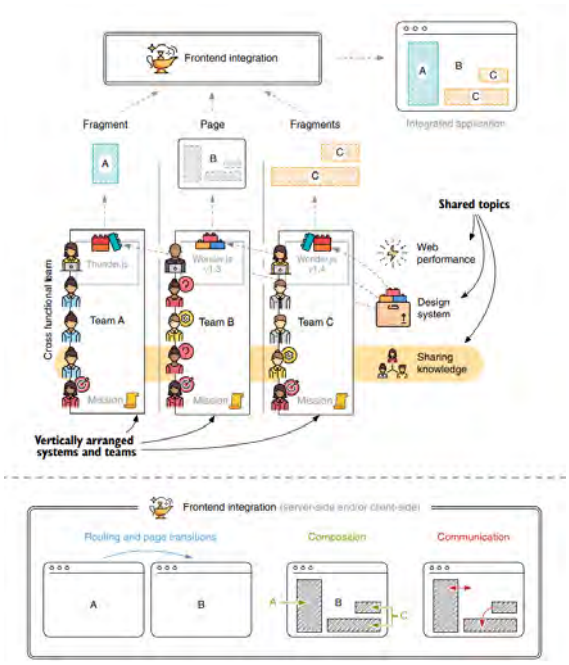

Abb. 1: Gesamtüberblick über das Konzept der Mikro-Frontends [1]

Zielstellung & Motivation

Das Ziel der Bachelorarbeit ist es ein Konzept für einen Marktplatz zu entwickeln, welcher unterschiedliche Micro-Frontends anbietet und in eine Anwendung dynamisch ein- bzw. ausladen kann. Diverse Nutzer sollen Micro-Frontends auf diesem Marktplatz anbieten können und durch eine Konfigurationsschnittstelle für andere Host-Applikationen benutzbar machen. Hierbei ist der Gedanke, dass innerhalb großer Strukturen (beispielsweise eines Unternehmens oder sogar des gesamten Internets) einige Sachverhalte redundant bearbeitet werden, obwohl eine mögliche Lösung bereits besteht. Nutzer haben durch die Verwendung des Marktplatzes somit die Möglichkeit vor dem Aufwenden eigener Gelder und Ressourcen ein passendes Micro-Frontend zu finden und in die Host-Anwendung einbinden zu lassen. Durch das Verwenden eines eignen Style-Themes können die einzelnen Komponenten innerhalb der bestehenden Anwendung an das bestehende Aussehen angepasst werden.
Durch die Verwendung eines solchen Marktplatzes können Synergieeffekte zwischen beispielsweise den Abteilungen eines Unternehmens genutzt werden, da einige Komponenten bereits erstellt sind und durch die Kollegen weiterhin gepflegt werden. Für den Nutzer eines solchen angebotenen Micro-Frontends ergibt sich daher der Vorteil, dass dieser die Kosten für die Entwicklung einsparen kann und das Modul durch den Ersteller weiterentwickelt wird. Zusätzlich können neue Innovationspotenziale ausgeschöpft werden, da eine Verbindung mehrerer Bereiche zustande kommt. Somit können Ideen und fachliche Kompetenzen besser diskutiert bzw. ausgetauscht werden.
Zusätzlich zur Konzeptionierung des Marktplatzes wird eine prototypische Implementierung erstellt, welche die grundlegende Funktionalität bestätigten und eine erste Grundlage für ein späteres Produkt liefern soll.

Umsetzung & Ausblick

Das Konzept sieht vor, dass der Marktplatz an sich als Micro-Frontend implementiert wird. Somit kann der Marktplatz als eigenständige Applikation, aber auch innerhalb jeder anderen Host-Applikation verwendet werden. Wodurch sich ein hohes Maß an Flexibilität ergibt und die Benutzerfreundlichkeit deutlich verbessert werden kann. Die Host-Applikation eines Nutzers kann durch ein Template erstellt werden, welches das Single-SPA Framework verwendet. Dieses Template kann durch ein Helm Container ausgeliefert und konfiguriert werden. Innerhalb des Templates wurde ein weiteres Micro-Frontend eingebunden, welches als eine Art Plugin fungiert, um mit ein Backend, welches ebenfalls als Helm-Container ausgeliefert wird zu kommunizieren. Dieses Backend wird zum Abspeichern einer Konfigurationsdatei verwendet, welche alle notwendigen Informationen für die eingebunden Micro-Frontends enthält. Das Plugin kann die Informationen der Konfigurationsdatei auslesen und lädt die notwendigen Micro-Frontends in die Host-Applikation ein. Alternativ können diese Informationen auf dem Backend des eigentlichen Marktplatzes abgespeichert werden. Der Nutzer hat daher die Möglichkeit durch das Hosten eines eigenen Backends die Abhängigkeit zu dem Marktplatz zu verringern. Innerhalb des Marktplatz-Backends werden alle Informationen der angebotenen Micro-Frontends, wie zum Beispiel Bewertungen, notwendige Konfigurationen, Profile oder auch Kommentare abgespeichert.
Entscheidet sich ein Nutzer dafür seine Anwendung erweitern zu wollen, so muss dieser lediglich den Marktplatz innerhalb seiner Host-Applikation, über eine URL aufrufen und ein passendes Micro-Frontend auswählen. Daraufhin werden notwendige Konfigurationsparameter, wie beispielsweise API-Pfade oder Token abgefragt und durch das Abspeichern des Verantwortlichen der Host-Applikation voll automatisch eingeladen. Dabei ist kein Neustart bzw. Update der Anwendung notwendig. Erweiterungen können nach demselben Prinzip entfernt bzw. deaktiviert werden. Die Änderungen sind sofort für alle Nutzer der zu erweiternden Webanwendung verfügbar.
Durch den bereits erstellten Prototypen konnte bestätigt werden, dass die Umsetzung nach dem obigen Beispiel realisierbar ist. Nachfolgend soll die prototypische Implementierung durch ein ausgereiftes Produkt abgelöst werden und innerhalb der Abteilung zum Test freigegeben werden. Dabei besteht der Use-Case darin, bereits vorliegende Komponenten aufzuarbeiten und innerhalb des Marktplatzes für zukünftige Aufträge abzulegen. Für zukünftige Kundenprojekte können die Komponenten verwendet werden, ohne erneuten Entwicklungsaufwand aufwenden zu müssen. Sobald die produktiven Tests abgeschlossen sind, wäre es denkbar den Marktplatz für mehrere Abteilungen innerhalb des Konzerns zur Verfügung zu stellen und somit das volle Potenzial auszuschöpfen.

Literatur und Abbildungen

[1] Michael Geers. *Micro Frontends in Action*. Manning Publications Co., 2020.

[2] Eberhard Wolff. *Microservices: Grundlagen flexibler Softwarearchitekturen*. dpunkt.verlag, 2 edition, 2018.

Einsatzuntersuchungen für eine RFID-gestützte Bauteilerfassung in der Fahrzeugmontage

Lorena Braendle **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Dr. Ing. h.c. F. Porsche AG, Stuttgart-Zuffenhausen

Einleitung

RFID-Systeme (Radio Frequency Identification) gehören zu den Auto-ID-Systemen. Sie umfassen Techniken zur Identifizierung, Datenerhebung, Datenerfassung und Datenübertragung. Abbildung 1 zeigt eine zusammenfassende Übersicht der wichtigsten Auto-ID-Verfahren. RFID ist eine Technologie zur automatischen Identifikation von Objekten und zeigt einzigartige Potentiale für deren Einsatz in der Automobilindustrie auf. Sie gilt als Schlüsseltechnologie im Bereich Logistik- und Produktionstransparenz [4]. Der Vorteil von RFID gegenüber den anderen Auto-ID Verfahren ist, dass der Datenaustausch unter Verwendung magnetischer oder elektromagnetischer Felder erfolgt. Dies ermöglicht eine kontaktlose und sichtkontaktfreie Identifikation von Objekten [3].

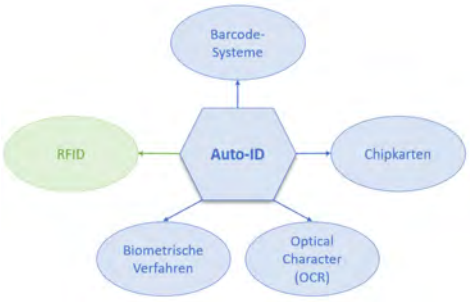

Abb. 1: Die wichtigsten Auto-ID-Verfahren [1]

Zielsetzung der Arbeit

Diese Abschlussarbeit fokussiert sich auf die RFID-Technologie und untersucht deren Nutzen und Potenziale für die Bauteilerfassung in der Fahrzeugmontage der Dr. Ing. h.c. F. Porsche AG. Dafür wird untersucht, an welchen Positionen im Fahrzeug abhängig vom Produktionsfortschritt ein Bauteil mittels RFID prozesssicher erfasst werden kann. Im Fokus der Bachelorarbeit steht dabei die Untersuchung der technischen Realisierbarkeit.

Technische Grundlagen der RFID-Technologie

Um ein RFID-System einsetzten zu können muss vorerst betrachtet werden, welche Komponenten ein RFID-System bilden und wie diese miteinander interagieren. Ein RFID-System besteht aus zwei Komponenten: dem Transponder und einem Schreib-/Lesegerät (RFID-Reader), wobei beide mit einem eigenen Koppelelement in Form einer Antenne ausgestattet sind. Die Abbildung 2 stellt die Grundbestandteile eines RFID-Systems exemplarisch dar. Ein Schreib-/Lesegerät ist ein Steuergerät, das die Steuerung der Antenne und die Logik zur Übertragung der empfangenen Daten an die Applikation übernimmt. Ein Lesegerät umfasst ein Hochfrequenzmodul (Sender und Empfänger), eine Kontrolleinheit sowie ein Koppelelement zum Transponder. Zusätzlich kann ein Lesegerät über eine Schnittstelle mit einem weiterverarbeitenden IT-System verbunden sein, um die erhaltenen Daten weiterzuleiten. Antennen in Lesegeräten oder in Transpondern dienen zur Erzeugung des elektromagnetischen Feldes, das zur Spannungsversorgung des Transponders und zur Nachrichtenübertragung zwischen Lesegerät und Transponder eingesetzt wird [2]. Ein Transponder, auch Tag genannt besteht aus einem elektronischen Mikrochip, der den eigentlichen Datenträger darstellt und einem Koppelelement in Form einer Antenne. Auf einem Transponder können bauteilspezifische Daten, die für das Unternehmen relevant sind, hinterlegt werden. Dies kann bspw. die Serien-, Herstellnummer oder die Artikelbezeichnung sein. Der Transponder wird fest an dem zu identifizierenden Objekt angebracht und kontaktlos über Funktechnologie ausgelesen. Ein Transponder der in der Regel keine eigene Spannungsversorgung (Batterie) besitzt verhält sich Außerhalb des Ansprechbereichs eines Lesegeräts passiv. Er wird erst aktiviert, wenn er sich innerhalb des Ansprechbereichs

eines Lesegeräts befindet. Alle Transponder, die sich im Lesefeld befinden, empfangen die vom Lesegerät ausgesendeten Befehle und Daten und schicken die entsprechenden Antwortdaten an das Lesegerät zurück. Durch die Kopplungseinheit des Lesegeräts wird die zum Betrieb des Transponders benötigte Energie ebenso wie die Daten (kontaktlos) zum Transponder übertragen [2].

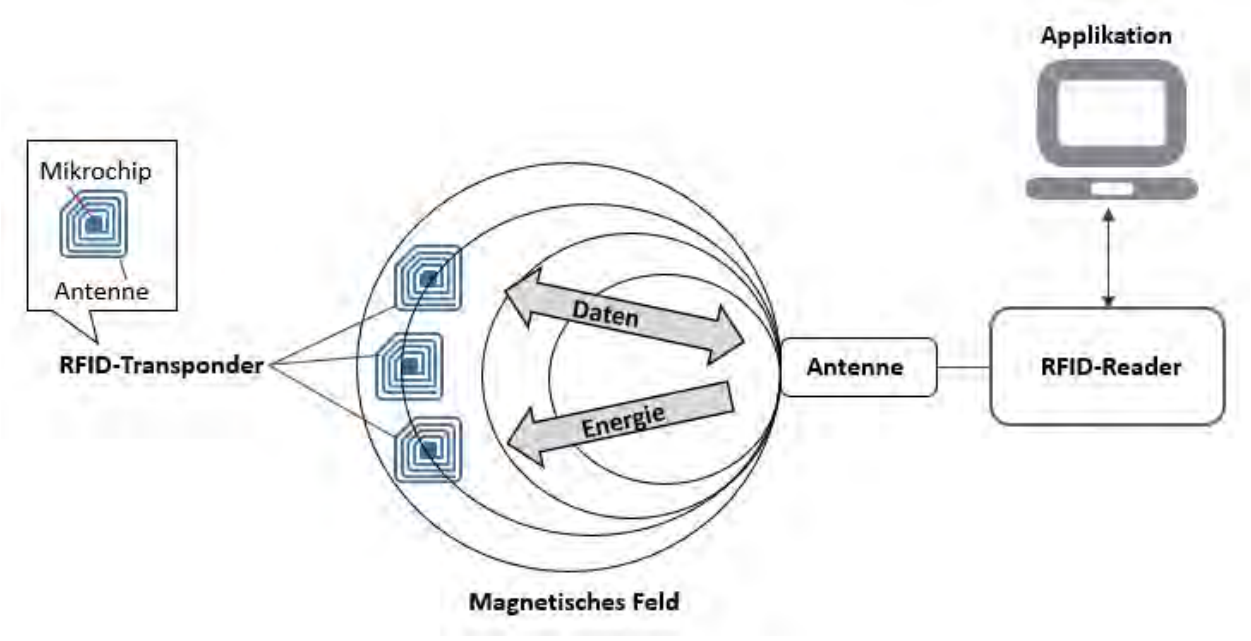

Abb. 2: Grundbestandteile eines RFID-Systems [1]

Zwischenstand

Ersten Ergebnissen zufolge ist ein RFID-System anfällig gegenüber bestimmten Störgrößen, die für einen erfolgreichen Einsatz im Montageband beachtet werden müssen. Herausforderungen oder Einschränkungen, die sich für das RFID-System ergeben, können durch das kennzeichnende Objekt selbst oder die Umgebungsbedingungen des Erfassungsprozesses hervorgerufen werden. Ein Einflussfaktor ausgehend von dem Bauteil selbst ist das Oberflächenmaterial des Bauteils. Es beeinflusst die Lesequalität und die Reichweite zwischen Reader und Transponder. Das Oberflächenmaterial bestimmt die Wahl der Transponderart, beispielsweise benötigt man für metallische Bauteiloberflächen spezielle Metall-Transponder sogenannte On-Metall Transponder. Sie besitzen ein zusätzliches Abstandsmaterial, das sich zwischen der Antenne und der Applikationsoberfläche befindet und somit einen Kurzschluss ausschließt. Auch die durch den Produktionsprozess oder die Umgebungsbedingungen hervorgerufenen Einflussfaktoren müssen bei der technischen Machbarkeitsuntersuchung beachtet werden. Mögliche Einflussfaktoren, die von dem Montageprozess ausgehen, sind die Zugänglichkeit der einzelnen Bauteile am Fahrzeug, sowie der Ort an der Montageline.

Ausblick

Die bereits gewonnenen Erkenntnisse über die Umgebungsbedingungen und Bauteileigenschaften, die für den Einsatz eines RFID-Systems erfüllt sein müssen, bilden die Grundlage um anschließend Positionen im Montageband zu identifizieren, um dort eine technische Erprobung unter Serienbedingungen durchzuführen. Dabei werden unterschiedliche Bauteile in verschiedenen Stufen des Produktionsfortschrittes des Fahrzeugs untersucht, ob sie sich mittels RFID prozesssicher erfassen lassen. Zu erwarten ist, dass Faktoren wie das Material des Bauteils und die Entfernung und Ausrichtung des Transponders zum Lesegerät die Qualität der Erfassung erheblich beeinflussen. Auf Grundlage der erzielten Ergebnisse wird anschließend eine Handlungsempfehlung ausgesprochen, welche Bauteile abhängig von der Position im Fahrzeug mittels RFID prozesssicher erfasst werden können. Und an welchen Positionen im Montageband eine RFID-Lesestation realisiert werden kann.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Klaus Finkenzeller. *RFID Handbuch: Grundlagen und praktische Anwendungen von Transpondern, kontaktlosen Chipkarten und NFC*. Hanser, 6 edition, 2012.

[3] Manfred Helmus et al. *RFID in der Baulogistik*. Vieweg + Teubner, 1 edition, 2009.

[4] Walter Huber. *Industrie 4.0 in der Automobilproduktion*. Springer Vieweg, 2016.

Erkennung von Topics in Jira Tickets mit Natural Language Processing Algorithmen

Dominik Buecher **Clemens Klöck**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Einleitung

Da die Softwareprojekte heutzutage immer größer werden, ist es sinnvoll ein Tool für das Managen dieser Projekte zu nutzen. In vielen Firmen wird für das Projektmanagement Jira genutzt, dabei handelt es sich um eine Software, welche entwickelt wurde, um das Planen, Verfolgen und Releasen von Software oder Produkten, innerhalb eines Scrum-Teams zu ermöglichen [4]. Eine wichtige Funktion in Jira ist die Erstellung von Tickets, diese können beispielsweise für das Zuweisen einer Aufgabe, z.B. die Lösung eines Fehlers, erstellt werden. Bei der Vector Informatik GmbH entstehen die Tickets auf der Grundlage von Problemen der Kunden, welche als Support-Anfrage gesendet werden. Diese Tickets wurden über einen gewissen Zeitraum abgespeichert und in einen separaten Datensatz zusammengefasst. Da die Tickets meist sehr unterschiedlich sind, fällt es selbst dem Menschen schwer, die Nachrichten den richtigen Gruppen zu zuordnen. Für dieses Problem soll ein Algorithmus implementiert werden, der die Zuordnung automatisch übernimmt, und damit Zeit und Kosten einspart.

Zielsetzung

Das Ziel der Bachelorarbeit ist es, mit dem Datensatz, welcher die Jira Tickets beinhaltet, ein geeignetes Model zu trainieren. Dieses soll dann möglichst automatisiert die Tickets den richtigen Bearbeitungsgruppen zuordnen, somit würde die Bearbeitung der Tickets noch effizienter vonstattengehen. Da es sich bei den Tickets zum Großteil um ungelabelte Daten handelt (es sind ca.16% der Daten gelabelt, jedoch sind diese Label sehr ungleich verteilt und ihre Qualität ist unbekannt, daher werden sie nur verwendet, um einen visuellen Eindruck über das Model zu erhalten), muss ein Algorithmus verwendet werden, welcher ein unsupervised Ansatz verfolgt. Mit diesem sollen dann Cluster gebildet werden, welche für die verschiedenen Bearbeitungsgruppen stehen. Im Bereich des Natural Language Processing (NLP) gibt es eine Unterkategorie, dass *Topic Modelling* welches hierfür in Frage kommt.

Topic Modelling

Topic Modelling Verfahren sind Algorithmen welche mit Wahrscheinlichkeitsrechnungen größere Textansammlungen, Verstehen, Suchen und Organisieren können [6]. Dabei ist es besonders interessant, dass die Verfahren für das Training der Models keine Labels benötigen. Die Models clustern die Daten in *Topics* welche für die verschiedenen Labels stehen. In diesem Bereich des NLP gibt es zahlreiche Algorithmen, welche mit unterschiedlichen Ansätzen an die Problemstellung heran gehen. Der bekannteste Algorithmus, ist die *Latent Dirichlet Allocation* (LDA). Dieses Verfahren stellt die Dokumente als Zufällige Mischung über latente Themen dar, dabei werden die Themen durch eine Verteilung über die Wörter charakterisiert [5]. In der folgenden Abbildung (Abb. 1) sieht man den groben Prozessablauf des LDA-Verfahrens, zur Bestimmung der Topics.

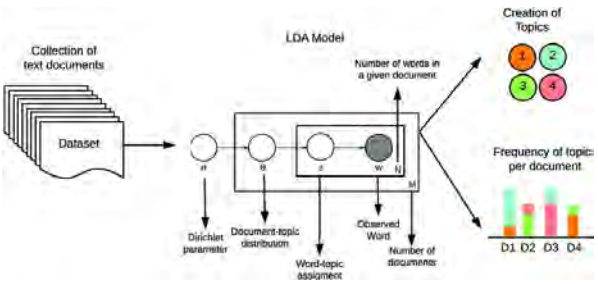

Abb. 1: Prozessablauf des Latent Dirichlet Allocation [2]

Dieses und noch weitere Verfahren werden im Laufe des Bachelorarbeit implementiert und untereinander verglichen, um so ein Verfahren zu finden, welches die Tickets in möglichst gut zusammenpassende Topics, einteilt.

Konzept

Um ein Modell trainieren zu können, müssen als erstes die Daten in eine passende Form gebracht werden. Hierfür wird der gesamte Datensatz vorverarbeitet. Das heißt, der Text wird in Kleinbuchstaben dargestellt, es werden verschiedene Zeichen und *Stop-Words* entfernt. Bei *Stop-Words* handelt es sich um Wörter, welche dem Satz nicht viel Bedeutung verleihen und daher ignoriert werden können. Nach dem das erledigt ist, werden die Tickets in ein Dataframe geschrieben (dabei handelt es sich um ein Datentype für Tabellen in Python), um die weitere Verarbeitung zu erleichtern. Im Anschluss an das Training eines Modells mit einem Algorithmus, wird dieses dazu verwendet, im Training nicht verwendete Daten, zu clustern. Bei diesen Testdaten handelt es sich um den kleinen Teil des Datensatzes, welcher mit einem Label versehen ist. Die Dabei gebildeten Cluster können verwendet werden, um die Funktionsweise des Modells grafisch darzustellen (siehe Abb. 2).

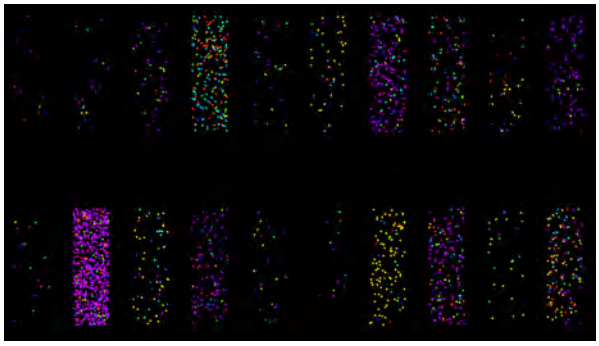

Abb. 2: Die Verteilung zeigt die Zuordnung der Tickets mit Label, durch das antrainierte LDA Modell. Wobei ein Punkt für ein Ticket steht und die Farbe für ein Label. Die vom Modell gebildeten 20 Cluster werden durch die angedeuteten Rechtecke dargestellt. [3]

In der Abbildung 2 kann man die Verteilung der Tickets auf die Topics erkennen, jedoch ist das nur eine Darstellung, um ein Eindruck über das Model zu bekommen. Die richtige Auswertung des Ergebnisses muss anderweiter vorgenommen werden. Da am Ende der Thesis die Tickets möglichst automatisch gelabelt werden sollen, werden die Algorithmen in eine *DVC* Pipeline implementiert. Eine DVC-Pipeline ermöglicht, mehrere zusammenhängende Python Skripte mit nur einem Befehl durchlaufen zu lassen [1]. Somit werden alle Modelle direkt nacheinander trainiert und getestet, dabei werden für jeden Algorithmus die gleichen Parameter gewählt, um so vergleichbare Ergebnisse zu erzeugen.

Ausblick

Nachdem drei Algorithmen aus unterschiedlichen Gebieten, von klassischen statistischen Ansätzen bis zu Neuronalen Netzen, implementiert wurden, liegt der Fokus im weiteren darauf Vergleichs Metriken für die Ergebnisse dieser Verfahren zu finden. Hierbei liegt, da es nur sehr wenige Label gibt und diese auch noch von unbekannter Qualität sind, der Schwerpunkt auf Metriken die keinen „Ground Truth“ benötigen.

Literatur und Abbildungen

[1] Amine Barrak, E. Ellis Eghan, and Bram Adams. On the Co-evolution of ML Pipelines and Source Code - Empirical Study of DVC Projects. *IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2021.

[2] Diego Buenano-Fernandez. Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *SPECIAL SECTION ON ADVANCED DATA MINING METHODS FOR SOCIAL COMPUTING*, 2020.

[3] Eigene Darstellung.

[4] Fisher John, D. Koning, and A.P. Ludwigsen. Utilizing Atlassian Jira For Large-Scale Software Development Management. *ICALEPCS*, 2013.

[5] John Lafferty. Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 2003.

[6] Tong Zhou and Zhang Haiyi. A TEXT MINING RESEARCH BASED ON LDA TOPIC MODELLING. In *Computer Science & Information Technology*, pages 201–210. Jan Zizka, 2016.

Vergleichsanalyse der Sicherheitsfunktionen zwischen den Ladekommunikationsstandards ISO 15118-2 und ISO 15118-20 in der Elektromobilität

Martin Dell **Tobias Heer**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Motivation und Problemstellung

Die jährliche verkehrsnahe Stickstoffdioxid-Belastung sinkt [6]. Dies ist den Innovationen in der Automobilindustrie, aber auch gesetzlichen und gesellschaftlichen Vorgaben zu verdanken. Viele Regionen setzen nun auf einen Wandel zur Elektromobilität. Um ein Elektrofahrzeug zu laden, wird eine spezielle Infrastruktur benötigt. Die Spezifizierung dieser Infrastruktur wurde bereits in mehreren Standards definiert. Ein Mangel wurde jedoch bei der Standardisierung der Kommunikation zwischen Fahrzeug und Ladesäulen festgestellt, weshalb die Norm ISO 15118 ins Leben gerufen wurde [4]. Part 2 der Norm (ISO 15118-2) beschäftigt sich mit den Anforderungen an die Netzwerk- und Anwendungsprotokolle. In der ISO 15118-2 wird auch die *Plug & Charge* (PnC) Funktion definiert. Plug & Charge erlaubt das Laden des Fahrzeugs, ohne weitere Nutzerinteraktionen, außer dem Anschließen des Ladekabels. Das Fahrzeug übernimmt die Authentifizierung gegenüber der Ladesäule mit einem Vertragszertifikat (Ladevertrag), welches im Fahrzeug gespeichert wird [4]. Dieser Nutzerkomfort senkt die Hemmschwelle zum Umstieg auf Elektrofahrzeuge. Abbildung 1 zeigt vereinfacht den Plug & Charge Use Case.

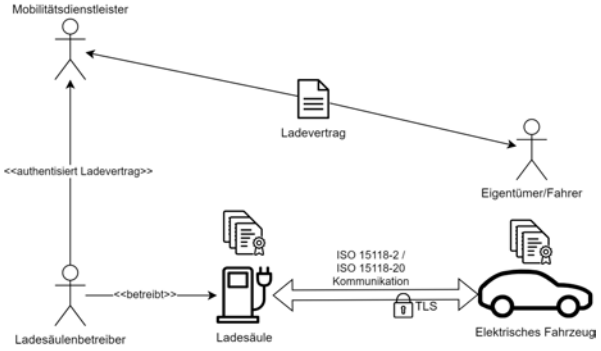

Abb. 1: Plug & Charge Use Case (vereinfacht) [3]

Voraussetzung für Plug & Charge ist, dass der Eigentümer des Fahrzeugs einen Ladevertrag mit einem Mobilitätsdienstleister abgeschlossen hat. Die Ladesäule und das Fahrzeug kommunizieren mit TCP/IP-Protokollen und dem V2G-Protokoll aus der ISO 15118-2. Die Verbindung ist mit dem Transport Layer Security (TLS) Protokoll abgesichert, welches die zertifikatsbasierte Authentifizierung, den Integritätsschutz sowie die Verschlüsselung der Nachrichten übernimmt [4].
Auf Basis der ISO 15118-2 wurde der Standard ISO 15118-20 entwickelt. Neben vielen neuen Funktionen wie Bidirectional und Wireless Power Transfer (BPT und WPT) werden auch folgende sicherheitsrelevante Anforderungen neu definiert, die in späteren Abschnitten noch mal aufgegriffen werden:

- Obligatorischer Einsatz von TLS 1.3 mit Zwei-Wege-Authentifizierung
- Kryptografische Algorithmen mit höherem Security-Level
- Sekundäre kryptografische Algorithmen, falls die Primären als unsicher eingestuft werden
- Geänderte Zertifikatsprofile
- Einführung von Cross-Zertifizierung

Zu beachten ist hier jedoch auch, dass die ISO 15118-20 nicht mit der ISO 15118-2 kompatibel ist und somit alle entwickelten ISO 15118-2 Tools nicht mit der neuen Norm funktionieren [5].

Ziele der Arbeit

Die ISO 15118-20 bringt Änderungen im Bezug der Sicherheitsfunktionen mit sich. Im Rahmen dieser Arbeit sollen diese Funktionen denjenigen des Vorgängers ISO 15118-2 gegenübergestellt werden. Die Betrachtung der funktionalen Sicherheit (Safety) von beiden Normen ist nicht Bestandteil. Es soll

anhand des Fallbeispiels von einem bereits existierenden Validierungstool für die ISO 15118-2 aus der Arbeit von Bogner [1] gezeigt werden, wie sich diese Änderungen auf eine mögliche Umstellung zur ISO 15118-20 auswirken. Dabei soll das in .NET Framework 4.7 geschriebene Validierungstool die Funktionen der neuen Norm mit unterstützt. Ebenfalls ist Teil der Arbeit, TLS 1.2 und TLS 1.3 im Kontext der Normen gegenüberzustellen und möglich Auswirkungen auf Sicherheit und Kompatibilität zu beleuchten. Es soll die Unterstützung von kryptografischen Algorithmen sowie auch die möglichen Auswirkungen von TLS 1.3 in Ladesäulen und Fahrzeugen analysiert werden. Es wird nicht Bestandteil sein, TLS 1.3 in dem Validierungstool zu implementieren. Entsprechend sind die Ziele der Arbeit:

1. Vergleich von Zertifikatsprofilen der ISO 15118-2 und ISO 15118-20
2. Vergleich von TLS 1.2 und 1.3 im Kontext der Standards
3. Analyse der bekannten Schwachstellen aus ISO 15118-2 im Kontext der ISO 15118-20
4. Analyse der Auswirkungen auf das Validierungstool
5. Unterstützung für ISO 15118-20 im vorhandenen Tool hinzufügen

Zertifikate

Für die Authentifizierung und den Integritätsschutz auf Transportebene und darüber des OSI-Schichtmodells werden mehrere verschiedene Zertifikate benötigt. Wie ein Zertifikat aussehen muss, definiert ein Zertifikatsprofil. Mehrere logisch miteinander verknüpfte Zertifikatsprofile nennt man *Cluster* [4]. ISO 15118-2 hat für die verschiedenen Beteiligten der Infrastruktur Cluster definiert. Außerdem wird jeweils nur ein kryptografischer Algorithmus für die Signatur des Zertifikats und die Schlüssel erlaubt. Mit der ISO 15118-20 ändern sich die Zertifikatsprofile inhaltlich und es wird ein weiteres Cluster eingeführt. Zusätzlich gibt es nun zwei kryptografische Algorithmen, die parallel unterstützt werden müssen. Einer der Algorithmen ist dabei nicht sehr verbreitet [5].
Zur Authentifizierung werden Signaturen verwendet. Signaturen werden mit Zertifikaten geprüft. Jedoch besitzen wiederum auch Zertifikate Signaturen. Falls diese nicht selbst signiert wurden, benötigt man somit wieder Zertifikate zur Prüfung. So wächst rekursiv die Zertifikatskette, bis sie bei einem Stammzertifikat, welches sich selbst signiert, endet. Ein Zertifikat kann nur dann vollständig validiert werden, wenn die gesamte Kette validiert wurde.

Cross-Zertifizierung

Eine Neuerung, die mit der ISO 15118-20 eingeführt wird, ist die Unterstützung von Cross-Zertifizierung, wodurch größere Unterschiede in den Zertifikatsketten entstehen können [5]. Die Norm besitzt wenig Einschränkungen bezüglich Cross-Zertifizierung und erlaubt ebenfalls aufgrund mangelnder Spezifikation mehrere Interpretationsmöglichkeiten, wie eine Cross-Zertifizierung aussieht. Die folgende Abbildung 2 stellt eine der möglichen Interpretationsmöglichkeiten dar.

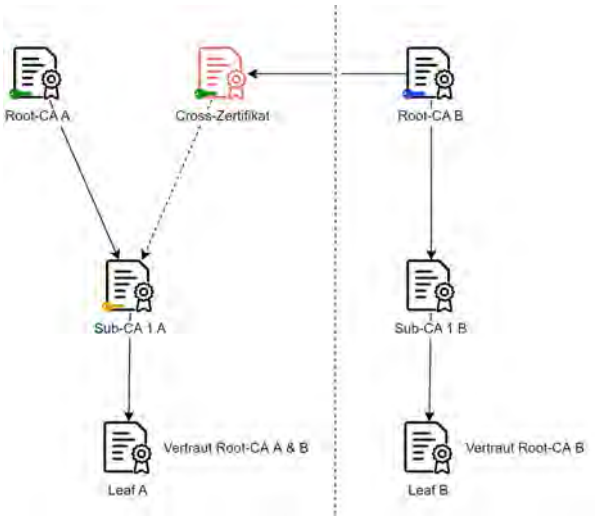

Abb. 2: Interpretation für Cross-Zertifizierung in der ISO 15118-20 [3]

Das Cross-Zertifikat ist ein Zertifikat mit unter anderem dem Public-Key des Zertifikats, welches cross-zertifiziert werden soll (hier *Root-CA A*). Die jeweilige Zertifizierungsstelle der anderen Public-Key-Infrastruktur (hier *Root-CA B*), signiert dann das Cross-Zertifikat. Dadurch können nun alle von *Root-CA A* abgeleiteten Zertifikate auch mit dem *Root-CA B* geprüft werden. Damit vertrauen Zertifikate von *Root-CA A* auch *Root-CA B*, jedoch nicht umgekehrt. Es gibt jedoch noch weitere Interpretationen in der ISO 15118-20. Es kann beispielsweise auch zwei Zertifikate von Zwischenzertifizierungsstellen (Sub-CA) geben, wovon eins dann das Cross-Zertifikat ist. Mehrere Interpretationsmöglichkeiten schaden der Interoperabilität und müssen später noch mal von einem Institut genauer spezifiziert werden.

Transport Layer Security

Zur Absicherung der TCP-Kommunikation zwischen Ladesäule und Fahrzeug ist in der ISO 15118-2 TLS 1.2 nur bei der Benutzung von Plug & Charge verpflichtend [4]. In der ISO 15118-20 ist TLS 1.3 immer notwendig [5]. Mit TLS 1.3 haben sich auch

die kryptografischen Algorithmen geändert. Wie bei den Zertifikaten gibt es zwei Algorithmen, wobei der sekundäre Algorithmus hauptsächlich zum Einsatz kommt, wenn der primäre Standardalgorithmus als unsicher eingestuft wird. Damit wird eine Krypto-Agilität sichergestellt [5].

Kryptografische Bibliotheken

Das Validierungstool nutzt für kryptografische Funktionen der ISO 15118-2 die nativ erhältliche Bibliothek in .NET Framework 4.7. Diese verwendet intern die kryptografische Bibliothek von Windows [2]. Da sich die kryptografischen Algorithmen in der ISO 15118-20 geändert haben, müssen Bibliotheken auf die neuen Anforderungen geprüft werden. Nach einer Analyse stellte sich heraus, dass ausschließlich die Bibliothek BouncyCastle in der Version 1.9.0 alle Anforderungen erfüllt und demnach in dem Validierungstool für Funktionen der ISO 15118-20 verwendet werden kann.

Ausblick

Die ISO 15118-20 gewährt besseren Schutz vor Angriffen. Dies wurde im Rahmen dieser Arbeit untersucht. Sicherheit ist jedoch immer ein Schutz auf Zeit. Kryptografische Algorithmen und Verfahren bleiben nicht ewig sicher. Standards müssen regelmäßig erneut auf ihre Sicherheit bewertet werden. Es darf nicht vergessen werden, dass die ISO 15118-20 noch sehr jung ist. Dementsprechend ist unklar, wie genau die eingesetzten Technologien praktisch im Feld umgesetzt werden.

Literatur und Abbildungen

[1] Danny Bogner. ISO 15118 - Security Analysis, Cryptographic Verification and Development of a Crypto Validation Tool, 2019.

[2] Microsoft Corporation. .NET cryptography model. https://docs.microsoft.com/en-us/dotnet/standard/security/cryptography-model, 12 2021.

[3] Eigene Darstellung.

[4] International Organization for Standardization. *ISO 15118-2:2014-04 Road vehicles - Vehicle to grid communication interface - Part 2: Network and application protocol requirements.* International Organization for Standardization, 2014.

[5] International Organization for Standardization. *ISO 15118-20:2022-04 Road vehicles - Vehicle to grid communication interface - Part 20: 2nd generation network layer and application layer requirements.* International Organization for Standardization, 2022.

[6] Umweltbundesamt UBA. Stickstoffdioxid-Belastung. https://www.umweltbundesamt.de/daten/luft/stickstoffdioxid-belastung#belastung-durch-stickstoffdioxid, 10 2021.

Konzeption und Implementierung einer ‚AI as a Service' (AIaaS) Plattform

Philipp Dobler **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen

Einleitung

Das Thema Künstliche Intelligenz (KI) gewinnt weltweit immer mehr an Bekanntheit, indem die Medien zunehmend davon berichten. Auch von 79 Prozent der deutschen Unternehmen werde es als maßgebender oder kritischer Erfolgsfaktor angesehen [1]. Die Ursache hierfür ist die enorme Vielfältigkeit einer KI und die mit dessen Einsatz erreichbaren beachtlichen Möglichkeiten. Ein damit einhergehendes Problem bildet jedoch die Tatsache, dass KI-Systeme sehr schnell komplex werden können, wodurch eine Implementierung mit Herausforderungen verbunden ist. Insbesondere wenn ein Unternehmen nicht in der Lage ist, das benötigte Knowhow oder die notwendigen Ressourcen aufzubringen, kann ‚Artificial Intelligence as a Service' (AIaaS) Abhilfe schaffen, da hier Ressourcen und Entwicklungswerkzeuge cloudbasiert zur Verfügung gestellt werden. Die Nachfrage nach dem Service wird voraussichtlich stark wachsen. Bis zum Jahr 2025 werde der weltweite Markt für diesen Service auf 77 Milliarden US-Dollar ansteigen, woraus sich anhand des Vergleichs zu den im Jahre 2017 erreichten 2,4 Milliarden US-Dollar eine jährliche Wachstumsrate von 56,7 Prozent ergebe [5].

Problemstellung

Bei der IT-Designers GmbH setzen sich vermehrt Projekte durch, bei denen das Thema Künstliche Intelligenz ein wichtiger Bestandteil bildet. Jedoch hat die beste KI keinen Mehrwert, wenn sie aufgrund unzureichender Ressourcen nicht hinreichend verarbeitet werden kann. Somit ist das Aufsetzen einer KI auf einzelnen, kleineren Computern ausgeschlossen. AIaaS stellt daher einen geeigneten Lösungsansatz dar. Um über verschiedene Künstliche Intelligenzen einen Überblick zu behalten und eine zentrale Anlaufstelle zu schaffen, ist eine Plattform notwendig, die über mehrere KI-Systeme verfügt und diese in organisierter Weise bereitstellt.

Zielsetzung

Das Ziel der Bachelorarbeit besteht nicht nur darin, eine soeben beschriebene Plattform zu entwickeln und den Aufwand hinsichtlich der Hinzufügung eines neuen KI-Modells zur Plattform zu minimieren, sondern es soll auch der Erstellungsvorgang einer dazugehörigen Benutzeroberfläche vereinfacht werden. Darüber hinaus existiert das Ziel, dass die letztendliche Plattform dem Entwickler den Prozess vom KI-Modell zum Frontend vereinheitlicht und dem Endbenutzer verschiedene Funktionalitäten sowie die Möglichkeit zur Speicherung von Ergebnissen anbietet. Dabei soll trotz Einbindung verschiedener KI-Systeme ein stimmiges anwenderfreundliches Design resultieren. Um eine möglichst zweckmäßige Plattform zu entwickeln, besteht eine wesentliche Aufgabe darin, verschiedene Lösungsansätze in Betracht zu ziehen.

Lösungsansätze und Technologien

Die IT-Designers GmbH verwendet bei KI-gestützten Themen hauptsächlich die Programmiersprache Python. Daher erweist es sich als sinnvoll, direkt hier anzusetzen, um den Entwickler in seinem Workflow nicht zu unterbrechen. Als passende All-in-one-Lösung ist das Open-Source-App-Framework Streamlit hinsichtlich der Funktionen vielversprechend, bekannt und auch beliebt. Auf GitHub hat in den letzten Jahren kein anderes Framework dieser Art einen derartig starken Beliebtheitszuwachs bekommen [6]. Streamlit erlaubt eine schnelle Erstellung eines Frontends direkt aus einem Python-Modul heraus. Demnach ist bei diesem Ansatz keine Separierung zwischen Backend und Frontend notwendig, wodurch auch keine Schnittstelle zur Kommunikation zwischen den eigentlich vorhandenen beiden Schichten ausgearbeitet und gepflegt werden muss. Die wichtigsten Frontend-Komponenten werden vom Framework zur Verfügung gestellt und können mit eigen erstellten Komponenten ergänzt werden. Daher ließ sich anfangs der Bachelorarbeit die Hypothese

aufstellen, dass sich die Plattform ideal mithilfe von Streamlit entwickeln lässt. In der Praxis deutete alles auf die Verifizierung der Hypothese hin, da sich das Framework, wie zu erwarten, hervorragend für ein schnelles Erstellen einfacher Benutzeroberflächen für einzelne KI-Modelle eignet. Mit der Zeit stellte sich jedoch immer mehr heraus, dass Streamlit für die Umsetzung der Ziele der Bachelorarbeit nicht infrage kommt. Hauptgründe hierfür sind unausweichliche Performanzprobleme und Einbußen sowohl in der Funktionalität als auch Flexibilität. Diese Defizite würden sich zwar mithilfe von Workarounds herabmindern lassen, haben sich jedoch im Rahmen der Bachelorarbeit als unpraktisch, ineffizient oder sogar als unsicher herausgestellt.
Aus dem gegebenen Anlass fand eine Umwidmung zu einem Backend-Frontend-Lösungsansatz statt, da hier ein Entwickler bei Bedarf die volle Kontrolle behält und somit Anforderungen besser umsetzen kann. Im Backend wird weiterhin auf die Programmiersprache Python gesetzt, um mit dem Workflow eines KI-Entwicklers zu harmonieren. Dabei fiel die Entscheidung auf FastAPI – ein Webframework zum Entwickeln von RESTful-APIs. Ein Grund für dieses Framework stellt die als sehr nützlich angesehene automatisch generierte interaktive API-Dokumentation dar. Des Weiteren weist FastAPI laut mehrerer Quellen und Benchmarks von TechEmpower eine sehr hohe Performanz auf. Ein beispielhafter Benchmark vom August 2021, der die beliebtesten Python-Backend-Frameworks beinhaltet, wird in Abbildung 1 dargestellt.

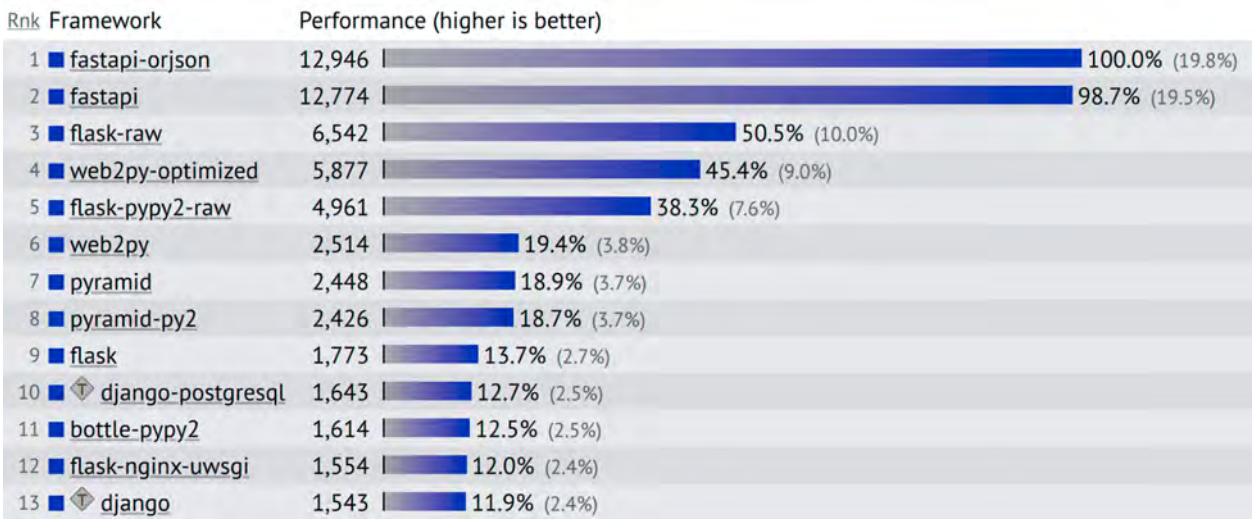

Abb. 1: Leistungsbenchmark basierend auf Antworten pro Sekunde bei 20 Abfragen pro Anfrage [3]

Beim Frontend wird auf die JavaScript-Softwarebibliothek React gesetzt. Hierfür gibt es einige Gründe. Zu den wichtigsten gehören dessen größere Verbreitung und Beliebtheit. Die Anzahl der Downloads der npm-Pakete bestätigen dies und können Abbildung 2 entnommen werden. Des Weiteren spricht für React dessen angenehme Lernkurve sowie dessen Leichtgewichtigkeit gegenüber Angular [2].

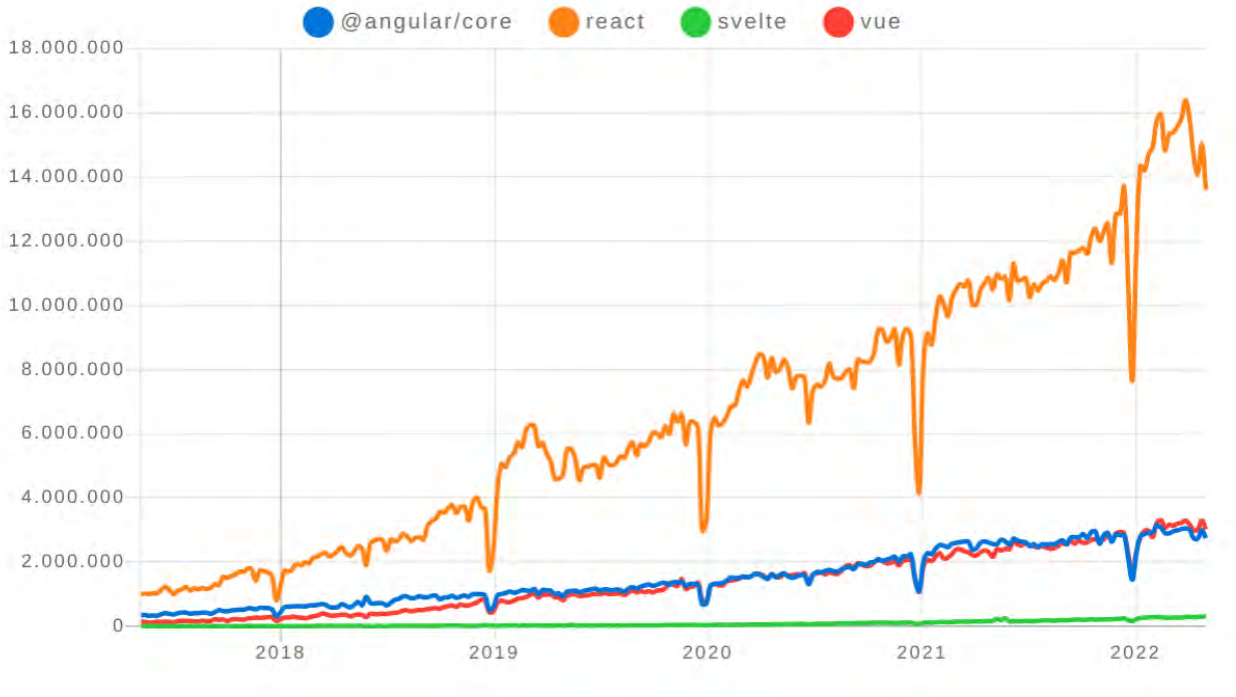

Abb. 2: Vergleich der Anzahl der npm-Paket-Downloads von Angular, React und Vue [4]

Ausblick

Mit der Zuwendung zum FastAPI-Backend und React-Frontend eröffnen sich zahlreiche Möglichkeiten bezüglich der Funktionalität und der Benutzerfreundlichkeit. Mit diesem Schritt nimmt jedoch auch die allgemeine Komplexität und somit auch der Aufwand für einen Entwickler zu, wenn neue KI-Modelle und neue Benutzeroberflächen der Plattform hinzugefügt werden. Diese Komplexität soll wieder auf einen akzeptablen Grad reduziert werden, indem fertige Komponenten zur Verfügung gestellt und Generatoren entwickelt werden, die den Prozess vereinfachen, vereinheitlichen und automatisieren. Des Weiteren sollen auch all die im Kapitel ‚Zielsetzung' genannten Ziele umgesetzt werden.

Literatur und Abbildungen

[1] Deloitte Deutschland. KI-Studie 2020: Wie nutzen Unternehmen Künstliche Intelligenz? https://www2.deloitte.com/de/de/pages/technology-media-and-telecommunications/articles/ki-studie-2020.html, 2021.

[2] Krusche und Company GmbH. Welches JavaScript-Framework ist im Jahre 2022 das Richtige für Sie? https://kruschecompany.com/de/angular-vue-jquery-react-oder-ember/, 01 2022.

[3] TechEmpower Incorporated. Round 20 results - Web Framework Benchmarks. https://www.techempower.com/benchmarks/#section=data-r20&hw=ph&test=query&l=zijzen-sf&a=2&f=zhb2t3-yyku7z-v2qcjj-zik0zj-tys5bz-zik0zj-zik0zj-hr89a7-zik0zj-v2qiv3-zik0zj-cn3, 08 2021.

[4] John Potter. npm trends - @angular/core vs react vs svelte vs vue. https://www.npmtrends.com/@angular/core-vs-react-vs-svelte-vs-vue, 05 2022.

[5] Allied Market Research. AI as a Service Market Size, Technology | AIaaS Market Forecast - 2025. https://www.alliedmarketresearch.com/artificial-intelligence-as-a-service-aiaas-market, 09 2018.

[6] Markus Schmitt. Streamlit vs. Dash vs. Shiny vs. Voila vs. Flask vs. Jupyter. https://www.datarevenue.com/de-blog/streamlit-vs-dash-vs-shiny-vs-voila-vs-flask-vs-jupyter, 2021.

Implementierung verschiedener Modelle der Zeitreihenanalyse als Prognose-Web-Anwendung

Marcel Englisch **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Datinex GmbH, Göppingen

Motivation

Die Vorhersage der finanziellen Entwicklung und auch der Lagerentwicklung ist vermutlich jedem Unternehmen ein wichtiges Anliegen. Sie dient dazu einen Überblick zur Situation über das nächste Jahr oder auch über mehrere Jahre im voraus zu bekommen. Bekannte Planwerte sind üblicherweise Umsatz, Gewinn und auch Ausgaben. Eine möglichst genaue Vorhersage dieser Werte ermöglicht nicht nur Sicherheit sondern zeigt den möglichen Spielraum für unternehmerische Entscheidungen. Während Wirtschaftskrisen, Pandemien oder Kriege nicht oder kaum vorhersehbare Einflussfaktoren auf die wirtschaftliche Lage sind, lassen sich in einer üblich voranschreitenden Situation die eigenen Planwerte durchaus gut prognostizieren.
Wie kann nun ein Unternehmen ohne Statistik-Fachpersonal eine Vorhersage aus den Verläufen der Vergangenheit treffen?
Um eine Prognose der zukünftigen Entwicklung zu errechnen gibt es mehrere Möglichkeiten. Eine davon ist nur auf Basis der in der Vergangenheit gemessenen Werte zu arbeiten. Am Beispiel des Umsatzes bedeutet das nur aus den vergangenen Umsatzwerten die Zukünftigen zu prognostizieren. Durch die Anwendung und Vergleich verschiedener Verfahren kann so das Optimale für den eigenen Bedarf gefunden werden.

Ziele

Die Bachelorarbeit befasst sich mit der Entwicklung einer einfach zu bedienende Webanwendung mit deren Hilfe es möglich sein soll, auch ohne tiefe Mathematische oder Statistischen Kenntnisse eine Prognose zu erstellen. Die daraus entstehende Software dient als Grundlage für die Entwicklung eines eventuell neuen Produkts der Datinex GmbH. Für die Umsetzung wird Python für die Serverseite und die üblichen Webtechnologien HTML, CSS und JavaScript für das Frontend verwendet. Das Backend auf dem Server übernimmt dabei die Verarbeitung der Daten und prognostiziert bis zu einem gewünschten Horizont voraus. In der Arbeit werden nur eindimensionale Daten behandelt. So werden monatlich, wöchentlich oder täglich erhobene Daten ohne Beachtung weiterer Faktoren betrachtet. Dies wird auch als skalare Zeitreihe bezeichnet [2]. Diese Vorhersage der Software kann danach manuell durch Informationen entsprechend angepasst werden. So könnte Beispielsweise der Verkaufsstart eines neuen Produkts den Umsatz steigern, die bevorstehende Marktsättigung ihn schmälern oder der Kauf neuer Maschinen die Ausgaben erhöhen und den Gewinn mindern. Diese planbaren Faktoren wird die Software nicht miteinbeziehen.

Zeitreihenanalyse

Zeitreihen sind eine zeitliche Abfolge von Daten wie beispielsweise der monatliche Umsatz. Bevor aus einer Reihe von Daten eine Prognose für den zukünftigen Verlauf erstellt werden kann braucht es die Analyse. Bei der Analyse geht es darum die Zeitreihe in ihre einzelnen Komponenten bestehend aus Trend, Konjunktur, Saison und den Rest aufzuteilen [3]. Aus einem einzelnen Messwert lässt sich keine Information gewinnen in welchen Verhältnissen diese Komponenten vorhanden sind. Erst wenn die einzelnen Messwerte zusammen als Zeitreihe betrachtet werden lassen sich diese Bestandteile herausrechnen.

Verfahren

Die für die Implementierung ausgewählten Verfahren sind *Exponentielle Glättung*, *Auto Regressive Integrated Moving Average* (kurz *ARIMA*) und das in 2017 von Facebook erstmals veröffentlichte *Prophet*.
Dabei ist Keine per se besser als ein Anderes. Wie passend ein Verfahren ist hängt immer auch von der Form der Zeitreihe selbst ab. Die Qualität eines Modells zu einer Zeitreihe kann mittels verschiedener Fehlerindikatoren ermittelt werden. Ein solcher Qualtätsindikator ist die *mittlere quadratische Abweichung* (kurz *MSE* für das Englische *mean squared error*) [2].

Die Prognose wird auf die Vergangenheit angewendet, sogenannten *In-Sample-Prediction*, und ermöglicht so den Vergleich zwischen ihr und den echten Daten.
Grundsätzlich ist die Software für Firmen konzipiert. Da die Methoden für andere Datensätze die Gleichen sind, sind sie auch für andere Messdaten (zum Beispiel Temperatur, Anzahl der Regentage pro Monat, etc.) geeignet um Prognosen zu erstellen. Die Software ist also nicht an Unternehmenstypische Daten gebunden.

Realisierung

Die Rahmenbedingungen der Implementierung sind durch das Unternehmen vorgegeben. Die Serverseite ist, aufgrund bereits vorhandenen Wissens über die Sprache, in Python mit der Verwendung des Django-Frameworks umgesetzt. Für die Verfahren ARIMA und Exponentielle Glättung ist die Programmbibliothek *statsmodels* geeignet, da sie beide Verfahren zur Verfügung stellt. Bei Prophet gibt es aktuell nur die Implementierung von Facebook selbst, weshalb diese dafür verwendet werden muss. Speziell für ARIMA in statsmodels wird über die Wrapper-Bibliothek *pmdarima* aufgerufen. Vereinfacht ausgedrückt testet dieser Wrapper verschiedene Parameter die für das ARIMA-Modell möglich sind und wählt nach einem Kriterium das am besten passende Modell aus.
Alle diese Programmbibliotheken geben das Modell beziehungsweise die Vorhersage in unterschiedlichen Formaten zurück. Damit keine Konvertierung im Frontend auf der Client-Seite nötig ist, werden die Daten auf dem Server in ein einheitliches Format gebracht. Der Client erhält also unabhängig vom ausgewählten Verfahren die Daten immer im gleichen JSON-Format.
Die Darstellung im Webbrowser muss dort per JavaScript realisiert werden. Dort werden die JSON-Daten nochmals leicht für die verwendete Bibliothek *billboard.js* angepasst, die dann wiederum mittels *D3.js* das Diagramm erzeugt.

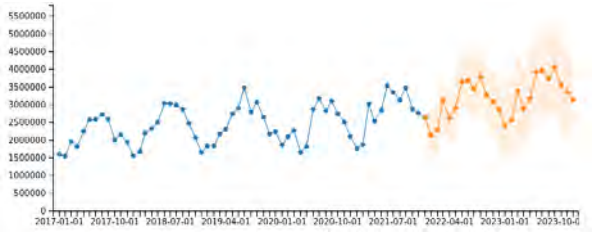

Abb. 1: Beispieldaten und Prognose mittels Exponentieller Glättung [1]

Abbildung 1 zeigt ein Diagramm mit Messwerten und einer Zweijahresprognose, die von der entwickelten Software generiert worden ist. Die dort verwendet Daten enthalten sowohl eine Saison als auch einen leichten Aufwärtstrend.

Ausblick

Die grundlegendsten Funktionen sind implementiert und die Software kann bereits über die Benutzeroberfläche Daten einlesen um daraus eine zukünftige Werte eines beliebig weiten Zeitraums prognostizieren. Die Voraussetzung ist momentan eine im CSV-Format vorliegende Datei, in der die Daten als monatliche Abfolge gespeichert sind. Geplant ist noch eine flexiblere Importroutine, sodass auch täglich oder wöchentlich erhobene Daten analysiert werden können. Des weiteren soll die Oberfläche auch für Personen nutzbar sein, die sich nicht oder nur sehr wenig mit Zeitreihenanalyse auskennen, aber trotzdem damit arbeiten möchten. Hierzu soll die Auswahl des Modells automatisiert werden. Die benutzende Person importiert nur die Daten, woraufhin die Software aus allen drei Verfahren das Passende zurückgibt. Dazu rechnet die Software mit allen implementierten Modellen und wählt anhand eines noch zu definierenden Wertes - zum Beispiel den *Mean Square Error* - das Modell mit dem geringsten Fehler aus.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Manfred Deistler and Wolfgang Scherrer. *Modelle der Zeitreihenanalyse*. Birkhäuser, 2018.

[3] Josef Puhani. *Statistik : Einführung mit praktischen Beispielen*. Springer Gabler, 2020.

Der Digitale Zwilling als 3D Karte dynamischer Umgebungen für die Lokalisierung

Kenan Ercan **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Virtual Automation Lab, Hochschule Esslingen, Sascha Röck

Motivation

Flexible Fertigungen der Zukunft erlauben eine beliebige Abfolge von Produktionsschritten und erfordern einen gleichermaßen flexiblen Teiletransport zwischen Maschinen und Anlagen. Zur Begegnung dieser Herausforderung untersucht das Virtual Automation Lab (VAL) der Fakultät Maschinen und Systeme den Einsatz von autonomen Indoor-Flugrobotern. Diese können sich in dem meist ungenutzten Luftraum der Fertigungsstätte bewegen und ermöglichen einen Teiletransport mit hoher Dynamik. Um Mensch und Maschine nicht zu gefährden, muss eine kollisionsfreie Navigation sichergestellt werden. Voraussetzung hierbei ist eine zuverlässige Indoor-Lokalisierung. Im VAL wird eine Lokalisierungsmethode auf Grundlage eines Digitalen Zwillings sowie wenigen Distanzmessungen erforscht.

Ein Digitaler Zwilling ist eine digitale Repräsentation eines Objekts aus der realen Welt. Im eingesetzten Fall basiert dieser auf dreidimensionalen Objektrepräsentationen aus dem Entwicklungsprozess der Produktionsanlage. Bewegte Maschinenobjekte werden kinematisiert und deren Position mit Maschinen- und Sensordaten im Digitalen Zwilling in Echtzeit aktualisiert. Der Digitale Zwilling dient in der Lokalisierung als detaillierte und stets aktuelle a priori Karte. Damit ein mit minimaler Sensorik ausgestatteter Flugroboter in einer a priori Karte lokalisiert werden kann, wird die probabilistische Monte Carlo Lokalisierung (MCL) eingesetzt. Dieser Algorithmus schätzt die Pose des Flugroboters anhand von Bewegungs- und Sensordaten in einer gegebenen Karte. Als Pose wird die Kombination von Position und Orientierung des Flugroboters bezeichnet. Die MCL repräsentiert die Pose als Partikelwolke, in der jedes Partikel eine mögliche Pose darstellt. Zur Bewertung der Pose eines Partikels werden Distanzmessungen, entsprechend den Distanzsensoren auf dem Flugroboter, im Digitalen Zwilling ausgehend von der jeweiligen Partikelposition durchgeführt. Der Algorithmus gewichtet die Partikel nach der Wahrscheinlichkeit, mit der sich der Roboter an der Pose eines Partikels befindet. Ist die Wahrscheinlichkeit zu gering, wird das Partikel verworfen und durch eine wahrscheinlichere Pose ersetzt. Dieser Vorgang wird mit jeder Distanzmessung durchgeführt, wodurch die Partikelwolke mit der Zeit zur tatsächlichen Position des Flugroboters konvergiert. [9] Abbildung 1 zeigt den Digitalen Zwilling und eine Partikelwolke nach mehreren Distanzmessungen.

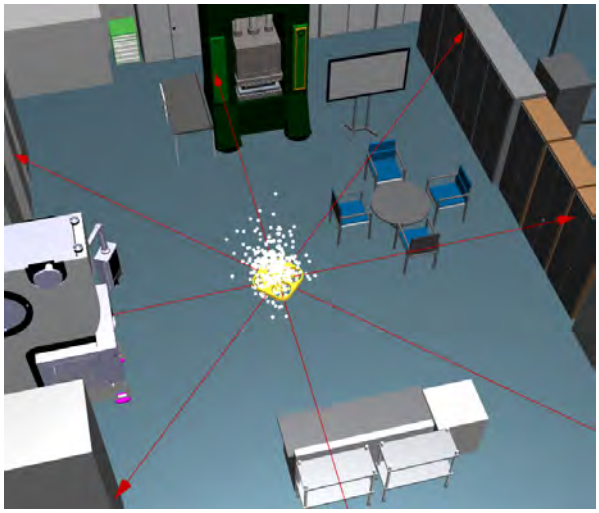

Abb. 1: Digitaler Zwilling mit Partikelwolke der MCL [3]

Um Lasersensoren zur Abstandsmessung im Digitalen Zwilling zu simulieren, werden Raycaster verwendet. Raycaster sind Strahlen, die Schnittpunkte mit der Umgebung berechnen. Die Karte der Umgebung stammt aus dem Engineering und liegt in Form eines CAD-Modells vor. Aufgrund der hohen Detaillierung des CAD-Modells ist die Berechnung der Schnittpunkte mit der Umgebung sehr rechenaufwendig. Um eine hohe Updatefrequenz der Partikelwolke für die Indoor-Lokalisierung zu erreichen, muss die Schnittpunktberechnung weiter beschleunigt werden.

Im Rahmen dieser Arbeit wird der Stand der Forschung und Technik im Bereich der Kartenrepräsentationen untersucht. Weiter soll der Digitale Zwilling um Algorithmen zur Beschleunigung des Raycastings erweitert werden, sodass eine hohe Updatefrequenz für eine echtzeitfähige Lokalisierung ermöglicht wird.

Kartenrepräsentationen in der Robotik

In der Robotik werden hauptsächlich topologische und geometrische Karten eingesetzt. Topologische Umgebungsrepräsentationen werden typischerweise als Graphen mit Beziehungen zwischen Objekten dargestellt. Sie speichern keine metrischen Distanzinformationen und eignen sich daher nicht für die Lokalisierung. Geometrische Modelle hingegen speichern die Position von Umgebungsmerkmalen in einem gegebenen Koordinatensystem und ermöglichen somit Abstandsmessungen innerhalb der Karte. [8] Gängige geometrische Modelle sind Punktwolken und Occupancy Grid Maps. Punktwolken visualisieren Endpunkte von Abstandsmessungen und kommen hauptsächlich im Simultaneous Localization And Mapping (SLAM) Verfahren zum Einsatz. Eine wesentliche Herausforderung dieser Kartenrepräsentation ist der Umgang mit dynamischen Objekten und Sensorrauschen. [2]
Bei Occupancy Grid Maps muss im Gegensatz zu den Punktwolken das Ausmaß der Karte bekannt sein. Occupancy Grid Maps teilen die Umgebung in ein vordefiniertes zwei- oder dreidimensionales Gitternetz ein und bilden die Belegung einzelner Zellen ab. Die Belegung wir probabilistisch abgebildet. Dies ermöglicht eine Berücksichtigung von Sensorungenauigkeiten bei der Kartenerstellung. [4] Elevation Maps sind eine weitere Form von Grid Maps. Hierbei wird in einem zweidimensionalen Gitter die Höhe der Hindernisse gespeichert. Die daraus resultierende Karte modelliert nur eine Ebene der Umgebung und wird daher auch als 2,5-dimensionale Karte bezeichnet. Elevation Maps werden zur Repräsentation von Terrain verwendet. [1]
OctoMaps sind dreidimensionale Occunpancy Grid Maps, die auf einer Octree-Baumstruktur basieren. [6] In der Computergrafik werden Octrees eingesetzt, um dreidimensionale Daten hierarchisch zu unterteilen. Diese Baumstruktur ermöglicht es, Objekte bei der Schnittpunktberechnung frühzeitig zu eliminieren. Schneidet ein Strahl einen Knoten nicht, müssen dessen Kindknoten nicht auf Schnittpunkte überprüft werden. Eine weitere Baumstruktur zur Darstellung geometrischer Objekte ist die Bounding Volume Hierarchy (BVH). Bei dieser Baumstruktur bilden geometrische Objekte die Blätter des Baums und werden von Hüllkörpern umgeben. Diese Hüllkörper werden iterativ zu größeren Hüllkörpern zusammengefasst, sodass eine Baumstruktur entsteht. Abbildung 2 zeigt, wie der BVH Algorithmus aus einer vereinfachten Szene eine Baumstruktur bildet. [7] Baumstrukturen eignen sich für die Repräsentierung von Karten, die aus dem Engineering stammen. CAD-Modelle modellieren Objekte mit einem Polygonnetz. Je nach Detailgrad und Größe kann ein Modell aus Tausenden bis Millionen Polygonen bestehen. Die frühzeitige Eliminierung von Objekten, unter der Verwendung von Beschleunigungsstrukturen wie die BVH, kann in diesem Fall die Anzahl der notwendigen Schnittpunktberechnungen deutlich reduzieren. [5]

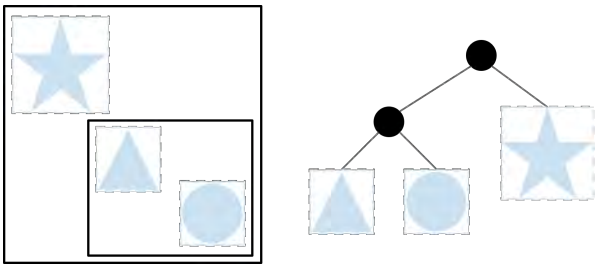

Abb. 2: BVH einer simplen Szene [3]

Ausblick

Diese Abschlussarbeit befindet sich aktuell in den ersten Wochen der Bearbeitung. In einem nächsten Schritt werden analoge Probleme in der Umsetzung von Videospielen untersucht, die eine hohe Anzahl von Kollisionen in komplexen Umgebungen berechnen. Nach Abschluss der Recherche zum Stand der Forschung und Technik sollen verschiedenen Beschleunigungsstrukturen verglichen werden. Der Fokus soll dabei auf der Rechendauer des Raycastings sowie dem Umgang mit polygonbasierten geometrischen Repräsentationen liegen.
Umgebungskarten enthalten meist nur die statischen Elemente der Umgebung. Während in statischen Umgebungen aus Sicht des Roboters nur dieser selbst die Pose ändert, können sich in dynamischen Umgebungen auch weitere Hindernisse bewegen. Die ausgewählte Kartenrepräsentation und die dazugehörigen Beschleunigungsstrukturen sollen effizient um dynamische Objekte erweitert werden.

Literatur und Abbildungen

[1] Sunglok Choi, Jaehyun Park, Eulgyoon Lim, and Wonpil Yu. Global path planning on uneven elevation maps. In *2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 49–54. IEEE, 2012.

[2] D.M. Cole and P.M. Newman. Using laser range data for 3D SLAM in outdoor environments. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 1556–1563. IEEE, 2006.

[3] Eigene Darstellung.

[4] A Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22:46–57, 1989.

[5] Bernd Fröhlich. Beschleunigungsstrukturen für echtzeitfähiges Ray-Tracing auf aktueller Hardware-Infrastruktur. https://www.uni-weimar.de/de/medien/professuren/medieninformatik/vr/research/real-time-rendering/beschleunigungsstrukturen-fuer-echtzeitfaehiges-ray-tracing/, 2022.

[6] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. OctoMap: an efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 34:189–206, 2013.

[7] Thomas Larsson and Tomas Akenine-Möller. A dynamic bounding volume hierarchy for generalized collision detection. *Computers & Graphics*, 30:450–459, 2006.

[8] Roland Stenzel. Steuerungsarchitekturen für autonome mobile Roboter, 2002.

[9] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. The MIT Press, 2005.

Analyse von Machine Learning Verfahren zur Anomalie-Erkennung in Log-Daten für eine automatisierte Überwachung von Software-Systemen

Simone Falzone **Gabriele Gühring**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma AEB SE, Stuttgart

Einleitung & Motivation

Eine hohe Verfügbarkeit moderner Software-Systeme wird immer wichtiger und ist ein kritischer Faktor, um konkurrenzfähig zu bleiben. Hierfür ist es notwendig, die Systeme sorgfältig zu überwachen und Probleme frühzeitig zu erkennen. Eine Möglichkeit, Software-Systeme zu überwachen, bieten Log-Daten.

Log-Daten dienen hauptsächlich dazu, die inneren Systemzustände zu beschreiben und wichtige Ereignisse zu protokollieren, um die Fehlersuche bei Systemausfällen zu erleichtern und gegebenenfalls eine Ursachenanalyse durchzuführen [1]. Damit bieten Log-Daten eine wichtige Grundlage, um das Verhalten und den Zustand eines Software-Systems zu analysieren. Durch immer komplexer werdende Software-Systeme steigt die Anzahl an Log-Einträgen rapide an. Die große Anzahl an Log-Einträgen erschwert eine manuelle Analyse der Log-Daten, um Fehler oder Störungen frühzeitig erkennen zu können. Eine automatisierte Auswertung durch Machine Learning Verfahren, kann es ermöglichen, die Log-Daten in Echtzeit zu überwachen und so anomales Verhalten von Software-Systemen frühzeitig zu erkennen.

Die AEB SE entwickelt Software im Bereich Versand, Logistik und Außenhandel. Die Produkte werden als Software as a Service angeboten (SaaS) und im hauseigenen Rechenzentrum betrieben. Durch immer strengere Service-Level Agreements (SLA) mit Kunden ist es für die AEB SE essenziell, Ausfallzeiten zu reduzieren und System-Fehler frühzeitig zu erkennen. Für die Überwachung der Software setzt die AEB SE ein zentralisiertes Logging-System ein, welches alle Applikations-Logs aggregiert. Das Logging -System basiert auf dem EFK-Stack. Das Akronym EFK steht für die drei Komponenten Elasticsearch, Fluentd und Kibana. Fluentd dient der Aggregation und Normalisierung der Log-Daten der einzelnen Systeme, Elasticsearch dient der Persistierung und Kibana der Analyse und Visualisierung. Dieses Logging-System dient derzeit für statistische Analysen der Log-Daten. Des Weiteren dient es der manuellen Fehlersuche, was aufgrund der wachsenden Anzahl an Log-Daten immer zeitaufwendiger wird.

Problemstellung & Zielsetzung

Durch die stetig steigenden Datenmengen und auch die stetig wachsende Anzahl an betriebenen Software-Systemen stellt sich die Frage, ob und in welchem Rahmen Machine Learning Verfahren den Betrieb bei der Analyse der Log-Daten unterstützen kann. Hierfür müssen die Log-Daten analysiert werden und geeignete Machine Learning Modelle ausgewählt und untersucht werden. Die Kernaufgabe der Arbeit ist es, geeignete Verfahren zu finden, die es ermöglichen den Betrieb frühzeitig auf anomales Verhalten der Software-Systeme zu informieren.

Ansatz & Grundlagen

Die Grundidee in der Anomalie-Erkennung von Log-Daten besteht darin, zu versuchen, das normale Verhalten von Software-Systemen durch Log-Daten zu lernen. In Abbildung 1 sind die dafür notwendigen Schritte aufgelistet.

Log Collection: Die Log Collection ist der erste Schritt und stellt das Sammeln der erstellten Log-Einträge dar. Durch das oben beschriebene zentralisierte Logging-System der AEB SE ist dieser Schritt schon erfüllt.

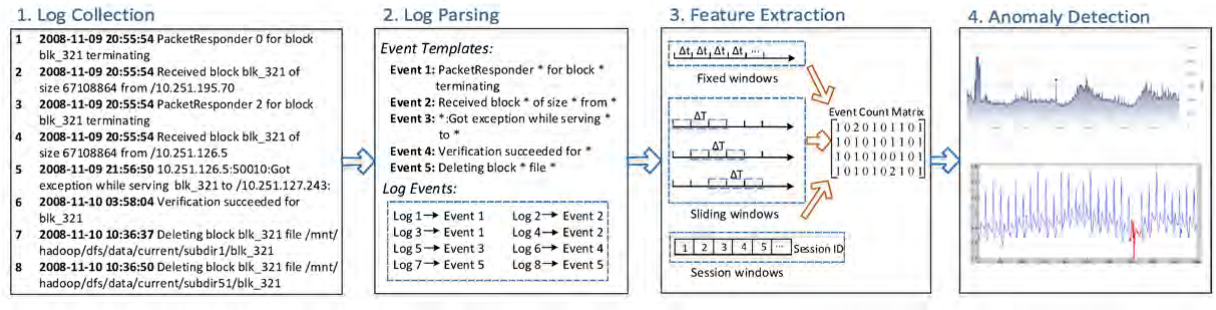

Abb. 1: Schritte für die Anomalie-Erkennung in Log-Daten [2]

Log Parsing: Log-Einträge sind in der Regel unstrukturierte Daten in Textform. Dabei bestehen sie meist aus einem fixen und einem variablen Teil. Der fixe Teil eines Log-Eintrags wird von einem Entwickler im Quellcode definiert und verändert sich somit nicht. Der variable Teil eines Log-Eintrags hingegen wird dynamisch zur Laufzeit eines Systems generiert, dies können beispielsweise Port-Nummern oder IP-Adressen sein. Ziel des Log Parsing ist es, den variablen Teil von dem fixen Teil zu trennen. Bei diesem Vorgang entsteht ein Log-Template, welches den fixen Teil der Log-Nachricht darstellt und eine Event-Id. Ein Beispiel kann man der Abbildung 1 entnehmen, dort wird aus dem ursprünglichen Log-Eintrag:
Received block blk_321 of size 67108864 from /10.251.195.70
Folgendes Log-Template generiert:
*Received block * of size * from ** mit der Event-Id *Event 2*
Durch das Log-Parsing wird jedem Log-Eintrag die passende Event-Id zugeordnet, zu sehen in Abbildung 1.
Feature Extraction: Nachdem die Log-Einträge geparst sind, müssen sie noch in eine maschinenlesbare Repräsentation überführt werden, sodass sie von Machine Learning Modellen verarbeitet werden können. Dafür werden die Log-Einträge in Log-Sequenzen unterteilt. Anschließend wird für jede Log-Sequenz ein Event-Count-Vektor erstellt. Der Event-Count-Vektor zählt, wie oft jeder Log-Eintrag innerhalb einer Log-Sequenz vorkommt. Diese Event-Count-Vektoren werden zu einer Event-Count-Matrix zusammengeführt und stellen somit die Eingabe für die Machine Learning Modelle dar [2].
Anomaliy Detection: Die im vorherigen Schritt konstruierte Event-Count-Matrix wird dazu verwendet, verschiedene Machine Learning Modelle [4], [3] zu trainieren, um so ein Modell zur Anomalie-Erkennung zu erhalten. Das trainierte Modell ist dann in der Lage zu erkennen, ob es sich bei einer neuen Log-Sequenz um eine Anomalie handelt oder nicht [2].

Ausblick

Durch die stetig steigende Anzahl an Log-Daten wird es immer schwieriger und unpraktikabeler, diese manuell zu auszuwerten. Deshalb soll mit dieser Arbeit analysiert werden, ob sich Machine Learning Verfahren eignen Log-Daten zu analysieren und automatisiert Anomalien zu erkennen. Hierfür werden verschiedene geeignete Modelle ausgewählt und trainiert. Da keine gelabelten Daten vorliegen, werden die Ergebnisse der Modelle in Zusammenarbeit mit Mitarbeitern der AEB SE ausgewertet, um die Qualität der Modelle zu bewerten. Durch dieses Vorgehen können die Daten auch sukzessiv gelabelt werden, wodurch das Training und die Evaluation zukünftiger Modelle vereinfacht werden können.

Literatur und Abbildungen

[1] Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar. DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1285–1298. Association for Computing Machinery, 2017.

[2] Shilin He, Jieming Zhu, Pinjia He, and Michael Lyu. *Experience Report: System Log Analysis for Anomaly Detection*. IEEE Computer Society, 2016.

[3] Qingwei Lin, Hongyu Zhang, Jian-Guang Lou, Yu Zhang, and Xuewei Chen. Log clustering based problem identification for online service systems. In *Proceedings of the 38th International Conference on Software Engineering Companion*. Association for Computing Machinery, 2016.

[4] Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael I. Jordan. Detecting Large-Scale System Problems by Mining Console Logs. In *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*. Association for Computing Machinery, 2009.

Konzeption und prototypische Implementierung einer echtzeitfähigen Datenverarbeitungsarchitektur im Kontext von Verkehrsflussdaten

Patrick Fauth **Jörg Nitzsche**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Viele Verkehrsteilnehmer kennen die Problematik, dass man oft im Verkehr im Stau steht oder fast immer bei jeder Ampel anhalten muss. Diese Bachelorarbeit soll mithelfen, dass in Zukunft bei solchen Verkehrsbedingung eine Verkehrsflussoptimierung stattfinden kann.
Mit der Entwicklung der Mobilfunktechnologie 5G in Deutschland sowie mit dem voranschreitenden Ausbau der entsprechenden Mobilfunkabdeckung werden neue Einsatzmöglichkeiten dieser Technologie realisierbar.
Im Straßenverkehr beispielsweise bietet die hohe Datenübertragungsrate mittels 5G neue Möglichkeiten, den Verkehrsfluss zu optimieren. Dazu werden alle relevanten Verkehrsdaten durch eine Vielzahl verschiedenartiger Sensoren, wie beispielsweise Videokameras und Bewegungsmelder, erfasst.
Die technische Herausforderung besteht dabei darin, die so gesammelte Masse an Daten möglichst schnell und fehlerfrei zu verarbeiten, sodass die ermittelten Informationen beispielsweise an eine Ampelsteuerung rechtzeitig weitergegeben werden können, um den Verkehrsfluss auf Grundlage dieser Daten zu optimieren. Wichtig ist, dass die Daten fehlerfrei und rechtzeitig an die Ampelsteuerung übermittelt werden, sodass eine falsche Reaktion die zu Unfällen führen kann, verhindert wird.
Dies soll mithilfe von echtzeitfähigen Datenverarbeitungs-Technologien realisiert werden, die unter Open Source Lizenz verfügbar sind. Unter der Echtzeitfähigkeit eines Systems versteht man die Fähigkeit, in einer gegebenen Betriebsumgebung alle anstehenden Aufgaben und Funktionen, unter allen Betriebszuständen, immer rechtzeitig und ohne Ausnahme erledigen zu können [2] . Das kann zeitlich in wenigen Sekunden, Minuten oder auch in Stunden erfolgen und ist dennoch in Echtzeit. Die Bearbeitungsgeschwindigkeit von Echtzeitverarbeitungssysteme ist anwendungsabhängig. Die meisten Echtzeitverarbeitungssysteme verarbeiten die Daten in einer sehr kurzen Zeitspanne.

Zielsetzung

Das Ziel dieser Bachelorarbeit ist es:

- Verschiedene Open Source Echtzeitdatenverarbeitungstools im Hinblick auf ihre Verarbeitungsfähigkeit von Verkehrsflussdaten zu analysieren.
- Die Tools dabei insbesondere auf ihre Geschwindigkeit und Ausfallsicherheit zu evaluieren, da dies zentrale Faktoren für einen möglichen, zukünftigen Einsatz der Technologie im Straßenverkehr sind.
- Basierend auf der durchgeführten Analyse soll eine prototypische Implementierung erfolgen, die aufzeigen soll, inwiefern eine Echtzeitdatenverarbeitung von Verkehrsflussdaten in der Praxis umsetzbar ist und welchen Nutzen diese erzielen kann.

Vorgehensweise

Als erstes muss recherchiert werden, welche der Open Source Echtzeitverarbeitungstools zur Verfügung stehen. Danach müssen Kriterien für einen Vergleich ausgearbeitet werden. Die grundlegenden Anforderungen an die Echtzeitverarbeitungstools ist eine schnelle und fehlerlose Datenverarbeitung sowie eine Skalierbarkeit, im Hinblick auf eine spätere Erweiterung des Systems. Danach werden die Architekturen miteinander verglichen und geeignete Frameworks für diese ausgewählt. Die Architektur eines solchen Echtzeitverarbeitungssystems lässt sich annähernd wie in Abbildung 1 gezeigt darstellen.

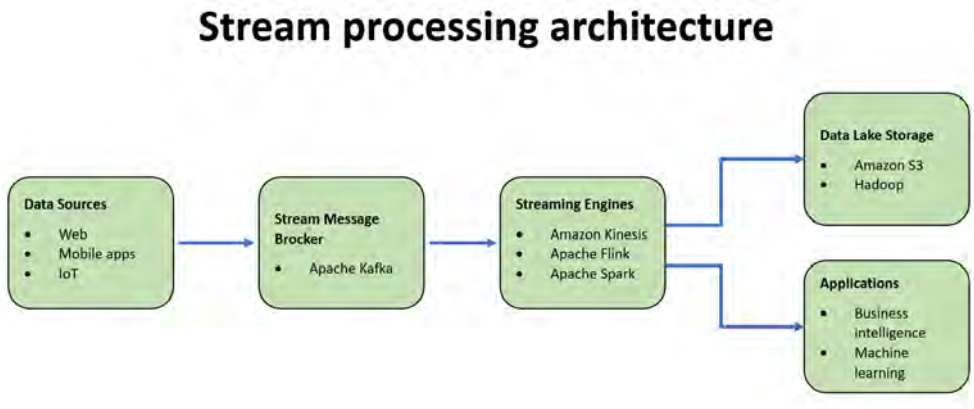

Abb. 1: Beispiel einer echtzeitfähigen Datenverarbeitungsarchitektur [1]

Die Daten werden hierbei über das 5G Mobilfunknetz gesendet und anschließend, beispielsweise durch Apache Kafka und Apache Flink, in Echtzeit analysiert und weiterverarbeitet. Im hier dargestellten Falle werden diese verarbeiteten Daten zur Ampelsteuerung weitergeleitet. Des Weiteren werden die Daten zur Durchführung nachträglicher Analysen archiviert. Der Prototyp soll in diesem Zusammenhang aufzeigen, welche der Open Source Technologien für eine solche Echtzeitdatenverarbeitung geeignet sind.
Was noch zu berücksichtigen ist, wird in Abbildung 2 erkennbar. An einer Kreuzung sind viele Sensoren installiert, was bei hohem Verkehrsaufkommen zu einer dementsprechend hohen Datenmenge führt, die in diesem Umfang bei unzureichenden technischen Voraussetzungen nicht rechtzeitig analysierbar wäre.

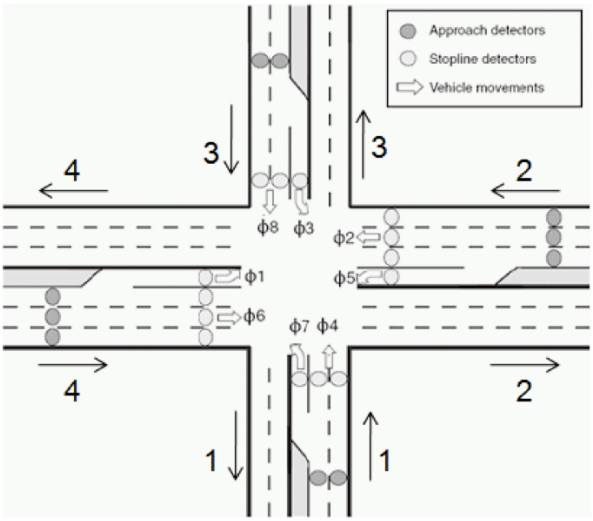

Abb. 2: Straßenkreuzung mit Sensoren [3]

Dies verdeutlicht, dass die Machbarkeit einer Analyse des Straßenverkehrs (z.B. einer ganzen Stadt) in Echtzeit, zum Zwecke der Optimierung des Verkehrsflusses, davon abhängig ist, ob das für diese Aufgabe eingesetzte System eine solch große Menge an gewonnenen Daten zuverlässig verarbeiten kann.

Ausblick

Da sich die Bachelorarbeit zum Zeitpunkt des Verfassens dieses Artikels noch in seiner Anfangsphase befindet und dementsprechend noch wenige Verkehrsflussdaten in Echtzeit zur Verfügung stehen, werden die zu analysierenden Datensätze zu Testzwecken eigenständig simuliert.
Mit dem Ergebnis dieser Arbeit soll eine Plattform implementiert werden, welche die tatsächlichen gesammelten Verkehrsdaten in Echtzeit verarbeitet, analysiert und anschließend damit den Verkehrsfluss optimiert.

Literatur und Abbildungen

[1] Angelehnt an George Lawton. Big data streaming platforms empower real-time analytics. https://www.techtarget.com/searchdatamanagement/feature/How-to-build-an-effective-streaming-data-architecture, 05 2020.

[2] Peter Scholz. *Softwareentwicklung eingebetteter Systeme Grundlagen, Modellierung, Qualitätssicherung*. Springer, Berlin, Heidelberg, 2005.

[3] Pravin Varaiya. PointQ model of an arterial network: calibration and experiments. https://www.researchgate.net/figure/Standard-intersection-with-four-approaches-and-four-departures_fig1_280589911, 2015.

Umsetzung der topografischen Landschaft und des Straßennetzes der Stadt Esslingen in VR

Ben Feucht **Reinhard Schmidt**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Durch den Hype in der Medienbranche hat sich die Virtual Reality-Technologie in den letzten Jahren stark entwickelt. So werden VR-Systeme nicht mehr nur zur Unterhaltung eingesetzt, sondern auch in der Industrie, Wissenschaft, Bildung und Medizin. Es können Welten von weit vergangenen oder weit entfernten Orten erschaffen und besucht und reale Orte virtualisiert werden [2]. Das gleiche Ziel verfolgt auch die hier vorgestellte Arbeit. Sie ist Teil des Projekts „Esslingen 3D“, in welchem ein Großteil der Stadt Esslingen virtuell erstellt wird und über VR zugänglich sein soll. Die Aufgabe dieser Arbeit ist die Erstellung der Umgebung ohne Gebäude sowie des Straßennetzes von Esslingen. Zusätzlich soll eine Methode entwickelt werden, bei welcher Teile der Welt ausgeblendet werden um die Rechenleistung zu reduzieren.

Realistische Umsetzung von Esslingen

Die virtuelle Stadt komplett manuell zu erstellen, würde den zeitlichen Rahmen dieser Arbeit deutlich übersteigen. Deswegen sollte so viel wie möglich von der Stadt mithilfe von realen Daten generiert werden. Dieser Ansatz wurde anfangs für die Höhenkarte sowie das Straßennetz der Stadt gewählt. Um mit den Höheninformationen von Esslingen arbeiten zu können, mussten sie früher oder später in Unity importiert werden. Nach einiger Recherche war kein Datentyp für öffentlich zugängliche Höheninformationen mit Unity kompatibel. So musste auf ein Add-on von Blender zurückgegriffen werden, mit welchem die Höhenkarte und das Straßennetz von ausgewählten Ausschnitten der Weltkarte erstellt werden können. Nach kleinen Verbesserungen konnte die Landschaft dann in Unity importiert werden. Das Straßennetz, welches durch das Add-On erstellt wurde, war zwar recht genau positioniert, jedoch wurde jede Straße nur als flache Ebene erstellt (siehe Abbildung **??**). Da jede Straße in der realen Welt verschiedenes Material, verschiedene Breite und Anzahl an Spuren hat, war es ungeeignet, dieses Objekt unverändert für ein immersives VR-Erlebnis zu benutzen. Die Umgestaltung der Straßen würde einen zu hohen Aufwand mit sich bringen, wodurch eine andere Lösung gesucht wurde. Hier wurde klar, um die Details und Qualität zu kreieren, die eine virtuelle und möglichst immersive Welt benötigt, darf sich nicht voll auf jeden zu findenden Automatisierungsprozess verlassen werden. Stattdessen müssen bestimmte Objekte manuell platziert werden, um der Realität möglichst nahe zu kommen. Als Referenz werden hier jegliche öffentlich zugängliche Satellitenbilder wie zum Beispiel Google Maps verwendet.

Gaia / GeNa

Um nicht komplett auf die automatische Generierung von Landschaft und Stadt verzichten zu müssen, wird das Unity Add-On GeNa von Procedural Worlds benutzt. GeNa ist ein Level-Design System, welches die Erstellung von großen Welten vereinfachen soll. Das funktioniert, indem die zeitaufwändige Arbeit, die bei der Virtualisierung einer ganzen Stadt entsteht, vom Add-On übernommen wird. Mit einem extra Tool für Straßen und Flüsse, können diese ganz einfach mithilfe von Splines auf die Welt gezeichnet werden. Im nachhinein können dann Merkmale wie Größe oder Textur der Straßen und Flüsse geändert werden.

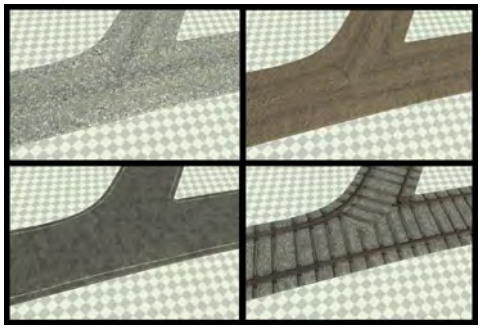

Abb. 1: Straße mit verschiedenen Texturen [1]

Mit dem Carve Tool hat man die Möglichkeit, sie dann optimal in die Umgebung zu integrieren, indem die Landschaft mit verschiedenen Einstellungsmöglichkeiten entlang der Straßen und Flüsse erhöht, herabgestuft oder anderweitig angepasst werden kann. Bei dieser Arbeit sorgt GeNa dafür, dass trotz gewissen Automatisierungsprozessen, die Genauigkeit erhalten bleibt und der gesamte Prozess deutlich verschnellert werden kann. In GeNa enthalten, sorgt Das Landschafts-Design-System Gaia für eine realistische Landschaft, um die von GeNa generierten Objekte optimal in die Welt einzubringen. Die Generierung von Gaia funktioniert, indem der Nutzer einen Bereich in der Landschaft auswählt. Alle Naturobjekte, die später in der Generierung enthalten sein sollen, sowie Einstellungsmöglichkeiten wie Dichte der Bäume, kann der Nutzer hier ebenfalls ändern. Mit diesen Informationen generiert Gaia dann eine realistische Landschaft. Mit Gaia könnte bei dieser Arbeit zum Beispiel die entfernte Umgebung von Esslingen ohne viel Aufwand schnell generiert werden können, da der Nutzer diese nur von weitem sieht und sie deshalb nicht so genau dargestellt sein muss wie der Rest der Stadt.

Abb. 2: Mit GeNa generierte Straße [3]

Handhabung einer großen Welt

Um einen möglichst hohen Wiedererkennungswert zur realen Stadt und eine hohe Immersion zu haben, sollte die virtuelle Welt sehr detailliert und hochauflösend sein. Damit die detailreiche Welt technisch umgesetzt werden kann, kann nicht die ganze Landschaft auf einmal in Hochauflösung angezeigt werden. Teile der Karte, die weit weg vom Charakter sind, werden komplett ausgeblendet. Um das zu erreichen, wird die Karte in mehrere Teile geschnitten. Für jede Teilmitte wird der Abstand zum Nutzer gemessen. Wenn der Nutzer eine bestimmte Entfernung zum Teil hat, wird das Teil ausgeblendet. So wird nie die komplette Karte angezeigt und die zu erbringende Rechenleistung reduziert.

Ausblick

Auch wenn die Karte schon teilweise ausgeblendet wird, könnte dieser Prozess noch erweitert werden, indem zum Beispiel die komplette Landschaft, die nicht in der Kamera zu sehen ist, sich also hinter dem Nutzer befindet, zusätzlich ausgeblendet wird. Ein weiteres Problem der aktuellen Vorgehensweise ist, dass weit entfernte hohe Punkte auf der Karte auch ausgeblendet werden, obwohl sie durch die Höhe in der echten Welt zu sehen wären. Eine Lösung dafür wäre, immer eine niedrig aufgelöste Version dieser hohen Punkte einzublenden obwohl sie weit weg sind. Für das ganze Projekt „Esslingen 3D“ bedeutet diese Arbeit eine Grundlage für mögliche Erweiterungen der Welt mit anderen nahen Städten, sowie zusätzliche Funktionen wie Verkehr oder mehrere Nutzer.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Philip Hammer. Virtual Reality: Die Erschaffung neuer Welten. https://www.zukunftsinstitut.de/artikel/virtual-reality-die-erschaffung-neuer-welten/, 09 2016.

[3] Procedural Worlds. GeNa Pro. https://www.procedural-worlds.com/products/professional/gena-pro/, 2022.

Entwicklung einer Lebenszeichenüberwachung mit Hilfe eines FMCW Radarsensors unter Berücksichtigung der Alarmnorm IEC 60601-1-8

Thomas Gaenzle **Clemens Klöck**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma seleon GmbH, Heilbronn

Einleitung

Aufgrund der aktuellen demografischen Entwicklung in vielen westlichen Gesellschaften erhöht sich das Durchschnittsalter der Bevölkerung. Die beiden Hauptgründe hierfür sind eine rückläufige Geburtenrate bei gleichzeitig steigender Lebenserwartung aufgrund des medizinischen Fortschrittes. [4] Der dadurch steigende Anteil an Senioren in der Gesellschaft verursacht eine höhere Belastung des Pflegesektors. Eine der wichtigsten Aufgaben der Pflegekräfte ist hierbei die Überwachung der Vitalfunktionen ihrer Patienten. Dafür werden in den meisten Fällen Sensoren verwendet, die am Körper getragen werden müssen, d.h. entweder muss der Patient selbst den Sensor anlegen und für eine ausreichende Batterieladung sorgen oder es entsteht eine zusätzliche Belastung für die Pflegekraft. Mit Hilfe von modernen Radarsensoren aus dem Automobilbereich, sogenannte *Frequency Modulated Continouus Wave* (FMCW) - Radare, lassen sich Bewegungen berührungslos im mm-Bereich messen. Durch die Messung der Brustkorbbewegung, welche durch Atmung und Herzschlag verursacht wird, lässt sich mit entsprechender Signalverarbeitung auf die Frequenzen dieser beiden Größen schließen. Im Gegensatz zu anderen Messverfahren besteht hier weder die Notwendigkeit, den Sensor am Körper zu tragen, noch müssen regelmäßig Batterien gewechselt werden, denn das Radar kann direkt an das Stromnetz angeschlossen werden. Da bei einer Radarmessung keine Bilddaten, sondern maximal Punktwolken ausgegeben werden, kann damit eine Patientenüberwachung erfolgen, ohne die Privatsphäre der Patienten zu verletzen.

Ziel der Arbeit

Das Ziel dieser Arbeit ist die Konzeption und prototypische Entwicklung einer Überwachung der Vitalzeichen einer schlafenden Person in Echtzeit. Über- oder Unterschreiten die Lebenszeichen der zu überwachenden Person fest definierte Grenzwerte, so erfolgt eine optische und akkustische Alarmausgabe nach der Norm *DIN EN 60601-1-8* für Alarmsysteme. Dafür soll in dieser Arbeit eine Signalverarbeitung implementiert werden, die eine Erkennung von Atem- und Herzrate aus den Radardaten ermöglicht. Bei der Sensorik, die hierfür genutzt wird, handelt es sich um den FMCW Radarsensor *AWR1642* der Firma Texas Instruments.

FMCW Radar

Das Grundprinzip der Radartechnologie basiert auf elektromagnetischen Wellen, welche von Objekten reflektiert und mit Hilfe der Radarantennen empfangen werden. Beim FMCW Radar handelt es sich um ein aktives Radar, d.h. es werden Radarwellen ausgesendet und die reflektierten Wellen werden wieder empfangen.

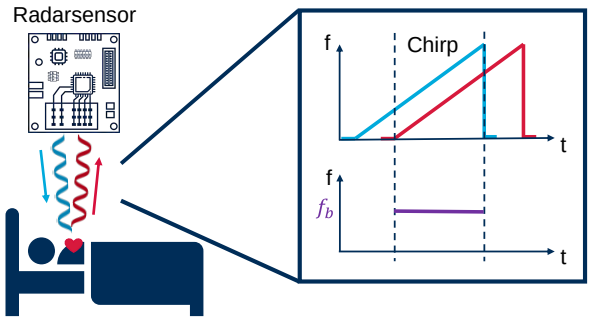

Abb. 1: Schematischer Messaufbau - Messung einer schlafenden Person mit einem FMCW Radarsensor [2]

Die Besonderheit dieses Radartyps ist die kontinuierliche Modulation der Frequenz der Radarwellen, sodass über der Zeit betrachtet, deren Frequenz linear ansteigt. Diese sogenannten *Chirps* werden dann vom Radar ausgesendet und mit den reflektierten Wellen gemischt. Die Frequenz f_b (beat frequency) des gemischten

Signals ist konstant und proportional zum Abstand des Objektes, von dem die Welle reflektiert wurde. Typischerweise wird das empfangene Signal für jeden ausgesandten Chirp mit Hilfe eines Analog-Digital-Wandlers abgetastet. Anschließend wird eine Fast Fourier Transformation über die Chirp Samples ausgeführt, um die dominanten Frequenzen bestimmen zu können. Diese Frequenzen entsprechen den Objekten vor dem Radar und ihre Amplituden sind abhängig vom Rückstrahlquerschnitt des Objektes sowie vom Abstand zum Radar. Die Distanzauflösung des Radars wird durch die Lichtgeschwindigkeit und die Bandbreite des konfigurierten Chirps vorgegeben. Bewegt sich das Objekt von einer Messung zur nächsten innerhalb dieses Auflösungsfensters (*Range Bin*), bleibt die zugehörige Frequenz f_b dieselbe. Die Phase des Signals verändert sich jedoch abhängig von der Position des Objektes innerhalb des Range Bins. Diese Eigenschaft wird für die Erfassung der Lebenszeichen genutzt, da hiermit Bewegungen im mm-Bereich zuverlässig erkannt werden können. Bei der Konfiguration des Radars muss hier ein Mittelweg gefunden werden, sodass die Distanzauflösung groß genug ist, damit die Person im gleichen Range Bin bleibt, da sonst die Phasenwerte nicht mehr übereinstimmen. Außerdem muss die Distanzauflösung klein genug sein, um über die Phase eine ausreichende Auflösung zu erhalten, damit die Bewegung durch den Herzschlag noch zuverlässig erkannt wird.

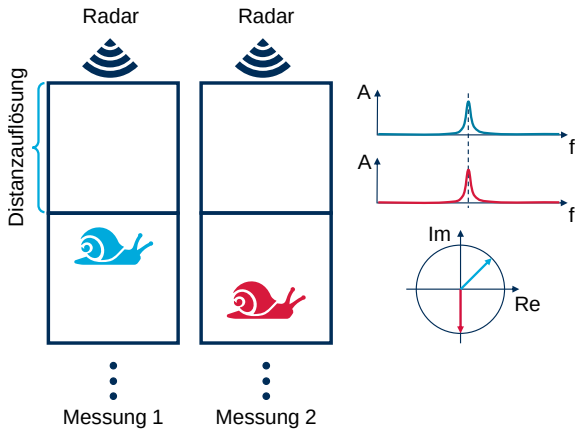

Abb. 2: Auswirkung von Bewegungen innerhalb der Distanzauflösung des Radars auf Frequenz und Phase [2]

Erkennung der Lebenszeichen

Sowohl Atmung als auch Herzschlag verursachen eine Körperbewegung. Die größte Amplitude lässt sich hierbei am Brustkorb einer Person feststellen: [5]

- Atmung: 1 - 12 mm
- Herzschlag: 0.1 - 0.5 mm

Zur Bestimmung der Vitalzeichen aus dem Radarsignal wird in einem ersten Schritt der Range Bin ermittelt, in dem sich die Person befindet. Dafür wird innerhalb festgelegter Grenzwerte der Range Bin mit der maximalen Amplitude ausgewählt. Die gesamte restliche Auswertung bezieht sich nur noch auf den Phasenwert dieses Range Bins über die Zeit, d.h. aus jedem ankommenden Signal wird der Phasenwert dieses Range Bins extrahiert. Die durch die Atmung verursachte Amplitude ist im Vergleich zur Amplitude durch den Herzschlag deutlich größer und lässt sich entsprechend einfacher detektieren. Dafür wird die ermittelte Phasenänderung mit einem Bandpass gefiltert um anschließend im gefilterten Signal die Anzahl der Signalspitzen zu zählen. Die Werte für die untere und obere Frequenz des Bandpassfilters betragen 0.1 Hz und 0.6 Hz.
Für die Ermittlung der Herzfrequenz wurde ein anderer Filteransatz gewählt. Hier kommt eine Kombination aus Tiefpass, Hochpass und gleitendem Mittelwert zum Einsatz. Die Grenzwerte des Tief- und Hochpassfilters beruhen auf dem elektrischen Signal, das für die Muskelkontraktion des Herzens und damit auch für die Bewegung des Brustkorbes verantwortlich ist, dem sogenannten *QRS - Komplex*.

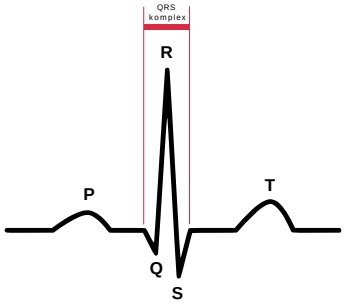

Abb. 3: QRS Komplex [1]

Die Dauer dieses Signals liegt zwischen 0.06 s und 0.12 s, woraus sich die Grenzen für die Filter von 8 und 17 Hz ergeben. Als erstes werden mit dem Tiefpassfilter alle Störfrequenzen unterhalb 8 Hz herausgefiltert und vom Phasensignal abgezogen. Die Frequenzen oberhalb 17 Hz werden mit dem Hochpass gefiltert und ebenfalls vom Signal abgezogen. Anschließend wird das gefilterte Signal mit einem gleitenden Mittelwertfilter geglättet und stabilisiert. Um aus diesem Signal die Herzfrequenz zu bestimmen, werden die Stellen mit der größten Steigung gesucht, da diese der *R-Zacke* im QRS-Komplex entsprechen. Aus der Anzahl der R-Zacken im Herzsignal bzw. der Anzahl der Signalspitzen

im Atemsignal werden dann abschließend die Atem- und Herzfrequenz bestimmt.

Alarmnorm 60801-1-8

Die Norm *60801-1-8* beschreibt den Aufbau und die Anforderungen an das Alarmsystem medizinischer Systeme. Mit Hilfe einer Risikomatrix wird eine Risikoanalyse durchgeführt und die Alarmbedingungen im System werden festgelegt. Eine Alarmbedingung beschreibt den Grund, aus dem ein bestimmter Alarm ausgelöst werden soll, wie z.B. das Unterschreiten eines Grenzwertes der Herzfrequenz für einen vorgegebenen Zeitraum. Jede dieser Bedingungen ist abhängig von ihrer Einordnung in der Risikomatrix an einen Alarm einer bestimmten Priorität gebunden. Wird eine der Alarmbedingungen erfüllt, so wird ein optisches und/oder akkustisches Signal am Gerät ausgegeben. [3]

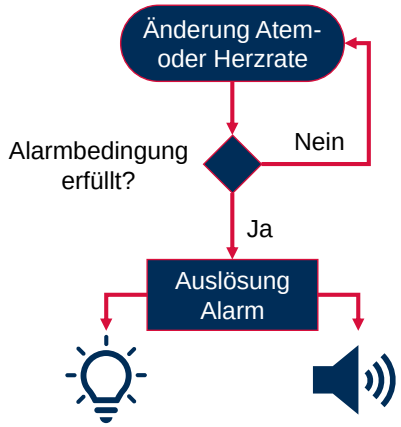

Abb. 4: Ablaufplan der Alarmausgabe [2]

Hierfür soll ein allgemeingültiges Framework implementiert werden, welches Projekt-unabhängig eingesetzt werden kann.

Ausblick

Nach Fertigstellung der Signalverarbeitung erfolgt ein Vergleich mit den Signalverarbeitungsmethoden von Texas Instruments in Hinblick auf die Genauigkeit. Für die Ermittlung der tatsächlichen Herzrate wird hier eine Pulsuhr von Garmin verwendet, die nach dem Prinzip der Photoplethysmographie aufzeichnet. Im Anschluss daran werden mögliche Optimierungen der Signalverarbeitung untersucht, um die Genauigkeit noch weiter zu verbessern. Mit der Anbindung an das Alarmframework können erste Tests als ganzheitliches System durchgeführt werden.

Literatur und Abbildungen

[1] Anthony Atkielski. Schematic diagram of normal sinus rhythm for a human heart as seen on ECG. https://commons.wikimedia.org/w/index.php?curid=7875780, 2009.

[2] Eigene Darstellung.

[3] Verband der Elektrotechnik Elektronik Informationstechnik. *DIN EN 60601-1-8*. IEEE, 2021.

[4] Rina Goldenberg. Geburtenrate sinkt, Deutschland überaltert. https://p.dw.com/p/3gDA8, 07 2020.

[5] Texas Instruments. Driver Vital Signs - Developer's Guide. https://dev.ti.com/tirex/explore/node?node=AAZ.3LdjgB9ICpyXqcY3zA__AocYeEd__LATEST, 2017.

Redesign infolge veränderter Marktanforderungen am Beispiel eines elektronischen Überwachungsgeräts für Kälteanlagen

Simon Gaubatz **Hermann Kull**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Leitenberger GmbH, Kirchentellinsfurt

Einleitung

Kein Produkt und keine Leistung, die eine Firma anbietet, ist bei Markteinführung perfekt. Durch die Nutzung des Produkts ergeben sich Verbesserungen und Ideen, die im Vorfeld nicht abzusehen sind. Je länger ein Produkt auf dem Markt ist, desto mehr Kunden geben Feedback und wollen ein für den Kunden angepasstes Produkt. Deshalb muss jede Firma ihre Produktpalette ständig weiter entwickeln oder für einzelne Produkte ein Redesign machen, damit das Produkt oder die angebotene Leistung weiter attraktiv bleibt. Verpasst die Firma den Zeitpunkt zur Weiterentwicklung so wird das Produkt von anderen Firmen, die ein ähnliches Produkt anbieten vom Markt verdrängt, da es die Kundenwünsche besser abdeckt. Die Herausforderung eines Redesigns ist, alle konkurrierende Anforderungen zu betrachten und die relevanten Anforderungen auszuwählen und umzusetzen. Es entsteht dabei ein Gestaltungsraum, der zugleich auch ein Entscheidungsraum darstellt [2]. Um das Potential eines Produkts und die an es gestellten Anforderungen zu ermitteln, gibt es das Anforderungsmanagement. Dabei gibt das Anforderungsmanagement Methoden und Modelle an die Hand, um Anforderungen zu finden, sie anschließend zu formulieren und zu validieren. Die Firma Leitenberger ist seit vielen Jahren erfolgreich in der Mess- u. Regeltechnik vertreten. Ein Teil davon ist die Überwachungstechnik an Kälteanlagen. Das Smart Pressure Gate ist das am Markt erfolgreich Gerät in dieser Sparte. Es vereint das Hoch- und Niederdruckmanometer, den Doppeldruckschalter zur Steuerung des Kompressors und den Hochdruckeinzeldruckschalter zur Steuerung von Ventilatoren, in einem Gerät. Dadurch wird das Risiko von Undichtigkeit an den einzelnen Messeinrichtungen und die gesamte Montagezeit reduziert.

Zielsetzung

Das Ziel dieser Bachelorarbeit ist es das Smart Pressure Gate, das zur Überwachung von Kälteanlagen zum Einsatz kommt neu zu entwickeln. Für das Redesign des Smart Pressure Gates werden mithilfe des Anforderungsmanagements alle Anforderungen an das Gerät ermittelt. Anschließend werden die relevanten Anforderungen ausgewählt und das Gerät wird überarbeitet.

Vorgehen

Für die Entwicklung eines Produkts in der Kältetechnik ist es wichtig die Grundlagen zu kennen. Deshalb ist eine Einarbeitung in das Thema notwendig, um beurteilen zu können, was dort überwacht werden kann, wo Gefahren erkannt und reduziert werden können. Auch eine Einarbeitung in das Thema Anforderungsmanagement ist nötig, um herauszufinden woher Anforderungen kommen, welche Quellen zur Verfügung stehen, wie sie ermittelt und präzise formuliert werden können. Im Anforderungsmanagement teilt sich das Ermitteln von Anforderungen in vier Phasen Abb. 1 auf.

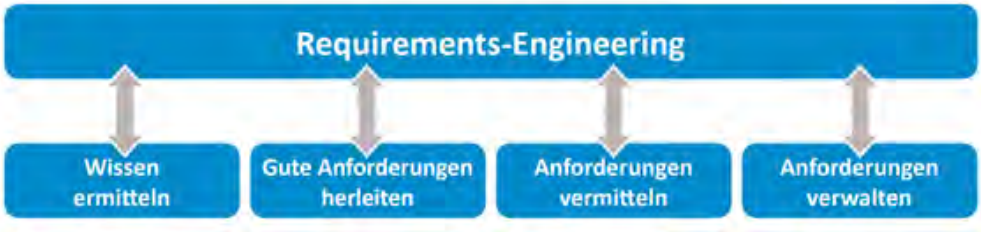

Abb. 1: Haupttätigkeiten im Anforderungsmanagement [1]

In der ersten Phase werden verschiedene Quellen gesucht aus denen Anforderungen ermittelt werden können. Das Smart Pressure Gate dient dabei als Hauptquelle, da es viele Funktionalitäten beinhaltet, die auch für das neue Gerät relevant sind. Deshalb wird die Funktionalität und der Stand der Technik umfassend dokumentiert. Es wird eine kleine Marktanalyse durchgeführt die darauf abzielt, die verschiedenen Konkurrenzprodukte zu ermitteln und zu vergleichen. Auch eine Kundenumfrage wird durchgeführt in der nach den Wünschen der Kunden gefragt wird. Aus dem gesammelten Wissen werden Anforderungen für das Display, das Gehäuse, die Schnittstelle, die Sensorik, die Elektronik und die Bedienung ermittelt. Anschließend werden diese dokumentiert und im Nachgang validiert. Es wird eine Auswahl getroffen welche Anforderungen umsetzbar sind und welche verworfen werden. So entsteht ein Anforderungsdokument, das die Kundenwünsche und ihre Umsetzbarkeit beinhaltet. Sind die Anforderungen fertig dokumentiert betrachtet die Bachelorarbeit einige Technologien, die für Umsetzung der Anforderungen in Frage kommen. Die Bachelorarbeit schließt mit einer Empfehlung welche Technologien für die genannten Bereiche verwendet werden sollen, damit die Anforderungen am besten umgesetzt werden können.

Ausblick

Die Arbeit ist der erste Schritt im Redesign eines Produkts und legt den Grundstein zur Weiterentwicklung des Smart Pressure Gates. Im nächsten Schritt kann mit der Realisierung des Smart Pressure Gates gestartet werden. Mit Hilfe des Anforderungsdokuments und der Empfehlungen, können alle Funktionen entwickelt und überprüft werden. Nach der Entwicklungs- und der Testphase kann das Gerät dann auf den Markt gebracht werden.

Literatur und Abbildungen

[1] Grande Marcus. *100 Minuten für Anforderungsmanagement. Kompaktes Wissen nicht nur für Projektleiter und Entwickler, 2., aktualisierte Auflage*. Springer Vieweg Springer Fachmedien, 2014.

[2] Max Reinecke, Ludger Overmeyer, Rouven Nickel, and Anna M. Schwibode. Global Redesign – eine Methode zur globalisierungsgerechten Produktgestaltung. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 10:655–658, 2007.

Neuartige Schlupfregelung für ein autonom fahrendes Fahrzeug

Pakize Goekkaya **Reiner Marchthaler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Hochschule Esslingen, Esslingen

Einleitung

Um die einwandfreie Funktion eines autonom fahrendes Fahrzeug zu gewährleisten, muss dieses in jeder Situation kontrollierbar und steuerbar sein. Dies gilt insbesondere bei Notfallsituationen, wie zum Beispiel bei einer Notbremsung. Darüber hinaus muss das Fahrzeug auch während der Beschleunigung das Durchdrehen der Reifen und damit einen Kontrollverlust vermeiden. In der Industrie existieren für die beiden Anwendungsfälle bereits die passenden Lösungen. Für die gezeigten Situationen wird ASR (Antriebsschlupf-) und ABS (Antiblockiersystem)-Reglungen verwendet. [2]
Beide Regelungen versuchen den Schlupf zu minimieren. Der Schlupf ist die Geschwindigkeitsabweichung mechanischer Elemente oder Flüssigkeiten, die unter Belastung in Reibungskontakt miteinander stehen. In diesem Fall ist der Schlupf der Geschwindigkeitsunterschied zwischen Reifen und Fahrbahn. In anderen Worten ist der Schlupf die Verknüpfung der Geschwindigkeit des angetriebenen Rads zur Geschwindigkeit des nicht angetriebenen Rads. Der Schlupf entsteht generell, sobald sich Antriebs- oder Bremskräfte auf das Rad wirken. Sobald sich das Rad über die Haftgrenze hinaus angetrieben oder gebremst wird, wächst der Schlupf, bis das Rad unkontrolliert durchdreht oder durchrutscht/blockiert. Durch das Vorzeichen (»0 bzw. «0) des Schlupfwertes ist zu erkennen, ob es sich hierbei um einen Antriebsschlupf oder um einen Bremsschlupf handelt. Der Antriebsschlupf wird von der ASR-Regelung reguliert. Der Bremsschlupf dagegen wird mit der ABS-Regelung reguliert. In Abb. 1 ist die Beziehung zwischen dem Reibwert und dem Schlupf dargestellt.

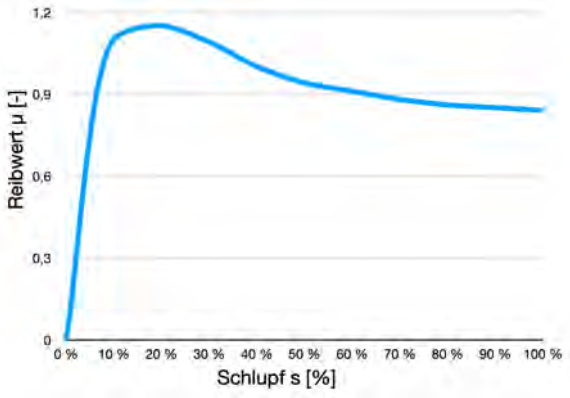

Abb. 1: Reibwert-Schlupfwert [1]

Die Antriebsschlupfregelung (ASR), die auch als Traktionskontrolle bezeichnet wird, verhindert, dass die Räder beim Beschleunigung aus dem Stand eines Fahrzeuges durchdrehen. Die Regelung berechnet kontinuierlich den Schlupfwert. Registriert die ASR-Regelung einen zu großen Schlupf an einem oder mehreren Rädern, reduziert die Regelung das Antriebsmoment auf die Räder auf zwei verschiedene Arten. Im unteren Drehzahlbereich greift sie in das Bremsmanagement ein und im vollen Drehzahlbereich greift sie in die Motorsteuerung ein. Das ASR-System kann jedes Rad einzeln abbremsen. Durch das Abbremsen eines durchdrehenden Rades, kann das Achsdifferenzial dazu gebracht werden, Antriebsmoment auf das jeweils gegenüberliegende Rad zu verlagern. Generell kann das ASR-System nicht verwendet werden, wenn das Fahrzeug schneller fährt, da es die Bremsen zu stark belastet. Stattdessen wird die ASR bei höheren Geschwindigkeiten abgeschaltet.
Das Antiblockiersystem (ABS) wird dafür verwendet, um bei einer Gefahrenbremsung das Blockieren der Räder zu vermeiden. Hierfür wird dafür gesorgt, dass während einer plötzlichen Vollbremsung des Fahrers, die Reifen des Fahrzeugs nicht blockiert werden und stattdessen gleichmäßig abgebremst werden. Die ABS-

Regelung überwacht zunächst wie die ASR-Regelung den Schlupf der einzelnen Reifen. Falls der Schlupf zu groß wird, sodass eines der Räder zu stark belastet wird und zum Blockieren neigt, wird die Bremskraft dieses Rades sofort automatisch reduziert. Der ABS regelt sozusagen den Bremsdruck auf Basis der Haftfähigkeit des Untergrundes. Durch das Antiblockiersystem, kann während einer Vollbremsung, Hindernissen ausgewichen werden, da die Lenkfähigkeit erhalten bleibt. Sobald bei einer Vollbremsung das Bremspedal pulsiert, kann davon ausgegangen werden, dass der ABS eingegriffen hat. Autos ohne ABS können ins Rutschen kommen, wodurch das Fahrzeug unsteuerbar werden kann. [3]

Vorteile

ASR:

- Stabileres Fahrverhalten beim Anfahren, Beschleunigung und Kurvenfahrten auf glatten Fahrbahnen
- Unfallgefahr wird reduziert
- Erhaltung von Vortriebskräfte und Seitenführungskräfte
- Verringerung des Reifenverschleiß

ABS:

- Reduzierung des Reifenverschleiß
- Verkürzung des Bremsweges
- Unterbindung des Einknickens von Fahrzeugkombinationen
- Stabiles Bremsverhalten auf jeder Art von Fahrbahnen [4]

Ziel der Arbeit

Das Ziel dieser Bachelorarbeit liegt darin, eine neuartige Anti-Schlupfregelung für ein autonom fahrendes Fahrzeug zu implementieren. Diese Anti-Schlupfregelung soll eine ASR- und ABS-Regelung enthalten. Beim autonom fahrendes Fahrzeug handelt es sich um ein Versuchsfahrzeug von it:movES. Das Fahrzeug ist in der Abb. 2 zu sehen. Das Fahrzeug verfügt ausschließlich über einen Elektromotor, über welches es beschleunigt und gebremst wird. Der Elektromotor treibt über ein Differenzial die Hinterräder an. Die zentrale Steuerung des Fahrzeugs erfolgt über ein System-on-Modul vom Typ NVIDIA Jetson AGX Xavier. Die Ansteuerung der Sektoren und Aktoren erfolgt über zwei Teensy Mikrocontroller, die über USB mit dem System-on-Modul kommunizieren. Die Anti-Schlupfregelung soll auf einer dieser Teensy Mikrocontroller implementiert werden. Das System-on-Modul schickt Soll-Geschwindigkeitswerte an die Anti-Schlupfregelung. Die Regelung soll anschließend versuchen, die gewünschte Geschwindigkeit zu erreichen, ohne dabei die Haftung der Reifen zu verlieren.

Umsetzung

Der Aufbau der Anti-Schlupfregelung umfasst zwei Komponenten. Als Erstes die Berechnung der Schlupfwerte und als Zweites ein PI-Regler, um die Geschwindigkeit auszuregeln. Für die Berechnung der Reifengeschwindigkeiten verfügt das Versuchsfahrzeug über einen „Wheelspeed-Sensor“ an jedem Rad. Für die Implementierung der Regelung wurde davon ausgegangen, dass die Geschwindigkeit der Vorderreifen stets ungefähr die Geschwindigkeit des Fahrzeugs entspricht. Da in dem Versuchsfahrzeug lediglich die Hinterachse durch einen Elektromotor angetrieben wird, muss lediglich der Schlupf der beiden Hinterräder berechnet werden. Die Berechnung des Schlupfs erfolgt durch den Vergleich der Geschwindigkeiten der Vorder- und Hinterreifen. Der Aufbau des PI-Reglers ist modular gestaltet. Durch den modularen Aufbau des PI-Reglers können die verschiedenen Stufen der ABS- und ASR-Regelungen implementiert werden. Zunächst wurde ein einfacher Regler implementiert, der lediglich die Geschwindigkeitsänderung verlangsamt. Der Nachteil dieses Reglers ist eine schlechte Ausnutzung der maximalen Brems- und Beschleunigungskraft. Außerdem kann der Regler nicht feststellen, ob genügend Haftung noch vorhanden ist oder ob der Schlupf an den Antriebsrädern bereits zu groß geworden ist. Anschließend wurden die Messdaten aufgezeichnet und ausgewertet. Als Letztes wird ein richtiger Regler implementiert. Dieser Regler betrachtet zusätzlich den aktuellen Schlupf und regelt anschließend das Antriebs- und Bremsmoment.

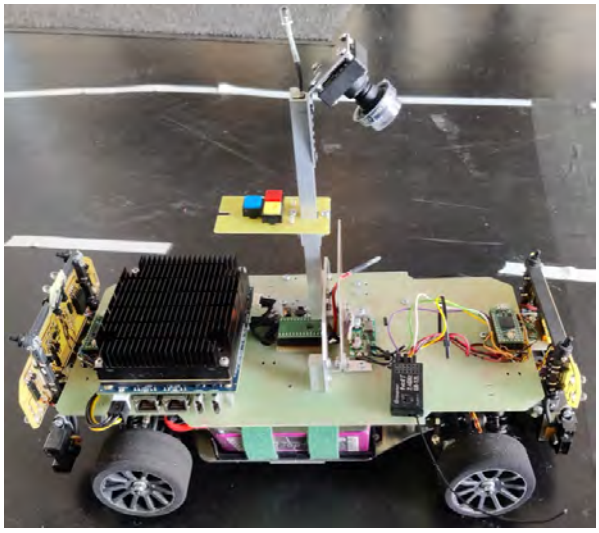

Abb. 2: Versuchsfahrzeug von it:movES [1]

Ausblick

Für die Arbeit sind verschiedene Verbesserungen denkbar. Da die Versuchsfahrzeuge über keine Bremsen verfügen, ist nur die Implementierung einer groben Anti-Schlupfregelung möglich. Mit einzelnen regelbaren Bremsen, wäre die Entwicklung eines effektiveren ABS und ASR möglich. Diese wäre beispielsweise auch auf anspruchsvolleren Straßenverhältnisse, wie zum Beispiel bei Nässe und Schnee verwendbar.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Erich Schindler. *Fahrdynamik: Grundlagen des Lenkverhaltens und ihre Anwendung für Fahrzeugregelsysteme*. expert Verlag, 2007.

[3] Florian Umhöfer. ABS- und ASR-Implementierung in ein ferngesteuertes Modellauto. https://fbmk.h-da.de/fileadmin/documents/Fachbereiche/MK/Forschung/im2s/Projekte/IFP_DJ_03_ABS_und_ASR_Implementierung.pdf, 11 2015.

[4] Vehicle Control Systems WABCO. Das Anti-Blockier-System (ABS) und die AntriebsSchlupf-Regelung (ASR). https://www.wabco-customercentre.com/catalog/docs/8150201943.pdf, 02 2011.

Entwicklung von neuronalen Netzen zur Zustandsbestimmung des Maschinentisches einer Werkzeugmaschine auf Basis von synthetischen Daten

Gabriel Goldschmitt **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Trumpf Werkzeugmaschinen GmbH + Co. KG, Ditzingen

Einleitung

Der Zustand des Maschinentisches einer Trumpf-Lasermaschine beeinflusst Funktion und Schnittergebnis der Maschine. Um eine höhere Zuverlässigkeit und Qualität zu gewährleisten, soll der Zustand des Maschinentisches stetig überwacht werden, um rechtzeitig Gegenmaßnahmen ergreifen zu können.

Zielsetzung

Ziel der Arbeit ist es, auf Basis von synthetisch generierten Daten ein neuronales Netz zu trainieren, welches erkennen kann, in welchem Zustand sich der Maschinentisch befindet. Dabei sollen einzelne Bauteile erkannt und mit den Kategorien „gut, mittel und schlecht“ bewertet werden. Neben guten Erkennungsraten der Bauteile soll auch der Einfluss von Verschleiß, Lichtverhältnissen, Kameraposition und Materialoberflächen auf die Erkennungsraten untersucht werden.

Methode und Vorgehen

Für die Erkennung sollen ‘Convolutional Neural Networks’ eingesetzt werden, die sich in der Bilderkennung etabliert haben. Diese Netze gehören zu der ‘supervised learning’ Kategorie und benötigen somit annotierte Daten [6]. Ein Annotieren muss manuell geschehen, was sowohl fehleranfällig, als auch zeitintensiv ist.
Um dieses Problem zu reduzieren, soll das Modell nicht mit Realdaten, sondern mit synthetischen Bilddaten trainiert werden. Synthetische Daten können prozedural erzeugt werden und haben gegenüber zu Realdaten den Vorteil, dass bekannt ist, in welchem Zustand sich der Maschinentisch befindet.
Dadurch kann der manuelle Aufwand des Annotierens für den Trainings- und Validierungsdatensatz minimiert werden. Die Ergebnisse des Trainings werden mithilfe von Metriken sowie mit ‘Explainable AI’ Techniken bewertet.

Generierung der Bilder

Die Generierung der Bilder wird über bereitgestellte CAD-Modelle und eine 3D-Grafiksuite fotorealistisch ermöglicht (siehe Abbildung 1). In dieser Simulationsumgebung sollen weitere Parameter wie Lichteinfall, Helligkeit, Metalloberfläche oder Verschleiß des Maschinentisches verändert und simuliert werden können. Dies erhöht die Variabilität der Daten, was dem Training zugutekommen soll.

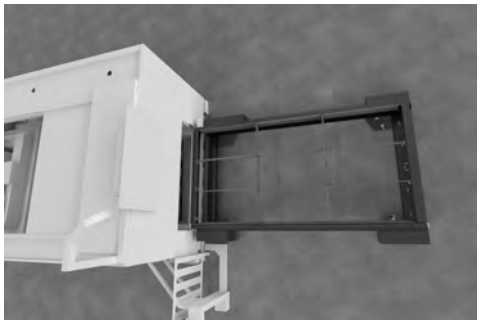

Abb. 1: 3D-Modell einer Trumpf Laserschneidmaschine [2]

Schließen der ‘Domain Gap’ von synthetischen Daten

Neuronale Netze, die ausschließlich auf generierten Daten trainiert werden, haben auf synthetischen Testdaten eine höhere Erkennungsrate als auf Realbildern. Dieser Unterschied wird in der Literatur als ‘Domain Gap’ bezeichnet. [4]
Mit verschiedenen Ansätzen kann versucht werden, diese Lücke zu schließen. Zum Beispiel kann das

zu erkennende Objekt in möglichst verschiedenen Umgebungen dargestellt werden, damit das Netzwerk lernt, Objekte und Hintergründe, die nicht von Interesse sind, auszublenden. Auch das Objekt, das erkannt werden soll, kann mit Texturen dargestellt werden, um das Netz besser auf die Struktur generalisieren zu lassen. Dabei kann teilweise in Kauf genommen werden, dass der Realismus darunter leidet - die Varianz der Daten überwiegt den Verlust des Realismus. [8]
Ein weiterer Fokus sollte auf die Oberfläche und Eigenschaften des Objektes gelegt werden. Diese bestimmen Lichtreflexion und Aussehen des Objektes. Eigenschaften wie Helligkeit, Verschattung und Spiegelungen der Szene beeinflussen diese. [3] Eine metallische, glatte Oberfläche reflektiert beispielsweise das einfallende Licht deutlich besser als eine matte, raue Oberfläche. Diese Materialeigenschaften und Lichtreflexionen können über verschiedene Methoden simuliert werden, die unterschiedlich akkurat sind. Als Beispiel ist die bidirektionale Reflektanzverteilungsfunktion (BRDF) zu nennen [5].
Auch die teilweise Verdeckung des Elementes durch andere Gegenstände oder Verrauschen ist sinnvoll. Dies kann helfen, teilweise verdeckte Elemente zuverlässig zu erkennen. [8]

Explainable AI

Häufig werden neuronale Netze als Blackbox verwendet. Durch 'Explainable AI' Techniken soll der Grund der Entscheidung durch das Neuronale Netz verständlicher gemacht werden. Dabei wird zwischen lokaler und globaler Verständlichkeit unterschieden. Lokal erklärt die Entscheidung anhand einer bestimmten Prädiktion, global das Modell als Ganzes. Die darüber erlangten Erkenntnisse sollen helfen das Netz zu erklären, verbessern und kontrollieren zu können. [1] In der Bilderkennung wird häufig versucht dies visuell darzustellen. Verschiedene Methoden wurden dazu entwickelt. Beispielsweise kann Grad-CAM [7] visualisieren, welche Bereiche in einem Bild für die Entscheidung relevant sind und diese entsprechend markieren (siehe Abbildung 2).

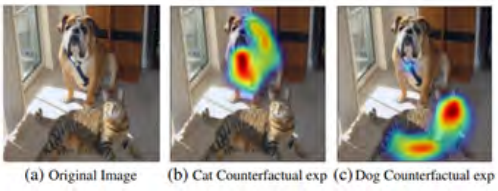

Abb. 2: Beispielbild einer Grad-CAM Auswertung [7]

Ausblick

Über die genannte Bildgenerierungspipeline können fotorealistisch Daten erzeugt werden, die für das Training des neuronalen Netzes geeignet sind. Dabei soll die Pipeline in den nächsten Schritten weiter verbessert werden, um die Realitätslücke im Modell möglichst komplett schließen zu können. In der Arbeit werden letztendlich erfolgte Verbesserungen gegenübergestellt, verglichen und bewertet.

Literatur und Abbildungen

[1] Amina Adadi and Mohammed Berrada. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, pages 52138–52160, 2018.

[2] Eigene Darstellung.

[3] Sebastian Hartwig and Timo Ropinski. Training Object Detectors on Synthetic Images Containing Reflecting Materials. https://arxiv.org/pdf/1904.00824.pdf, 2019.

[4] Christopher Mayershofer, Tao Ge, and Johannes Fottner. Towards Fully-Synthetic Training for Industrial Applications. https://mediatum.ub.tum.de/doc/1554870/1554870.pdf, 2021.

[5] Rosana Montes and Carlos Ureña. An Overview of BRDF Models. https://digibug.ugr.es/bitstream/handle/10481/19751/rmontes_LSI-2012-001TR.pdf, 2012.

[6] Myeongsuk Pak and Sanghoon Kim. A review of deep learning in image recognition. *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, pages 1–3, 2017.

[7] Ramprasaath R Selvaraju et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. http://arxiv.org/abs/1610.02391, 2016.

[8] Jonathan Tremblay et al. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. https://arxiv.org/abs/1804.06516, 2018.

Konzeption, Implementierung und Evaluation eines ML-Algorithmus zur automatischen Preisgestaltung von Mietfahrzeugen

Annette Grueber **Steffen Schober**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz Tech Innovation, Stuttgart

Einleitung und Motivation

Durch verschiedene Faktoren wie

- die über die Jahre stetig gewachsenen Datenmengen,
- die immer größer werdende Produktvielfalt und
- die Transparenz der Preise, die durch Vergleichsplattformen gewährleistet wird,

entsteht ein immenser Druck auf ein Unternehmen. Damit das Geschäft unter den geänderten Umgebungsbedingungen wettbewerbsfähig bleiben kann, muss es Gegenmaßnahmen ergreifen. Eines der wichtigsten Themen hierbei ist die Gestaltung des Preises eines Produkts oder einer Dienstleistung, da der Preis entscheidend ist für die Rentabilität des Unternehmens. [3] Deshalb beschäftigt sich die hier vorliegende Arbeit mit der Konzeptionierung, Implementierung und Evaluierung eines Machine Learning (ML) Algorithmus zur automatischen Preisgestaltung für Mietfahrzeuge von Vehicle Booking System (VBS). Dabei handelt es sich um eine Buchungsplattform, die von (End-)Kund*innen und Händlern für zwei Hauptanwendungsfälle genutzt wird:

1. Buchung von Testfahrten bei bestimmten Händlern und
2. Buchung von Mietfahrzeugen über einen gewissen Zeitraum bei bestimmten Händlern.

Die Entwicklung eines automatischen Preisalgorithmus ist ein möglicher Lösungsansatz zur Gestaltung des Preises. Diese Algorithmen passen den Preis kontinuierlich an die sich ändernden Anforderungen des Marktes an, sodass das Gewinnpotenzial ausgeschöpft und der Umsatz gesteigert wird. Dabei muss darauf geachtet werden, dass die Preise nicht zu stark angehoben werden, damit die Kundenzufriedenheit nicht darunter leidet. Durch all diese Rahmenbedingungen entsteht eine Preisspanne, in der sich die Unternehmen bewegen können, um den maximalen Gewinn für den Anbieter zu erzielen ohne die Kundenzufriedenheit zu schwächen. Des Weiteren reduziert die Integration eines automatischen Preisfindungsalgorithmus den Arbeitsaufwand und ist somit ressourcenschonender. [3]
Für diese Algorithmen stehen eine große Auswahl an Methoden und Technologien zur Verfügung wie z.B. ML-Algorithmen, worauf sich die hier vorgestellte Arbeit bezieht. Durch ML-Technologien wird die Leistung der Preisoptimierung im Vergleich zu traditionellen Preisgestaltungsmethoden erheblich angehoben. Die traditionellen Preisgestaltungsmethoden basieren auf Regeln und nutzen die vorhandenen Daten meist nicht, wogegen ML-Modelle viele verschiedene Faktoren miteinbeziehen, welche Preisänderungen beeinflussen wie die Nachfrage, das Angebot, die Produktcharakteristik oder den Wettbewerb. [3]

Zielsetzung und Vorgehensweise

In bestehenden Buchungssystemen wird für die Preisgestaltung von Mietfahrzeugen, wie bei vielen anderen Unternehmen, noch ein regelbasierter Preisalgorithmus verwendet. Dadurch wird z.B. ein Kunde in einigen Buchungssystemen nachteilig behandelt, wenn er vor mehreren Jahren in einen Unfall involviert war und sich seitdem nichts mehr zu Schulden kommen lassen hat. In VBS soll der traditionelle Preisalgorithmus nun durch einen ML-Ansatz ersetzt werden, um viele verschiedene Einflussfaktoren für den Mietpreis eines Fahrzeugs zu berücksichtigen. [3]
Für die Entwicklung der automatischen Preisgestaltung werden alle Schritte bis auf das Deployment des Prozessmodells CRISP-DM, das in Abbildung 1 veranschaulicht wird, durchlaufen. Der letzte Schritt fällt weg, da dies von der Geschwindigkeit des Genehmigungsprozesses der Firma abhängt und somit

möglicherweise nicht mehr im Bearbeitungszeitraum der Masterarbeit möglich ist.

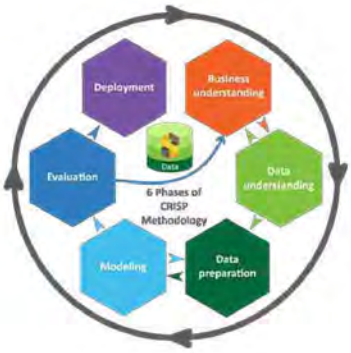

Abb. 1: CRISP-DM Prozess [1]

Zu Beginn wird in den Schritten „Bussiness understanding“ und „Data understanding“ anhand von Projekt- und Datenanalyse der produktiven Daten ein Ansatz zur automatischen Preisgestaltung mit ML-Algorithmen gesucht. [4]
Durch die Ergebnisse der beiden Schritte werden die Einflussfaktoren auf den Preis festgelegt. Die externen Daten werden vorerst nicht berücksichtigt, da dies über den Rahmen der vorliegenden Arbeit hinausgeht. Mögliche Einflussfaktoren auf den Preis sind beispielsweise der Fahrzeugtyp, Kraftstoffart, Abhol- und Abgabestandort des Fahrzeugs sowie der Buchungszeitraum.
Der Schritt „Data preparation“ behandelt die Bereinigung der Daten auf Basis der Datenanalyse und des gewählten Modells. [4]
Nun wird beim „Modelling“ der ausgewählte Ansatz implementiert und mit verschiedenen Parametereinstellungen trainiert. [4]
Zum Schluss werden die einzelnen Modelle, die durch die unterschiedlichen Parametrisierungen entstanden sind, im Schritt „Evaluation“ anhand geeigneter Metriken miteinander verglichen. So kann bewertet werden, welches der Modelle sich am besten zur Festlegung des Preises eignet. Die Wahl der Metriken ist davon abhängig, was genau verglichen werden soll. Hier stehen Metriken wie z.B. Mean Squared Error, Mean Absolute Error, R Squared und Root Mean Squared Error zur Verfügung. [4]

Konzept für die automatisierte Preisgestaltung

Auf Basis einer Literaturrecherche zu dem Thema automatische Preisgestaltung im ML-Bereich, den Projekt- und Datenanalyseergebnissen und Gesprächen mit Fachpersonen wird zur Lösung der Aufgabe der in der Abbildung 2 dargestellte Ansatz verwendet. Bei diesem Konzept wird der Preis basierend auf der Nachfrage optimiert. [5]

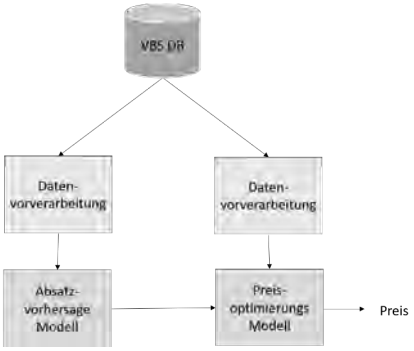

Abb. 2: Konzept für die automatische Preisgestaltung [2]

Dabei besteht die automatische Preisgestaltung aus drei Schritten:

1. Der Vorbereitung der Daten,
2. einer Vorhersage der Kundennachfrage zu einem bestimmten Zeitpunkt t und
3. einer Preisoptimierung.

Bevor die Daten von dem Absatzvorhersage- und Preisoptimierungsmodell verarbeitet werden können, müssen sie aufbereitet werden. Hier werden die Daten so vorbereitet, dass der Datensatz direkt von dem entsprechenden ML-Modell verwendet werden kann.
Der vorbereitete Datensatz wird zur Vorhersage der Buchungsanzahl an das Absatzvorhersagemodell übergeben. Für die Vorhersage wird ein einfaches Neuronales Netzwerk verwendet.
Danach wird der vorhergesagte Wert an das Preisoptimierungsmodell weitergegeben, um mit dessen Hilfe und den vorbereiteten historischen Daten einen Preis für den zu buchenden Zeitpunkt zu bestimmen. Den historischen Daten wird der Zusammenhang zwischen Buchungsanzahl und bezahltem Preis zu früheren Zeitpunkten entnommen. Dadurch besteht die Möglichkeit, anhand eines Optimierungsalgorithmus, wie dem Partikel-Schwarm-Algorithmus [6], die Gewinnfunktion

$$G_t = (Preis_t - Kosten_t) * Absatz_t \qquad (1)$$

zum Zeitpunkt t zu maximieren und hieraus den optimalen Preis abzuleiten.

Fazit und Ausblick

Ob der hier umgesetzte nachfragebasierte Preisgestaltungsansatz in der Realität brauchbare Ergebnisse liefert, wird sich erst im Laufe der Zeit zeigen. Das

lässt sich darauf zurückführen, dass der Preisoptimierungsalgorithmus erst durch die Kundenreaktion auf den gesetzten Preis lernt. Die Kundenreaktion schließt der Algorithmus aus dem Rückgang oder Anstieg der Buchungen. Somit werden in dieser Ausarbeitung hauptsächlich die Ergebnisse des Absatzmodells evaluiert.
Zur Verbesserung der Ergebnisse können weitere Änderungen vorgenommen werden wie:

- Hinzufügen weiterer Merkmale für die Vorhersage wie Feiertage oder spezielle Events
- Austauschen der ML-Modelle: Das einfache Neuronale Netzwerk kann beispielsweise durch ein Long-Short Term (LSTM) Netzwerk ausgetauscht werden, um auch seltene Datenpunkte wie z.B. Buchungen an Weihnachten zu berücksichtigen. [7]
- Erweiterung des einfachen neuronalen Netzes zu einem Bayesian Netzwerk, um Unsicherheiten in der Vorhersage feststellen zu können. Dadurch wird das Modell für das Unternehmen transparenter. Das ist ein wichtiger Faktor bei Preisgestaltungen, da ML-Algorithmen meist sehr schwer nachvollziehbar sind und ein Unternehmen bei der Preisgestaltung wissen möchte, wie sich dieser zusammensetzt. [7]

Literatur und Abbildungen

[1] Zenah Yaser Alzubaidi. A Comparative Study on Statistical and Machine Learning Forecasting Methods for an FMCG Company, 2020.

[2] Eigene Darstellung.

[3] Z. Elraffah. Was ist Dynamic Pricing? https://7learnings.com/de/blog/was-ist-dynamic-pricing/, 2020.

[4] Neha Kaul. Applications of Data Mining in Engineering, Management and Medicine. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2013894&site=ehost-live, 2019.

[5] Ajitesh Kumar. Pricing Optimization & Machine Learning Techniques. https://vitalflux.com/pricing-optimization-machine-learning-techniques/, 10 2021.

[6] Yushan Liu. Partikelschwarmoptimierung für diskrete Probleme. In *Ferienakademie14*. Fakult¨at f¨ur Mathematik TU München, 2014.

[7] L. Zhu and N. Laptev. Deep and Confident Prediction for Time Series at Uber. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 103–110, 2017.

ÖPNV Verkehrssimulation mit SUMO basierend auf GTFS Daten zur Analyse intermodaler Reiseketten

Arber Guri **Jörg Nitzsche**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Der ÖPNV spielt für die Verkehrswende eine entscheidende Rolle. Bis zu zehn Mrd. Fahrgäste sind in Deutschland mit dem ÖPNV jedes Jahr unterwegs [3]. Um den Fahrgästen eine möglichst komfortable Anreise und immer mehr Menschen den Umstieg vom Auto auf dem ÖPNV zu ermöglichen, ist die Verkehrsplanung ein wichtiger Bestandteil der Verkehrswende. Verkehrsimulationen geben den Verantwortlichen im Bereich Verkehrsplanung ein mächtiges Werkzeug in die Hand um unterschiedliche Szenarien zu untersuchen. SUMO (Simulation of Urban Mobility) stellt so eine Simulationsumgebung dar.

Ziel der Arbeit

Ziel dieser Arbeit ist es, ein funktionsfähiges Simulationsmodell für die Stadt Esslingen mit realen ÖPNV-Fahrplandaten zu erstellen und anhand dieses Modells einige exemplarische Analysen durchzuführen. Insbesondere die Analyse unterschiedlicher intermodaler Reiseketten steht hierbei im Vordergrund.

Simulation of Urban Mobility

SUMO ist eine kostenlose Open-Source Verkehrssimulationssoftware. Sie wird vom Deutschen Zentrum für Luft- und Raumfahrt (DLR) entwickelt und ist seit 2001 verfügbar. Sie ermöglicht die Modellierung intermodaler Verkehrssysteme mit unterschiedlichen Fahrzeugarten, öffentlichen Verkehrsmitteln und Fußgängern. In SUMO ist eine Fülle von unterstützenden Tools enthalten, die Kernaufgaben für die Erstellung, Ausführung und Auswertung von Verkehrssimulationen, wie Netzimport, Routenberechnung, Visualisierung und Emissionsberechnung automatisieren. SUMO kann mit benutzerdefinierten Modellen erweitert werden und bietet verschiedene APIs zur Fernsteuerung der Simulation.

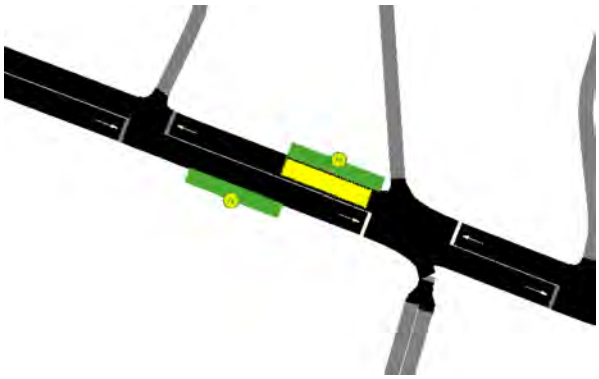

Abb. 1: Haltestelle mit vorbeifahrenden Bus in SUMO [2]

OpenStreetMap Daten

OpenStreetMap (OSM) [4] ist eine kostenlose, editierbare Weltkarte, die größtenteils von Freiwilligen von Grund auf neu erstellt und unter einer Open-Content-Lizenz veröffentlicht wird. Mithilfe von OSM kann das SUMO-Tool OsmWebWizard verwendet werden, um jede Stadt oder Region auszuwählen, in der die Simulation ausgeführt werden soll [6]. Mit nur wenigen Klicks können entsprechende Einstellungen, wie z. B. die Dauer der Simulation oder die Verkehrsdichte für jeden Verkehrstyp, ausgewählt werden. Außerdem besteht die Möglichkeit den Öffentlichen Verkehr mit zu simulieren. Allerdings wurde festgestellt, das die Haltestellen oft an falschen Positionen angezeigt werden. Auch der eigentliche Fahrplan wird nicht richtig importiert.

General Transit Feed Specification

Die General Transit Feed Specification (GTFS) ist eine Datenspezifikation, die es öffentlichen Verkehrsunternehmen ermöglicht, ihre Verkehrsdaten in einem Format zu veröffentlichen, das von verschiedenen Softwareanwendungen verwendet werden kann [5].

Heute nutzen Tausende von ÖPNV-Anbietern das GTFS-Datenformat. Die Datensätze werden in einer Reihe von Textdateien dargestellt, die in eine ZIP-Datei komprimiert werden und Informationen wie feste Fahrpläne, Routen und Daten zu Bushaltestellen enthalten. Der Verkehrs- und Tarifverbund Stuttgart (VVS) stellt regelmäßig aktualisierte GTFS-Daten für unsere Arbeit zur Verfügung. Deutschland verfügt mit über 20.000 Routen, über 500.000 Haltestellen und fast zwei Millionen Linienfahrten über einen der größten GTFS-Datensätze der Welt [1].
Um das zuvor erwähnte Problem zu beheben, bei dem die Haltestellen nicht korrekt angezeigt werden, stellt SUMO ein Python-Skript bereit, das die Daten von GTFS verwendet, um zusätzliche Dateien zur korrekten Anzeige der Haltestellen zu erstellen. Zusätzlich wird auch der reale Fahrplan berücksichtigt. Der Nutzer muss das Skript „gtfs2pt.py“ ausführen, wie in Abb. 2 gezeigt.

```
python3 sumo/tools/import/gtfs/gtfs2pt.py -n osm.net.xml --gtfs
gtfs.zip --date 20220423 --osm-routes osm_ptlines.xml
--repair --modes bus
```

Abb. 2: Ausführung von gtfs2pt.py [2]

TraCI Schnittstelle

Das Tool Traffic Control Interface (TraCI) dient zur Steuerung und zum Abrufen von Informationen über laufende Simulationen. TraCI verwendet eine TCP-basierte Client/Server-Architektur, um den Zugriff auf SUMO bereitzustellen. Hier fungiert SUMO als Server, der mit zusätzlichen Befehlen gestartet wird. Die Client-Anwendung steuert den Simulationsprozess, beeinflusst das Verhalten einzelner Fahrzeuge und sendet Befehle an SUMO, um Details der Umgebung abzufragen. SUMO antwortet mit einer Statusantwort auf jeden Befehl und zusätzlichen Ergebnissen basierend auf dem angegebenen Befehl [7].

Intermodale Reiseketten

Um für die Stadt Esslingen einige exemplarische Analysen durchzuführen, werden mit Hilfe von TraCI einzelne Personen hinzugefügt und kontrolliert. Die zu vergleichenden Personen haben den gleichen Start- und Endpunkt, jedoch benutzen sie verschiedene Verkehrsmittel um die intermodale Reise durchzuführen. Ziel ist es, zu vergleichen, wie sich die Ergebnisse bei der Nutzung von unterschiedlichen Verkehrsmitteln verändern.

Literatur und Abbildungen

[1] Patrick Brosi. GTFS.DE - GTFS für Deutschland. https://gtfs.de/, 2019.

[2] Eigene Darstellung.

[3] Verband Deutscher Verkehrsunternehmen. VDV Statistik 2019. https://www.vdv.de/vdv-statistik-2019.pdfx, 10 2020.

[4] OpenStreetMap Foundation. OpenStreetMap. https://www.openstreetmap.org/, 2022.

[5] Deutsches Zentrum für Luft-und Raumfahrt und andere. GTFS. https://sumo.dlr.de/docs/Tutorials/GTFS.html, 07 2021.

[6] Deutsches Zentrum für Luft-und Raumfahrt und andere. OSMWebWizard. https://sumo.dlr.de/docs/Tutorials/OSMWebWizard.html, 02 2022.

[7] Deutsches Zentrum für Luft-und Raumfahrt und andere. TraCI. https://sumo.dlr.de/docs/TraCI.html, 02 2022.

Regressionstestwerkzeuge für Low-Code-Entwicklung im Bereich Machine Vision

Manuel Haerer **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Kontext

Industrielle Bildverarbeitung ist heutzutage eine wichtige Komponente von *Industrie 4.0*. Das Anwendungsgebiet erstreckt sich von der Qualitätssicherung im Produktionsumfeld bis hin zu Positionierungsaufgaben bei Robotersystemen [2]. Da die Bildverarbeitung grundsätzlich spezielle, technische Herausforderungen mit sich bringt, verwenden Anwender üblicherweise bereits verfügbare Softwarebibliotheken oder -systeme. Beispiele hierfür sind zum einen Open-Source-Lösungen wie *OpenCV*, *SimpleCV*, *Keras*, oder *BoofCV* [10] und zum anderen kommerzielle Lösungen wie *Halcon* [5] oder *Matrox Imaging* [9]. Diese Kernkomponenten sind, je nach Applikation und Kontext, meist in übergeordnete Systeme integriert. Abbildung 1 zeigt die typische Gesamtarchitektur solcher Systeme.

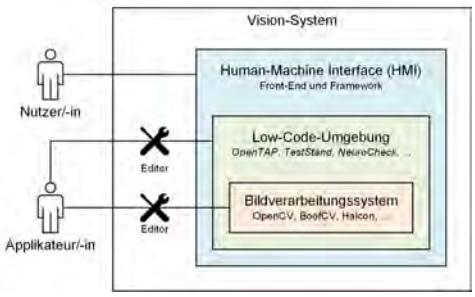

Abb. 1: Architektur typischer Vision-Systeme. [4]

Damit sich Applikationsingenieure und -ingenieurinnen auf das Bildverarbeitungsproblem an sich konzentrieren können, besteht oftmals die Anforderung, das applikationsspezifische Verhalten der Vision-Systeme möglichst einfach und unkompliziert definieren und ändern zu können. Aus diesem Grund werden auf dieser Ebene *Low-Code-Umgebungen* eingesetzt. Auch hierbei sind sowohl Open-Source-Lösungen wie *OpenTAP* [8] als auch kommerzielle Lösungen wie *TestStand* [3] oder *NeuroCheck* (speziell für den Vision-Kontext) [6] verfügbar. Je nach System wird ggf. noch ein zusätzliches HMI-Front-End verwendet, damit die Anlage im Feld einfach durch Maschinenführer und -führerinnen bedient werden kann. Somit erhält man insgesamt ein System, das aus vielen Komponenten besteht, die jeweils verschiedenste Aufgaben erfüllen.

Problemstellung

Vision-Systeme sollten, wie andere Softwaresysteme auch, getestet werden. In diesem Kontext sind insbesondere Regressionstests (wiederholt durchgeführte Tests nach Änderungen [7]) auf Systemebene nützlich und wirtschaftlich. Hierfür notwendige Testwerkzeuge sind jedoch, aufgrund der heterogenen und firmenspezifischen Softwarearchitektur, nicht ohne Weiteres verfügbar. Dementsprechend beschäftigt sich diese Arbeit mit der Konzeptionierung und dem Entwurf von passenden Regressionstestwerkzeugen. Schließlich soll auch eine prototypische Implementierung erfolgen.

Analyse

Um die Anforderungen an derartige Testwerkzeuge vollständig zu ermitteln, wurden im ersten Schritt die folgenden Kernanwendungsfälle identifiziert:

- Softwareverifikation: Korrekte Funktionsweise der Software nach einer Änderung nachweisen.
- Maschinentest: Korrekte Funktionsweise der gesamten Anlage (SW und HW) nachweisen.
- Parameteroptimierung: Optimierung von Bildverarbeitungsparametern auf Grundlage ausgeführter Tests und deren Ergebnisse.
- Fehlerteilanalyse: Fehlersuche bei unstimmigen Ergebnissen; z. B. bei falscher Klassifizierung.

Aus diesen Anwendungsfällen leiten sich funktionale Anforderungen an Werkzeuge ab: U. a. soll das einfache Definieren von Testfällen, Ausführen von Tests und Analysieren der Ergebnisse ermöglicht werden. Jene

Anforderungen sind maßgeblich durch den Kontext des Sondermaschinenbaus und der dort üblichen Arbeitsabläufe bestimmt. Zusätzlich ergeben sich nichtfunktionale Anforderungen: Beispielsweise sollen die Regressionstestwerkzeuge möglichst flexibel und über den gesamten Lebenszyklus einer Vision-Applikation hinweg (von der Entwicklung über die Inbetriebnahme bis zur Produktion) einsetzbar sein.

Konzept

Die für diese Arbeit betrachteten Vision-Systeme arbeiten grundsätzlich nach einem Job-Prinzip. Dabei verharrt das System in einem Bereitschaftsmodus, bis eine Jobanfrage von der zentralen Maschinensteuerung eingeht; damit wird die Bildaufnahme und -verarbeitung angestoßen. Die Ergebnisse werden schließlich zentral abgelegt und der Maschinensteuerung zurückgemeldet. Ein solcher Durchlauf ist in Abbildung 2 auf der rechten Seite dargestellt. Er soll die Entität darstellen, gegen welche die gewünschten Regressionstests ausgeführt werden.

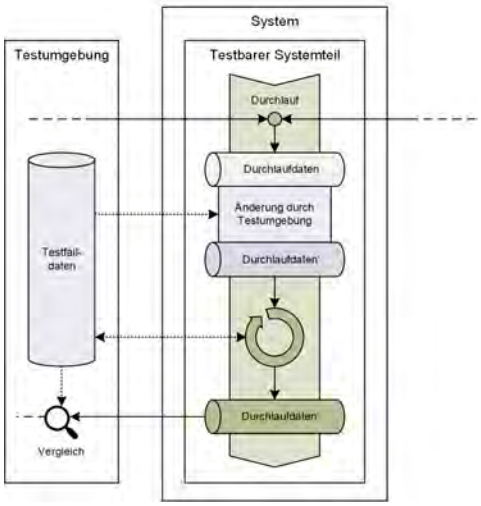

Abb. 2: Konzept für Systemtests. [4]

Um alle Eingaben und Ausgaben bezüglich eines Durchlaufs einheitlich kontrollieren zu können, wird eine durchlaufbezogene, hierarchische Datenstruktur verwendet. Diese *Durchlaufdaten* halten Handles, die einen Zugriff auf verschiedene Ressourcen abstrahieren, sowie tatsächliche Ein- und Ausgabedaten wie z. B. Bilder. Im Falle eines Tests muss somit lediglich zu Beginn des Durchlaufs eine Modifikation der Durchlaufdaten erfolgen, indem alle Testeingaben des jeweiligen Testfalls hinzugefügt werden. Im weiteren Ablauf verwendet das Vision-System die zugeführten Handles und Daten; die Ergebnisse werden in den Durchlaufdaten abgelegt. Nach Ende des Durchlaufs kann der resultierende Datensatz anhand von Kriterien, die ebenfalls Teil des jeweiligen Testfalls sind, bewertet werden. Daraus ergibt sich schließlich das Testergebnis.

Entwurf

Die in Abbildung 2 dargestellte *Testumgebung* wird durch ein Regressionstestwerkzeug realisiert. Jenes muss den beschriebenen Ablauf für die Testausführung umsetzen. Damit eine möglichst hohe Kompatibilität zu verschiedenen Systemarchitekturen gegeben ist, sollte die Interaktion zwischen Vision-System und Werkzeug auf der untersten Architekturebene (Bildverarbeitungsbibliothek) realisiert werden. Dadurch wird erreicht, dass das Testwerkzeug für alle Vision-Systeme verwendet werden kann, die auf der gleichen Bildverarbeitungsbibliothek basieren. Neben der Ausführung von Tests ist die Definition der Testfälle sowie die Analyse der Ergebnisse notwendig. Es bietet sich an, all diese Funktionen in einem einzigen Programm zu vereinen. Abbildung 3 zeigt eine mögliche Benutzeroberfläche jener Software.

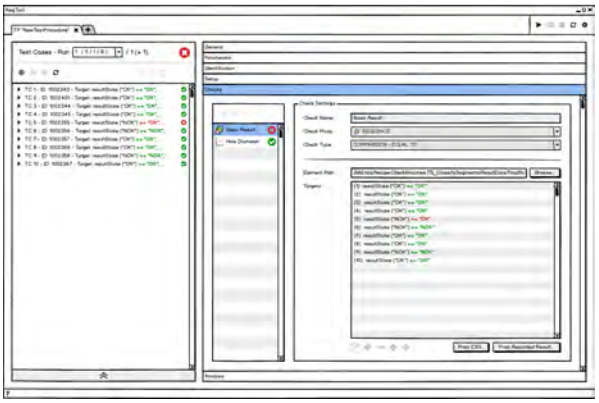

Abb. 3: Benutzeroberfläche eines möglichen Regressionstestwerkzeugs. [4]

Der Entwurf der Oberfläche orientiert sich an Testwerkzeugen, die bei integrierten Komplettlösungen (Bildaufnahme und -verarbeitung in der Kamera) wie z. B. *Cognex* [1] zum Einsatz kommen. Auf der linken Seite werden auszuführende Testfälle aufgeführt; rechts finden sich Reiter zur Konfiguration und Analyse.

Ausblick

Im weiteren Verlauf der Arbeit soll ein Prototyp umgesetzt werden, der die Kernanforderungen abdeckt und die technische Realisierbarkeit des Konzepts demonstriert. Zuletzt erfolgt eine kritische Bewertung der erzielten Ergebnisse.

Literatur und Abbildungen

[1] Cognex Corporation. In-Sight® Explorer Help - TestRun™ - Documentation | Cognex. https://support.cognex.com/docs/is_611/web/EN/ise/Content/User_Interface/TestRun.htm?tocpath=EasyBuilder%20View%7CPalette%7CTestRun%E2%84%A2%7C_____0, 2020.

[2] Cognex Corporation. Was ist industrielle Bildverarbeitung | Cognex. https://www.cognex.com/de-de/what-is/machine-vision/what-is-machine-vision, 2022.

[3] National Instruments Corporation. Was ist TestStand? - NI. https://www.ni.com/de-de/shop/electronic-test-instrumentation/application-software-for-electronic-test-and-instrumentation-category/what-is-teststand.html, 2022.

[4] Eigene Darstellung.

[5] MVTec Software GmbH. HALCON – The power of machine vision: MVTec Software. https://www.mvtec.com/de/produkte/halcon, 2022.

[6] NeuroCheck GmbH. Anwendungssoftware für anspruchsvolle Prüflösungen. https://www.neurocheck.de/software/was-ist-neurocheck/, 2022.

[7] ISO IEC IEEE. Part 1: General Concepts. In *ISO/IEC/IEEE 29119-1 - Software and systems engineering - Software testing*. ISO IEC IEEE, 2022.

[8] Keysight Technologies Incorporated. About | OpenTAP. https://opentap.io/about, 2022.

[9] Matrox Electronic Systems Limited. Matrox Imaging Library (MIL) | Software development kit | Matrox Imaging. https://www.matrox.com/en/imaging/products/software/sdk/mil?utm_medium=ppc&utm_source=adwords&utm_campaign=Machine+Vision+Software&utm_term=machine%20vision%20software, 2022.

[10] SuperAnnotate LLC. Top 15 computer vision libraries. https://blog.superannotate.com/computer-vision-libraries/, 2022.

Ausfallprognosemodell für die Analyse von Garantiefällen zur Kostenoptimierung

Daniel Haerer **Steffen Schober**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Andreas Stihl AG & Co. KG, Waiblingen

Motivation und Ziel

"Jede Motorsäge ist nur so gut wie ihr Service." Dies ist das Motto von STIHL im Bezug auf die Zufriedenheit der Kunden. Zu diesem Service gehört natürlich auch die Zuverlässigkeit der Geräte. Die Herstellung von Motorsägen, Freischneidern und anderen Geräten ist in der heutigen Zeit sehr komplex geworden. Das Zusammenspiel aus Software und Hardware war nie größer. Trotz hoher Qualitätsanforderungen und Dauererprobungen kann es vorkommen, dass Bauteile oder Produkte doch aufgrund ihrer Komplexität trotzdem ausfallen. Die Zeitdauer bis zur Erkennung von Problemen sollte deswegen minimiert werden. Falls bei den vielzähligen Dauererprobungen keine Probleme festgestellt werden, kann es teilweise mehr als 1 Jahr dauern bis Kunden die vorhandenen Probleme bemerken. Diese Fälle sind ärgerlich für das Unternehmen, aber noch viel ärgerlicher für den Kunden, da dieser das Produkt ja verwenden wollte. Für das Unternehmen sind diese Fälle mit hohen Garantiekosten, reduzierter Kundenzufriedenheit und möglicherweise Nachentwicklungen verbunden. Aus diesen Gründen wird eine Methode gesucht, um möglichst frühzeitig potenziell kritische Produkte zu identifizieren. Diese Produkte können dann intensiv geprüft werden, um mögliche Probleme aufzudecken.

Methoden und Vorgehen

In einem ersten Schritt wird die bisherige Vorgehensweise auf ihre Limitierungen und Probleme untersucht. Aufgrund dieser Analyse werden eigene Key-Performance-Indexe (KPIs) erstellt, welche eine Aussage über die Qualität des Produkts ermöglichen. Einer dieser KPIs wäre die Verteilungsfunktion der Produkte mit Garantiefall, die in den letzten 2 Jahren verkauft wurden. Der in Abbildung 1 dargestellte KPI eines fiktiven Produkts zeigt, dass 1 von 2000 Geräten ab Verkauf defekt wären und dass es einen deutlichen Sprung der Ausfälle ab dem 10. Monat gibt. In einem nächsten Schritt müsste nun der Grund für diese Ausfälle gesucht werden. Eines der Probleme dieser KPI ist, dass im Moment zwar die Anzahl der Garantiefälle genau bestimmt werden kann, es jedoch nicht klar ist wie viele Produkte sich im Garantiezeitraum befinden. Diese Anzahl der Garantiefähigen Geräte ist jedoch von enormer Wichtigkeit, um die absolute Anzahl der Garantiefälle in einen Verhältnis setzen zu können. Die Absatzzahlen an die Großhändler sind sehr genau bekannt. Die tatsächlichen Verkäufe an Kunden sind nicht alle bekannt, da die Registrierung dafür freiwillig ist.

Abb. 1: Verteilungsfunktion der Garantiefälle - Beispiel [1]

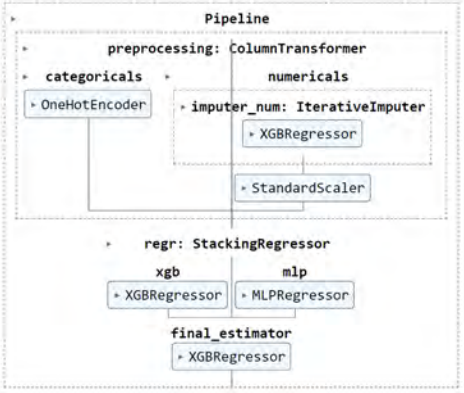

Abb. 2: Modelldiagramm der Verkaufsvorhersage [3]

Mithilfe der vorhandenen Daten kann jedoch ein Maschine-Learning-Modell trainiert werden, welches die tatsächlichen Verkaufsdaten basierend auf Datum, Land und Maschinentyp vorhersagt. So verbringt der GTA 26, der eine Warteliste von mehreren Monaten besitzt eine geringere Zeit im Lager als eine ältere Maschine, die kaum nachgefragt wird. Hierbei wird das CRISP Prinzip verwendet und das Modell ist in Abbildung 2 dargestellt. Für numerische Features wird ein Iterativer Imputer und eine Standardisierung verwendet, um fehlende Werte zu ersetzen [3]. Für kategorische Werte ohne Ordnung wird ein One-Hot-Encoder verwendet. Mit dieser Vorverarbeitung werden mehrere Modelle trainiert, die von einem Stacking Regressor [3] gewichtet werden. Dabei werden die besten Hyperparameter mit einer Grid-Search und Kreuzvalidierung bestimmt [3]. Die vorhandenen Produktregistrierungen können nun mit den Vorhersagen ergänzt werden, um Verkaufszahlen abzuschätzen.
Mithilfe der Anzahl an Geräten im Garantiezeitraum, können nun die vorhandenen Garantiefälle bewertet werden. Dabei soll auch die Schwere des Defekts eine Rolle spielen. Abschließen wird ein Maschine-Learning-Modell trainiert, das Aussagen über Ausfälle nach x-Monaten sowie die Unsicherheit der Vorhersage geben soll. Dafür werden mehrere Algorithmen getestet, die mit Zeitreihen umgehen können. Beispiele hierfür sind VARIMAX, kNN, XGBoost, Neurale Netze, RNN und LSTM, die im Buch *Advanced Forecasting with Python* näher erläutert werden [2]. Es muss auch geprüft werden wie viele Datenpunkte benötigt werden um sichere Vorhersagen treffen zu können. Zuletzt soll verglichen werden, ob die Modellvorhersage die Erfahrung von Fachexperten unterstützen kann und welcher Zeitgewinn durch eine frühzeitige Vorhersage entsteht.

Ergebnisse und Ausblick

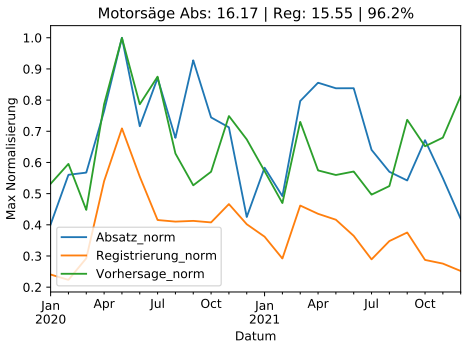

Abb. 3: Beispiel der Verkaufsdatenvorhersage [1]

Die vorhandenen Meldungen der Garantiefälle wurden intensiv untersucht, um Besonderheiten, Fehler und Unstimmigkeiten zu finden. Die Vorhersage der Verkaufszeitpunkte besitzt bei ersten Tests einen mittleren Fehler (MAE) von 60 Tagen und eine Wurzel der mittleren Fehlerquadratsumme (RMSE) von 80 Tagen. Ein durchschnittlicher Fehler von zwei Monaten ist ausreichend, um die tatsächlichen Verkaufszahlen abschätzen zu können. Abbildung 3 zeigt die tatsächlichen Produktregistrierungen (gelb), die gesamten vorhergesagten Verkäufe (grün) und die Absatzzahlen an die Großhändler (blau) jeweils mit einer Normalisierung. Dabei ist die Anzahl der Verkäufe 4% geringer als die, der Lieferungen an den Großhändler. Die Lager haben sich laut Prognose also gefüllt. Die Vorhersage der Garantiefälle und eine KPI Implementierung wurden zum Zeitpunkt der Einreichung dieses Artikels noch nicht durchgeführt. Die Resultate der Forschungsarbeit können möglicherweise verwendet werden, um frühzeitig Produktweiterentwicklungen oder Analysen und Tests der Produkte anzustoßen. Im besten Fall kann somit die Zeitdauer bis zur Reaktion um mehrere Monate reduziert werden.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Joos Korstanje. *Advanced Forecasting with Python*. Apress Berkeley, CA, 2021.

[3] F. Pedregosa et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Einbindung eines Secure Elements in einen IIoT-Sensor

Daniel Hartung **Tobias Heer**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Leuze electronics GmbH, Owen

I. MOTIVATION UND PROBLEMSTELLUNG

Das Themengebiet der IT-Security spielt in der IoT eine wichtige Rolle, welche jedoch bei der Entwicklung des Produktes häufig vernachlässigt wird. Oftmals werden Sicherheitsfunktionen nachträglich in das System implementiert. Da diese IoT-Geräte in der Regel in einem Netzwerk kommunizieren, wird eine gewisse Anforderung an Sicherheit erwartet, damit diese Geräte nicht manipuliert werden können. Dadurch sollen Gefahren, wie beispielsweise eine Bildung eines Botnets, eliminiert werden.
In Embedded Geräten ist in der Regel ein System on a Chip (SoC) eingebaut, welches ein Secure Element (SE) in der CPU besitzt. Das SE kümmert sich um das Speichern von sensiblen Daten, wie zum Beispiel private Schlüssel oder andere Daten zur Authentifizierung des Benutzers. Die Daten im SE sind manipulationssicher, wodurch diese ausschließlich über die serielle Schnittstelle oder sogar gar nicht ausgelesen werden können. Mithilfe von kryptographischen Funktionen, wie zum Beispiel Hash-Funktionen, Symmetrischen und asymmetrischen Verschlüsselungen, werden die Daten im Speicher des SEs gesichert. SEs werden in zahlreichen Gebieten verwendet. Überwiegend profitieren Zahlungssysteme von den Funktionen. Bankkarten, Hardware Kryptowährung-Wallets und sogar der neue Personalausweis speichern deren Informationen in ein SE. In älteren SoC's ist meistens kein SE mit eingebaut. Dies ist der Fall bei Geräten der Leuze electronics GmbH. Dadurch können die Schutzziele, Confidentiality, Integrity und Availability (CIA), verletzt werden. Sobald der Hersteller die Security Principles nicht sicherstellen kann, ist das Gerät anfälliger auf Angriffe, wodurch entweder Daten ausgelesen, geändert oder unzugänglich werden. Da diese älteren SoC's ohne einen integrierten SE weiterhin im Markt verbreitet sind, gibt es die Möglichkeit ein SE als Hardwaremodul am Mikrocontroller nachträglich zu implementieren. Dafür wird jedoch ein Footprint an der Geräteleiterplatte benötigt, welches nicht immer der Fall ist. Dadurch können aufwendige Hardware Re-Designs nötig sein. Im Vergleich zu den integrierten SEs bietet ein Embedded SE nicht das volle Potential eines SEs, jedoch ist es für ein System ohne integrierten SE die einzige und beste Lösung. Durch die Einbindung und Nutzung der Funktionen eines SEs in einem Embedded System können bestimmte Sicherheitslücken entweder geschlossen, oder zu einer größeren Herausforderung für Angreifer werden. Darunter fallen folgende Use Cases zur erhöhten Sicherheit in einem Embedded Gerät:

- Sicheres Speichern von Schlüssel oder Zertifikaten zur mit TLS verschlüsselten Kommunikation.
- Verifizierung der Bootsoftware und Gerätefirmware mithilfe von Kryptographischen Funktionen, bevor diese auf dem Gerät ausgeführt werden.
- Ausführen von Kryptographischen Funktionen innerhalb des SEs, anstatt auf der CPU vom SoC.

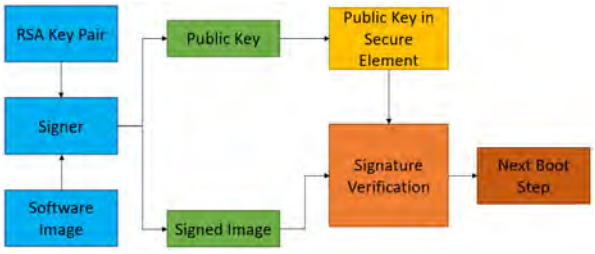

Abb. 1: Verifikation von Software Signaturen [1]

Das sichere Speichern von Schlüssel oder Zertifikaten ist in der heutigen Zeit von großer Bedeutung, besonders bei IoT Geräten, die in einem Netzwerk kommunizieren. Die größte Gefahr sind Angriffe aus dem Netzwerk, da dadurch die Angreifer auf einem Schlag eine große Anzahl von IoT Geräten mit der identischen Sicherheitslücke manipulieren können. Der Fokus dieser Abschlussarbeit wird die Implementierung einer Secure Boot Funktion sein. Wie schon vorhin aufgegriffen, wird beim Secure Boot die Bootsoftware auf ihre Integrität überprüft. Dies sollte ebenfalls auf

einem gewissen Abstraktions-Layer entwickelt werden, damit die Kommunikation zum SE auf anderen SoC's vereinfacht wird. Hierfür wird die Software vom Hersteller signiert. Diese Signatur wird an die zu verifizierende Firmware angehängt. Somit kann die Hardware beim Bootvorgang mithilfe eines Hashverfahrens und die Entschlüsselung der Signatur verifiziert werden.

II. DESIGN

Zur Auswahl eines Secure Elements muss bedacht werden, welche Use Cases man für das Embedded Gerät benötigt, und wie sicher das SE heutzutage ist. Bei älteren Modellen ergaben sich potenzielle Schwachstellen, wie zum Beispiel Informationen mithilfe eines Power Analysis Attack auslesen. Probleme wie diese sind bei neueren Modellen schwieriger auszuführen. Es werden keine speziellen Funktionen an das SE erfordert, da man für die Secure Boot Funktion lediglich Signaturen und Schlüssel speichert, und grundlegende kryptographische Funktionen zur Verifizierung benötigt. Somit hat man sich für das Secure Element ATECC608B vom Hersteller Microchip entschieden. Dieser verfügt über die Möglichkeit zur Speicherung von 16 Schlüssel, Zertifikaten oder sonstige Informationen. Ebenfalls unterstützt es die Verifizierung von Signaturen und Hashes auf dem Chip, wodurch die vertraulichen Schlüssel nicht an die CPU gesendet werden müssen. Zum Secure Boot muss ein Startpunkt definiert werden, ab dem die Verifizierung der gesamten Bootsoftware beginnt. Dieser wird auch der Root of Trust genannt. Da das SE als Modul am Mikrocontroller angebunden ist, kann es nicht im ersten Bootschritt in der CPU beginnen, da das SE zum Zeitpunkt nicht initialisiert ist. Zur Verdeutlichung wird ein Linux-Bootprozess auf ARM Prozessoren folgend beschrieben:

- Im ersten realisierbaren Schritt beginnt das Power-on Reset (PoR) des SoC's, welches den ROM-Bootloader startet.
- Der ROM-Bootloader sucht nach den Secondary Program Loader (SPL) und lädt diesen in den statischen RAM.
- Der SPL initialisiert das externe DRAM und weitere Module am SoC, wie zum Beispiel das SE. Anschließend wird der Bootloader zum Starten vom Kernel, in unseren Fall Das U-Boot, in das RAM geladen und ausgeführt.
- U-Boot ist zuständig für das Starten des Kernels, welcher dann das Root Filesystems aufsetzt.

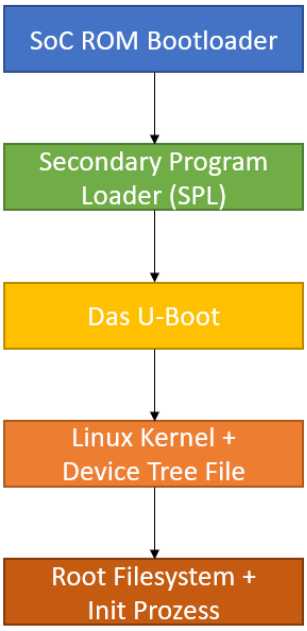

Abb. 2: Boot Prozess in Embedded Geräten [1]

Somit wäre es für den SoC erst ab dem SPL möglich die folgenden Bootschritte zu verifizieren, wodurch es zum Root of Trust für das System wird. Zur Verifizierung der Bootsoftware wird bereits die Vorgehensweise vom SE gegeben. Es wird die Bootsoftware des aktuellen Bootschritts basierend auf SHA-256 gehasht. Daraufhin wird die Signatur der Bootsoftware mithilfe des im SE gespeicherten Public Keys entschlüsselt und mit dem Hashwert verglichen. Diese sollen beim Einhalten der Integrität der Software identisch sein. Bei einer Manipulation bricht der Bootvorgang ab, oder es wird für eine bestimmte Anzahl erneut versucht. Die Kommunikation zwischen dem SE und der CPU könnte sich als Herausforderung darstellen, da eventuell Probleme mit der Kompatibilität auftreten können. Hierfür bietet der Hersteller Microchip eine Library namens CryptoAuthLib an. Ebenfalls ist es anhand einer Dokumentation zur Kommunikation möglich, diese Kommunikation selbstständig zu entwickeln. Da Angreifer oftmals hohes Interesse an der Manipulation des Kernels haben, kann man innerhalb des Bootloaders Das U-Boot zusätzliche Konfigurationen beifügen. Dadurch werden die Kernels und deren Device Tree Binaries, ebenfalls mit der vorherigen Logik, verifiziert.

III. EVALUATION

Microchip bietet neben dem Secure Element ATECC608B noch weitere an, die den Anforderungen entsprechen können. Hierfür wird das SE mit den

anderen Modellen von Microchip verglichen. In diesem Vergleich wird die Performance des Bootvorgangs, Sicherheit der gegebenen Kryptographischen Funktionen und Potential für zukünftige Sicherheitsfunktionen untersucht. Zu der Beurteilung der kryptographischen Funktionen ist zu beachten, welcher Grad von Sicherheit mehr als ausreichend ist für Embedded Geräte. Dabei muss ebenfalls der Einkaufspreis des SEs, aufgrund einer hohen Produktionsmenge, niedrig gehalten werden. Im Punkt Performance des Bootvorgangs muss in Betracht gezogen werden, welche zusätzliche Verzögerung für den Benutzer bemerkbar sein wird. Dabei werden zusätzlich die Performanceunterschiede zwischen ECC und RSA verglichen, welches Teil der Entscheidung des SEs ist. Es muss entschieden werden, ob der definierte Root of Trust im SPL wirksam sein wird vor potenziellen Angriffen auf den Softwarecode. Die Implementierung des SEs und der Secure Boot Funktion sollte für andere Mikrocontroller verwendbar sein, die ebenfalls keinen Integrierten SE im Prozessor haben. Dadurch würde es möglich sein, die bestmögliche Lösung ohne großen Aufwand in andere IoT Geräte einzubauen, um die Allgemeine Sicherheit der IoT-Welt zu erhöhen.

IV. VERWANDTE ARBEITEN

Die wissenschaftlichen Arbeiten „Secure element based framework for sensors anomaly detection in industry 4.0“ [3] und „A secure element and blockchain stratagem for securing iot“ [2] werden Secure Elements in ein System implementiert, um andere Probleme in der IoT-Sicherheit zu lösen. In der Arbeit von Guilley et al. [4] wird der Unterschied zwischen einem Integrierten und Embedded SE verdeutlicht. Bei der wissenschaftlichen Ausarbeitung von Sanwald et al. [6] werden die allgemeinen Herausforderungen bei der Implementation der Secure Boot Funktion genauer beschrieben. Die Performance Auswirkungen der Secure Boot Funktion untersuchen Profentzas et al. [5], jedoch ohne Secure Element, welches eventuell zusätzliche Rechenzeit benötigen kann.

V. AUSBLICK

Durch die Secure Boot Funktion kann das System mit verifizierter Firmware verwendet werden. Dies führt zu einer längeren Lebenszeit für IoT-Geräte, da deren Software die Integrität beibehält. Mithilfe des implementierten Secure Elements können weitere Sicherheitsfunktionen für Embedded Geräte verwirklicht werden, die keinen manipulationssicheren Speicher integriert haben.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Varun Deshpande et al. SEBS: A Secure Element and Blockchain Stratagem for Securing IoT. In *2019 Global Information Infrastructure and Networking Symposium (GIIS)*. Engineers, Institute of Electrical and Electronics, 2019.

[3] Varun Deshpande, Laurent George, and Hakim Badis. Pulsec: Secure element based framework for sensors anomaly detection in industry 4.0, 2019.

[4] Sylvain Guilley, Michel Le Rolland, and Damien Quenson. Implementing Secure Applications thanks to an Integrated Secure Element, 2021.

[5] Christos Profentzas et al. Performance of secure boot in embedded systems. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. Engineers, Institute of Electrical and Electronics, 2019.

[6] Steffen Sanwald et al. Secure boot revisited: challenges for secure implementations in the automotive domain. In *Int. J. Transp. Cyber. & Privacy*. SAE International, 2020.

Konzeption eines Frameworks zur Messung von Datenqualität und prototypische Umsetzung.

Dennis Herzog **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Problemstellung

Im Zeitalter von Big Data und IoT werden immer mehr Daten generiert. Diese gewonnenen Daten bilden die Grundlage für verschiedenste Projekte in den Bereichen Data Sciene, Künstliche Intelligenz und Business Intelligence. Fehlerhafte oder ungenaue Daten haben daher einen maßgeblichen Einfluss auf die Ergebnisse der Projekte. Laut einem Bericht von Gartner aus dem Jahr 2021 verlieren Unternehmen jährlich zwölf Millionen Dollar aufgrund von schlechter Datenqualität. Einer der Gründe für diesen Verlust ist, dass Entscheidungsfindungen im Unternehmen durch Daten untermauert werden. Daher ist die Grundlage für eine gute datenbasierte Entscheidung eine hohe Datenqualität [4]. Weitere Treiber für das Thema Datenqualitätsmanagement können Abbildung 1 entnommen werden.

Abb. 1: Treiber für Datenqualitätsmanagement [2]

Ziel der Arbeit

Ziel der Arbeit ist es, ein Framework zu entwickeln, mithilfe dessen die Datenqualität einer Datenquelle gemessen und falls nötig auch verbessert werden kann. Dieses Framework umfasst beispielsweise Schritte, die notwendig sind, um Anforderungen an Datenqualität zu definieren. Des Weiteren beinhaltet es notwendige Informationen, um sogenannte Datenqualitätsregeln aus diesen Anforderungen zu definieren und auch Templates für deren technische Implementierung. Dieses Framework soll dann im Anschluss prototypisch umgesetzt und abschließend evaluiert werden. Abbildung 2 zeigt eine grobe Übersicht über die Schritte, die zum Erreichen des Ziels durchgeführt werden sollen.

Abb. 2: Strukturierung der Arbeit [1]

Im Folgenden wird das Thema Datenqualität kurz angerissen sowie vereinzelte User Stories dargestellt, welche aus den Interviews ermittelt wurden.

Grundlagen

Das Thema Datenqualitätsmanagement findet im Jahr 1988 seinen Ursprung. In diesem Jahre begann Professor Richard Wang am MIT in Boston Daten als Produkte aus der Fertigungstechnik anzusehen. Dieses Produktdenken, wird in weiteren Arbeiten aufgegriffen und bildet eine Grundlage für Anforderungen an Datenqualität [8]. Diese sind sehr wichtig, da Datenqualität oft nach dem Ansatz ‚fitness for use' gemessen wird [7]. Das bedeutet, dass die Datenqualität höher ist, wenn die Anforderungen für den vorhergesehenen Einsatz der Daten erfüllt sind. Falls dies nicht zutrifft, ist die Datenqualität dementsprechend geringer [3]. Für die Messung der Datenqualität werden sogenannte Datenqualitätsregeln aufgestellt und diese einer Datenqualitätsdimension zugeordnet [6]. Zudem wird noch ein entsprechender Schwellenwert benötigt, welcher mit dem Ergebnis der Messung abgeglichen wird [3]. In der Literatur finden sich viele verschiedene Definitionen für Datenqualitätsdimensionen. Abbildung 3 zeigt die 15 Dimensionen, welche Wang und Strong im Jahre 1996 definiert haben.

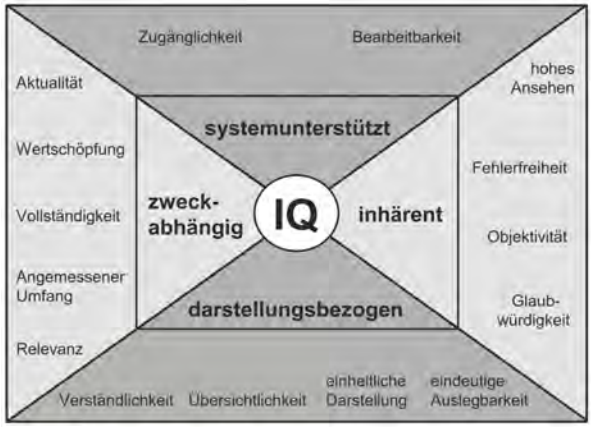

Abb. 3: Datenqualitätsdimensionen aus dem Framework von Wang und Strong [5]

Diese Datenqualitätsdimensionen sind in vier Kategorien unterteilt. Diese Kategorien unterscheiden sich beim Untersuchungsgegenstand. Dimensionen der Kategorie ‚Systemunterstützt' können nur untersucht werden, indem das Verarbeitungssystem oder das User Interface betrachtet wird. Dimensionen, welche der Kategorie ‚Inhärent' zugeordnet sind, befassen sich mit dem Inhalt der Daten. Dem Namen der Kategorie ‚Darstellungsbezogen' ist zu entnehmen, dass die Darstellung der Daten untersucht wird. Dimensionen, welche unter dem Begriff ‚Zweckabhängig' zusammengefasst werden, werden unter Berücksichtigung der Nutzung in Geschäftsprozessen gemessen [8].
Nun sind bereits zwei wichtige Bausteine des Datenqualitätsmanagements erwähnt worden. Hierbei handelt es sich um die Definition der Anforderung an die Datenqualität sowie die entsprechende Messung.
Neben diesen beiden Bausteinen gibt es weitere Aufgaben, die häufig in einem Datenqualitätsprojekt durchgeführt werden müssen. Dazu zählt unter anderem die Ursachenanalyse. Ziel hierbei ist es, den Ursprung eines fehlerhaften Datensatzes zu ermitteln und die Gründe für den Fehler zu finden. Sobald die Ursache identifiziert ist, ist es wichtig zu prüfen, ob eine Behebung des Problems einen Nutzen bringt, welcher größer als der Aufwand zum Lösen des Problems ist. Wenn dies der Fall ist, muss das Problem an der Quelle behoben werden, sodass dieses nicht mehr auftreten kann. Gegebenenfalls müssen zudem die entsprechenden Datensätze korrigiert, standardisiert oder formatiert werden [3].

Rollen und User Stories

Im Unternehmenskontext lassen sich verschiedene Rollen identifizieren, die Anforderungen an Datenqualität haben. Die erste Rolle stellt der Data Governor dar. Dieser ist unter anderem für die rechtliche Absicherung von Datenprojekten im Unternehmen verantwortlich. Eine weitere Rolle ist der Data Owner. Dieser ist für eine bestimmte Datenquelle verantwortlich und damit auch für deren Qualität. Wenn eine Entscheidung bezüglich der Daten getroffen werden muss, wird dies vom Data Owner getan. Zudem spielen auch die Nutzer der Daten eine wichtige Rolle beim Thema Datenqualität, da diese Anforderungen an die Qualität der Daten stellen. Bei den Nutzern der Daten lässt sich unter anderem der Data Scientist identifizieren. Der Data Scientist hat ausgeprägte Fähigkeiten in den Bereichen Statistik und Machine Learning und nutzt diese, um Erkenntnisse aus den Daten zu ziehen und einen Mehrwert zu generieren. Stakeholder sind in diesem Falle alle Rollen, welche ein Interesse an den Daten haben. Weitere Stakeholder neben den genannten Rollen sind der Product Owner oder der Data Analyst. In Abbildung 4 ist für jede dieser Rollen eine User Story dargestellt.

Rolle	User Story
Data Governor	Als Data Governor ist es mir wichtig, dass die Data-Governance-Regeln beim Thema Datenqualität berücksichtigt werden.
Data Owner	Als Data Owner möchte ich, dass meine Anforderungen an die Datenqualität (verwendete Daten, Schwellenwerte, Risiken) berücksichtigt werden.
Data Scientist	Als Data Scientist möchte ich Transparenz über die Datenqualität erhalten, um zu bewerten, wie Aussagekräftig meine Prognosen sind.
Stakeholder	Als Stakeholder ist es mir wichtig, dass die Datenqualität durch ein Monitoring überwacht wird.

Abb. 4: Rollen und User Stories [1]

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Marco Geuer. Datenqualitätsmanagement – Data Quality Management. https://www.haufe-akademie.de/blog/themen/controlling/datenqualitaetsmanagement-engl-data-quality-management/, 2021.

[3] Deborah Henderson and Susan Earley. *DAMA-DMBOK Data management body of knowledge*. Data Administration Management Association, 2017.

[4] Gartner Incorporated. 12 Actions to Improve Your Data Quality. https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality, 2021.

[5] Jan P. Rohweder, Gerhard Kasten, Dierk Malzahn, Andrea Piro, and Joachim Schmid. Informationsqualität - Definition, Dimensionen und Begriffe. In *Daten- und Informationsqualität*. Hildebrand, Knut; Gebauer, Marcus; Mielke, Michael; Hinrichs, Holger, 2018.

[6] Laura Sebastian-Coleman. *Measuring Data Quality for Ongoing Improvement*. Elsevier Science, 2012.

[7] Richard Y. Wang and Diane M. Strong. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 1996.

[8] Niels Weigel. Datenqualitätsmanagement – Steigerung der Datenqualität mit Methode. In *Daten- und Informationsqualität*. Hildebrand, Knut; Gebauer, Marcus; Mielke, Michael; Hinrichs, Holger, 2018.

Digitalisierung der Anlageberatung und Vermögensverwaltung – Trends sowie Erfolgsaussichten

Tom Junghanns **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Das Internet und die Digitalisierung betreffen wie viele andere Wirtschaftsbereiche auch die Finanzbranche und stellen eine wesentliche Herausforderung dar. Dabei ist zum einen die gesamte Branche betroffen und zum anderen einzelne Geschäftsmodelle und Geschäftsbereiche [3]. Auch in ihrem traditionellen Vermögensverwaltungs- und Beratungsgeschäft befindet sich die Finanzbranche in einem großen Wandel. In den letzten Jahren sind neue, innovative Teilnehmer in ihr Geschäft eingetreten, angetrieben von der Weiterentwicklung der Informationstechnologie sowie der Notwendigkeit, Transparenz und Zugänglichkeit in die seit langem etablierte und traditionelle Anlageberatung und Vermögensverwaltung zu bringen [4].

Ziel der Arbeit

In dieser Arbeit werden sowohl aktuelle Herausforderungen der Digitalisierung in der Finanzbranche als auch theoretische Grundlagen des Themas Anlageberatung und Vermögensverwaltung beschrieben. Dabei wird auf aktuelle Entwicklungen in den Bereichen FinTechs und Robo-Advice eingegangen sowie deren Vor- und Nachteile erörtert.

Anlageberatung und Vermögensverwaltung

Der Hauptzweck der Finanzbranche besteht darin, Einzelpersonen oder Institutionen dabei zu helfen, angemessene Investitionen zu tätigen und die individuellen Ziele von Anlegern unter Berücksichtigung lang- oder kurzfristiger Anlagehorizonte zu erreichen [4]. Laut Definition der Anlageberatung in § 1 Abs. 1a Satz 2 Nr. 1a KWG ist dies eine einmalige Empfehlung hinsichtlich der zu tätigenden Investitionen bzw. zu kaufenden Finanzprodukten. Dabei wird die persönliche Situation des Kunden berücksichtigt und auf seine Bedürfnisse eingegangen [7]. Viele Anleger vertrauen ihren Beratern bei Finanzentscheidungen, da sie meist selbst keinen ausreichenden Überblick über die wirtschaftlichen Zusammenhänge sowie über die am Markt vorhandenen Finanzprodukte haben. Als Resultat erwarten sie eine fachkundige und auf ihre persönlichen Bedürfnisse zugeschnittene Beratung. Die Verantwortung der Käufe bzw. Investitionen bleibt auf Seiten des Kunden [7]. Verglichen zur Anlageberatung hat die Vermögensverwaltung allerdings diverse Bedeutungen und ist weder in der Theorie noch in der Praxis eindeutig definiert. Da die Art und der Umfang des betreuten Vermögens je nach Kunde unterschiedlich ausgeprägt sind, kann der Begriff Vermögensverwaltung unterschiedlich interpretiert werden [1]. Nach der Definition von Spremann steht die Geld- oder Kapitalanlage und das Management des Portfolios im Vordergrund der Vermögensverwaltung [9]. Doch obwohl die Anlageberatung als auch die Vermögensverwaltung den Kunden bei der Vermögensanlage unterstützen, können die beiden Dienstleistungen hinsichtlich der rechtlichen Befugnis des Anbieters voneinander abgegrenzt werden. Während der Anbieter der Vermögensverwaltung auch ohne Rücksprache mit dem Kunden und nach eigenem Ermessen über das Anlagevermögen verfügen kann, werden bei der Anlageberatung nur Kauf- oder Investitionsempfehlungen vorgeschlagen [5].

FinTechs

Bei der Analyse des Wettbewerbs innerhalb des Finanzdienstleistungssektors darf jedoch der Eintritt von Technologieunternehmen (meist als ‚FinTech' bezeichnet) in die Erbringung von Finanzdienstleistungen nicht vernachlässigt werden [2]. Eine einheitliche Definition des Begriffes _FinTech_ existiert bislang nicht, jedoch sind unter dem Akronym Unternehmen bekannt, welche den Kunden technische Lösungen im Finanzsektor anbieten. Dieser Oberbegriff setzt sich aus den Wörtern _financial services_ und _technology_ zusammen und fand seinen Ursprung in den USA [5]. Die Startup-ähnlichen Unternehmen zeichnen sich durch disruptive und ausschließlich digitale Geschäftsmodelle aus, die

sie im Internet als kostengünstige Alternativen zu den Dienstleistungen traditioneller Banken anbieten. Dabei wollen sie bisherige Geschäftsmodelle oder IT-Systeme anhand von neuen Ideen revolutionieren und konventionelle Geschäftsmodelle vom Markt verdrängen. Das Herzstück dieser digitalen Alternativen sind meistens komplexe Algorithmen oder patentierte Software-Anwendungen [5]. Im Bereich der traditionellen Bankdienstleistungen werden insbesondere der Zahlungsverkehr, Finanzierungsangebote und Geldanlagen von FinTechs in Form von Online-Bezahldiensten, Crowdfunding oder Robo-Advisors abgedeckt. Dabei wird auf die maximale Nutzerfreundlichkeit, Übersicht und Transparenz der Dienstleistung geachtet, sodass der Kunde diese schnell und einfach in Anspruch nehmen kann [5].

Robo-Advisors

Einer dieser bereits erwähnten disruptiven Geschäftsmodelle sind die Robo-Advisors, die innerhalb des FinTech-Ökosystems entstanden sind. Eine einheitliche Definition lässt sich in der Literatur nicht finden, da es mehrere Geschäftsmodelle zu geben scheint, die sich durch den Grad des passiven Managements, der Tiefe der Anlageautomatisierung, Selbstbewertungsmechanismen und der Zielkundschaft unterscheiden [8]. Jedoch werden unter dem Begriff Robo-Advisor technische Systeme zusammengefasst, die einer Person bei der Anlage von Kapital oder der Vermögensverwaltung helfen [5]. Anders ausgedrückt sind Robo-Advisors digitale Plattformen, die interaktive und intelligente Benutzerassistenzkomponenten umfassen und Kunden mithilfe von Informationstechnologie durch einen automatisierten Finanzberatungsprozess führen [4]. Der Kontakt zwischen Kunde und Anbieter erfolgt dabei ausschließlich digital. Die Anlagestrategie bzw. Empfehlung wird durch Entscheidungsbäume oder computerbasierte Algorithmen generiert [5]. Durch den hohen Automatisierungsgrad der Robo-Advisors kann die Beratung und Verwaltung günstiger angeboten werden als es bei einer klassischen Anlageberatung oder Vermögensverwaltung der Fall ist [5].
#Ausblick In der heutigen Zeit empfinden immer mehr Kunden den technischen Fortschritt der Banken als zu langsam. In einer Zeit, in der die ständige Verfügbarkeit und eine reibungslose Nutzbarkeit der Minimalstandard geworden ist, bleibt traditionellen Banken keine andere Wahl als eine Veränderung des Geschäftsmodells [6]. Es gibt einige Hinweise darauf, dass die dynamische Expansion der FinTechs zu bahnbrechenden Veränderungen in der Marktstruktur führen kann. Diese Veränderungen können erhebliche Auswirkungen auf den Wettbewerb in der Finanzbranche sowie für etablierte Institutionen haben [2]. In Deutschland könnten die FinTech-Unternehmen in den nächsten Jahren rund ein Drittel aller Bankeinnahmen gefährden. Das disruptive Potenzial von Robo-Advisors, das durch verschiedene Schlüsselaspekte hervorgerufen wird, hat mehrere Implikationen: Eine Folge des vollautomatisierten Kundenprofilierungs- und Anlageprozesses ist die erheblich niedrige Gebührenstruktur und Mindestanlage. Als Resultat führte der Aufstieg von Robo-Advisors zu einer neuen Klasse von Low-Budget-Investoren, die zuvor nicht von traditionellen Finanzberatern bedient wurden [4].

Abb. 1: Prognose verwaltetes Vermögen durch Robo-Advisors in Deutschland [6]

Ursprünglich sollten die Robo-Advisors einen Durchbruch in einem durch Banken unterversorgten, aber stark digitalisierten Marktsegment bewirken. Allerdings begannen sie jedoch sehr schnell ein breites Publikum an Investoren anzusprechen, wie z.B. wohlhabende und vermögende Privatpersonen, die eigentlich im Mittelpunkt der Strategien etablierter Unternehmen stehen [8]. Wie in der Abbildung 1 verdeutlicht wird, nimmt das durch Robo-Advisor verwaltete Vermögensvolumen stark zu und wird bis zum Jahr 2023 auf über 30 Milliarden Euro geschätzt [6].

Literatur und Abbildungen

[1] J. Gulden. *Automatisierte Geldanlage – Determinanten und Einflussbedingungen der Akzeptanz von Investment Management FinTechs*. Springer Gabler, 2018.

[2] J. Harasim. FinTechs, BigTechs and structural changes in capital markets. In *The Digitalization of Financial Markets – The Socioeconomic Impact of Financial Technologies*, chapter 5, pages 80–100. Adam Marszk und Ewa Lechman, 2021.

[3] J. Hastenteufel and F. Ganster. *Einflussfaktoren auf die Akzeptanz von Robo Advisors – Digitale Kommunikation in der Anlageberatung*. Springer Gabler, 2021.

[4] D. Jung, F. Glaser, and W. Köpplin. Robo-Advisory: Opportunities and Risks for the Future of Financial Advisory: Recent Findings and Practical Cases. In *Advances in Consulting Research*, pages 405–427. Volker Nissen, 2019.

[5] T. B. Madel. *Robo Advice – Aufsichtsrechtliche Qualifikation und Analyse der Verhaltens- und Organisationspflichten bei der digitalen Anlageberatung und Vermögensverwaltung*. Nomos Verlagsgesellschaft Baden-Baden, 2019.

[6] P. Petre and J. Wunderlich. Analyse des Robo-Advisor Marktes für die Anlageberatung und Vermögensverwaltung. *Landshuter Arbeitsberichte zur Wirtschaftsinformatik*, 2021.

[7] B. Schloz. *Künstliche Intelligenz im Finanzdienstleistungssektor – Evaluierung des Meinungsbildes von Privatkunden zu Robo-Advice*. MA Akademie Verlags- und Druck- Gesellschaft mbH, 2020.

[8] P. Sironi. *FinTech Innovation – From Robo-Advisors to Goal Based Investing and Gamification*. Wiley, 2016.

[9] K. Spremann. *Vermögensverwaltung*. De Gruyter Oldenbourg, 1999.

Entwicklung von Bewertungskriterien zur Erfolgsmessung des Einsatzes von agilen Projektmanagementmethoden bei SAP-AddOn Projekten

Alexander Kaiser **Thomas Rodach**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma AFI Solutions GmbH, Stuttgart

Einleitung

Projekte im SAP-Bereich sind hoch komplex und besitzen das Problem, dass die Dauer der Durchführung oft sehr lang sein kann. AFI Solutions bietet für SAP ERP und SAP S/4HANA sogenannte Add-ons an, die Kunden erwerben können und die Funktionalitäten der SAP-Software erweitern können. Diese Add-ons können relativ einfach, ohne das produktive System zu verändern, installiert werden. Ähnlich wie SAP selbst, kann die Software in einer Standardkonfiguration ausgeliefert werden, oder es kann davor noch aufwendiges Customizing durchgeführt werden. Viele Kunden, die SAP nutzen, besitzen keine standardisierten Prozesse, sondern sind häufig hoch individuell, sodass das Add-on entsprechend angepasst werden muss. Außerdem bietet AFI den Kunden bei Bedarf an, zusätzliche Funktionen zu entwickeln. Bis jedoch alle Anforderungen vom Kunden geklärt werden und das Pflichtenheft fertiggestellt wird, vergeht eine lange Zeit, in der noch nichts entwickelt werden kann. Agile Projektmanagement Methoden sollen bei diesen Problemen Hilfe schaffen und sogar SAP selbst hat in ihren eigenen Methoden agile Praktiken integriert. Die AFI behilft sich seit 3 Jahren schon mit dem Scrum Framework für die Durchführung ihrer Projekte.

Zielsetzung

Mit der Bachelorarbeit soll nun bewertet werden, wie sich der Erfolg der Projekte bei AFI nach der Einführung von Scrum entwickelt hat. Zunächst sollen die Prozesse und Praktiken von AFI, rund um Scrum erfasst werden, um es in einen Kontext zu bringen mit den gewonnen Erkenntnissen aus der Theorie. Das bringt eine qualitative Bewertung und es soll auf die Abweichungen zu bewährten Praktiken in Scrum aufmerksam gemacht werden. In diesem Rahmen werden Handlungsempfehlungen zur Verbesserung der Arbeitsstruktur mit den Best Practices aus der Praxis formuliert. Darüber hinaus sollen die Daten aus vergangenen Projekten analysiert werden und anhand dieser soll bewertet werden, in welchen Fällen die Scrummethode erfolgreicher war als das klassische Vorgehen. Die gewonnenen Ergebnisse sollen dann auch dabei helfen in Zukunft anhand dieser Kriterien bewerten zu können, welche Projekte mit Scrum abgewickelt werden sollten und welche nicht.

Projektmanagement

Ein Projekt ist ein einmaliges Vorhaben, um ein Ziel zu erreichen und diese haben ihren Ursprung aus alten Bauprojekten wie z. B. die Pyramiden in Ägypten oder der Turmbau zu Babel. Damit solche Projekte erfolgreich durchgeführt werden können, benötigt es effektives Projektmanagement, um die Arbeit im Projekt zu strukturieren. [1]
Das klassische Projektmanagement besitzt ein sequenzielles und lineares Vorgehen. Dabei existieren in den klassischen Methoden Phasen, die nacheinander durchlaufen werden und jede Phase besitzt wiederum eigene Aufgaben und Ergebnisse. ASAP, eine Methode von SAP, die für die Einführung von ERP-Software entwickelt wurde, basiert auf genaue dieses Phasenmodell. [4]

Abb. 1: ASAP Projektphasen [4]

Die Problematik ist, dass bei den klassischen Projektmethoden vorausgesetzt wird, dass alle Anforderungen an das Ergebnis schon definiert sein müssen. Sie sind

nicht in der Lage, Anforderungen effektiv umzusetzen, die erst im Laufe des Projektes entstehen. [1]
Der IT-Sektor steht im ständigen Wandel und ist stark von disruptiven Technologien geprägt. Das stellt ein großes Problem für Projekte in der IT dar, denn Anforderungen, die zu Beginn eines Projektes definiert wurden, könnten bei Projektabschluss schon veraltet sein. Dieses Risiko solle also immer beachtet werden, wenn IT-Projekte mit einem klassischen linearen Projektmanagement durchgeführt werden. Für diesen Einsatzzweck glänzt das agile Projektmanagement, denn dieses bietet ein Rahmenwerk, das Veränderungen im Laufe des Projektes zulässt. [5]
Die agilen Projektmanagementmethoden werden durch ihre inkrementelle und iterative Vorgehensweise gekennzeichnet. Das bedeutet, dass das Ziel eines Projekts in mehrere kleinere Inkremente aufgeteilt wird, die alle die Zielbedingungen erfüllen sollen. Dafür werden diese Inkremente in kurzen Iterationen in Scrum als Sprints bezeichnet, entwickelt, getestet und dem Kunden präsentiert. Das Inkrement sollte schon Bestandteile des Ziels erfüllen können, sodass der Kunde rückmelden kann, ob das Projekt in die richtige Richtung geht. Unter Umständen können Anpassungen am Ziel vorgenommen werden oder neue Anforderungen gestellt werden. Durch diese Methode lässt sich das Risiko des Projektes reduzieren und es wird verhindert, dass etwas entwickelt wird, das nutzlos ist. Gemeinsam mit der Präsentation der neuen ERP-Software S/4HANA, wurde auch eine neue Methode angekündigt, die auch das Scrum Framework beinhalten soll. Diese Methode nennt sich SAP Activate und löst die ASAP Methode ab, weil diese für alle ERP-Projekte angewendet werden kann. [3]

Abb. 2: SAP Activate Projektablauf [2]

Ausblick

Bei AFI ist zu unterscheiden zwischen zwei Formen von Scrum. Denn das Unternehmen nutzt Scrum zum einen als inkrementelle und iterative Methode, um Projekte durchzuführen. Das bedeutet, dass dem Kunden zuvor mehrere Inkremente vorgestellt werden und diese nach und nach in das Produktivsystem eingesetzt werden. Die andere Form von Scrum ist die interne Organisation des Entwicklerteams. Das bedeutet, dass die Werte, Artefakte und Ereignisse von Scrum genutzt werden, um die Zusammenarbeit im Team zu verbessern und die Selbstorganisation zu fördern. Hier werden auch Sprints durchgeführt, jedoch gibt es keine inkrementelle Auslieferung von Produkten an den Kunden. Diese beiden Formen werden getrennt voneinander analysiert, um zuerst zu prüfen, wie sich die inkrementelle Auslieferung auf den Erfolg auswirkt und dann, ob die internen Scrumpraktiken die Produktivität des Teams steigert.

Literatur und Abbildungen

[1] Jörg Preußig. *Agiles Projektmanagement : Agilität und Scrum im klassischen Projektumfeld*. Haufe-Lexware GmbH & Co. KG, 2020.

[2] Jan Simmonds and James Olcott. Ten Tips for Successful SAP Leadership. https://eursap.eu/2020/03/05/blog-ten-tips-for-successful-sap-leadership/, 2020.

[3] Vinay Singh. *Manage Your SAP Projects with SAP Activate: Implementing SAP S/4HANA*. Packt Publishing, 2017.

[4] Aigner smartplm. Accelerated SAP. https://smart-plm.com/glossary/accelerated-sap/#:~:text=Accelerated%20SAP%20(ASAP)%20stellt%20eine,den%20gesammelten%20Erfahrungen%20vorausgegangener%20Projekte., 2020.

[5] Cornelia Zehbold and Melanie Chowanietz. Digitalisierung des Design Thinking. *Technische Hochschule Ingolstadt*, 2021.

Generische Ansätze zur Datenanalyse bei kleinen und mittelständigen Unternehmen

Ilhan Kasumovic **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Die Zeiten in der die Unternehmen von Intuitionen geleitet wurden sind vorbei. In der heutigen Zeit wird jedes Unternehmen von umfangreichen Datenerhebungen und Datensammlungen geprägt. Aus den gewonnenen Daten wird mittels einer effizienten und automatischen Datenanalyse durch die Anwendung Künstlicher Intelligenz Mehrwert erzeugt. Hierdurch werden Zusammenhänge und individuelle Muster erkannt, um kommende Verbrauchertrends zu erkennen geben oder mögliche Risiken minimieren. Dies ermöglicht Vorhersagen zu treffen oder neue Geschäftschancen zu kreieren. Ein Beispiel wären Prognosen über das Kundenkaufverhalten zu generieren, Kundengruppen nach dem Kundenwert zu identifizieren oder automatisierte Prozesse zur Maschinenwartung zu erzeugen. Die Kernaufgabe von Datenanalysen ist es durch möglichst automatisierte Prozesse den Aufwand zu verringern und den Gewinn von Daten zu maximieren [4] [3].

Problematik von kleinen und mittleren Unternehmen

Kleine und mittelgroße Unternehmen (auch KMU genannt) sind oft eher zögerlich oder überfordert bei der Implementierung von Datenanalysen. Laut einer IDC-Studie nutzt nur jedes vierte KMU automatisierte Datenanalysen, um damit neue Businessideen zu entwickeln. Ein Grund dafür ist das fehlende Know-How und vor allem das fehlende dafür aber notwendige Personal nicht im Unternehmen vorrätig ist, da dieses Personal überall benötigt wird und eher die großen Unternehmen fündig werden. Des Weiteren ist das Budget der KMU oft nicht hoch genug um eine qualitative Datenerhebung durchzuführen. Dadurch können sich Komplikationen einspielen, indem aus den schlechten Datenmengen eine ungenaue Auswertung erzeugt wird, welche die KMU falsche Schlüsse ziehen lässt. Doch durch die Technologien und Methoden kann dies vermieden werden, um auch im Mittelstand den Unternehmenserfolg zu verbessern [4].

Möglichkeiten von KMU durch Datenanalysen

Durch die Visualisierung der Daten ist es möglich unterschiedliche Quellen in einem Dashboard zusammenzufassen, beispielsweise Leistungen verschiedener Filialen deren Entwicklung sowie Verkäufe. Business Insights können Hintergründe für bestimme Muster sichtbar machen, z.B. weshalb hat sich die Lieferung verspätet oder Grund für eine Beschwerde. Datenanalysen ermöglichen Erkenntnisse über den Markt und Wettbewerb, so wird den KMU deutlich gemacht, was sich die Kunden wünschen und was die Konkurrenten bieten. Predictive Maintenance versichert dem Unternehmen notwendige Wartungen oder Reparaturen vorherzusagen, um dadurch einen effizienteren Prozess und weniger Ausfallzeiten zu gewährleisten [4].

Abb. 1: Datenanalyse-Prozess [2]

Zielsetzung

Ziel dieser Arbeit ist es zu zeigen welche Erkenntnisse für kleine und mittelständige Unternehmen aus Daten zu gewinnen sind. Anhand der Nutzung aller Prozesse und IT-Technologien zur Analyse, wie auch zum erheben und sammeln von Daten. Effiziente Analysen sind

mit verschiedenen Methoden des Datenmanagements möglich. Data Mining untersucht Muster und Beziehungen mit großen Datenmengen. Das Data Cleaning bereinigt die aufgezeichneten Daten und eliminiert redundante oder doppelte Informationen. Mit Text Mining können KMUs Details aus Text extrahieren, um beispielsweise herauszufinden, wie oft ein Wort vorkommt. Descriptive Analytics wertet aktuelle und historische Daten aus, die den Ist-Zustand beschreiben oder sucht nach Mustern. Beispielsweise geht es um Daten aus der Vergangenheit die helfen den aktuellen Zustand zu verstehen. Diagnostic Analytics zielt darauf ab, die Ursache für den Ist-Zustand zu finden und um die Frage zu beantworteten "Warum tritt dieses Muster auf?". Mit der Predictive Analytics wird es ermöglicht, basierend auf den Ergebnissen von deskriptiven und diagnostischen Analysen, Vorhersagen zu treffen. Dies wird durch ausgefeilte Algorithmen ermöglicht, jedoch sind es oft nur Schätzungen aufgrund der vorhanden Daten. Die Frage mit der man sich hier beschäftigt ist: "Was wird in Zukunft passieren?". Prescriptive Analytics sagt voraus, welche Maßnahmen getroffen werden sollten, um ein bestimmtes Ziel zu erreichen oder das volle Potential auszuschöpfen zu können.
Gemäß dem Analytics-Reifegradmodell von Gartner gibt es vier Möglichkeiten, Daten zu analysieren, wobei von der einfachsten bis zur anspruchsvollsten unterschieden wird. Je komplexer die Methode, desto mehr Wert (Wettbewerbsvorteile), kann sie bringen [4] [1].

Abb. 2: Methoden der Datenanalyse [1]

Zusammenfassung und Ausblick

Einem Zitat von Microsoft COO Bob Herbold zufolge, lässt sich draus schließen wie wichtig Datenanalysen für die heutigen Unternehmen sind. Seiner Meinung nach lag der Erfolg großer Organisationen immer in der Datenauswertung und Datenanwendung. Doch in der Neuzeit bestehen grenzenlose Möglichkeiten zu Verbesserung der Abläufe und Ergebnisse. Sieht man sich die Entwicklung der Unternehmen an, macht sich ein zunehmender Trend bei der Nutzung von IT-Technologien bemerkbar, welcher in naher Zukunft weitere Methoden liefern wird, um Geschäftsprozesse zu vereinfachen oder genauere Analysen zu erstellen. Bob Herbold beendete sein Zitat in dem er sagte: "Diejenigen, die verstehen, wie Daten funktionieren und was sie leisten können, werden in der neuen Wirtschaft einen enormen Vorteil haben."[3].

Literatur und Abbildungen

[1] Alex Bekker. 4 Methoden der Datenanalyse: ein Überblick für tiefere Einblicke, ScienceSoft. https://www.scnsoft.de/blog/4-methoden-der-datenanalyse, 2021.

[2] Schütz Bianca. Schritt für Schritt von Rohdaten zu Erkenntnissen (Analyse-Fahrplan). https://bianca-schuetz.com/datenanalyse-prozess-kreislauf/, 2022.

[3] Minhaj Rais. Data analysis: Benefits and challenges for small and medium businesses, Der Kolabtree-Blog. https://www.kolabtree.com/blog/de/data-analysis, 2017.

[4] Spezialist Techdata. Was ist Data Analytics und wie können KMUs davon profitieren? https://dach.techdata.com/ibm-cloudpaks/news/was-ist-data-analytics-wie-konnen-kmus-profitieren/?it=news/was-ist-data-analytics-wie-konnen-kmus-profitieren, 2021.

Entwicklung eines ROS2-basierten Moduls zur Trajektoriengenerierung für eine Pick & Place Anwendung

Aaron Kiani **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Adolf Würth GmbH & Co. KG, Künzelsau

Einordnung

Eine Industrie ohne Roboter ist heutzutage kaum noch denkbar. Schon seit mehr als einem Jahrhundert beherrschen die Maschinen viele Fabrik- und Produktionshallen. Die unvergleichbare Präzision und die Geschwindigkeit der Bewegungen machen die maschinellen Arbeiter ihren menschlichen Entwicklern in vielen Feldern überlegen. Aus diesem Grund wächst die Anzahl an Robotern in der Industrie seit mehreren Jahrzehnten stetig an und die Anwendungsfelder werden immer vielfältiger.

Abb. 1: Stetiges Wachstum von operationsfähigen Robotern weltweit [1]

Einordnung der Arbeit mit der Firma Würth

Gerade für sich wiederholende Arbeiten wie Verpacken, Aufheben, Greifen, Platzieren, Sortieren, etc. eignen sich Industrieroboter hervorragend. Da die aufgezählten Tätigkeiten vor allem in der Logistik eine beachtliche Anzahl an Personalressourcen benötigen, ist auch die Würth-Gruppe mit ihren riesigen Logistikzentren und Produktionshallen an einer effizienteren Kommissionierung und Abwicklung der Arbeitsprozesse interessiert. Das aktuelle Bestreben des Unternehmens liegt in vielen Geschäftsbereichen darin, weniger Produkte einzukaufen und stattdessen selber zum Produzenten zu werden. Das gilt auch für den Bereich Robotik und dessen Einsatz im Konzern, weshalb dafür sogar eigens ein Teilbereich für die Forschung und Entwicklung gegründet wurde. Das magische Stichwort dafür ist "Logistik 2030ünd schließt viele Ideenansätze für Digitalisierung und Automatisierung ein. Ziel ist es, selber Expertise über die Anwendungsmöglichkeiten aktueller Industrieroboter anzuhäufen. Dazu gibt es verschiedene Ansätze, die von Computer Vision und Objekterkennung bis zu der direkten Programmierung eines Roboters reichen. Hier setzt auch die Abschlussarbeit an, welche sich mit dem Einsatz verschiedener Bewegungsalgorithmen befasst. Ziel ist es Objekte aus einer Quell Box A aufzuheben und in eine Ziel Box B über eine möglichst präzise, schnell berechnete und natürlich kollisionsfreie Trajektorie zu bewegen. Dafür wurde ein Greifarmroboter vom Typ UR10e der Firma Universal Robots eingekauft und mit einem Vakuum-Sauger ausgestattet, der mit dem programmierten Modul angesteuert wird.

Abb. 2: Greifarm Roboter UR10e von der Firma Universal Robots [1]

Aktueller Stand

Um wirklich unabhängig von anderen Softwareherstellern zu sein, wurde in der Abteilung F&E Robotik bei der Programmierung wirklich bei 0 angefangen. Ein vorangehender Bachelorand arbeitete die ersten Anforderungen an die Architektur heraus und implementierte das erste Fundament einer Softwarearchitektur für das Konzept eines typischen Kommissionier-Roboters. Die Implementierung basiert auf ROS2 und dem Planungs-Framework Moveit2. Mithilfe mehrerer, überwiegend in der Programmiersprache C++ geschriebener Module, werden vorprogrammierte Aufträge mit Artikeln (und natürlich deren Positionen) weitergegeben bis Moveit letztendlich basierend auf Zielkoordinaten auf unterer Abstraktionsebene eine Trajektorie generiert. Dieser Zustand beinhaltete jedoch lediglich das feste Abfahren einer vordefinierten Bahn ohne Kollisionsprüfung.

Ziele der Arbeit

Ziel der aktuellen Abschlussarbeit ist es, willkürliche Positionen von Artikeln angeben zu können, die dann aus der Quell Box entnommen und in die Ziel Box bewegt werden. Dazu muss der Roboter seine Umgebung kennen, mit Hindernissen umgehen können und Objektpositionen dynamisch verändern können. Dafür ist zuerst ein Modell des Roboters und allen anderen potenziellen Kollisionsobjekten zu erstellen. Im Anschluss sollen verschiedene Bewegungsalgorithmen werden und durch ein eigenes ROS2 Modul optimiert werden. Randbedingungen dafür sind: ~~* ~ die Erstellung eines User Interfaces, um dem Nutzer zu ermöglichen, Koordinaten einzugeben und Planungsparameter anzupassen. ~ die Bewegung soll kollisionsfrei sein. ~ die Generierung der Trajektorie soll möglichst schnell erfolgen. ~ die Geschwindigkeit, sowie Beschleunigung des Roboters sollen abhängig vom Gewicht des aufgehobenen Artikels sein. ~ es soll die Möglichkeit feste Punkt für die Trajektorie vorzugeben, durch die der Roboter sich bewegt. ~~~ Die technische Randbedingungen bestehen darin, dass die Steuerung des Roboters, sowie die Generierung der Trajektorie über ROS2, sowie dem Planungsframework Moveit2 basieren.

Abb. 3: Ein experimenteller UR10e beim Kommissionieren [1]

Ausblick

Sobald die Hauptziele der Arbeit erreicht sind, gilt es die Funktionalität der Architektur zu erweitern. Dabei sind potenzielle Themen die Packplanung (wie können Artikel platzsparend in einer Box angeordnet werden), die Anbindung an ein SAP-System oder auch die Erkennung von Objekten mittels eigener Kameras. Was die Abschlussarbeit angeht, besteht die Möglichkeit ein graphisches User Interface zu implementieren. Mit dessen Hilfe könnte ein Nutzer per Drag & Drop die Quell- und Zielposition von Artikeln in den Boxen angeben und dann die Planungsparameter anpassen. Was die Adolf Würth AG GmbH & Co. KG und die Vision "Logistik 2030ängeht, wird in Zukunft noch mehr Forschung im Teilbereich Robotik stattfinden und viel Potenzial für neue Ideen bieten.

Literatur und Abbildungen

[1] Eigene Darstellung.

Performance-Analyse von WebAssembly bei Frontup-Loading

Pascal Kneisel **Harald Melcher**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen

Einleitung

Mit der Erscheinung im Jahre 1995 entwickelt sich JavaScript bis heute zur führenden Programmiersprache für Webbrowser. Die Vorteile gegenüber objektorientierten Sprachen wie Java liegen leicht auf der Hand. So wird auf einen Compiler verzichtet und stattdessen ein Interpreter inklusive dynamischer Typisierung eingesetzt, der den Quellcode während der Laufzeit ausführt. Dies erleichtert dein Einstieg in die Programmiersprache, da hierfür keine besonderen Werkzeuge benötigt werden, außer dem Webbrowser. Die Auslieferung erfolgt direkt im Quellcode. Der fehlende Compiler ist hierbei auch eine der größten Schwachstellen JavaScripts. Der Quellcode kann hierbei nicht durch einen Compiler optimiert werden, die durchschnittliche Performance leidet hierunter. Mögliche Syntaxfehler werden ohne weitere Werkzeuge, wie zum Beispiel eines Linters, nicht erkannt. Aus diesem Grund hat Microsoft im Jahr 2017 WebAssembly veröffentlicht.

Was ist WebAssembly

Es ist nicht WebAssemblys Ziel, JavaScript vollständig zu ersetzen, sondern um es zu erweitern. [2] So wird es für performance- und speicherkritische Programmabschnitte an der Seite von JavaScript eingesetzt, um die Nachteile, die JavaScript mit sich bringt, auszugleichen. Hierbei wird der Quellcode für das WebAssembly-Modul in einer unterstützten Sprache wie C/C++, Rust oder C# geschrieben und mit dem WASM-Compiler kompiliert. Das kompilierte Programm wird daraufhin über JavaScript geladen und vom Webbrowser ausgeführt. Durch die Browser-Engine wird das WebAssembly-Programm auf Maschinenebene ausgeführt, wodurch im Vergleich zu JavaScript deutliche Performancesteigerungen zu erwarten sind. Die Abbildung 1 zeigt das Konzept WebAssemblys, vom Quellcode bis zum Ausführen im Webbrowser des Benutzers.

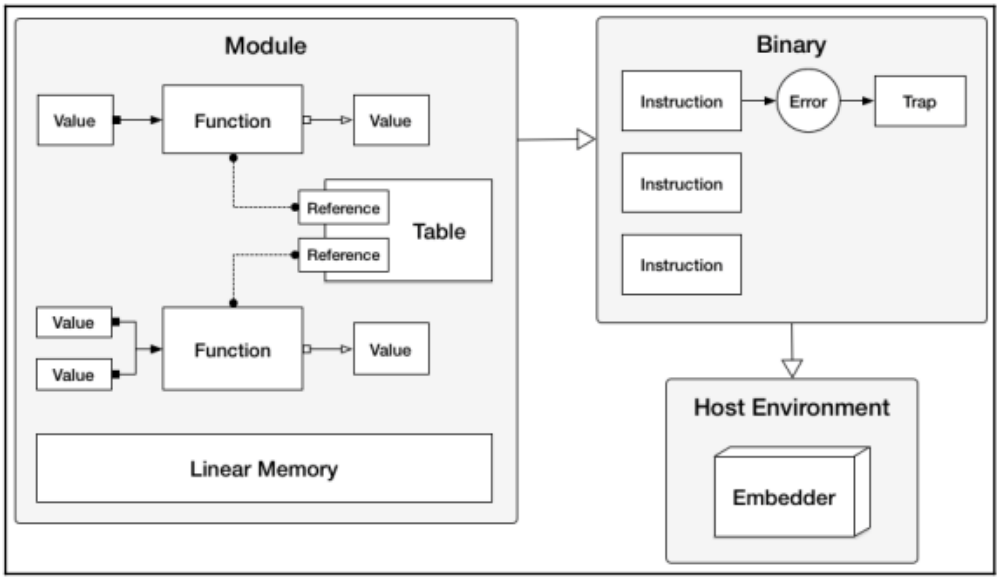

Abb. 1: Sprachkonzept von WebAssembly [2]

Problemstellung und Umsetzung

Moderne Webseiten laden Datenmengen im Hintergrund nach, während ein Teil der Webseite selbst schon ersichtlich ist. Dieser Teil kann „im Zweifelsfall auch erstmal nur Text [sein], während die Bilder noch laden." [3] Die Abschlussarbeit untersucht die Möglichkeiten von WebAssembly den Seitenaufbau zu unterstützen und somit die Ladezeiten von Webseiten zu verringern. Als Ausgangstechnologie wird hierfür ein React.js-Frontend verwendet, welches beim Seitenaufbau eine bestimmte Menge an Ausgangsdatensätzen erhalten und diese darstellen soll. Die native JavaScript-Implementierung wird hierbei um eine WebAssembly-Implementierung in Rust und eine weitere C# erweitert um den Seitenaufbau bei wachsenden Datensätzen zu verglichen.
Als nützliche Werkzeug-Sammlung zur Performance-Analyse von Webseiten ist seit dem Jahr 2012 sitespeed.io auf dem Markt und hat sich bis heute mit mehr als 4.000 Sternen auf GitHub (Stand 05.2022) als beliebtestes Werkzeug seiner Art etabliert. [1] Für die Analyse der Implementierungen werden hierbei Seitenaufrufe zweier URLs mittels verschiedener Metriken wie der Aufbaugeschwindigkeit, der gesamten Datengröße oder der nachladenden Datengröße verglichen.

Ausblick

Anhand der Implementierungen soll abseits der Performance-Analyse, der gesamte Entwicklungsprozess miteinander verglichen werden.
Ein wichtiges Kriterium sind die Debugging-Unterschiede, da die Browser-Engine nicht auf den Quellcode der WebAssembly-Programme zugreifen kann. Hierbei wird untersucht, was in der Praxis häufig verwendet und wie es angewendet wird um WebAssembly-Programme zu debuggen.
Außerdem werden die sprachlichen Einschränkungen von WebAssembly untersucht, im Detail vor allem ob unterschiede zwischen den verschiedenen Programmiersprachen in Form von Unterstützung der speziellen Features oder der Geschwindigkeit existieren.
Das abschließende Kriterium des Entwicklungsprozesses ist das Deployment-Verfahren, vom Kompilieren bis zur Ausführung im Browser des Endnutzers. Besondere Aspekte hierbei sind die Anbindung an die JavaScript-Schnittstelle und die Veröffentlichung auf dem Webserver.

Literatur und Abbildungen

[1] Peter Hedenskog, Tobias Lidskog, Jonathan Lee, et al. sitespeed.io. https://www.sitespeed.io/, 2022.

[2] Mike Rourke. *Learn WebAssembly*. Packt Publishing, 2018.

[3] Robert Weller and Ben Harmanus. *Content Design*. Carl Hanser Verlag, 2 edition, 2021.

Integration von Seamless Payment in eine event-getriebene Microservice-Architektur

Jonas Koringer **Harald Melcher**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Gebit Solutions GmbH, Stuttgart

Einleitung

Aus einer im Jahr 2021 von dem EHI Retail Institute publizierten Studie geht hervor, dass sich die Anzahl der Einzelhandelsgeschäfte mit Self-Checkout-Lösung zwischen 2019 und 2021 fast verdoppelt hat. Die Anzahl der mobilen Self-Scanning-Lösungen hat sich im selben Zeitraum sogar fast verzehnfacht. Zudem hat unter Anderem die Corona-Pandemie zu veränderten Zahlungsgewohnheiten der Menschen geführt, konkret wurden weniger Barzahlungen getätigt [3]. Ein weiterer Schritt Richtung Vollautomatisierung des Einkaufsprozesses stellt ein sogenannter Seamless Checkout dar. Dabei muss der Kunde im Laden gar keinen Zahlprozess mehr ausführen. Vielmehr wird das Verlassen des Ladens registriert und eine Zahlung wird im Hintergrund ausgeführt. Dieser Bezahlprozess wird in diesem Kontext als Seamless Payment bezeichnet.

Ziel der Arbeit

Ziel dieser Arbeit ist es, einen solchen Seamless Checkout-Prozess in eine bereits bestehende Self-Scanning-Anwendung zu integrieren. Die Anwendung besteht aus einer App für mobile Endgeräte. Die Backend-Funktionalitäten werden durch eine Microservice-Landschaft bereitgestellt. Bei der Implementierung soll evaluiert werden, inwiefern sich die Event-getriebene Architektur der Microservices eignet, eine fachliche Anforderung wie einen Seamless Payment Prozess abzubilden.

Ausgangslage

Die angesprochene bereits existierende Self-Scanning-Applikation bietet Nutzern die Möglichkeit sich zu registrieren und anschließend eine Supermarkt-Filiale zu betreten. Im Markt lassen sich über das Scannen von Produkt-Barcodes gewünschte Artikel in den virtuellen Warenkorb legen. Der Bezahlvorgang kann beim Verlassen des Ladens entweder Bar an einer Kasse, oder durch ein Online-Payment (z.B. durch PayPal) erfolgen.
Das Backend, welches dem Client die Funktionalität zur Verfügung stellt, besteht aus diversen fachlich geschnittenen Microservices. Beispielsweise existiert ein Service, welcher den digitalen Warenkorb verwaltet, ein weiterer ist für das Abwickeln der Zahlungen zuständig. So wird aus einer Vielzahl an Teilfunktionalitäten, welche die einzelnen Services bereitstellen, die gesamte Funktionalität der Anwendung aggregiert.

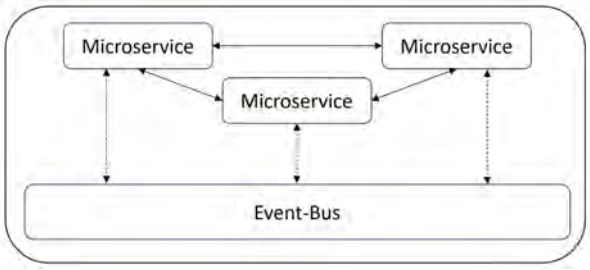

Abb. 1: Exemplarische Darstellung einer event-getriebenen Architektur [1]

In Abb. 1 ist exemplarisch eine eventgetriebene Microservice-Landschaft abgebildet. Die direkte synchrone Kommunikation zwischen den einzelnen Services findet anhand von RESTful-Schnittstellen statt. Zusätzlich können die Services auch asynchron anhand von Events kommunizieren. Dabei kommt ein sogenannter Event-Bus zum Einsatz. Bei den Microservices kann man zwischen Event Emittern und Event Consumern unterschieden. Dabei publiziert ein Event Emitter Events. Dies geschieht indem der Service das Event beim Auftreten gewisser fachlicher Auslöser auf den Event-Bus schreibt. Ein Event Consumer empfängt dieses Event und kann es bei Bedarf entsprechend verarbeiten [4]. Ein Beispiel für diese Form der asynchronen Kommunikation ist das Anlegen eines neuen Kunden. Bei erfolgreichem Registrierprozess publiziert der Microservice der für die Kundenverwaltung zuständig ist ein Event. Andere Services, die

ebenfalls Kundeninformationen benötigen empfangen dieses Event und verarbeiten es entsprechend ihrer internen Geschäftslogik.
Eine solche Architektur bietet verschiedene Vorteile: Zum einen erlaubt sie eine sehr lose Kopplung zwischen Services, was Änderungen und funktionale Weiterentwicklungen begünstigt. Zudem ist die Bearbeitung der Events zeitlich entkoppelt, sie muss also nicht synchron stattfinden [4].

Implementierung

Es bieten sich mehrere Möglichkeiten an, das Verlassen des Kunden aus dem Store in der App abzubilden. Die einfachste Lösung wäre es, den Kunden einen QR-Code scannen zu lassen, welcher dann als Checkout-Signal interpretiert wird. Eine weitere Möglichkeit wäre es, mit Geofencing das Verlassen des Ladens auf Basis eines GPS-Signals zu erkennen (vgl. Abb. 2).

Abb. 2: Einsatz von Geofencing um das Verlassen von Örtlichkeiten zu erkennen [2]

Eine andere Möglichkeit wäre die Anbringung von NFC-Tags am Ladenausgang. Im weiteren Verlauf der Arbeit wird evaluiert, welche der genannten Methodiken am besten geeignet ist die funktionalen Anforderungen umzusetzen. Die passendste Variante wird anschließend in die Applikation integriert.
Das Backend muss dahingehend erweitert werden, dass es das Event, welches das Verlassen des Kunden aus dem Laden darstellt erkennt, den digitalen Warenkorb abschließt und anschließend eine Zahlung gegen einen sogenannten Payment-Service-Provider (PSP) initialisiert. Im Zuge dessen soll evaluiert werden, inwiefern die im Backend gewählte Event-getriebene Architektur die Umsetzung einer solchen fachliche Anforderung begünstigt. Des Weiteren sollen eventuelle Vor- und Nachteile dieser Methodik gegenüber einer rein synchronen Kommunikation zwischen einzelnen Services identifiziert und bewertet werden.

Problemstellungen und Ausblick

Bei der fachlichen Anforderung an einen Seamless Checkout Prozess drängen sich mehrere Problemstellungen auf. Konkret stellt sich die Frage inwiefern mit Reklamationen (z.B. nach fehlerhaftem bezahltem Betrag) umgegangen werden kann. Dieser Fall kann beispielsweise auftreten, wenn ein Artikel versehentlich oder mehrfach gescannt wurde. Ein Lösungsansatz hierbei wäre die Integration eines Service-Portals, in welchem der Kunde fehlerhafte Rechnungen reklamieren kann.
Eine technische Weiterentwicklung des Konzepts besteht theoretisch darin, dass Scannen der Produkt-Barcodes durch eine automatisierte KI-gesteuerte Bilderkennung zu ersetzen. Dabei filmen Kameras durchgehend den Supermarkt, um zu erkennen, wenn ein Kunden einen bestimmten Artikel in seinen Warenkorb legt. Das System ordnet dem passenden digitalen Warenkorb dann das entsprechende Produkt zu. Eine Kombination aus diesem Verfahren und dem in dieser Arbeit vorgestellten Seamless Checkout Prozess würde einen vollständig kassenlosen Einkaufsprozess ermöglichen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Mathijs Hoek. Making mobile apps smarter with geofence technology. https://medium.com/mobgen/making-mobile-apps-smarter-with-geofence-technology-36c6442ab406, 10 2016.

[3] EHI Retail Institute. Self-Checkout: Markterhebung 2021. https://www.self-checkout-initiative.de/markterhebung-2021/, 2021.

[4] Eberhard Wolff. *Microservices - Grundlagen flexibler Softwarearchitektur*. dpunkt.verlag, 2 edition, 2018.

Effektive neuronale Netze zur Anomaliedetektion

Mara Alena Lehmann **Gabriele Gühring**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Das Netzwerkdesign neuronaler Netze wird in der Regel anhand eines vorhandenes Rechnersystems optimiert und ist somit von diesem abhängig. Die Verwendung eines Designraums ermöglicht die Definition von generalisierten Designprinzipien, anhand derer die Struktur neuronaler Netze unabhängig vom vorhandenen System optimiert wird.

Ziel der Arbeit

Ziel der Arbeit ist es, geeignete Designprinzipien herauszuarbeiten, anhand derer neuronale Netze Anomalien in Zeitreihen detektieren können. Designräume für Long Short-Term Memory, Gated Rekurrent Unit basierende und Spiking neuronalen Netze und Anomaliedetektion werden detektiert. Durch die Analyse von Unterräumen werden schrittweise Relationen herausgearbeitet, Verbesserungen abgeleitet und der Raum für effektive Modelle definiert.

Anomaliedetektion

Nach Aggarwal [1] werden Ausreißer und Anomalien in der Datenanalyse zwar häufig synonym verwendet. Hier gilt: Ein Ausreißer unterscheidet sich deutlich von den restlichen Datenpunkten und ist zu erwarten. Anomalien sind im Gegenzug unerwartete Abweichungen. In Zeitreihen- bzw. Sequenzanalysen verändern Anomalien durch unerwartete Sequenzen.

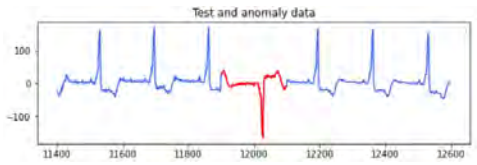

Abb. 1: Anomalie in Testdaten des Elektrokardiogrammdatensatzes des UCR Time Series Classification Archive [5]

Es gibt Supervised und Unsupervised Anomalieerkennung, dabei werden verschiedene Methoden verwendet, wie z. B. statistische oder abstands-/dichtebasierte.

Daten

Wu und Keogh [9] beschreiben Schwachstellen, die viele Zeitreihendatensätze aufweisen. So seien z. B. Anomalien unrealistisch in den Daten verteilt oder falsch gelabelt. Das *UCR Time Series Classification Archive* [5] hat eine Reihe von Datensätzen zusammengestellt, die diese genannten Schwachstellen nicht aufweisen. Es handeln sich um Zeitreihen unterschiedlicher Themengebiete, wie z. B. aus dem medizinischen Feld, Handbewegungen oder Erdbeben.
Die univariaten Daten sind in Trainings- und Testdaten aufgeteilt. Die einzige Anomaliesequenz in den Testdaten ist gekennzeichnet, die Trainingsdaten enthalten keine Anomalien.

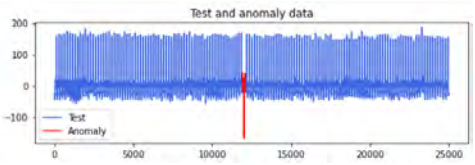

Abb. 2: Test- und Anomaliedaten des Elektrokardiogrammdatensatzes des UCR Time Series Classification Archive [5]

Für multivariate Zeitreihenanalysen werden Anomaliedatensätze aus gelabelten Daten erstellt, indem in den Testdaten, die aus einer Klasse bestehen, eine Sequenz aus einer anderen Klasse hinzugefügt wird.

Designraum

Ein Designraum beschreibt eine Menge an Modellarchitekturen. Das Facebook AI Research Team verfolgt den Ansatz, einen Designraum vorzugeben, in dem das neuronale Netz anhand aller Variationen der Parameter in diesem Raum trainiert wird. Die Fehlerverteilung

wird analysiert und der Designraum wird angepasst. Schrittweise wird so der Designraum optimiert [8]. Dieser Ansatz unterscheidet sich vom Suchen einer einzigen optimalen Lösung, bei der lediglich eine optimale Modellarchitektur gesucht wird.

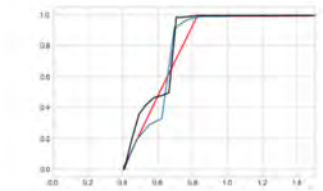

Abb. 3: Empirischen Verteilungsfunktion der Loss. Weiter links liegende Kurven stellen eine Verbesserung dar [4]

Verwendete Modelle und Anomaliedetektionsmethoden

Rekurrente neuronale Netze sind für die Verarbeitung von Zeitreihen geeignet. Long Short-Term Memory (LSTM) Schichten speichern Informationen und gibt sie an nachfolgende Schichten weiter [6]. Modelle können einzelne Werte vorhersagen bzw. als Autoencoder/-decoder ganze Sequenzen verarbeiten. Gated Recurrent Unit (GRU) basierte Schichten sind simpler und weniger rechenaufwändig [3]. Bäßler et. al [2] verwenden echtzeitfähige Spiking bzw. gepulste neuronale Netze, die Anomalien mit selbst-überwachtem Lernen erkennen.
Zur Anomaliedetektion wird ein Fehlervektor anhand der Differenz des vom neuronalen Netz vorausgesagten Wert mit dem wahren Wert bestimmt. Mit der Maximum-Likelihood-Schätzung wird die Gaußsche Verteilung des Fehlervektors analysiert (s. Abb. 4). Seltene Fehlerwerte, die stark vom Erwartungswert abweichen, stellen in der Regel Anomalien dar [7].

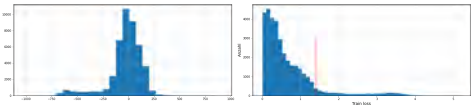

Abb. 4: Verteilung Training-Loss, obere Standardabweichung bildet Grenzwert [4]

Für multivariate Zeitreihen wird die Mahalanobis Distanz zur Anomaliedeketion verwendet, die Abstände zu den Mittelwerten berechnet und somit die Verteilung der Daten berücksichtigt [1].

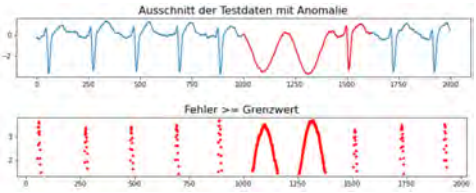

Abb. 5: Exakte Werte und Werte mit Abweichung zur Prediction, die größer als der Grenzwert sind [4]

Der Facebook-Ansatz wird dahingegen erweitert, dass die Bewertung eines Verfahrens nicht nur von der Genauigkeit eines Modells abhängt, sondern auch von dem Erfolg der Anomaliedetektion wie z. B. den Raten von wahr oder falsch positiven Ergebnissen.

Zusammenfassung und Ausblick

Die Verwendung eines Designraums und dessen automatisierte Auswertung erweitert die manuelle Optimierung der neuronalen Netze und Anomaliedetektion von Zeitreihen. Weitere Modelle und Anomaliedetektionsmethoden sowie weitere Bewertungsmaßstäbe für den Erfolg einer Anomaliesuche können integriert und ausgewertet werden.

Literatur und Abbildungen

[1] Charu C. Aggarwal. *Outlier Analysis*. Springer, 2017.

[2] Dennis Bäßler, Tobias Kortus, and Gabriele Gühring. Unsupervised anomaly detection in multivariate time series with online evolving spiking neural networks. *Machine Learning*, 2022.

[3] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv*, 2014.

[4] Eigene Darstellung.

[5] Hoang Anh Dau et al. The UCR Time Series Classification Archive. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/, 2018.

[6] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9:1735–1780, 1997.

[7] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. *Long short term memory networks for anomaly detection in time series*. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2015.

[8] Ilija Radosavovic, Raj Kosaraju Prateek, Ross Girshick, Kaiming He, and Piotr Dollar. Designing Network Design Spaces. *arXiv*, 2020.

[9] Renjie Wu and Eamonn Keogh. Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress. *Institute of Electrical and Electronics Engineers (IEEE)*, 2021.

Digitale Dezentrale Identität – Analyse und exemplarische Bewertung entsprechender Open Source Software im Vergleich zu existierenden Applikationen

Polina Liepelt **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Bosch.IO GmbH, Ludwigsburg

Problemstellung

Das Internet ist in den letzten Jahrzehnten ein untrennbares Element des modernen Lebens geworden. Allerdings wurde das Internet von Anfang an als eine reine Kommunikationsmöglichkeit entwickelt, um Daten digital transferieren zu können. Erst später, durch die weite Verbreitung der Technologie, wurde festgestellt, dass dem Internet die Identitätsschicht fehlt, was zu vielen monetären und Sicherheitsproblemen geführt hat. Diese Probleme können heute mithilfe von Self-Sovereign-Identity (SSI oder auch als Digitale Dezentrale Identität bekannt) gelöst werden.

Self-Sovereign-Identity (SSI) stellt eine Vision dar, wie unsere Identität in der digitalen Welt behandelt werden soll. Die Grundidee ist, dass jeder seine Identität selbst besitzt und verwaltet, was einen bewussten und sparsamen Umgang mit personenbezogenen Informationen bedeutet. [7] SSI soll nicht nur die Kontrolle über die eigenen Daten anbieten, sondern auch eine bessere User Experience (UX) bspw. durch die Anmeldung ohne Passwort. Das Föderative Identitätsmodell basierend auf Identity Providers (IDPs), wie z.B. Facebook oder Google, soll mit der neuen Technologie ersetzt werden. Damit werden heutige „Honeypots“ für die Cyberkriminalität eliminiert. Die Sicherheit und das gegenseitige Vertrauen haben bei der SSI die entscheidende Priorität.

Die Basis-Technologie, welche die SSI-Implementierung ermöglicht, wurde durch die Verbreitung von Blockchain seit 2008 [5] geschaffen. Dezentralität spielt bei SSI eine bedeutende Rolle, da die volle Kontrolle über die eigenen Daten nur durch eine Architektur ohne Zentralinstanz behalten werden kann. Grundsätzlich ist SSI als Technologie nicht an Blockchain gebunden und kann auch anders implementiert werden, allerdings sind die meisten Anwendungen heutzutage Blockchain bzw. Distributed Ledger Technology (DLT) basiert. [11] Die SSI-Technologie ist sehr jung und existiert in der beschriebenen Form erst seit 2016. In diesem Jahr hat Christopher Allen die 10 Prinzipien von SSI [1] veröffentlicht und den Anstoß für die weitere Verbreitung der Idee gegeben. Einen zweiten Schub hat die Technologie der COVID-19 Pandemie zu verdanken, da sie von der WHO als eine Alternative für die Umsetzung der COVID-Pässe betrachtet wurde.

Anwendungsmöglichkeiten und konzeptioneller Aufbau der SSI-Technologie

Bosch.IO GmbH, eine Tochtergesellschaft der Robert Bosch GmbH mit Spezialisierung auf IT-Technologien, ist eines der führenden Unternehmen, welches sich mit konzerninternen SSI-Implementierungen beschäftigt. Das Unternehmen wirkt intensiv auf mehreren Ebenen mit, um SSI als Grundtechnologie für die fehlende Identitätsschicht im Internet durchzusetzen. Bosch.IO sieht Potential für vielfältige Anwendungsmöglichkeiten der SSI Technologie bei verschiedenen Use Cases innerhalb und außerhalb des Konzerns und nimmt an zahlreichen externen Kollaborationen, wie z.B. Catena-X [12], teil.

Warum SSI eine sichere Lösung für die Identitätsverwaltung darstellt, wird anhand des konzeptionellen Aufbaus der Technologie deutlich. Konzeptionell basiert SSI auf dem in Abbildung 1 dargestellten Vertrauens Dreieck, welches von der Trust over IP Foundation [6] erarbeitet wurde. Das Vertrauens-Dreieck besteht aus drei interagierenden Seiten: Issuer – Aussteller, Holder – Besitzer, Verifier – Verifizierer. Das Herz der SSI sind so genannte Verifiable Credentials (VCs) – Berechtigungsnachweise. VCs können digitale Abbildungen von bspw. Personalausweis, Führerschein, Bankkarte, Fitnessstudio-Mitgliedsausweis etc. sein. Vereinfacht sind Verifiable Credentials digitale Kopien von allem, was wir in unseren Portemonnaies aufbewahren. Digitale Portemonnaies heißen bei SSI-Anwendungen Wallets und werden u.a. dafür benutzt, um Credentials zu speichern.

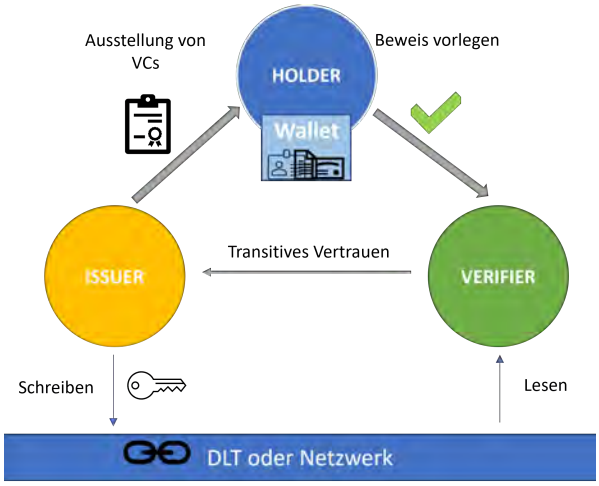

Abb. 1: Vertrauens-Dreieck [2]

Wie in Abbildung 1 zu sehen, funktioniert das Vertrauens-Dreieck vereinfacht wie folgt:
Der Issuer erstellt einen VC für einen bestimmten Holder, welcher diesen VC in seinem Wallet speichert. Beispielsweise erstellt ein Impfzentrum einen digitalen Impfpass für die geimpfte Person und die Person speichert den Pass digital auf dem Handy. Der Holder möchte seinen VC einem Verifier zeigen, um bestimmte Leistungen erhalten zu können. In dem Beispielfall wird die geimpfte Person seinen digitalen Impfpass beim Betreten des Restaurants vorzeigen.
Der Verifier bestätigt die Gültigkeit des VCs und die Dienstleistung kann vollbracht werden, d.h. die Person erhält Eintritt in das Restaurant. Der Verifizierungsprozess verläuft sicher und schnell auf Basis von DLT. Der Verifier ist dabei nicht gezwungen den Issuer direkt zu kontaktieren, da die wichtigsten Daten für die Verifizierung auf dem Netzwerk gespeichert sind.
Die Verifizierung bei SSI funktioniert über decentralized Identifiers (DIDs) [10] – kryptographisch verifizierbare und dezentrale Uniform Resource Identifiers (URIs), wie in Abbildung 2 dargestellt.

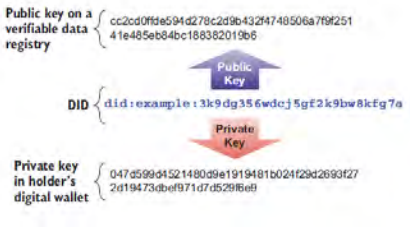

Abb. 2: DID mit dazugehörigen Schlüsseln [9]

Diese Adressen sind unikal und beinhalten asymmetrische kryptographische Beweise der Zugehörigkeit – Public und Private Keys. Dank dem Public Key, welcher auf dem Netzwerk gespeichert ist, ist es möglich den Holder, Issuer oder Verifier sicher zu identifizieren, sowie die Zugehörigkeit des VCs nachzuweisen. Dies wird durch einen entsprechend dazugehörigen Private Key – der zweite Teil des Schlüssels, welcher geheim ist – realisiert. Private Keys sind sicher in dem Benutzer Wallet abgelegt und der Zugriff auf diese Daten ist ausgeschlossen.
Wallet-Anwendungen erlauben dem Besitzer seine kryptographische Schlüssel oder andere sensible Daten zu generieren, verwalten, halten und zu schützen. [8] Für die Interaktionen mit anderen Parteien ist allerdings eine Agenten-Software verantwortlich, welche im Namen des Benutzers agiert.

Zielsetzung und Ergebnisse der Arbeit

Diese Teile des SSI-Ökosystems sind von bedeutendem Interesse für die Bachelorarbeit zum Thema „Digitale dezentrale Identität – Analyse und exemplarische Bewertung entsprechender Open Source Software im Vergleich zu existierenden Applikationen“, welche im Auftrag von Bosch.IO durchgeführt wird.
Bosch.IO ist ein aktiver Teilnehmer in dem Linux Foundation Projekt Hyperledger Aries [4] und trägt zur Entwicklung der Open Source Software Business Partner Agent (BPA) [3] bei. Die SSI-Software von Bosch.IO ist eine generische Lösung, welche zurzeit Use Case unabhängig aufgebaut ist. Diese Aries-Basierte Software wird Enterprise Wallet genannt und bietet eine skalierbare und sichere Server-Wallet-Implementierung für Unternehmen an.
Die technologische Entwicklung von SSI hat in der jüngsten Vergangenheit einen Durchbruch verschiedenster Implementierungslösungen angestoßen. Damit die Lösungen in der Zukunft weltweit kompatibel sind, wird aktuell intensiv an der Standardisierung der Prozesse gearbeitet. Welche Standards sich durchsetzen werden, steht noch aus. In der Phase der dynamischen Entwicklung konkurrierender Anwendungen möchte Bosch.IO den Überblick über die Wettbewerber erhalten. In erster Linie besteht das Ziel, die Wettbewerber kennenzulernen und technologische Trends durch den Vergleich der Produkte zu identifizieren. Die Bachelorarbeit soll auch grundlegend die Stärken und Schwächen der verfügbaren Software-Lösungen aufzeigen. Dies ist ein Schlüsselelement für weitere strategische Entscheidungen bezüglich der Entwicklung einer eigenen Applikation.
Im Rahmen der Bachelorarbeit wird eine Recherche von möglichen Wettbewerbern durchgeführt. Die am besten passenden Lösungen werden anhand von internen und externen Anforderungen durch einen iterativen Filtervorgang identifiziert. Mit Hilfe einer gewichteten eindimensionalen SWOT-Analyse werden die Anwendungen bewertet. Im Anschluss wird eine zweidimensionale SWOT-Analyse in Matrix-Form vi-

sualisiert und für die strategische Entscheidungsfindung benutzt.
Als Ergebnis der Arbeit werden auf Basis der durchgeführten analytischen Untersuchungen Handlungsempfehlungen angeboten.

Literatur und Abbildungen

[1] C. Allen. Self-Sovereign Identity Principles 1.0. https://github.com/WebOfTrustInfo/self-sovereign-identity/blob/master/self-sovereign-identity-principles.md, 2016.

[2] Eigene Darstellung.

[3] P. Etschel et al. Business Partner Agent. https://github.com/hyperledger-labs/business-partner-agent#:~:text=Business%20Partner%20Agent%20.%20Short%20Description.%20The%20Business,a%20machine-readable%20and%20tamper-proof, 2020.

[4] N. George. Announcing Hyperledger Aries, infrastructure supporting interoperable identity solutions. https://www.hyperledger.org/blog/2019/05/14/announcing-hyperledger-aries-infrastructur, 2019.

[5] G. Iredale. History Of Blockchain Technology: A Detailed Guide. https://101blockchains.com/history-of-blockchain-timeline/, 2020.

[6] J. Jordan et al. Introduction to Trust Over IP. https://trustoverip.org/wp-content/uploads/Introduction-to-ToIP-V2.0-2021-11-17.pdf, 2021.

[7] A. Kudra et al. What is self-sovereign Identity? https://sovrin.org/faq/what-is-self-sovereign-identity/, 2018.

[8] D. O'Donnell. Digital wallets and digital agents. In *Self-Sovereign-Identity*, pages 191–219. Manning Publications Co., 2021.

[9] D. Reed, R. Joosten, and O. Deventer. The basic building blocks of SSI. In *Self-Sovereign-Identity*, pages 21–38. Manning Publications Co., 2021.

[10] D. Reed, M. Sporny, M. Sabadello, D. Longley, and C. Allen. Decentralized Identifiers (DIDs) v1.0. https://www.w3.org/TR/did-core/, 2021.

[11] D. Reed, M. Sporny, M. Sabadello, D. Longley, C. Allen, et al. DID Specification Registries. https://www.w3.org/TR/did-spec-registries/#did-methods, 2021.

[12] S. Schindler-Le Huray et al. Cartena-X. https://catena-x.net/de/, 2021.

Entwicklung digitaler Assistenten für Cluu

Mathis Ludwig **Rainer Keller**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Softwarehelden GmbH & Co. KG, Stuttgart

Einführung

Die Entwicklung einer Software wird aus zwei Richtungen beeinflusst. Zum einen sind die Entwickler selbst, die durch ihre Implementierung vorgeben, wie die Software genutzt werden kann und soll. Zum anderen haben Nutzer Anforderungen an die Software. Mit der Weiterentwicklung kommt es so zu Wechselwirkungen, bei denen Nutzer auf die Entwickler einwirken und die Entwickler wiederum durch neue Funktionen Raum für neue Anforderungen schaffen. Durch diese Wandel in der Nutzung ergeben sich neue Herausforderungen an die Anwendung. In dieser Arbeit wird eine dieser Herausforderungen angegangen.

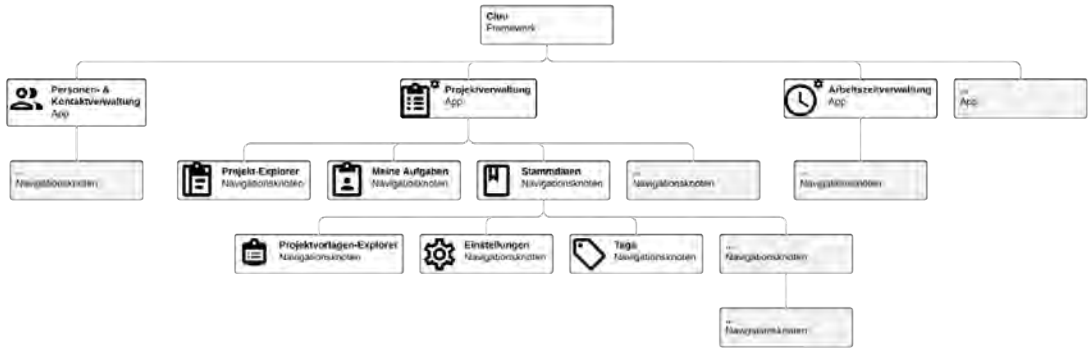

Abb. 1: Beispiel der Cluu-Navigationsstruktur [2]

Cluu

Ein gutes Beispiel hierfür ist die Plattform Cluu. Diese wird von der Firma Softwarehelden GmbH & Co. KG entwickelt und vertrieben. Begonnen hat die Entwicklung mit dem Gedanken, einen baumartigen Explorer (vgl. Windows Dateiexplorer) zu erstellen, über den auf verschiedenste Datenquellen zugegriffen werden kann. Abbildung 1 zeigt beispielhaft einen kleinen Ausschnitt der Navigationsstruktur von Cluu. Auf der ersten Ebene von Cluu befinden sich die verschiedenen Apps. Darunter können selbsthierarchisch Navigationsknoten angelegt werden. Diese können auf in tiefere Ebenen der Navigation führen oder auf Daten verweisen.
Cluu ist eine Low-Code/No-Code-Plattform. Das bedeutet, dass Anwendungen mit nur wenig oder komplett ohne eigenen Code schreiben zu müssen, entwickelt werden können [1]. In Cluu können neben der UI auch CURD- und weitere vorgefertigte Aktionen konfiguriert werden, ohne eine Zeile Code schreiben zu müssen. Spezielle Business-Logik lässt sich mit nur geringem Programmieraufwand einbinden. Das ermöglicht sehr schnelle und dennoch hochwertige Entwicklung neuer Anwendungen. Dieser Vorteil bringt jedoch mit sich, dass es auf Cluu Instanzen sehr viele Anwendungen geben kann. Dies kann dazu führen, dass Nutzern die Übersicht über die Funktionen fehlt und sie viel Zeit damit aufwenden müssen, durch Cluu zu navigieren. Über die jahrelange Entwicklung gab es immer neue Anforderungen an Cluu. Auch die Entwickler hatten immer neue Ideen, wie Cluu noch besser werden kann. Durch die neuen Möglichkeiten ergaben sich neue Use-Cases und damit wieder neue Anforderungen der Nutzer an Cluu. Auch wenn die Nutzung von Cluu sich mittlerweile geändert hat, ist die Bedienung noch stark an die eines Explorers angelehnt. Um mit den Daten arbeiten zu können, muss zunächst über die hierarchische Navigationsstruktur zu diesen Daten navigiert werden.

Zielsetzung

Die Nutzung von Cluu hat sich über die Jahre gewandelt. Nutzer arbeiten nicht lange auf Daten einer Klasse. Cluu hebt sich durch die Verbindung der verschiedenen Datenquellen und der Daten auf diesen Quellen hervor. Ziel ist es, dem Nutzer eine Bedienung von Cluu zu ermöglichen, wie dieser sie erwarten würde. Er soll nicht gezwungen sein, zu den Daten zu navigieren. Stattdessen soll dem Nutzer eine Bedienung ermöglicht werden, die seinem natürlichen Workflow entspricht.

Konzeptionierung

Die Idee ist, jedem Nutzer seinen eigenen digitalen Assistenten zur Verfügung zu stellen. Diese sollen wie reale Assistenten den Nutzer bei seiner Arbeit unterstützen und ihm alles so vorbereiten, dass der Nutzer bestmöglich arbeiten kann. Hierfür sollen die Aktionen nicht mehr nur aus dem Kontext der Daten ausgeführt werden können. Stattdessen soll es möglich sein, sogenannte Assistenten zu konfigurieren. Diese beinhalten alle Aktionen, die der Nutzer für bestimmte Arbeitsabläufe benötigt. Ein Beispiel wäre ein Buchhaltungsassistent. Dieser kann verschiedenste Funktionen aus Anwendungen wie der Arbeitszeitverwaltung oder der Projektverwaltung zusammenfassen. So müssen Mitarbeiter der Buchhaltung nicht durch die verschiedenen Anwendungen navigieren, sondern können alle Aktionen direkt von der ersten Ebene des Assistenten ausführen. Die Aktionen werden hierfür abstrahiert und in den Assistenten „Skill“ genannt. Neben den CURD-Aktionen können über Skills Apps, Navigationsknoten oder Listen mit Daten geöffnet werden. Außerdem kann jede Custom-Aktion als Skill konfiguriert und ausgeführt werden. Das Datenmodell sieht dann wie in Abbildung 2 dargestellt aus. Assistenten können mehrere Gruppen zugeordnet werden. Diese Gruppen fassen Skills zusammen. Somit lassen sich die Skills innerhalb der Assistenten gruppieren, um somit den Nutzern bessere Orientierung zu ermöglichen. Den Gruppen werden die Skills M-zu-N zugeordnet. Das ist notwendig, um ein Skill in mehreren Assistenten darstellen zu können. Eine M-zu-N-Zuordnung der Gruppen zu den Assistenten wäre nicht sinnvoll gewesen. Die Gruppen sollen spezielle Arbeitsabläufe thematisch zusammenfassen. Diese Arbeitsabläufe sollten jedoch nicht in mehreren Assistenten vorkommen.

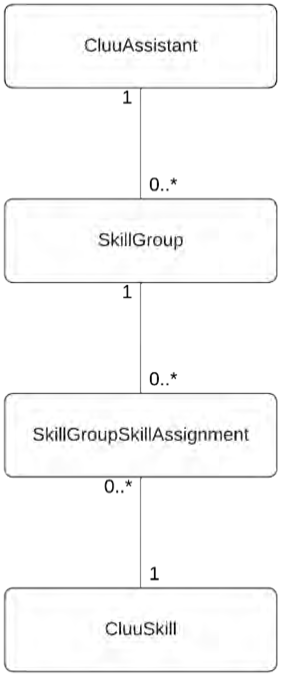

Abb. 2: Klassendiagramm der Cluu-Assistenten & -Skills [2]

Ausführen der Skills

Um diese Funktionen bereitstellen zu können, muss der Aktionsaufruf umgestellt werden. Dieser ist bisher so umgesetzt, dass die Aktionen aus dem Kontext der Daten aufgerufen werden kann. Beim Aufruf eine Aktion wird der Applikationskontext übergeben. Dieser enthält Daten wie den Klassennamen, eine Einschränkung der Daten oder Anweisungen zur Sortierung der angezeigten Daten. Jede Aktion zieht sich aus diesem Objekt dann die Daten, die sie benötigt. Diese Herangehensweise ist bisher nicht falsch, da die Aktionen ja immer im Kontext der Daten aufgerufen wurden. Mit der neuen Anforderung ist dieser Aufruf nicht mehr korrekt. Außerdem sorgt der Aufruf so für eine schlechte Lesbarkeit des Codes, da nicht direkt erkennbar ist, welche Daten die Aktion wirklich benötigt. Deshalb wird der Aktionsaufruf so umgestellt, dass die Aktionen als Parameter ein Objekt erwarten, das alle möglichen Argumente beinhalten kann, die die Aktion verarbeiten kann.

Ausblick

Die Nutzer können Cluu dank der Assistenten so konfigurieren, dass Cluu ihre Workflows abdecken kann. Im nächsten Schritt sollen die Assistenten den Nutzer noch tiefgreifender unterstützen. Mithilfe

maschinellen Lernens sollen Arbeitsabläufe erkannt werden. Die Assistenten sollen den Nutzern Vorschläge machen, welche Skills sie als Nächstes ausführen könnten. Zur Klassifizierung können beispielsweise die Uhrzeit, der Wochentag, der gewählte Assistent oder der zuletzt ausgeführte Skill herbeigezogen werden. Cluu kann diese Daten anonymisiert sammeln und den Algorithmus immer wieder neu trainieren. So werden die Vorhersagen der Assistenten immer besser, je mehr die Assistenten genutzt werden. Im Idealfall können Nutzer Cluu zukünftig fast nur über diese Vorschläge bedienen.

Literatur und Abbildungen

[1] Unbekannter Autor. What is low-code/no-code application development? | SAP Insights. https://www.sap.com/insights/what-is-low-code-no-code.html, 04 2022.

[2] Eigene Darstellung.

Lokalisierung von Personen in industriellen Indoor-Umgebungen

Matthias Machtolf **Reiner Marchthaler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Virtual Automation Lab, Hochschule Esslingen, Sascha Röck

Motivation

Zur Flexibilisierung von Produktionsabläufen sowie den Transport von Kleinteilen zwischen Bearbeitungsstationen wird am Virtual Automation Lab der Hochschule Esslingen der Einsatz von Flugrobotern, wie beispielsweise Multikoptern, untersucht. Diese erlauben die Nutzung des zumeist ungenutzten Luftraums in den Fertigungsstätten und ermöglichen dadurch eine hohe Flexibilität und Dynamik. Jedoch bringt der Einsatz von Flugrobotern zahlreiche Herausforderungen mit sich, wie die Inbetriebnahme in sich verändernden Umgebungen oder eine kollisionsfreie Bahnplanung. Als Teil dessen sind dabei ebenfalls Informationen hinsichtlich der Position von Menschen in der Umgebung notwendig. Diese müssen messtechnisch genau erfasst werden.

Während Global Navigation Sattelite System (GNSS) Lösungen, wie beispielsweise GPS oder GALILEO im Outdoorbereich unter Verwendung von Zweifrequenzempfängern oder sattelitengestützten Erweiterungssystemen, Genauigkeiten im Zentimeterbereich erreichen, bleibt die Frage der zu verwendenden Indoor-Lokalisierungstechnik weiterhin offen. Ziel dieser Arbeit ist es daher, zu untersuchen, inwiefern Personen als Hindernis in einer industriellen Umgebung mit dynamischen autonomen Flugrobotern zuverlässig repräsentiert werden können.

Dabei bieten sich neben optischen, akustischen oder mechanischen Lösungen, vor allem funkbasierte Technologien an. Aufgrund der hohen Genauigkeit im Bereich weniger Zentimeter, der hohen Updaterate, wie auch der hohen Reichweite, stellt die Ultrabreitband (UWB, engl. Ultra-Wideband) Technologie eine ideale Möglichkeit dar, Personen zu lokalisieren. UWB-Systeme weisen dabei eine absolute Bandbreite von 500MHz und höher sowie eine relative Bandbreite von mindestens 20% auf. Die hierbei erzeugten Impulse besitzen eine sehr kurze Dauer im Nano- bzw. Pikosekundenbereich. Hierdurch ergeben sich beispielsweise Vorteile im Falle von Mehrwegeempfang oder eine hohe Robustheit gegenüber Fading.

Aufgrund dessen hat man sich in dieser Arbeit für das auf dem IEEE802.15.4-2015 sowie IEEE802.15.4z Standard basierende DWM3000 Evaluierungsboard von DecaWave entschieden. Auf diesem ist ein DW3110 UWB IC sowie eine Keramik UWB Antenne verbaut, womit laut Datenblatt eine Ranging Genauigkeit von +/- 6cm unter Line-of-Sight-Bedingung erreicht werden kann [4]. Des Weiteren ist eine Kommunikation mit den neuesten UWB-Chips, wie beispielsweise dem U1-Chip von Apple, welcher seit dem iPhone 11 verbaut wird, möglich. Die zu lokalisierende Person trägt hierbei ein solches UWB-Modul, bestehend aus dem Evaluierungsboard sowie einem STM32-Nucleo Board, in diesem Fall als "Tag"bezeichnet, am Körper, der über mehrere Anker- bzw. Referenzstationen lokalisiert wird. Das STM32-Nucleo Board dient hierbei als Host Mikrocontroller. Zusammen können diese sowohl als Tag als auch als Ankerstation genutzt werden.

Klassische Probleme, welche sich bei der Positionsbestimmung ergeben können und welche vor allem in industriellen Umgebungen verstärkt auftreten, können Non-Line-of-Sight (NLOS) oder Mehrwegeempfang sein. Das 400 qm große Maschinenbaulabor am Standort Stadtmitte bietet dabei für die Untersuchung solcher Lokalisierungssysteme eine ideale Grundlage.

Lokalisierung

Um die absolute Position eines Tags innerhalb eines Bezugssystems zu bestimmen, findet das Prinzip der Lateration Anwendung, welches in Abbildung 1 veranschaulicht wird. Dieses beruht darauf, dass die Entfernung eines Tags zu mindestens drei Ankerstationen berechnet wird. Generell spricht man dabei im Falle von drei festen Ankerstationen von einer Trilateration, falls mehr Ankerstationen vorhanden sind von der Multilateration. Die Position des Tags ist dabei durch folgende Kugelgleichung definiert: [1]

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}.$$

Dabei beschreibt d_i die Distanz zwischen Tag und Anker, x, y, und z die Koordinaten des Tags und x_i, y_i sowie z_i die Koordinaten der Anker. Bei n Ankern ergeben sich entsprechend n Kugelgleichungen und damit für n>3 ein überbestimmtes Gleichungssystem, das mittels der Least-Squares Methode gelöst wird.

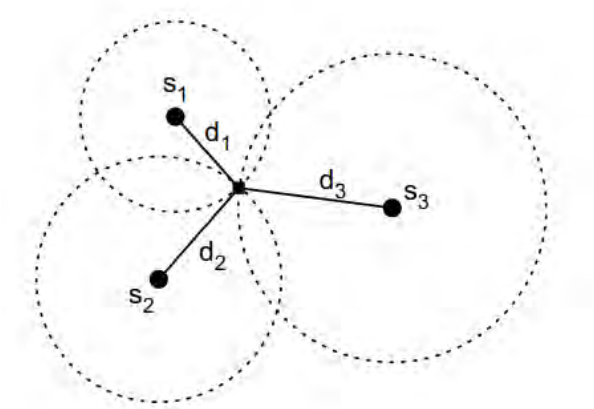

Abb. 1: Lateration [2]

Hinsichtlich der Distanzmessung findet im Rahmen dieser Arbeit sowohl das Single-Sided als auch das Double-Sided Two-Way Ranging Anwendung. Dieses basiert darauf, dass die Distanz des Tags zu den Ankerstationen proportional zur Ausbreitungszeit ist. So wird die Ausbreitungszeit gemessen und darüber auf die Distanz geschlossen. Im Fall des Single-Sided Two-Way Rangings sendet der Tag dabei eine so genannte Poll Nachricht an den Anker. Dieser wiederum antwortet mit einer Response Nachricht, welche sowohl den Empfangszeitstempel der Poll Nachricht als auch den Versandzeitstempel der Response Nachricht enthält. Mittels dieser Informationen sowie der Empfangs- und Versandzeitstempel der Response und Poll Nachricht, kann der Tag anschließend die Position berechnen. Eine Erweiterung dessen stellt das Double-Sided Two-Way Ranging dar, bei dem nach Empfang der Response Nachricht auf Seiten des Tags zusätzlich eine Final Nachricht an den Anker gesendet wird. Anschließend hat dieser die Möglichkeit die Distanz zu berechnen oder er sendet eine Report Nachricht mit Zeitstempeln an den Tag, der daraufhin die weitere Berechnung vornimmt. Die auf dem Tag berechneten Distanzen werden im Falle dieser Arbeit anschließend über einen virtuellen COM-Port an einen Rechner zur Ermittlung der Position gesendet.
Da es sich hierbei jedoch um Pseudoentfernungen handelt, muss auf Positionsschätzverfahren zurückgegriffen werden. Dabei findet der Least-Squares Algorithmus Anwendung. Diesbezüglich kann unterschieden werden zwischen dem linearen Least-Squares Algorithmus, bei welchem zunächst eine Linearisierung der Kugelgleichungen durchgeführt wird und anschließend die Position analytisch berechnet wird sowie dem nicht-linearen Least-Squares Algorithmus, bei welchem per iterativem Verfahren, wie etwa mittels dem Levenberg-Marquardt Algorithmus, eine numerische Lösung ermittelt wird.

Realisierung

Vor der eigentlichen Lokalisierung eines Tags über mehrere Ankerstationen, wurden zunächst Distanzmessungen sowohl unter Line-of-Sight als auch Non-Line-of-Sight Bedingungen im Maschinenbaulabor der Hochschule Esslingen durchgeführt. In Abbildung 2 ist der Aufbau zwischen zwei UWB-Modulen zu sehen. Dabei war zwar eine Sichtverbindung gegeben, jedoch durch verschiedene metallische Gegenstände, wie beispielsweise durch eine Umformpresse, welche sich am linken Bildrand befindet, Mehrwegeempfang vorhanden. Der Betrag des absoluten Fehlers für Distanzmessungen mittels Single-Sided Two-Way Ranging auf 3, 6, 9 und 12 Meter Entfernung ist dabei in Abbildung 3 dargestellt, wobei sich hier Genauigkeiten im Zentimeterbereich ergaben.

Abb. 2: Line-of-Sight Distanzmessung [3]

Weitere Untersuchungen umfassen die Lokalisierung mit vier bis acht Ankern. Einerseits findet diese statisch an mehreren im Labor verteilten Messpunkten statt, wobei hier sowohl die Anker als auch der Tag auf einem Lichtstativ montiert sind.
Zur Untersuchung des Verhaltens im dynamischen Fall wird der Tag auf einem Industrieschutzhelm montiert und anschließend unter Bewegung lokalisiert. Zur Ermittlung der Ground Truth wird dabei auf das von Valve entwickelte Lighthouse System zurückgegriffen.

Dabei handelt es sich um ein Infrarot-Trackingsystem, bestehend aus mehreren Basisstationen und einem Tracker. Der mit Photosensoren ausgerüstete Tracker, welcher ebenfalls auf dem Helm montiert wird, sendet die als wahr angenommene Position an den Rechner. Anschließend kann die mittels UWB ermittelte Position mit der wahren Position verglichen werden.

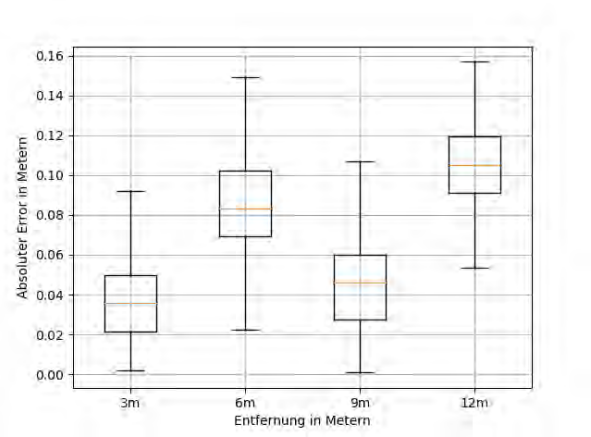

Abb. 3: Boxplots Line-of-Sight Distanzmessung [3]

Aus den Ergebnissen soll anschließend darauf geschlossen werden, wie Personen in einem Umgebungsmodell dargestellt werden können, damit diese sicher als Hindernis repräsentiert werden. Eine Möglichkeit hierfür ist die Repräsentation als geometrisches Objekt mit einem entsprechenden Radius. Für den Anwendungsfall der Flugrobotik dient dieses als No-Fly Zone zum Schutz der jeweiligen Person.

Literatur und Abbildungen

[1] Jörg Blankenbach et al. Indoor-Positionierung auf Basis von Ultra Wide Band. *Allgemeine Vermessungs-Nachrichten*, pages 169–178, 2007.

[2] Alejandro Correa et al. A Review of Pedestrian Indoor Positioning Systems for Mass Market Applications. *Sensors (Basel)*, 2017.

[3] Eigene Darstellung.

[4] DecaWave Ltd. DW3000 Data Sheet. https://www.decawave.com/wp-content/uploads/2021/03/DW3000_Datasheet.pdf, 2020.

Untersuchung zur Krypto-Agilität von eingebetteten Systemen im Nutzfahrzeug

Martin Mager **Dominik Schoop**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Daimler Truck AG, Stuttgart

Motivation

Der Automobilsektor durchläuft derzeit einen tiefgreifenden digitalen Wandel, getrieben durch die zunehmende Automatisierung und Vehicle-to-everyting-(V2X)-Vernetzung. Die damit verbundenen Technologien bringen erhebliche, wachsende Sicherheitsrisiken und Komplexität mit sich [8]. Cyber-Security gewinnt daher zunehmend an Bedeutung und ist mittlerweile nicht mehr nur ein unternehmensinterner Anspruch sondern auch eine wesentliche, gesetzliche Anforderung geworden [8]. Die Effektivität der Cyber-Security-Maßnahmen ist dabei maßgeblich abhängig von der Sicherheit der eingesetzten Verfahren und deren Implementierung [5]. Der deutliche Anstieg der kryptografisch-assoziierten Vorfälle in der Automobilindustrie der vergangenen Jahre zeigt, dass getroffene kryptografische Maßnahmen jeweils nur für einen begrenzten Zeitraum hinreichend schützen [5] [8]. Zusammen mit der langen Lebensdauer der Fahrzeuge sowie der eingeschränkten Zugriffsmöglichkeit im Feldbetrieb ergeben sich daraus große Herausforderungen für eine kontinuierliche Absicherung der Fahrzeuge [8]. Unternehmen sollten daher insbesondere im Bereich der Kryptografie agil aufgestellt sein, um auf Risiken unmittelbar und effizient reagieren zu können. Diese Eigenschaft bezeichnet man als Krypto-Agilität [9].

Ziel der Arbeit

Ziel der Arbeit ist es, die Anforderungen an ein kryptografisch agiles Gesamtsystem, bestehend aus Fahrzeug und zugehöriger Infrastruktur, zu definieren. Hierbei sollen aktuelle sowie zukünftige, fallspezifische Anforderungen analysiert werden. Abschließend soll eine Handlungsempfehlung für die Umsetzung eines kryptografisch agilen Gesamtsystems erfolgen.

Krypto-Agilität

Krypto-Agilität beschreibt die Fähigkeit eines Systems, in angemessener Zeit zwischen kryptografischen Algorithmen und Verfahren wechseln zu können, ohne die Funktionalität des restlichen Systems wesentlich zu beeinträchtigen [4]. Die Bedeutung der Krypto-Agilität kann mit Hilfe des in Abbildung 1 gezeigten Mosca's Timeline Model veranschaulicht werden [2] [7].

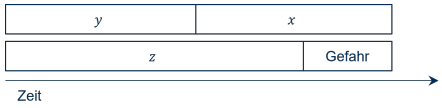

Abb. 1: Mosca's Timeline Model [7]

Der zeitliche Faktor X beschreibt die Dauer, wie lange die Verschlüsselung, bestehend aus Schlüssel und Daten, als sicher eingestuft werden soll. Die notwendige Zeit für die Migration auf ein neues, sicheres Verfahren wird mit dem Y-Faktor beschrieben. X und Y gegenübergestellt ist der Z-Faktor, welcher die Zeitspanne, bis ein groß angelegter Angriff möglich ist, beschreibt. Ein Security Risiko besteht, wenn $X+Y$ länger als Z andauert. In diesem Moment sind der Schlüssel und die Daten nicht länger sicher [2] [7].
Um die Krypto-Agilität eines Systems zu verbessern, muss somit der Y-Faktor minimiert werden. Die hierfür erforderlichen Maßnahmen betreffen den gesamten Lebenszyklus des Fahrzeugs, angefangen in der Entwicklung bis hin zur Außerbetriebnahme (siehe Abbildung 2).

Abb. 2: Lebenszyklus eines Fahrzeugs [1]

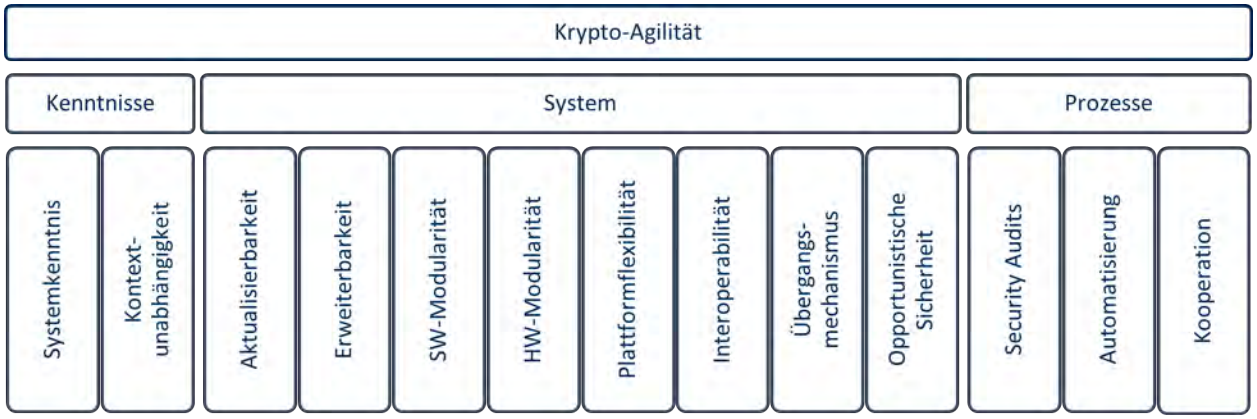

Abb. 3: Anforderungen - Krypto-Agilität [3]

Anforderungen: Die generischen Anforderungen an ein kryptografisch agiles System werden in Abbildung 3 dargestellt und können in die drei Bereiche *Kenntnisse*, *System* und *Prozesse* unterteilt werden [3].
Die Anforderungen im Bereich *Kenntnisse* verfolgen das Ziel, das Bewusstsein und tiefgehende Verständnis für Cyber-Security Mechanismen zu fördern. Ohne tiefgehendes Systemverständnis kann kein Lösungskonzept erarbeitet werden, welches im Idealfall kontextunabhängig ist, um den Aufwand zu reduzieren [3].
Die Anforderungen auf Software- und Hardwareebene werden im Bereich *System* betrachtet. Ein wichtiger Aspekt ist hier beispielsweise die Modularität, welche die Aktualisierbarkeit und Erweiterbarkeit eines Krypto-Systems erst ermöglicht [3]. Der Übergangsmechanismus zu beispielsweise neuen kryptografischen Verfahren ist einer der zentralsten und wichtigsten Aspekte bei der Optimierung des *Y*-Faktors. Hier ergeben sich insbesondere Anforderungen bezüglich der Interoperabilität zwischen den Steuergeräten und der Umsetzung einer opportunistischen Sicherheit, bei der stets das sicherste verfügbare Verfahren eingesetzt wird [3] [6].
Krypto-Agilität hat überdies einen Einfluss auf die produktbegleitenden *Prozesse*. Hier sollten in regelmäßigen Security Audits die Maßnahmen im Bereich der Krypto-Agilität bewertet und optimiert werden. Zudem sollte Krypto-Agilität auch in die Entwicklung einfließen und hier insbesondere in den Entwicklungs- und Bereitstellungsprozess, um die Abläufe automatisiert und damit effektiv umsetzen zu können. Des Weiteren ist eine enge und transparente Kooperation mit den Zulieferern erforderlich [3].
Weitere Anforderungen: Neben generischen Anforderungen werden auch fallspezifische Anforderungen im Rahmen der Arbeit betrachtet. Insbesondere die fahrzeuginterne und -externe Kommunikation sowie die für die verschiedenen Anwendungsfälle notwendige Infrastruktur und spezifische Prozesse werden detailliert betrachtet.
Ein Beispiel hierfür ist der Einsatz von Secure-Onboard-Communication-(SecOC) - ein Verfahren, welches für den Schutz vor Manipulation und Replay-Angriffen eingesetzt wird, bei dem ein Message-Authentication-Code-(MAC)-Tag am Ende der Nachricht angehängt wird. Für die Berechnung wird ein symmetrischer Schlüssel benötigt [10].
In Bezug auf Krypto-Agilität ergibt sich in diesem Anwendungsfall unter anderem die Anforderung nach einer variablen Schlüssellänge, um den gleichen Security Level auch in Zukunft aufrecht erhalten zu können.

Strategie & Vorgehensmodell

Auf Basis der erarbeiteten Grundlagen und Anforderungen soll nun eine Handlungsempfehlung mit den erforderlichen Maßnahmen für die Umsetzung eines kryptografisch agilen Gesamtsystems entwickelt werden. Die Handlungsempfehlung orientiert sich dabei an dem Lebenszyklus des Fahrzeugs (siehe Abbildung 2) und bietet Lösungsansätze innerhalb der verschiedenen Abschnitte für die zuvor definierten Anforderungen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Alan Grau. Quantum-Safe Cryptography—Surviving the Upcoming Quantum Cryptographic Apocalypse. https://sectigo.com/resource-library/quantum-safe-cryptography-surviving-the-upcoming-quantum-cryptographic-apocalypse, 08 2020.

[3] Julian Hohm, Andreas Heinemann, and Alexander Wiesmaier. Towards a maturity model for crypto-agility assessment. *Cryptography and Security*, 2022.

[4] AppViewX Incorporated. What is Crypto-agility? https://www.appviewx.com/education-center/what-is-crypto-agility/, 2022.

[5] Keyfactor Incorporated. *Crypto-Agile PKI for the Future*. Keyfactor, 2020.

[6] Hassane Mehrez and Othmane El Omri. The Crypto-Agility Properties. In *Proceedings of The 12th International Multi-Conference on Society, Cybernetics and Informatics*. International Institute of Informatics and Systemics, 2018.

[7] Michele Mosca. Cybersecurity in a Quantum World: will we be ready? https://csrc.nist.gov/csrc/media/events/workshop-on-cybersecurity-in-a-post-quantum-world/documents/presentations/session8-mosca-michele.pdf, 04 2015.

[8] Peggy Nahmany. End-to-End Cybersecurity for Connected Vehicles with Thales Trusted Key Manager. https://www.thalesgroup.com/sites/default/files/database/document/2020-09/iot-end-to-end-cybersecurity-for-connected-vehicules.pdf, 09 2020.

[9] David Ott and Christopher Peikert. Identifying Research Challenges in Post Quantum Cryptography Migration and Cryptographic Agility. *CCC Workshop Report*, 2019.

[10] AUTOSAR Standard. Specification of Secure Onboard Communication. https://www.autosar.org/fileadmin/user_upload/standards/classic/4-3/AUTOSAR_SWS_SecureOnboardCommunication.pdf, 2017.

Automatisches Maschinelles Lernen: Eine Evaluation von AutoML Lösungen

Robin Maurer **Steffen Schober**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma IT-Designers GmbH, Esslingen am Neckar

Einleitung

Arthur Samuel entwickelte im Jahr 1956 ein Programm, das mit Hilfe eines Suchbaumes den gegenwärtig besten Zug im Spiel „Checkers“ (deutsch: Dame) berechnen konnte. Drei Jahre später prägte er den Begriff „Machine Learning“ (ML) in seinem Beitrag „Some Studies in Machine Learning Using the Game of Checkers“ im „IBM Journal of Research and Development“. (siehe [8])
Seit einigen Jahren wird Maschinelles Lernen immer häufiger für kommerzielle Zwecke verwendet. Das hat zur Folge, dass „AI“ und somit auch „Machine Learning“ in jüngster Vergangenheit zu Schlagworten der technischen Innovation aszendiert sind. Gegenwärtig ist die Entwicklung von ML-Lösungen jedoch sehr kostspielig, denn nur wer das Budget hat Teams von Data-Scientists und Rechen-Cluster zu unterhalten, kann der Nachfrage des Marktes gerecht werden. Mit der steigenden Nachfrage nach ML-Lösungen steigt das Bedürfnis, die benötigte Zeit und Ressourcen für deren Entwicklung zu optimieren. Aus diesem Bedürfnis entstand automatisiertes maschinelles Lernen, ein Prozess, der auch als AutoML bezeichnet wird.

Automatisiertes Maschinelles Lernen

Im klassischen maschinellen Lernen ist es die Aufgabe, von Data Scientists, Input-Daten aufzubereiten und Hyperparameter-Tuning durchzuführen. Für diese Prozesse muss im klassischen ML viel Erfahrung vorhanden sein sowie viel Zeit aufgebracht werden. Daher verspricht der Ansatz des AutoML große Effizienzsteigerungen durch das automatisierte Erstellen von Modellen. Dies reduziert Kosten in Hinsicht auf Entwicklungs-, Rechen- und Arbeitszeit, da hierbei lediglich Trainingsdaten an ein Framework übergeben werden. Auch Teile des Feature Engineering, ein Prozess der viel Wissen erfordert, werden übernommen, sodass selbst weniger erfahrene Benutzer erfolgreich Modelle erstellen können. Aufgrund dessen ist es auch für kleine Unternehmen, ohne großes Budget für Spezialisten, möglich in ML einzusteigen. Obwohl AutoML viele Bereiche vereinfacht, werden weiterhin Spezialisten benötigt, um das Feature Engineering mit ihrem Wissen zu optimieren, die Modelle in die Arbeitsprozesse zu integrieren sowie für den wissenschaftlichen Transfer. [2]
Zu den bekanntesten AutoML kostenpflichtigen Frameworks gehören u. a. „AWS Sagemaker Autopilot“, „Google Cloud AutoML“ und „Azure AutoML“. Alternativ kann auf kostenlose Frameworks wie „autosklearn“, „AWS H2O AutoML“ für ML und „AutoPyTorch“ oder „AutoKeras“ für Deep Learning zurückgegriffen werden. Grundsätzlich funktionieren AutoML Frameworks als „Black Box“. Es werden Input-Daten an das Framework übergeben und dieses generiert als Output ein passendes Model. Um genauer zu beleuchten, wie Frameworks zu dem Modell gelangen, folgt nun eine Erklärung anhand Abbildung 1. Zunächst werden etwaige, zuvor gelernte Meta-Parameter gesetzt z. B. die Lern-Rate oder die Klassifizierung bzw. der Regressor mit den besten Erfolgen. Diese Meta-Parameter sind aufgabenübergreifend, sodass vorherige Erfolge anderer Aufgaben in die Modellierung des neuen Algorithmus einfließen können.

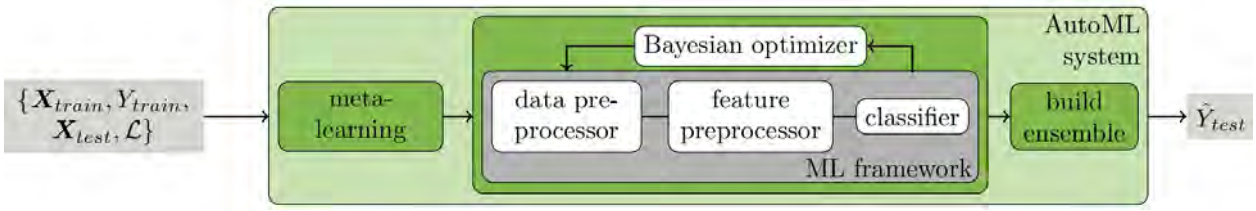

Abb. 1: Schematische Darstellung der AutoML-Pipeline des Frameworks autosklearn [1]

Hiernach startet die klassische ML-Pipeline mit Präprozessoren für den Datensatz sowie die Merkmale und der anschließenden Klassifizierung bzw. Regression. Mit Hilfe eines Bayes'schen Optimierungs-Algorithmus werden, nach dem Durchlauf der ML-Pipeline, die Ergebnisse dieser evaluiert und neue Parameter für den nächsten Durchlauf ausgewählt. Bei Erreichen des Abbruch-Kriteriums, z. B. maximale Modell-Anzahl oder gegebene Zeit, werden die erstellten Modelle in den letzten Schritt übergeben, dem „Ensemble". Ein Ensemble ist das Ergebnis der Kombination der n-besten Modelle durch „Voting" und „Stacking". „Voting: Trifft Vorhersagen auf Grundlage des gewichteten Durchschnitts der vorhergesagten Klassenwahrscheinlichkeiten (für Klassifizierungsaufgaben) oder auf Grundlage der vorhergesagten Regressionsziele (für Regressionsaufgaben). Stacking: Stacking kombiniert heterogene Modelle und trainiert ein Metamodell, basierend auf der Ausgabe der einzelnen Modelle. Die aktuellen Standardmetamodelle sind LogisticRegression für Klassifizierungsaufgaben und ElasticNet für [Regressionsaufgaben]" [2].

Bayes Optimierung / SMAC

Den meisten AutoML-Frameworks liegt die im vorherigen Abschnitt erwähnte Bayes'sche Optimierung zugrunde. Hierbei wird versucht aus gegebenen Messpunkten, in der Abbildung 2 als Datenpunkt bezeichnet, die Unbekannte Zielfunktion anzunähern, dargestellt als unterbrochene grüne Linie. Um dies zu erreichen, werden zunächst mehrere Regressoren trainiert und zusammengefasst. Die Unsicherheit der Modelle wird in der Abbildung als hinterlegter Bereich verschiedener Grautöne, welche nach Grad der Standardabweichung heller werden, dargestellt. Die blaue Linie bildet die Erwartete Funktion ab, das Mittel der Funktionen. Als Nächstes wird eine sogenannte „Akquisitions-Funktion", hier beispielsweise für die Darstellung des unteren Bands der Standardabweichung nach Formel 1, berechnet.

$$a(x) = f(x) - kappa * stdDev_{f(x)} (1)$$

„f(x)" bezeichnet die derzeitig angenommene Funktion, zunächst das Mittel der Modelle. "kappaïst hierbei ein Hyperparameter, durch den die Suche nach Minima entweder tendenziell lokal oder tendenziell global durchgeführt wird. Das Ergebnis, das untere Band der Standardabweichung, wird daraufhin auf Minima untersucht und dieser Wert „gesamplet". Mit diesem neuen Datenpunkt beginnt die Bayes'sche Optimierung von vorne, bis eine Abbruchbedingung, oft Zeit-Restriktionen oder Konvergenzen, erreicht wird. [6]

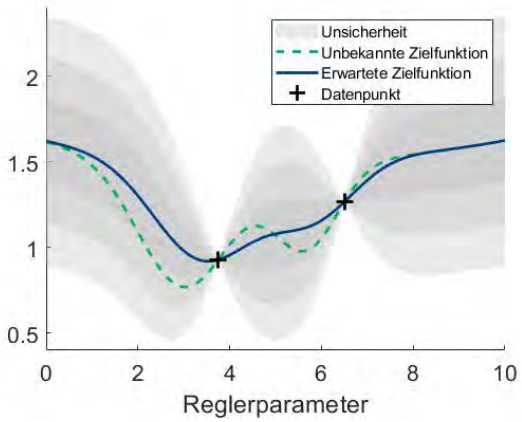

Abb. 2: Diagramm zur Erklärung der Bayes'schen Optimierung [5]

Genauer verwenden viele AutoML Frameworks „Sequential Model-based Algorithm Configuration", kurz „SMAC", ein Tool welches auf „Sequential Model-Based Optimization", kurz „SMBO", aufbaut, dem wiederum der Ansatz der Bayes'schen Optimierung zu Grunde liegt. Marius Lindauer und Frank Hutter erklären die Idee hinter SMAC in ihrem Paper wie folgt: „The core idea of sequential model-based algorithm configuration is to iteratively fit an EPM [Empirisches Performanz Modell] based on the cost data observed so far and use it to guide the search for well-performing parameter configurations" [7].

Evaluation

Für die Untersuchung der AutoML Frameworks wurde von der Firma „IT-Designers" ein GPU-Cluster sowie das Tool „Polyaxon" zur Verfügung gestellt.
Polyaxon ist eine Open-Source Machine Learning-Plattform, die Vorgänge der MLOps unterstützt und

beschleunigt. Dieses Tool ermöglicht ML-Pipelines und -Workflows zu erstellen und in Kombination mit Kubernetes, Google Cloud, AWS und Azure zu deployen. Es bietet eine eigene REST-API und unterstützt gängige Deep Learning-Frameworks nativ. Während des Forschens nach bewährten Datensätzen zur Bewertung von ML-Modellen stellte sich heraus, dass aktuell kein Konsens zu diesem Thema bestand. Diesem Problem stellten sich Studierende der „Eindhoven University of Technology“. Pieter Gijsbers und sein Team entwickelten eine Benchmarking Suite für AutoML Lösungen, mit Datensätzen von „OpenML.org“, unter dem Namen „automlbenchmark“ (amlb). Hierbei definierten sie, welchen Kriterien Datensätze entsprechen müssen, um als Benchmark aussagekräftige Ergebnisse zu erzielen. Zu diesen Kriterien gehören u. a. die Komplexität, der Bezug zur realen Forschung der Datensätze und die Diversität der Problemdomänen. Die Benchmark Suit ist Open Source, modular erweiterbar und verfügt über eine Anbindung zu AWS. Derzeit sind 25 verschiedene AutoML Lösungen implementiert. (siehe: [3], [4])
Durch die Kombination des amlb-Projektes und Polyaxon ist es möglich automatisiert Modelle zu trainieren und zu vergleichen.

Aussicht

Bisher hat sich die Erweiterbarkeit der Benchmark Suite bewahrheitet, da innerhalb weniger Stunden das Framework „AutoPyTorch“ eingebunden werden konnte. Jedoch bestehen in Hinsicht auf das amlb-Projekt Verbesserungsmöglichkeiten, dazu gehören Kompatibilität von Deep Learning-Frameworks wie „AutoKeras“, weitere Anbindungen z. B. zu Google Cloud oder Azure und die Verbesserung der Darstellung der Ergebnisse. Weiterhin ist die Funktionalität bisher auf tabellarische Klassifikation und Regression beschränkt.
Betrachtet man den Status quo der AutoML Frameworks, ist das Konzept AutoML noch weit entfernt von futuristischen Vorstellungen der selbstlernenden künstlichen Intelligenz. Jedoch konnte der Grundgedanke des AutoML eingehalten werden. So war es möglich, die Grundfunktionen der einzelnen Frameworks ohne tieferes Wissen zu diesen einzusetzen und zufriedenstellende Modelle zu generieren.

Literatur und Abbildungen

[1] Matthias Feurer et al. Efficient and Robust Automated Machine Learning. In *Advances in Neural Information Processing*, volume 28, pages 2962–2970. NIPS, 2015.

[2] Larry Franks et al. Was ist automatisiertes maschinelles Lernen (AutoML)? https://docs.microsoft.com/de-de/azure/machine-learning/concept-automated-ml, 05 2022.

[3] Pieter Gijsbers et al. Benchmark Datasets. https://openml.github.io/automlbenchmark/benchmark_datasets.html, 04 2022.

[4] Pieter Gijsbers et al. OpenML AutoML Benchmarking Framework. https://openml.github.io/automlbenchmark/about.html, 04 2022.

[5] Michael Hesse. BAYESSCHE OPTIMIERUNG FÜR DIE INBETRIEBNAHME VON REGELUNGEN. https://www.hni.uni-paderborn.de/nachwuchsgruppe-dart/forschung/bayessche-optimierung-fuer-die-inbetriebnahme-von-regelungen/, 04 2022.

[6] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. Sequential Model-Based Optimization for General Algorithm Configuration. In *LION'05: Proceedings of the 5th international conference on Learning and Intelligent Optimization*, volume 5, pages 507–523. Springer-Verlag, 2011.

[7] Marius Lindauer and Frank Hutter. Warmstarting of Model-based Algorithm Configuration, 2017.

[8] Arthur Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3:210–229, 1959.

Quantifizierbare Cybersicherheit in der Automotiveindustrie

Nader Meschi **Reinhard Schmidt**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ALTEN GmbH, Stuttgart

Motivation

Durch die Weiterentwicklung neuer Technologien und autonomer Fahrzeugkonzepte steigt damit auch das Potenzial neuer Angriffe auf Fahrzeugsysteme. Daher wird es umso wichtiger, dass die Fahrzeuge vom Start der Entwicklung bis hin zur Außerbetriebnahme, Wert auf die Cybersicherheit legen. Ursächlich hierfür ist die Kommunikation innerhalb der eigenen bordinternen Fahrzeugarchitektur, wodurch verschiedene Angriffsquellen basierend auf Ethernet, Mobilfunk, Internet bzw. Wi-Fi, Bluetooth und viele mehr entstehen. Um dem Problem etwas entgegenzukommen, werden eine Reihe von Sicherheitsmaßnahmen und Prinzipien angewendet, die den zahlreichen Angriffen Stand halten sollen. Außerdem gilt es hier die Risiken gemäß den Auswirkungen einzustufen, um einen besseren Einblick zu erhalten, worum es sich primär um diese Arbeit auch handelt. Dieses Problem wurde bereits gelöst, allerdings nur im Rahmen der funktionalen Sicherheit von elektrisch bzw. elektronischen Systemen gemäß der ISO 26262. Nun soll ebenfalls eine Lösung basierend auf dem Standard der ISO 21434, welches für die Cybersicherheit der Straßenfahrzeuge zuständig ist, vorgeschlagen werden [1].

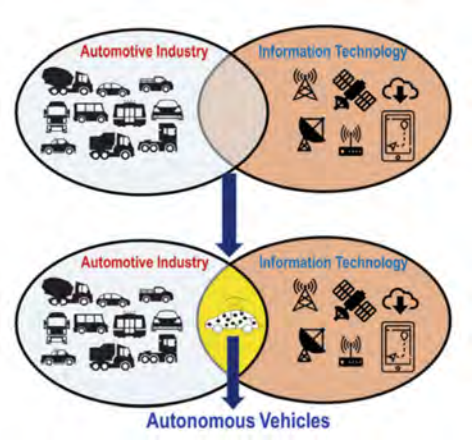

Abb. 1: Zusammenführung der Automobil- und Informationstechnologie [2]

Unterschied zwischen funktionaler Sicherheit und der Cybersicherheit

Die funktionale Sicherheit, definiert im Englischen als „Safety“ und die Cybersicherheit bzw. genauer gesagt, die Informationssicherheit beschreibt definitionsgemäß als „Security“. Bei der funktionalen Sicherheit geht es primär um die Sicherheit vor einem (internen) Versagen eines technischen Systems, wodurch primär der Mensch, die Umwelt und Maschinen bzw. Anlagen gefährdet sind. Der Fehler entsteht in diesem Fall innerhalb des technischen Systems und wird hauptsächlich durch schlechte Technik oder Manipulation verursacht.
Im Gegensatz dazu verfolgt die Cybersecurity das Ziel, die Sicherheit vor unberechtigten Zugriffen oder bösartigen Angriffen auf ein System der Informationsverarbeitung zu gewährleisten. In diesem Fall versucht sich ein Hacker von außerhalb des Systems Zugriff zu erzwingen. Dadurch könnten wichtige Sicherheitselemente wie die Bremse oder der Airbag manipuliert werden und somit zu einem Unfall führen. Daraus resultierend ist die Safety gefährdet, falls keine Security vorhanden sein sollte.

Bedrohungen der heutigen Generation

Obwohl Fahrzeuge auf der einen Seite sicherer werden, entstehen auf der anderen Seite potenzielle Angriffsvektoren sowohl für die Informationssicherheit als auch für die funktionale Sicherheit. Mit der Zeit häufen sich Bedrohungen, die aus unterschiedlichen Gründen entstehen. Ein kritischer Aspekt ist die Kommunikation unter den ECUs, basierend auf das von Bosch im Jahre 1986 vorgestellte CAN-Bussystem. Die Technologie wird demnach komplexer, baut aber trotzdem auf einfachen Protokollen auf und entspricht nicht den heutigen Sicherheitsrichtlinien in der Entwicklung.
Durch die On-Board Diagnostik bzw. OBD-Ports werden verschiedene Informationen wie Emissionen, dem Kilometerstand, der Geschwindigkeit und weiteren Daten gesammelt und überwacht. Darüber hinaus kann der Angreifer zum einen eine direkte, physische Verbindung zu allen CAN-Bussen herstellen, aber auch

zum anderen sogenannte Dongles angebracht werden. Dongles dienen als Funkadapter und können in die OBD-Ports gesteckt werden, um die Daten auch aus der Ferne dauerhaft abgreifen zu können. Mithilfe spezieller Tools können entstehende Fehlermeldungen leicht entziffert werden. Demnach ist der Angreifer in der Lage schädliche Nachrichten über den CAN-Bus an die Steuergeräte weiterzuleiten, was zu einer Gefährdung des Fahrers führt.
Außerdem beinhaltet jedes moderne Auto ebenfalls USB-Anschlüsse, die für Kommunikationszwecke mit GPS-Systeme, anderen USB-Zubehör und diversen Computern verwendet werden. Auch in diesem Fall ist die Installierung von Schadsoftware, die Manipulation der Netzwerkkarte sowie Angriffe auf das Betriebssystem nicht auszuschließen. Vor allem nahm die Nutzung der Elektrofahrzeuge in den letzten Jahren stark zu und muss im Vergleich zu herkömmlichen Fahrzeugen häufig aufgeladen werden. Hierbei besteht eine Gefahr eines Angriffs über die Ladeinfrastruktur während des Ladevorgangs, da der Angreifer durch einen potenziellen Angriff das Fahrzeug über die Ladestationen selbst kompromittieren könnte. Darüber hinaus existieren viele weitere Bedrohungen, bei der das Motiv des Angreifers trotz der bisherigen Erkenntnisse offenbleibt. Von Diebstahl bis hin zur Erpressung, Rufschädigung von Unternehmen oder sogar Mord und Terrorismus [1].

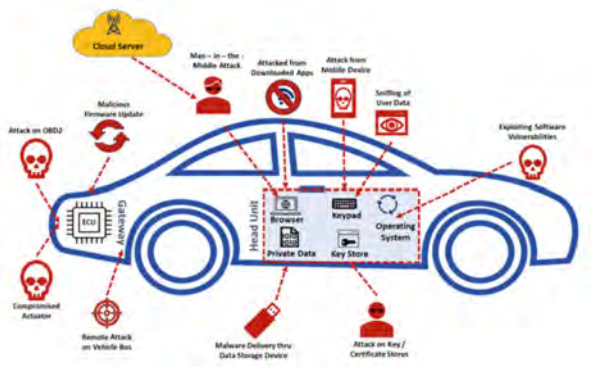

Abb. 2: Verschiedene Angriffsvektoren auf Straßenfahrzeuge [2]

Herausforderungen

Innerhalb der Automobilbranche sind viele Lieferanten an der Produktion beteiligt. Dies sorgt für eine erhöhte Wahrscheinlichkeit von Sicherheitslücken innerhalb der einzelnen Komponenten. Es ist somit nahezu unmöglich die Fahrzeuge zu sichern, da jedes ungesicherte Glied eine potenzielle Bedrohung darstellen kann. Außerdem ist es durch die aktuellen Lieferantenbeziehungen und Vereinbarungen nicht möglich, die End-to-End-Cybersecurity der Fahrzeugplattformen oder Technologiepaketen zu testen. Durch die einzelnen Komponenten der OEMs und Zulieferer wird das Entwickeln und Testen von Software in Bezug auf die effektive Informationssicherheit erschwert. Aus diesem Grund ist es entscheidend die Informationssicherheit während des gesamten Produktlebenszyklus zu berücksichtigen und nicht nur bis zum erfolgreichen Kauf des Kunden, da stets Sicherheitslücken mit der Zeit entstehen können. Aufgrund der riesigen Datenmengen, die in Form von eingebettetem Code zu verarbeiten sind, muss die Software im Gegenzug flexibler und leichter zu konfigurieren sein. Mithilfe von Over-the-air (OTA)-Updates gelingt es dem Hersteller ein entsprechendes Problem effizienter zu lösen. Dabei wird effektiv Zeit gespart und gleichzeitig eine kosteneffiziente Alternative bereitgestellt. Allerdings fordert der Prozess sowohl eine zuverlässige Umgebung innerhalb des Betriebszustandes als auch einen kontrollierbaren Datenfluss, damit der Schutz vor Manipulation und Missbrauch gewährleistet wird [1].

Ausblick

Das Ziel des Konzepts für die Quantifizierung ist es eine qualitative Risikobewertung anhand verschiedener Kriterien durchzuführen. Hierfür fließen zum einen die drei grundlegenden Sicherheitsziele Vertraulichkeit, Integrität und Verfügbarkeit mit ein aber auch zum anderen Ziele die sich bspw. auf die Widerstandsfähigkeit, der Zurechenbarkeit und Verlässlichkeit des Systems beziehen. Zusätzlich werden die vom Bundesamt für Sicherheit in der Informationssicherheit (BSI) definierten Schutzbedarfe in die Bewertungsskala integriert und mit den Reifegraden der TISAX verknüpft.

Literatur und Abbildungen

[1] Shiho Kim and Rakesh Shrestha. *Automotive Cyber Security*. Springer, 2020.

[2] Madhusudan Singh. *Information Security of Intelligent Vehicles Communication*. Springer, 2021.

Comparative analysis of image feature detection and matching algorithms

Hoang An Nguyen **Markus Enzweiler**

Department of Computer Science and Engineering, Esslingen University

Work carried out at Robert Bosch GmbH, Stuttgart

Introduction

One of the biggest hurdles of computer vision is the extraction of meaningful data from visual sensors. A possible solution to this problem are the feature detection and matching algorithms. These are crucial components of many computer vision applications, such as structure-from-motion, automated object tracking, 3D object reconstruction and more. With algorithms like for instance SURF, ORB, BRIEF, FAST or BRISK, which will be analyzed in the thesis, we are able to detect feature points in given images and determine their sets of correspondences across different images. Such correspondences can be used to align different images, e.g. when stitching image mosaics or performing video stabilization. They are also extensively used to perform object instance recognition [2]. These algorithms consist of three main components [3]:

1. Detection: Identification of feature points
2. Description: Find compact representations of a point's local neighborhood.
3. Matching: Use of descriptors to identify similar feature points across multiple images.

Motivation and Objective

Given the large number of feature detection/matching algorithms that have been developed in computer vision, it's hard to find the right methods that fit the best to one's use case. The aim of this thesis is to analyze various feature matching and detection algorithms, compare them to each other in regard to performance and robustness, especially against perspective changes. This will be useful for e.g. fully automated 3D-Modeling, camera pose estimation or real-time applications which use these methods, to get more performant and/or accurate results.

Feature Detection

Feature points are points which have expressive properties to them, that are relevant for solving a computational task in which they're used for. These properties include mountain peaks, building corners, street sign corners or more generally large contrast changes (gradients) preferably in at least two different orientations. They can be extracted with a multitude of feature detectors that have been implemented in recent times.

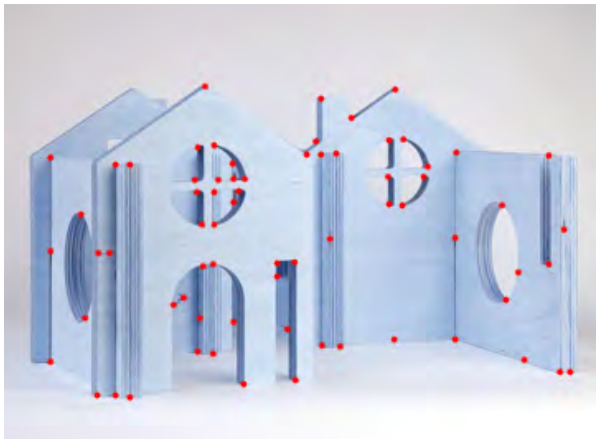

Fig. 1: Feature detection with Shi-Tomasi feature detector [1]

Feature Description

In our case, after extracting feature points, we need to match them to corresponding feature points in another image. In most cases, however, the local appearance of features in different images will change in orientation and scale, so the matching with feature points alone isn't feasible. As such, we use feature descriptors which encode interesting information around the feature point (e.g. SIFT encodes information about the local

neighborhood image gradients) which can be used to differentiate one feature point from each other.

Fig. 2: Feature Descriptors shown with SIFT (Scale Invariant Feature Transform) algorithm [1]

Feature Matching

After having extracted feature points and their descriptors from two or more images, the next step is to find correspondences between those images. The performance of those matching algorithms depends on the properties of interest points, the associated image descriptors, the matching approach that is taken [3] and the method how outliers are detected. Determining matching strategies depends on the context in which the matching is performed. Given two images that we know will overlap a fair amount (e.g. image stitching) we need to have a different approach as opposed to if we are trying to recognize objects from a database, where most of the features may not match. Additionally, a lot of objects must be loaded and searched in the database, which requires more efficient strategies [2]. In recent times there has been also an increase of machine learning algorithms like LoFTR and SuperGlue which use neural networks to find correspondences in images. These show promising results, even for low-texture areas, and sometimes also propose detector free methods.

Fig. 3: Feature matching between two different images of the same object [1]

Procedure

To give detailed performance comparisons of these feature detector and matching methods, we will perform benchmarks which will evaluate different combinations of these algorithms. The benchmark will include runtime, quality and robustness of the respective combination. The algorithm's quality is measured by the number of matches and outliers found given multiple input images, while the runtime will be measured by the time it takes to detect and match corresponding features in the images. Lastly, robustness will be measured by modifying the pictures with different effects, such as blur, scale, illumination and more. These input images will be captured by various cameras of a vehicle.

Further Work

Each feature detector, descriptor and matching algorithm's computational efficiency and robustness have a considerable impact on the time utilization and matching correctness of the respective method. Therefore, having a benchmark of multiple combinations of most state-of-the-art feature detection and matching algorithms will help tremendously in choosing the right methods for specific use cases in future works. This will be relevant for applications that need to figure out the right balance between getting the matching results in a timely manner, precision and robustness.

References and figures

[1] Own representation.

[2] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag, 2011.

[3] Deepanshu Tyagi. Introduction to Feature Detection and Matching. https://medium.com/data-breach/introduction-to-feature-detection-and-matching-65e27179885d, 01 2019.

Potentiale des Clouddatenmanagements in der Produktion

Anna Obenland **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ANDREAS STIHL AG & Co. KG, Waiblingen

Einleitung

Durch die immer schneller voranschreitende Digitalisierung stehen die Unternehmen vor zahlreichen Herausforderungen; moderne Unternehmensprozesse benötigen die passende IT-Unterstützung. Digitalisierungspotentiale finden sich unter anderem in den Bereichen intelligentes Produkt-Design, Entwicklung, Vernetzung, Informationstechnik, innovative Geschäftsmodelle und Produktion. Im Hinblick auf eine moderne Produktion werden intelligente und digital vernetzte Systeme benötigt, damit Informations- und Kommunikationstechniken verbunden werden können. Durch die dadurch verursachte Informations- und Datenflut können die Unternehmen an ihre Grenzen stoßen, was die internen Kapazitäten anbelangt. Auch mit Blick auf die Kosteneffizienz sollte nach anderen Lösungen, beispielsweise Cloud Computing, gesucht werden.

Zielsetzung

Die STIHL AG & Co. KG ist ein international aufgestelltes Unternehmen mit Sitz in Waiblingen im Bereich Motorsägen und Motorgeräte. Ziel der Bachelorarbeit ist es, für den Produktionsbereich der Firma STIHL aufzuzeigen, welche Daten und Services effizienter in einer Cloudumgebung gehalten bzw. betrieben werden können.

Cloud Computing

Cloud Computing ist die Bereitstellung von IT-Ressourcen wie Speicher, Netzwerkkomponenten und Software über das Internet. Charakterisiert wird Cloud Computing vor allem durch flexible Skalierbarkeit und nutzungsabhängige Kosten. Cloud-Dienste lassen sich in drei Kategorien unterteilen, die theoretisch betrachtet aufeinander aufbauen: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) und Software as a Service (SaaS). Während bei IaaS grundlegende Ressourcen wie Speicher gemietet werden, bezieht der Kunde bei SaaS eine komplette Software über die Cloud. Der Nutzer gibt also mit aufsteigender Ebene mehr und mehr seine Kontrollmöglichkeiten an den Cloud-Anbieter ab und erhält dafür aber mehr Leistungen. Die verschiedenen Arten, Dienste über die Cloud zu beziehen, unterscheiden sich im Nutzerkreis (vgl. Abb. 1): So teilen sich bei einer Public Cloud mehrere Kunden dieselbe IT-Infrastruktur, während eine Private Cloud exklusiv für einen Kunden bereitgestellt wird. Die Hybrid Cloud ist eine Mischform der zuvor genannten Cloudarten. Eine private Cloud, die speziell für eine Gemeinschaft von Organisationen entwickelt und betrieben wird, wird als Community Cloud bezeichnet [3].

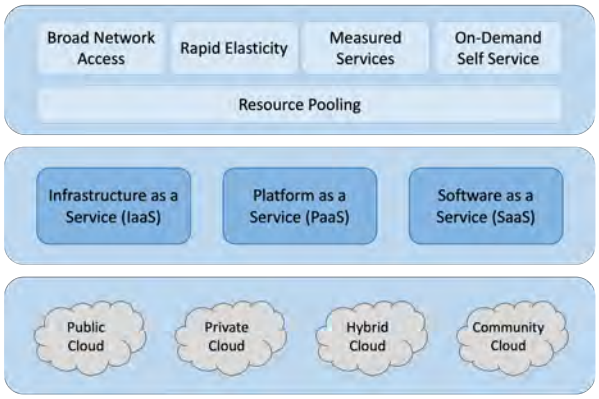

Abb. 1: Cloud Computing Grundlagen [4]

Die Nutzung von Cloud-Services kann für Unternehmen viele Vorteile bringen, u.a. reduzierte Kosten und erhöhte Agilität. Herausforderungen gibt es beispielsweise im Bereich Sicherheit und Abhängigkeit vom Cloud-Anbieter.

Daten in der Produktion

Die in der Produktion anfallenden Daten kann man in einem ersten Schritt grob in Maschinendaten und organisatorische Daten unterteilen (vgl. Abb. 2). Zu den organisatorischen Daten gehören unter anderem Personaldaten und Auftragsdaten; diese Daten werden im weiteren Verlauf der Bachelorarbeit nicht betrachtet. Die Maschinendaten gliedern sich in Betriebszustände, Störungen und Prozessdaten [2].

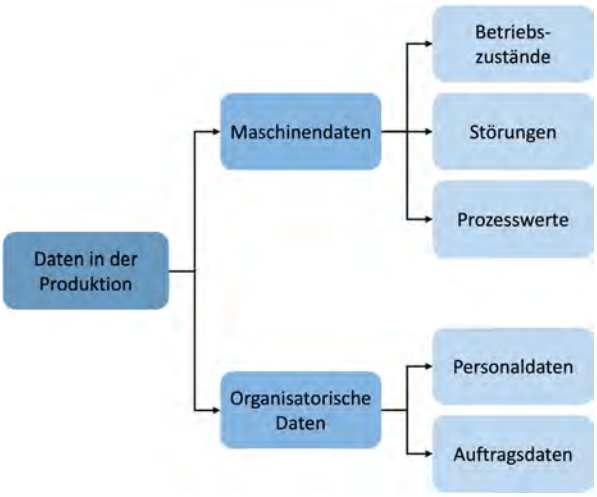

Abb. 2: Arten von Daten in der Produktion [1]

Die Daten unterscheiden sich in ihren Bedürfnissen und können so charakterisiert werden. So müssen Daten von Störungen in Echtzeit abgebildet werden, damit darauf umgehend reagiert werden kann. Prozessdaten, die einem speziellen Produkt zugeordnet werden, müssen zum Beispiel für Rückrufaktionen lange vorgehalten werden. Anhand dieser identifizierten Charakteristika wie Latenzzeit und Vorhaltungszeit kann für die verschiedenen Produktionsdaten eine Abwägung der Chancen und Herausforderungen durch die Auslagerung in eine Cloud vorgenommen werden. Hierbei ist auch festzulegen, welche Kategorie an Cloud-Diensten gewählt werden soll.

Ausblick

In dieser Bachelorarbeit soll ein Überblick über die Chancen und Risiken des Cloud Computing im Bereich Produktion aufgezeigt werden. Hierzu ist es erforderlich, nicht nur die Daten, die in der Produktion entstehen und benötigt werden, zu erfassen und zu bewerten, sondern auch noch die Produktionssteuerung und –prozesse näher zu betrachten. Daraus lassen sich Rückschlüsse auf die Anforderungen formulieren, die für eine Auslagerung in eine Cloud erforderlich sind. Zu berücksichtigen sind dabei die Wirtschaftlichkeit und der reibungslose Ablauf der Produktion.
Zu erwarten ist, dass Produktionsdaten zumindest teilweise ausgelagert werden können.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Helmuth Gienke and Rainer Kämpf. *Handbuch Produktion*. Carl Hanser Verlag, 2007.

[3] Peter Mell and Timothy Grace. The NIST Definition of Cloud Computing. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf, 2011.

[4] P Naveen et al. Cloud computing for energy management in smart grid - an application survey. https://iopscience.iop.org/article/10.1088/1757-899X/121/1/012010/pdf, 2016.

Entstehung von Digital Footprints und ihre Verwendung für unternehmerische Entscheidungen

Pinar Oezbey **Anke Bez**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Digitale Technologien sind heute in allen Bereichen des täglichen Lebens zu finden. Das gilt bei privater Nutzung, in der Menschen bevorzugt mit Smartphones über Social Media kommunizieren, aber auch im beruflichen Bereich, in der programmierte digitale Maschinen, wie z.B. Roboter, Produktionsprozesse und Softwaresysteme im täglichen Workflow unterstützen. Ein wesentlicher Treiber der zunehmenden Digitalisierung ist, dass durch die wachsende Verfügbarkeit des Internets auch eine Echtzeitkommunikation mit geografisch weit verteilten Stakeholdern möglich ist. Die Kommunikation und der Austausch von Daten sowie die Speicherung und Verarbeitung von Daten sind sehr wichtig geworden. Mit Hilfe großer Datenmengen und maschinellen Lernmethoden der künstlichen Intelligenz ist es möglich, das in diesen Daten enthaltene Wissen zu extrahieren und typische, wiederkehrende Verhaltensmuster von Maschinen oder Menschen zu analysieren. [2]

Ziel der Arbeit

Das Ziel dieser Bachelorarbeit ist es, anhand einer umfassenden Literaturrecherche erfassbare digitale Fußabdrücke von Online-Konsumenten aufzuzeigen sowie die folgenden Annahmen zu prüfen:

- Welche Kennzahlen und Daten sind für die Analyse des Online-Konsumentenverhaltens und deren Erkenntnisgewinn besonders hilfreich?
- Welche Implikationen können Unternehmen daraus ableiten, um bessere Entscheidungen zu treffen?
- Wie wird der wirtschaftliche Erfolg eines Unternehmens durch den Einsatz von Digital Footprints beeinflusst?

Dabei ist auch zu beachten, inwieweit Unternehmen digitale Footprints einzelner Personen laut der Datenschutz-Grundverordnung verwenden dürfen, um einen Mehrwert zu generieren.

Digital Footprints

Jeder von uns erzeugt bei der Nutzung digitaler Technologien, z. B. beim Surfen im Internet, Chatten, Posten und Telefonieren, Daten, die dann gespeichert und verarbeitet werden können. Diese digitalen Spuren (digital footprints), die jeder Nutzer im Internet hinterlässt, können nachverfolgt und in unterschiedlichem Umfang verwendet werden. Digitale Fußabdrücke können als passiv und aktiv klassifiziert werden und laufen parallel zum Leben ab (vgl. Abb. 1). Hauptsächlich unterscheiden sich die Typen dadurch, dass der aktive digitale Fußabdruck durch bewusste Entscheidungen, wie z.B. eine Instagram-Story oder durch die Annahme eines Cookies auf einer Webseite, entsteht. Passive digitale Fußabdrücke dagegen hinterlässt der Nutzer, ohne es beabsichtigt zu haben bzw. ohne es zu wissen, wie bspw. durch die Verbindung der IP-Adresse mit der Webseite. Der digitale Fußabdruck eines Babys beginnt somit bereits durch einen Instagram-Post einer Mutter mit ihrem Neugeborenem. [3]

Abb. 1: Passive und Aktive Digital Footprints [1]

Unternehmerische Umsetzung

Für Unternehmen ist es heutzutage äußerst wichtig, eine gute Kundenbeziehung aufzubauen und weiter zu pflegen. Daher ist der digitale Fußabdruck von großer Bedeutung, um Zielgruppen mit maßgeschneiderten Anzeigen zu definieren und anzusprechen. Für einen potenziellen Kunden wird dies dadurch sichtbar, dass bereits ein Klick auf einer Website auf ein neues Paar Schuhe ausreicht, um später dieselben oder ähnliche Artikel auf anderen Webseiten eingeblendet zu bekommen. Anhand verschiedener Tracking-Mechanismen erkennen Unternehmen die Fußabdrücke der Nutzer im Netz. Durch Analysetools werden die Daten verknüpft und sortiert und dienen danach zur Profilerstellung der Nutzer; zusätzlich können diese Daten an Dritt-Unternehmen weiterverkauft werden. Dadurch kann dem Nutzer personalisierte und individualisierte Werbung angezeigt werden, mit dem Ziel, dass dieser User das Angebot ansprechend findet und nutzt. Der digitale Fußabdruck jedes einzelnen Nutzers ermöglicht es somit den Unternehmen, wertvolle Informationen zu erhalten und den Kunden auf einer individuellen Ebene ansprechen zu können. Aus Unternehmenssicht können somit effizientere Marketing-Strategien umgesetzt werden, da die Zielgruppen im Netz erreichbar sind und mit gezielten Angeboten mehr verkauft werden kann. [4]
Die größte Schwierigkeit für die Unternehmen ist dabei, die richtige Balance zwischen dem Data Mining, den digitalen Fußabdrücken und der Privatsphäre der User zu finden.

Ausblick

Die digitale Transformation hat sich in den letzten Jahren professionalisiert und sich in die Planung und Umsetzung vieler Unternehmensprozesse erfolgreich etabliert. Die Trends in Bezug auf Collaboration-Tools, digitales Marketing und Cloud-Computing werden durch die dynamische und schnelllebige Online-Welt weiter zunehmen. Daher ist es umso wichtiger, als Unternehmen Prioritäten zu setzen und sich auf die digitale Transformation zu konzentrieren. Jedoch sollten Digital Footprints trotzdem kritisch betrachtet werden. Mit wachsender Datenspur und zunehmender Anzahl von gesammelten Informationen kann die Privatsphäre einer Person weitestgehend öffentlich zugänglich für verschiedenste Unternehmen, die Gesellschaft und Politik sein.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Gregor Engels. Der digitale Fußabdruck, Schatten oder Zwilling von Maschinen und Menschen. https://link.springer.com/article/10.1007/s11612-020-00527-9, 2020.

[3] University of Aberdeen IT Security Team. What is your Digital Footprint? https://www.abdn.ac.uk/news/12949/, 2019.

[4] Ofer Mintz. *The Post-Pandemic Business Playbook*. Palgrave Macmillan, 2021.

Evaluierung virtueller Daten zum Lernen einer Stixel Repräsentation

Kadir-Kaan Oezer **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Was ist ein Stixel?

Machine Vision und damit einhergehend Autonomes Fahren gehören zu den großen Themen unserer Zeit. Sie bieten uns Vorteile von kleinen Anwendungsbereichen, wie dem Verwenden vom Porträt Modus oder blitzschnellen HDR Algorithmen auf dem Mobiltelefon bis hin zur Erkennung der Objekte im Straßenverkehr eines Fahrzeuges. Eines der großen Probleme die damit einhergehen ist jedoch, die Menge an Daten, welche gesammelt und analysiert werden müssen. Um dieser großen Datenmenge entgegenzusteuern, wurden Stixel entworfen. Das Wort Stixel setzt sich dabei aus den englischen Wörtern stick und pixel zusammen und beschreibt ein vertikales aus Pixeln bestehendes Rechteck, welches eine feste Breite besitzt, in der Höhe jedoch variabel ist, um sich an das jeweilige Objekt anzupassen. Die Stixel sind dabei vertikal angesetzt, da in einer typischen Verkehrsszene die horizontale Ebene Dinge beherbergt, wie die Straße oder den Boden und im vertikalen dann die relevanten Objekte zu finden sind, wie Autos, Fußgänger oder Gebäude. [3] Stixel bieten durch diese Darstellung eine Möglichkeit Objekte zu erkennen, ohne verhältnismäßig stark an Genauigkeit zu verlieren. Auch die Tiefeninformationen bleiben erhalten, so sind in vielen Darstellungen Stixel, die eine geringe Distanz zur Kamera haben rot und weiter entfernte Stixel grün mit gelben Stixeln dazwischen, dabei werden die horizontalen Bereiche, wie die Straße, grau eingefärbt 1.

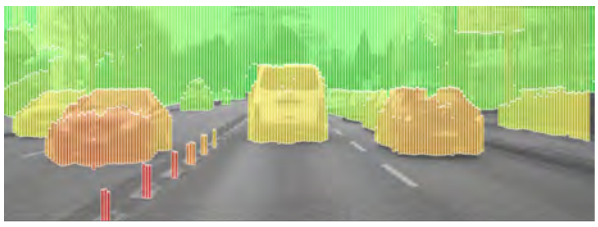

Abb. 1: Stixel-Welt mit angepasster Stixel Farbe, je nach Entfernung [3]

Als Endergebnis, hat man eine Objekterkennung, welche nur noch die Analyse weniger hundert Stixel erfordert, anstelle vieler Millionen Pixel.

Virtuelle Stixel

In dieser Arbeit, wurden die Stixel allerdings nicht in einer echten Verkehrsszene, sondern mithilfe des Open-Source Fahrsimulators Carla erstellt. Das bietet einem gleich mehrere Vorteile. Diese sind:

- Carla bietet einem die Möglichkeit die Verkehrsszenen mit Python zu bearbeiten
- Carla erlaubt es alle Objekte farblich mit einer Genauigkeit auf Pixel-Ebene voneinander zu trennen
- Carla benötigt keine Stereokamera zur Erstellung der Tiefenkarte
- Durch die Möglichkeit sehr genau positionierte Stixel zu erstellen, können die erstellten Daten als Ground-Truth dienen.
- Es ist theoretisch möglich unendlich Testdaten mit minimalem Aufwand zu erstellen

Innerhalb von Carla wurde dabei eine Kamera für die semantische Segmentierung als auch eine Kamera für die Tiefendaten am Fahrzeug angebracht. Die Bilder der semantischen Segmentierungskamera wurden dabei verwendet, um Masken in Python zu erstellen mit den jeweiligen RGB Farbcodes der zu erkennenden Objekte. Diese Farbcodes sind in der Dokumentation des Sensors auf der Carla Website zu finden und erlauben es jedes Objekt innerhalb der Simulation anzusprechen. Das wären Autos und Fußgänger, aber auch Gebäude, Verkehrszeichen und mehr 2.

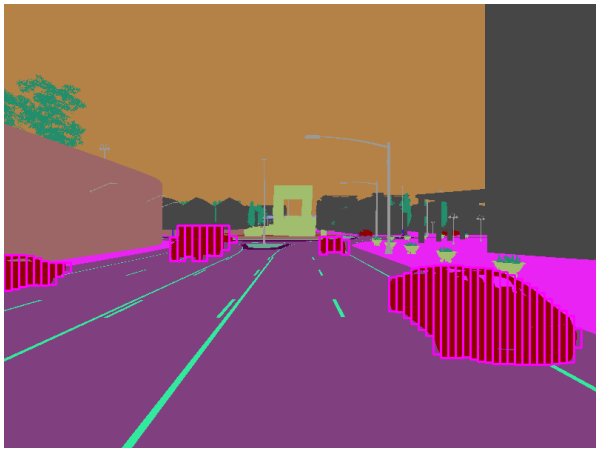

Abb. 2: Stixel auf Autos im Carla Simulator [1]

Wie hier zu sehen ist, sind die Stixel nur auf den Autos erstellt worden, man kann jedoch so viele verschiedene Objekte wie möglich im Code angeben und somit sehr gezielt Stixel erstellen lassen. Man sieht auch wie die Objekte im Simulator, mit der semantischen Segmentierungskamera, alle eine andere Farbe haben. Die Stixel sind allerdings nicht, wie oben zu sehen, entsprechend ihrer Entfernung eingefärbt. Das hat zwei Gründe. Zum einen soll das Endergebnis der erstellten Stixel innerhalb des Simulators in einer JSON Datei, mit den jeweiligen Pixelkoordinaten der Stixel gespeichert werden, um sowohl die Position, als auch die Tiefe ablesen zu können. Somit wäre eine Einfärbung zwar visuell hilfreich um die Richtigkeit der Tiefendaten einschätzen zu können, sie sind aber nicht für das Ergebnis erforderlich. Und noch dazu, ist das Abrufen der Tiefe der erstellten Stixel zum Zeitpunkt der Erstellung dieses Artikels noch nicht fertig implementiert. Das Vorgehen zum erstellen der Stixel ist dabei wie folgt: Zuerst speichert man die Bilder, welche von der Segmentierungskamera erstellt werden auf der Festplatte und wendet dabei auch den Parameter 'CityScapesPalette' auf sie an, um sie besser erkennbar zu machen. Im Anschluss werden die Bilder mit Python erneut im Skript aufgerufen und das anfängliche RGB Bild wird in das HSV Farbspektrum umgewandelt, da dieses Spektrum nicht nur resistenter gegen verschiedene Lichtstimmungen ist, sondern auch die Auswahl von Bereichen im Bild, welche eine bestimmte Farbe haben erleichtert. Das ist wichtig, um mithilfe der Python Bibliothek Skimage auf diese Bereiche zuzugreifen und Informationen über diese zu bekommen. Mit Skimage lassen sich einmal sehr genaue Bounding Boxen an den jeweiligen Stellen erstellen und noch dazu die Koordinaten aller Pixel in der jeweiligen Farbe zurückgeben. Diese Koordinaten sind dann je nach Region (in unserem Fall je Auto) in Arrays gespeichert. Die Bounding Box wird nun verwendet, um die Breite des Objektes zu bestimmen und damit die Anzahl der Stixel, welche benötigt werden. Im Anschluss werden die Arrays mit den Pixelkoordinaten danach gefiltert, wie breit die jeweiligen Stixel auf der Position sind. aus diesem gefilterten Bereich holt man sich dann die maximalen und minimalen y-Koordinaten, um den Anfang und das Ende eines Stixels zu definieren. Dadurch sind die Stixel exakt so hoch und tief, wie das Objekt an den jeweiligen Stellen. Das Skript selbst ist zum Zeitpunkt dieses Artikels noch nicht fertig. Es liefert bereits die Tiefendaten jedes einzelnen Punktes in der Simulation relativ zur Kamera, aber dieses Datenarray, wurde noch nicht je Stixel gefiltert und im Anschluss in eine JSON Datei geschrieben, für das Trainieren des neuronalen Netzes mit virtuellen Daten. Optional kann man die Stixel je nach Entfernung einfärben, wie oben zu sehen war 1.

Ziel der Arbeit

Das Ziel dieser Arbeit ist es zu evaluieren, ob die virtuell erstellten Testdaten mit tatsächlichen Testdaten mithalten können, wenn es darum geht ein neuronales Netz damit zu Trainieren. Dabei wird das Netz mithilfe der im Simulator erstellten Testdaten trainiert und verglichen mit den Ergebnissen des Netzes, wenn es mit echten Daten trainiert wird. Das Netz selbst ist für die Erstellung von Stixeln auf Bildern zuständig. Der Vorteil der Simulatordaten ist wie vorher erwähnt, dass sie durch ihre pixelgenaue Darstellung der Stixel eine Art Ground-Truth für die Stixel-Erstellung bieten. Das Unterscheidet sich von früheren Vorgehen, in welchen das neuronale Netz mit einer Point-Cloud Ground-Truth aus einem Lasersensor trainiert wurde. [2]

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Dan Levi, Noa Garnett, and Ethan Fetaya. StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation. In *Proceedings of the British Machine Vision Conference (BMVC)*. BMVA Press, 2015.

[3] David Pfeiffer and Uwe Franke. Towards a Global Optimal Multi-Layer Stixel Representation of Dense 3D Data. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2011.

Konzeption und Implementierung eines Dienstes zur automatischen Überwachung der eni.os Instanzen auf Kundensystemen

David Plattner **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Enisyst GmbH, Pliezhausen

Einleitung

Da die Anzahl von geschäftlichen Anwendungen mit dem Wachstum von Microservices und der Migration zu unterschiedlichen Cloud-Umgebungen zunimmt, ist es im Laufe der Zeit schwieriger geworden diese umfassend zu überwachen. Hierbei soll Application-Performance-Management (APM) Abhilfe verschaffen, welches Services, Prozesse, Hosts, Logs sowie Netzwerkdaten überwacht. Dies dient dem Ziel die Verfügbarkeit zu maximieren und den Endnutzer das beste Erlebnis zu bieten. Logging ist Teil von APM, aber Log-Management ist eine separate Funktion von APM (siehe Abbildung 1). Log-Management ist der Prozess der Behandlung von Logereignissen, die von allen Softwareanwendungen und der Infrastruktur generiert werden, auf der sie ausgeführt werden. Es umfasst Logerfassung, Logaggregation, Parsing, Speicherung, Analyse, Suche sowie Archivierung mit dem Ziel die Daten für die Fehlerbehebung und Einsicht in Geschäftsprozesse zu verwenden und gleichzeitig Verfügbarkeit und Sicherheit von Anwendungen sowie Infrastruktur zu gewährleisten.

	User Interaction	Data Enrichment	Data Management	Data Sources
Log Management	**All Alerts, Dashboard, Search** • Dashboards from any data • Alerts on any data • Search all data • Root-cause analysis • Threat hunting	**All Log and Event Correlation** • All Logs and events • Correlate any data	**Collect all machine data, Compression, Storage** • Collect any data • Live streaming (if available) • Indexing (if required) • Compression • Local of cloud storage • Long-term retention for exploration	**Logs, Metrics, Traces, Events, Machine Data** • All logs and event data • Traces, Metrics, and other machine data
APM	**App Alerts, Dashboards, Tracing, Search** • App reporting • App performance dashboard • App alerts • Root-cause analysis • Error and source maps	**Application Data & Transactions Correlation** • Transactions • Correlation with app performance, service, and system data	**Collect and Store App data** • App data ingest • Storage (7-14 days)	App data, Calls, Traces, Errors, Performance • App metrics and traces • App events • App changes • App exceptions and errors • Configuration data

Abb. 1: Log Management und APM [3]

Logging-Tools

Der Markt bietet eine Vielzahl von möglichen Tools zur Überwachung von Log-Daten wobei der ELK-Stack eine der weitverbreitesten Ansätze ist. Dieser besteht aus den drei Open-Source Projekten Elasticsearch, Logstash und Kibana. Elasticsearch ist eine Suchmaschine und Analytics-Engine. Logstash ist eine Datenverarbeitungspipeline, die Daten aus unterschiedlichen Quellen gleichzeitig verarbeiten sowie umwandeln kann

und anschließend an einen Speicherort sowie zum Beispiel Elasticsearch, sendet. Durch Kibana werden die Daten durch Diagramme und Tabellen als Visualisierung angezeigt. In letzter Zeit ist jedoch immer öfters vom sogenannten EFK-Stack im Zusammenhang mit Kubernetes zu hören. Hierbei ersetzt Fluentd Logstash. Fluentd ist Teil der Cloud-Native-Computer-Foundation (CNCF) und bietet viele Vorteile mit einer großen Anzahl an unterstützten Plattformen sowie der lightweight Variante fluentbit. Jedoch ist für das Monitoring anzunehmen, dass eine self-hosted Variante schnell zu einem Problem mit der Skalierbarkeit der Daten sowie zu weiterem Ressourcenaufwand führt. Außerdem würde zusätzliche Arbeitszeit für Pflege und Wartung aufkommen. Aufgrund dessen werden auch SaaS/LaaS-Anbieter überprüft. Um Fehler im Monitoring Ablauf zu vermeiden werden Ranum's-Logging-Gesetze angewand [2].

1. Niemals mehr Log-Daten sammeln, als geplant sind zu verwenden
2. Log-Daten so lange speichern, wie es denkbar ist diese zu verwenden - außer es ist gesetzlich länger Vorgeschrieben
3. Log alles was man kann, aber reagiere nur auf das nötigste
4. Zahle nicht um dein Monitoring verfügbarer zu machen als dein Geschäftssystem
5. Zahle nicht mehr für den Schutz deiner Log-Daten, als für den Schutz kritischer Geschäftsdaten
6. Log-Quellen, Log-Typen und Log-Nachrichten ändern sich

Inhalt der Arbeit

Durch die stets wachsende Anzahl an Kundensystemen ist es der Firma Enisyst nicht mit vertretbarem Aufwand möglich die Systeme laufend manuell zu überwachen. Dadurch werden Probleme und kritische Zustände nur in akuten Fällen entdeckt. Weiterhin ist es durch den begrenzten Speicher der Systeme nicht möglich Log-Daten die älter als einen Tag sind zu untersuchen. Deshalb ist es das Ziel dieser wissenschaftlichen Arbeit mit Hilfe eines Tools die vorhandenen Log-Einträge auf den Controllern an einen zentralen Cloud-Service weiterzuleiten und dort abzuspeichern, um diese später für Erkennung von generellen Problemen über Systeme und Kunden hinweg nutzen zu können. Dabei muss jedoch auf Kompatibilität der Kundensystemen, die auf Embedded-Linux auf ARM-Basis laufen, geachtet werden. Außerdem muss das Tool die Möglichkeit bereitstellen Log-Daten aus dem journald zu lesen, da dort die Log-Daten des eni.os eingetragen werden, in Abbildung 2 wird dies durch den Agent dargestellt. Gesucht wird nach einer technisch aktuellen Lösung welche mit geringem Aufwand in die bestehenden Systeme integriert werden kann. Das Monitoring-Tool sollte außer dem Speichern von Log-Daten noch die Funktion enthalten Log-Daten zu analysieren und gegebenenfalls Alerts an das Software-Team oder Kunden zu senden. Da nicht nur das Software-Team das Monitoring-Tool verwenden soll wird in der Auswahl die Möglichkeit nach Rollen basiertem Zugriff mit einbezogen. Da zahlreiche Kombinationen von Agents und Log-Monitoring-Tools in Frage kommen, soll mit Hilfe eines Kriterienkatalogs entschieden werden welche Lösung den Ansprüchen von Enisyst entspricht. Außerdem sollen Tests auf den Controllern der Hersteller AVNET und PHYTEC mit den unterschiedlichen Tools durchgeführt werden und ein konkreter Vergleich erstellt werden.

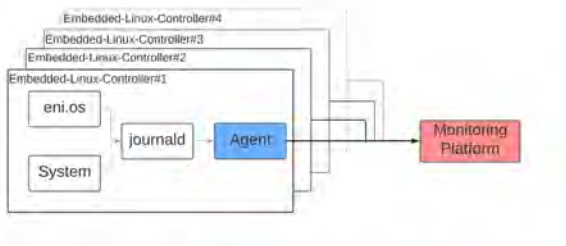

Abb. 2: Architekturkonzept [1]

Ausblick

Für die Zukunft stellt sich die Frage wie mehr Informationen aus Log-Nachrichten entnommen werden können. Derzeit existieren keine Regeln was genau geloggt wird was zu nicht ausschlaggebenden Nachrichten führt. Dagegen soll die Implementation von Log-Policies zu Klarheit führen wodurch besseres Monitoring sowie Fehlerbehebung entsteht. Außerdem befindet sich die Firma Enisyst momentan in der Planung sowie Gestaltung einer App für Ladestationen, wobei das Monitoring in Form von Real-User-Monitoring (RUM) ein Rolle spielen könnte. Ansonsten könnte das Monitoring bei der Serverinfrastruktur eingesetzt werden zur Überwachung des eni.serv.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Anton A Dr Chuvakin, Kevin J Schmidt, and Christopher Phillips. *Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management*. Syngress, 1 edition, 2012.

[3] Humio Staff. What is Application Monitoring? https://www.humio.com/glossary/application-monitoring/, 09 2021.

Entwicklung und Validierung eines Embedded Systems zum Monitoring von Druckluftsystemen

Lucas Rees **Clemens Klöck**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma WRS Energie + Druckluft GmbH, Obersulm

Einleitung

Das Unternehmen WRS Energie + Druckluft GmbH monitoret Druckluft Anlagen und gibt Handlungsempfehlungen zur Verbesserung des Energieverbrauchs. Eine Druckluftanlage in diesem Sinne besteht aus einem oder mehreren Kompressoren, einem Druckluftspeicher, dem Verteilernetz und den Anlagen, welche die Druckluft nutzen. Zur Aufnahme der benötigten Daten werden Druckluftsensoren, Durchflusssensoren und Leistungsmesser in Kombination mit einem Datenlogger benutzt.

Motivation und Ziel der Arbeit

Da der Datenlogger per Mobilfunk kommunizieren soll stehen nur eine begrenzte Anzahl an Modellen zur Auswahl. Alternative Systeme wären der ETM DeltaBlack [2] und der LineMetrics Datenlogger [4]. Beide verfügen über Analoge und Digitale Eingänge und kommunizieren über Mobilfunk. Der Datenlogger von LineMetrics wird aktuell eingesetzt und soll gegen das eigenentwickelte Systeme ersetzt werden. Grund dafür ist die Abhängigkeit von dem Hersteller in Bezug auf die Hardware aber auch das Cloud-Produkt, an welches man gebunden ist. Hinzu kommt, dass ein eigener Datenlogger genau auf die Anforderungen zugeschnitten werden kann und das Speichern der Daten auf eigenen Servern eine zusätzliche Sicherheit für Kunden bedeutet. Somit ergibt sich das Ziel der Arbeit als Entwicklung des Datenloggers und Implementierung eines Datenbank-Servers zum Speichern der Daten.

Abb. 1: Render des fertigen Datenlogger Designs [3]

Anforderungen

Die Hardwareanforderungen gehen von der Auswahl der Sensorbuchsen über die Auflösung d es ADCs bis hin zum benutzten Mobilfunk Standards. Die Sensorbuchsen zu gängigen Sensorkabeln kompatibel und direkt auf der Platine verlötbar sein. Damit aktive und passive Sensoren genutzt werden können benötigt jede Buchse einen 24V und Ground Anschluss. Alle 8 Buchsen sollen über 4-20mA Eingänge verfügen und 4 sollen zusätzlich auf einem zweiten Pin als digital Eingänge dienen. Um die Buchsen einheitlich zu halten, sollen deshalb für alle 8 Eingänge 4 Polige Buchsen zum Einsatz kommen. Damit die Sensoren mit Spannung versorgt werden können benötigt der Datenlogger einen 24V Eingang und interne Spannungswandler

auf 3.3V zur Versorgung der Hardwarekomponenten. Um die 4-20mA Signale aufnehmen zu können wird ein Analog-Digital-Converter mit entsprechender Stromwandlerschaltung benötigt. Die 4 Digitaleingänge sollen 24V tolerant sein und benötigen somit einen Optokoppler zur sicheren Ansteuerung der 3,3V toleranten Microcontroller Eingänge. Da die Daten exakt jede Sekunde mit Zeitstempel aufgenommen werden sollen wird eine Echtzeit Uhr, mit geringer Abweichung und Batterie zum Erhalten der Uhrzeit bei Spannungsverlust, benötigt. Wie schon in dem Ziel erwähnt, sollen die Daten über Mobilfunk übertragen werden. Zur Visualisierung des Status des Datenloggers sollen jeweils eine LED pro Sensor Port und 3 LEDs für andere Zwecke bereitstehen. Da viele Sensoren und Steuerungen Daten über Modbus-RTU [5] übertragen soll auch diese Schnittstelle als Bus Schnittstelle zur Verfügung stehen. Final stellt sich die Anforderung für den Microcontroller, dieser sollte über mindestens 11 digital Ausgänge für das Ansteuern der LEDs, 2 UART Ports für Modbus und das GSM Modul sowie ein i2c Port für die ADCs und RTC verfügen. Zusätzlich sollte der Microcontroller genug Arbeitsspeicher zum Puffern von Datenpunkten für 24h besitzen. Die Softwaretechnischen Anforderungen auf Seite des Datenloggers beschränken sich größtenteils auf zuverlässige Treiber für die Hardwarekomponenten. Es soll möglich sein die Eingänge und Modbus Schnittstell zu konfigurieren und diese Konfiguration abzuspeichern und bei erneutem Anschluss des Datenloggers zu laden. Die eingelesenen Analogsignale sollen in 4-20mA Einheitssignal umgewandelt werden und bei Unterschreiten der 4mA und überschreiten der 20mA soll dies am Blinken der Status LED erkennbar sein. Die Datenübertragung soll das MQTT [1] Protokoll nutzen und somit benötigt jeder Datenlogger eine Individuelle ID, welche auch zur Identifikation der einzelnen Box genutzt werden kann. Zusätzlich sollte die Datenübertragung verschlüsselt Stattfinden da das Passwort für den MQTT Server übermittelt werden muss. Auf Serverseite wird ein MQTT Broker, welcher ebenfalls Verschlüsselung unterstützt benötigt. Die eintreffenden Daten sollen in einer Datenbank gespeichert werden und auf Fehler überprüft werden.

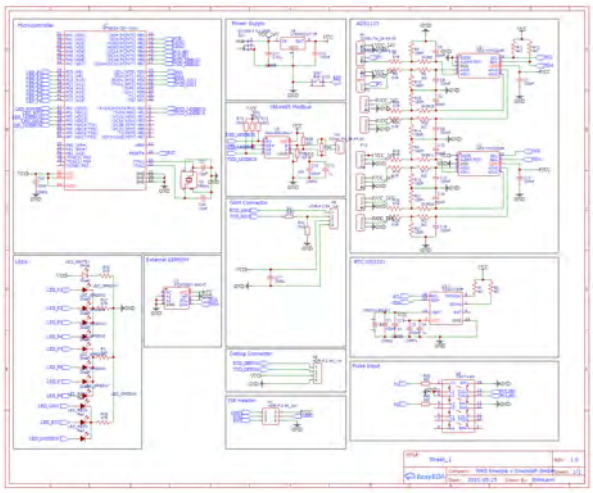

Abb. 2: Schema der Platine mit ATmega1281 [3]

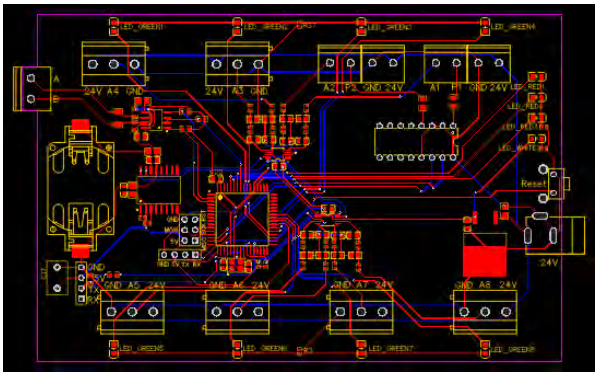

Abb. 3: Layout der Platine mit ATmega1281 [3]

Umsetzung

Zu beginn wurden die Anforderungen festgelegt. Diese waren zu Anfangs stark an dem Datenlogger von Line-Metrics orientiert und wurden im laufe der Entwicklung immer wieder angepasst. Anhand der festgelegten Anforderungen wurden dann die relevanten Hardwarekomponenten recherchiert und der erste Prototyp auf Basis eines Arduino Boards entwickelt. Dabei wurde im ersten Schritt die Schaltung auf mehreren Steckbrettern aufgebaut und nach testen von ADC und RTC eine erste Platine aus Lochrasterplatinen mit den Steckmodulen zusammengelötet. Beim Testen des ersten Firmenware Prototyps wurde festgestellt, dass der Arbeitsspeicher des Arduino Mega zu klein für das Programm ist und danach auf den Teensy 3.5 umgestiegen. Der Teensy 3.5 ist 5V kompatibel und verfügt über 256KB RAM. Neben den folgenden Tests zum Ansteuern des GSM Moduls SIM800L, wurde eine erste Platine auf Basis des Atmega1281 erstellt. Nach erhalten der Platine stellte sich heraus, dass der

Verwendetet Krystal Oszillator falsch beschaltet war und der Microcontroller deshalb nicht zu programmieren war. Danach folgte weitere Handgelötete Platinen sowie extern hergestellt und bestückte Prototypen, welche den Microcontroller als aufsteckbaren Header erhielten. Dies hat den Vorteil, dass bei Problemen mit den Microcontroller nur der Header geändert werden muss und die Hauptplatine weiter genutzt werden kann. Nach einigen Tests und Prototypen mit mehr Funktionen, stieß die TinyGSM Bibliothek, welche als Treiber für das SIM800L Modul dient, an die Grenzen und es kam zu Speicherfehlern. Da Debugger für Teensy und Arduino Boards aufgrund des Chipmangels entweder sehr teuer oder nicht verfügbar waren, konnte der Fehler in der GSM Bibliothek nicht gefunden werden. Nach abwägen des mehr Aufwands wurde auf das Raspberry Pi Pico Board umgestiegen. Dieses von der Raspberry Foundation entwickelte Board basiert auf dem RP2040 ARM Cortex-M0 Microcontroller und hat den Vorteil, dass es mit einem Zweiten Pico Board gedebugged werden kann. Da das Raspberry Pi Pico Board noch nicht vollständig in das Arduino Framework integriert ist, wurde auf die Entwicklung in C mit dem Pico-SDK umgestiegen. Die Möglichkeit des Inline-debugging und der sehr guten Dokumentation haben die Entwicklung stark beschleunigt, auch wenn alle Treiber neu geschrieben werden mussten.

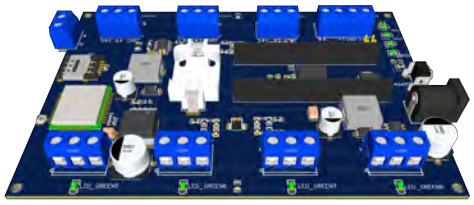

Abb. 4: Render der aktuellen Platine mit Raspberry Pi Pico Steckplatz [3]

Aktueller Stand und weiteres Vorgehen

Nach aktuellem Stand sind folgen Funktionen erfolgreich implementiert. Es ist möglich 8 Analogeingänge parallel aufzunehmen und es stehen zwei Digitaleingänge zur Verfügung. Der Datenlogger kann über MQTT konfiguriert werden und die Konfiguration wird bei einem Neustart aus dem EEPROM geladen. Die Modbus Schnittstelle ist hardwareseitig implementiert und verfügt auf Softwareseite über die Grundfunktionen zum Auslesen von Sensoren. Um die oben genannten Anforderungen zu erfüllen, muss das Modbus Protokoll vollständig implementiert werden, die Status LEDs durch ein Display ersetzt und die Datenübertragung verschlüsselt werden. Da der RP2040 Chip selbst nicht über ein Versschlüsselungsmodul oder einen true-random-generator verfügt, muss die Verschlüsselung im GSM Modul ausgeführt werden. Hierfür soll das SIM800L Modul durch das neuere SIM7020 Modul ersetzt werde. Zusätzlich nutzt das neue GSM Modul das NB-IOT Netzwerk und hat somit eine deutlich bessere Gebäudedurchdringung und Netzabdeckung als das alte Modul.

Literatur und Abbildungen

[1] Andrew Banks. Oasis MQTT Standard. http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html, 10 2014.

[2] Delta Black. ETM DeltaBlack - 4G Industriedatenlogger & SMS Alarm. https://etmiot.com/de/products/cellular-dataloggers-sms-alarms/mains-powered-dataloggers/etm-deltablack-4g-industrial-data-logger-sms-alarm/, 01 2019.

[3] Eigene Darstellung.

[4] Line Metrics. LineMetrics Box Datenlogger. https://www.linemetrics.com/de/plattform/linemetrics-box-datenlogger/, 08 2020.

[5] Olga Weis. Kommunikationshandbuch für Modbus RTU. https://www.virtual-serial-port.org/de/articles/modbus-rtu-guide/, 10 2019.

Entwicklung eines metrikenbasierten Sampling-Algorithmus für den szenariobasierten XiL-Test von Fahrerassistenzsystemen.

Tom Reichle **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Daimler Truck AG, Untertürkheim

Einleitung

Ein primäres Ziel der Entwicklung von Fahrerassistenzsystemen ist die Erhöhung der Sicherheit aller Verkehrsteilnehmer und Vermeidung von Unfällen. Dabei wird der Fahrer durch die Fahrerassistenzsysteme unterstützt und es können Aufgaben vom Fahrer übernommen werden. Dies steht im Einklang mit dem von der EU ausgerufenen Projekt Vision Zero mit dem Ziel, bis 2050 die Zahl der Verkehrstoten und schweren Verletzungen auf nahe null zu senken. [2]

Motivation

Durch die Entwicklung von immer weiter automatisierten Fahrzeugen nimmt auch der Aufwand, ein System abzusichern, immer weiter zu. Für die statistische Validierung des Systemverhaltens von hoch automatisierten Fahrzeugsystemen sind mehrere Milliarden Testkilometer notwendig. Das ist in der Praxis zeitlich und wirtschaftlich nicht umsetzbar. [1] Deshalb wird versucht, dem gesteigerten Aufwand durch den Einsatz von Simulationen entgegenzutreten. Um die Software und Hardware von Fahrerassistenzsystemen zu testen und zu bewerten, werden X-in-the-Loop Prüfstände eingesetzt. Die Software wird mithilfe von Software-in-the-Loop Prüfständen getestet. Dabei wird das reale Verkehrsgeschehen mithilfe von Umwelt-, Umgebungs- und Fahrzeugmodellen nachgebildet. Der Ansatz des szenariobasierten Testens dient dabei als Grundlage, um Testfälle bzw. Szenarien zu definieren. Die Simulation jedes einzelnen Szenarios bindet enorme Zeit- und Rechenkapazitäten. Für die Absicherung von Fahrerassistenzsystemen sind vor allem die kritischen Szenarien von großer Bedeutung. Ziel ist es daher, einen Algorithmus zu entwickeln, der aus den möglichen Szenarien die kritischen herausfindet und damit den Testaufwand erheblich reduziert. Um die Testressourcen effizient und ohne menschliches Zutun einsetzen zu können, spielt die Testautomatisierung eine wichtige Rolle.

Szenariobasiertes Testen

Beim szenariobasierten Testen wird anhand des Abstraktionsgrades zwischen funktionalen, logischen und konkreten Szenarien unterschieden. Bei funktionalen Szenarien wird sprachlich festgelegt, welche Eigenschaften erfasst werden müssen. Dabei werden den Eigenschaften keine Werte oder Wertebereiche zugewiesen. Logische Szenarien werden aus funktionalen Szenarien abgeleitet. Für die einzelnen Parameter wird ein Wertebereich festgelegt. Die Parameter können innerhalb dieses Wertebereichs variieren. Für die konkreten Szenarien werden logische Szenarien als Basis herangezogen. Jeder Parameter bekommt einen festen Wert aus dem Wertebereich zugewiesen. Die Anzahl der möglichen konkreten Szenarien entspricht aller möglichen Kombinationen der Parametervariationen. [4]

Kritikalitätsmetriken

Für die Bewertung der Kritikalität eines konkreten Szenarios steht eine große Auswahl an verschiedenen Kritikalitätsmetriken zur Verfügung. Im folgenden Abschnitt werden Beispiele geeigneter Metriken aufgeführt. [6]

- TTC (Time To Collision) gibt die minimale Zeit an bis zwei Akteure bei gleichbleibendem Kollisionskurs und gleicher relativer Geschwindigkeit kollidieren.
- BTN (Brake Threat Number) beschreibt das Verhältnis der erforderlichen longitudinalen Beschleunigung des Ego-Fahrzeuges, um eine Kollision zu vermeiden, zu der momentan verfügbaren longitudinalen Beschleunigung des Ego-Fahrzeuges.

- STN (Steer Threat Number) beschreibt das Verhältnis der erforderlichen lateralen Beschleunigung des Ego-Fahrzeuges, um eine Kollision zu vermeiden, zu der momentan verfügbaren lateralen Beschleunigung des Ego-Fahrzeuges.
- TIT (Time Integrated TTC) aggregiert die Differenz zwischen TTC und einem Schwellwert über die Dauer des Szenarios.
- THW (Time Headway) berechnet die Zeit bis Akteur A bei konstanter Geschwindigkeit die Position von Akteur B erreicht.

Machine Learning

Im Folgenden werden zwei Arten von Algorithmen vorgestellt, die für die Optimierung einer metrikenbasierten Kostenfunktion geeignet sind.

Genetischer Algorithmus:
Beim genetischen Algorithmus wird die natürliche Selektion basierend auf dem natürlichen Evolutionsprozess als Grundlage genutzt. Eine Anfangspopulation von Individuen wird in Richtung eines zuvor definierten Optimums entwickelt. Ein Individuum besitzt verschiedene Eigenschaften, die Chromosomen genannt werden. Jedes Chromosom wiederum besitzt verschiedene Gene. Ein Gen entspricht einem Wert. Zwischen den einzelnen Generationen können die Gene modifiziert werden. Das Funktionsprinzip ist wie folgt: Nach der ersten Testdurchführung erfolgt mithilfe einer im Vorfeld definierten Kostenfunktion die Bewertung der Individuen. Geeignete Individuen werden für die nächste Generation ausgewählt. Die Eigenschaften eines Individuums werden durch Kreuzung und Mutation modifiziert. Die Wahrscheinlichkeiten dafür werden im Vorfeld definiert. Diese vier Schritte werden in einer Schleife durchgeführt, bis ein zuvor definiertes Ziel erreicht ist oder eine bestimmte Anzahl an Generationen berechnet wurde. [3]

Reinforcement Learning Algorithmus: Reinforcement Learning Algorithmen basieren darauf, dass ein Agent eine Strategie erlernt, um numerische Belohnungen zu maximieren. Es gibt eine Menge von Zuständen und eine Menge von Aktionen. Der Agent muss durch Ausprobieren selbstständig herausfinden, welche Aktionen die größten Belohnungen bringen. Ein herausfordernder Fall ist, wenn eine Aktion nicht nur Einfluss auf die unmittelbare Belohnung hat, sondern auch Einfluss auf folgende Zustände und auf zukünftigen Belohnungen hat. Dabei ist es wichtig, die richtige Balance zwischen Erforschung und Ausbeutung herauszufinden. [5]

Ausblick

Um ein kritisches Szenario zu erhalten, muss eine geeignete Parameterkombination gefunden werden, die das konkrete Szenario beschreibt. Aus den Kritikalitätsmetriken wird eine Auswahl an Metriken getroffen, die für die Bewertung der einzelnen Fahrerassistenzsysteme geeignet ist. Daraufhin wird aus den gewählten Metriken eine Kostenfunktion erstellt. Im nächsten Schritt wird die Kostenfunktion mithilfe von einem der vorgestellten Machine Learning Algorithmen optimiert. Um zu bewerten, ob ein Szenario bzw. das System-Under-Test erfolgreich ist, werden Pass/Fail Kriterien definiert. Der gesamte Prozessablauf wird automatisiert. Am Ende folgt die Validierung des Sampling-Algorithmus und der Kostenfunktion.

Literatur und Abbildungen

[1] D. Baumann, R. Pfeffer, and E. Sax. Automatic Generation of Critical Test Cases for the Development of Highly Automated Driving Functions. In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE, 2021.

[2] European Commission and Directorate-General for Mobility and Transport. Next steps towards 'Vision Zero' : EU road safety policy framework 2021-2030. https://data.europa.eu/doi/10.2832/391271, 2020.

[3] F. Klück, M. Zimmermann, F. Wotawa, and M. Nica. Genetic Algorithm-Based Test Parameter Optimization for ADAS System Testing. In *2019 IEEE 19th International Conference on Software Quality, Reliability and Security (QRS)*. IEEE, 2019.

[4] T. Menzel, G. Bagschik, and M. Maurer. Scenarios for Development, Test and Validation of Automated Vehicles. In *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018.

[5] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning An Introduction second edition*. The MIT Press, 2018.

[6] Lukas Westhofen et al. Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art. https://arxiv.org/abs/2108.02403, 2021.

Einfluss von Module Federation auf den Arbeitseinsatz von Micro Frontends in Angular

Dennis Rittner **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma adesso SE, Stuttgart

Einleitung

Entwickler, die an Frontend Lösungen arbeiten, entscheiden sich zunehmend für den Ansatz der Entwicklung in Micro Frontends. Bekannte Unternehmen, wie beispielsweise IKEA, DAZN oder Spotify verwenden diese bereits in ihren Onlinediensten [4]. Micro Frontends sollen die Entwicklung großer komplexer Projekte vereinfachen. Viele Teams arbeiten dabei gleichzeitig an kleinen, voneinander getrennten Projekten, die zusammen eine Webanwendung ergeben. So arbeitet beispielsweise ein Team an einem Feature, welches die UI und Logik der Produktsuche enthält. Ein anderes Team entwickelt einen Warenkorb als separates Feature. Die Teams haben dabei die fachliche und technische Ende-zu-Ende Verantwortung für das eigene Projekt [3]. Darum ist es für Entwickler bedeutsam, diese kleinen Projekte richtig verwalten zu können. In Angular besteht die Möglichkeit, diese separaten Projekte zu einer Webanwendung zu vereinen. Ein Ansatz hierfür ist die Verwendung der neuen Technologie Module Federation.

Micro Frontend

Das Konzept der Micro Frontend Architektur basiert auf der Micro Service Architektur aus der Backendentwicklung [3]. Die Firma ThoughtWorks erwähnt den Begriff Micro Frontend zum ersten Mal im November 2016 in ihrem Technologieradar [5]. Ein Micro Frontend ist eine unabhängig, lauffähige Webanwendung. Mehrere Micro Frontends können zu einer zusammenhängenden Webwendung zusammengeschlossen werden. Jedes Micro Frontend verfügt dabei über eigene Funktionalitäten. Somit lassen sich Micro Frontends auch dafür verwenden, eine monolithische Webanwendung in isolierte Anwendungen aufzuteilen [3]. Die Anwendung soll dabei vertikal aufgeteilt werden. Das bedeutet eine Einteilung in die unterschiedlichen Funktionalitäten einer Webanwendung. Jedes Micro Frontend stellt eine solche Funktionalität dar und wird dabei von einem eigenen Team entwickelt und getestet.

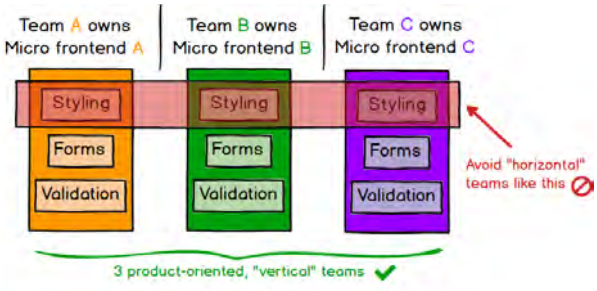

Abb. 1: Beispiel für die vertikale Aufteilung einer Anwendung [6]

Umsetzung

Für die Implementierung von Micro Frontends stehen Entwicklern verschiedene Methoden zur Verfügung. Diese werden in die Kategorien Build-Time und Runtime Integration eingeteilt [6]. Des Weiteren besteht die Möglichkeit Micro Frontends mithilfe von Module Federation in eine Webanwendung einzugliedern. Eine Gemeinsamkeit der verschiedenen Ansätze ist die Verwendung einer sogenannten Containeranwendung [6]. Mithilfe einer Containeranwendung werden die Micro Frontends zusammen auf einer Seite dargestellt und in eine gemeinsame Webanwendung geladen [6].

Build-Time Integration

Für diese Integration, stehen einer Anwendung ein oder mehrere Micro Frontends als Pakete, wie zum Beispiel Node Module zur Verfügung [2]. Häufig wird dafür ein Monorepo verwendet. Der Quellcode aller verwendeten Micro Frontends befindet sich dabei zusammen in einem gemeinsamen Repository. Die einzelnen Pakete werden innerhalb der Containeranwendung als Abhängigkeiten eingebunden [1].

Runtime Integration

In dieser Integration werden Micro Frontends zur Laufzeit in eine Anwendung geladen. Dies kann durch unterschiedlich Ansätze realisiert werden. Eine Möglichkeit ist das Einfügen eines Micro Frontends über ein iFrame HTML Element. Ein weiterer Ansatz ist das Erstellen von Web Components. So kann ein Micro Frontend als Web Component in eine HTML Datei eingebunden und über den Browser ausgegeben werden [1]. Des Weiteren besteht die Möglichkeit Server-Side Composition zu verwenden. Dafür kann beispielsweise die Webserver-Software NGINX verwendet werden. Ein NGINX Server befindet sich zwischen einer Anwendung und den Micro Frontends. Dieser Server kümmert sich dabei um das Routing zu dem jeweiligen Micro Frontend, wie in Abbildung 2 zu sehen ist.

Abb. 2: Routing mit NGINX [1]

Module Federation

Module Federation ist ein Plugin, welches in Version 5 von Webpack erschienen ist [7]. Je nach Bedarf, lassen sich entweder ein komplettes Projekt oder bestimmte Teile eines Projekts, zu einer Gesamtanwendung hinzufügen. Gleichzeitig werden benötigte Abhängigkeiten, die in den Projekten verwendet werden, nicht für jedes Projekt gesondert geladen. Dies bedeutet, dass eine Abhängigkeit, die in mehreren dieser Teilprojekte verwendet wird, einmalig in das Gesamtprojekt geladen wird und im Anschluss von jedem Teilprojekt verwendet werden kann. Module Federation verspricht, einen neuen Weg Micro Frontends zu entwickeln, indem Abhängigkeiten zwischen unterschiedlichen Modulen geteilt werden können [8].

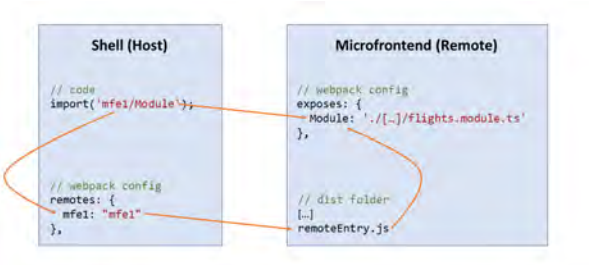

Abb. 3: Beispiel für das Laden von Micro Frontends in Module Federation [8]

Aufgabenstellung und Ausblick

Anhand einer kleinen Beispielanwendung soll untersucht werden, wie Module Federation funktioniert. Die Erkenntnisse, die sich daraus ergeben, sollen aufzeigen, welchen Einfluss Module Federation auf Angular Projekte hat. Darüber hinaus wird die Beispielanwendung mit weiteren Ansätzen der Micro Frontend Integration entwickelt und mit Module Federation verglichen. Durch die Verwendung der verschiedenen Ansätze sollen die jeweiligen Vor- und Nachteile aufgezeigt werden.

Literatur und Abbildungen

[1] Barghav Bachina. 6 Different Ways To Implement Micro-Frontends With Angular. https://medium.com/bb-tutorials-and-thoughts/6-different-ways-to-implement-micro-frontends-with-angular-298bc8d79f6b, 2020.

[2] Rany ElHousieny. Micro Frontends: What, why and how. https://levelup.gitconnected.com/micro-frontends-what-why-and-how-bf61f1f0a729, 2021.

[3] Michael Geers. Micro Frontends - extending the microservice idea to frontend development. https://micro-frontends.org/, 2017.

[4] Entando Inc. 7 Successful Companies Using Micro Frontends. https://www.entando.com/page/en/7_successful_companies_using_micro_frontends_en_1?contentId=BLG2303&modelId=25, 2020.

[5] Thoughtworks Holding Inc. Micro Frontends - Technology Radar. https://www.thoughtworks.com/radar/techniques/micro-frontends, 2016.

[6] Cam Jackson. Micro Frontends. https://martinfowler.com/articles/micro-frontends.html, 2019.

[7] Manfred Steyer. The Microfrontend Revolution: Module Federation in Webpack 5. https://www.angulararchitects.io/aktuelles/the-microfrontend-revolution-module-federation-in-webpack-5/, 2020.

[8] Manfred Steyer. The Microfrontend Revolution: Module Federation with Angular. https://www.angulararchitects.io/aktuelles/the-microfrontend-revolution-part-2-module-federation-with-angular/, 2020.

Konzeptionierung und Implementierung eines Road-Generators für die Unity-Simulation

Felix Rudolf **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Simulationen spielen allgemein eine große Rolle in der Entwicklung und beim Testen. Vor allem im Bereich des autonomen Fahrens sind Simulationen sehr wichtig, da es nur schwer möglich ist, autonome Fahrzeuge große Distanzen zurücklegen zu lassen, ohne vorher eine gewisse Sicherheit zu besitzen, dass dies auch sicher möglich ist. Durch eine realitätsnahe Simulation kann dies sichergestellt werden, außerdem können Änderungen ohne großen Aufwand getestet werden und gefahrlos das Verhalten des Fahrzeugs in gefährlichen Situationen untersucht werden. Gleichzeitig sind Simulationen auch in Form von Computerspielen beliebt, z.B. in Form von Rennspielen und Fahrsimulatoren. Die Werkzeuge und Programme, mit denen Computerspiele erstellt werden, können somit auch für Simulationen genutzt werden. Das Team it:movES der Hochschule Esslingen, welches autonom fahrende Modellfahrzeuge im Rahmen des Carolo Cups entwickelt, verwendet aus diesem Grund zurzeit eine selbstentwickelte Unity-Simulation zum Testen der Software und zum Trainieren neuronaler Netze in der Bildverarbeitung. Das simulierte Fahrzeug kann in mehreren verschiedenen Umgebungen in Form von in sich geschlossenen Rundkursen getestet werden. Diese Rundkurse müssen momentan jeweils von Hand erstellt werden, was natürlich einen nicht unerheblichen zeitlichen Aufwand bedeutet. Aus diesem Grund soll ein „Road-Generator“ entwickelt werden, um automatisch zufällige Strecken generieren zu können, auf denen das Fahrzeug getestet werden kann.

Ziel der Arbeit

Das Ziel dieser Arbeit ist die prozedurale Generierung einer kontinuierlichen validen Strecke innerhalb der Unity-Simulation während der Laufzeit. Die generierte Strecke soll deterministisch auf einem Startwert (einem sogenannten „Seed“) basieren und damit reproduzierbar sein. Des Weiteren muss die Strecke die Regeln und Einschränkungen des Carolo Cups einhalten, auf dessen Spezifikationen die Simulation beruht.

Prozedurale Generierung

Prozedurale Generierung bezeichnet die automatische Erstellung von Inhalten zur Laufzeit. Dabei werden die Inhalte nicht von Hand erstellt, sondern über Algorithmen generiert. Nahezu alle Inhalte z.B. einer Simulation können hierbei generiert werden: Texturen, Musik, Modelle, gesamte virtuelle Welten (z.B. eine befahrbare Welt für einen Fahrsimulator oder ein Rennspiel). Dabei werden deterministische Algorithmen verwendet, um reproduzierbare Ergebnisse zu erhalten. Ein bekanntes Beispiel dafür ist das Spiel „Minecraft“, in welchem die gesamte Spielwelt deterministisch aus einem einzelnen Seed generiert wird. Dadurch müssen Inhalte nicht oder nur teilweise von Hand erstellt werden, was Entwicklungszeit sparen und für größere Spielwelten und Vielfalt sorgen kann. Außerdem kann auch Speicherplatz gespart werden, da anstelle z.B. einer gesamten Spielwelt lediglich der Seed sowie evtl. die grundlegenden Objekte, aus denen die Welt aufgebaut ist, gespeichert werden müssen.

Carolo Cup

Der Carolo Cup ist ein jährlich abgehaltener Wettbewerb zwischen Hochschulen, bei denen Teams der teilnehmenden Hochschulen mit autonomen Modellfahrzeugen im Maßstab 1:10 gegeneinander antreten. Ausgerichtet wird er von der Technischen Universität Braunschweig. Der Cup besitzt zwei Schwierigkeitsstufen: Basic Cup und Master Cup. Beide Cups umfassen jeweils statische und dynamische Disziplinen. Im Basic Cup bestehen die statischen Disziplinen aus einer Präsentation der Konzepte und der technischen Umsetzung der Modellfahrzeuge, welche von einer Jury bewertet werden. Die dynamischen Disziplinen umfassen verschiedene Fahrszenarien, welche möglichst fehlerfrei absolviert werden müssen. Im Basic Cup sind

dies die freie Fahrt ohne Hindernisse mit Längs- und Querparkvorgängen und eine Fahrt mit statischen und beweglichen Hindernissen. [1] Im Master Cup enthalten die statischen Disziplinen zusätzlich eine Präsentation des Projektmanagementprozesses und die dynamischen Disziplinen des Basic Cups sind um zusätzliche Regeln und Elemente wie z.B. Überholverbotszonen erweitert und beinhaltet zusätzlich ein Fahrszenario in einer Vorstadtgegend. [2] Eine mögliche Strecke des Carolo Master Cups ist auf Abb. 1 zu sehen.

Unity

Unity ist eine sogenannte Spiel-Engine, eine kombinierte Laufzeit- und Entwicklungsumgebung für Computerspiele. Entwickelt vom Unternehmen Unity Technologies, dient Unity als Tool zur Erstellung von 2D- oder 3D-Grafikumgebungen und Computerspielen. Bis zu einem jährlichen Umsatz von 100.000$ ist die Nutzung der Engine, auch kommerziell, kostenlos. Unity umfasst neben der Grafikdarstellung unter anderem auch eine Physikengine zur physikalischen Simulation. Zusätzlich zu den Grundfunktionen und -werkzeugen der Engine gibt es den Unity Asset Store. Hier können Entwickler eigene Werkzeuge, Modelle, Frameworks usw. kostenfrei oder -pflichtig anbieten. Diese können einfach als Plug-Ins in Unity verwendet werden. Dadurch gibt es effektiv eine große Auswahl an (jedoch hauptsächlich kostenpflichtigem) zusätzlichem Inhalt und Werkzeugen.

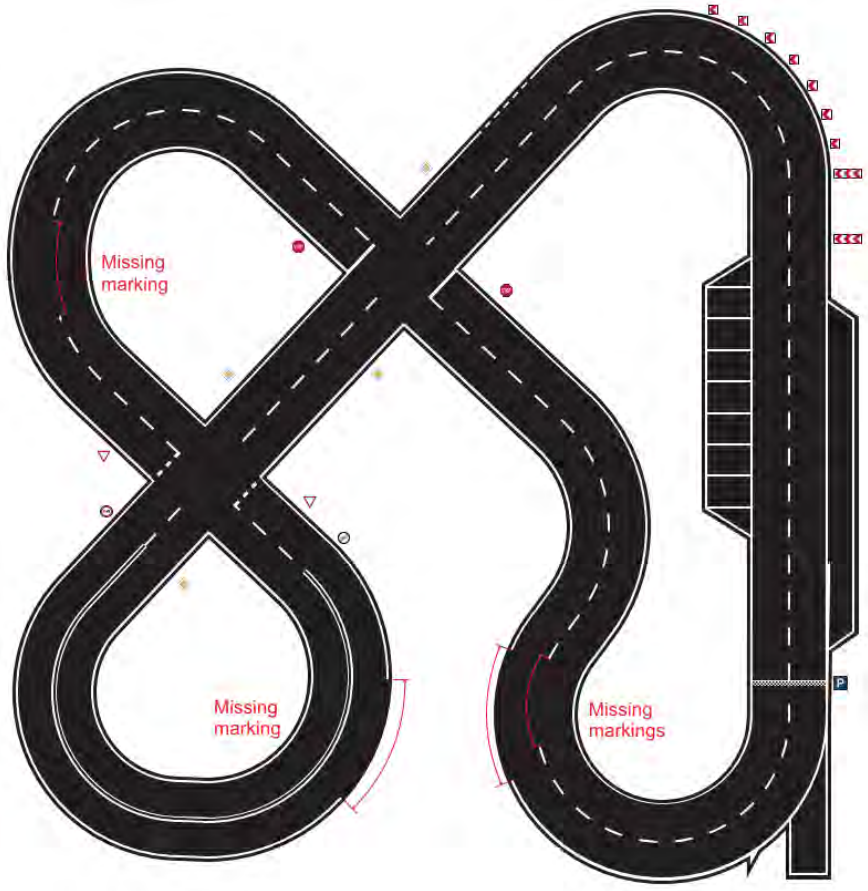

Abb. 1: Beispielstrecke für den Carolo Master Cup [2]

Literatur und Abbildungen

[1] TU Braunschweig. Carolo-Basic-Cup@Home Regulations 2021. https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Institute_Fakultaet_5/Carolo-Cup/Basic-Cup_Regulations_210120.pdf, 01 2021.

[2] TU Braunschweig. Carolo-Master-Cup@Home Regulations 2022. https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Institute_Fakultaet_5/Carolo-Cup/Master-Cup_Regulations_210120.pdf, 12 2021.

Development of a framework for interactive control of IBM Z virtual machines running Linux

Julian Ruess **Rainer Keller**

Department of Computer Science and Engineering, Esslingen University

Work carried out at IBM Germany Research & Development GmbH, Böblingen

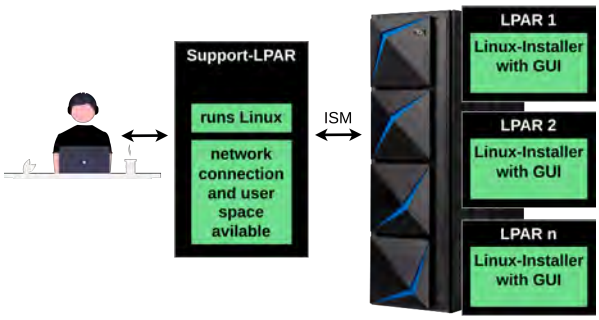

Fig. 1: A framework to interactively control IBM Z virtual machines running Linux should use ISM for I/O communication. [4]

Introduction

IBM Z/LinuxONE servers can be partitioned in multiple logical partitions (LPAR) as seen in figure 1. Each of these partitions is a subset of the machine's hardware resources with its own operating system installed.
When installing Linux on an LPAR partition, the user at the administrator's workstation should be able to see the graphical output of the installation process. Also, the input devices of the user should be forwarded to the operating system installer on the LPAR.
Because there is no established network connection available at boot time of the Linux operating system installer on the LPAR, typical remote access and remote-control applications like Virtual Network Computing (VNC) or X Window Virtual Framebuffer (Xvfb) cannot be used to remote control the installation process directly from the administrator's workstation. However, a newly started Linux installer on IBM Z/LinuxONE servers detects a Peripheral Component Interconnect (PCI) device at boot time. This device can be used to transmit data from LPAR to LPAR over a connection based on shared memory. This connection is called Internal Shared Memory (ISM).

The Support-LPAR on an IBM Z/LinuxONE server already has a readily established network connection and therefore can act as a gateway. Users can connect to it with a commonly used protocol like e.g. VNC. This gateway can be used to show the administrator the graphical output of the Linux installer that runs on a new LPAR. Also, the user's input devices can be forwarded over this gateway to the Linux installer. Data between the gateway and the Linux installer must be transmitted over ISM.
This thesis analyzes different open-source building blocks and novel approaches to realize the behavior described above. For graphics output, virtual framebuffer drivers are analyzed. Also, a QEMU/KVM virtio solution is described that can be used to forward existing virtio devices in a unified way over ISM.

The Linux graphics stack

To understand how the graphics output of a new Linux installer can be forwarded over the ISM connection, the Linux graphics stack must be analyzed.
Compositors like X.Org or Wayland composite multiple windows and write the actual screen image in memory. In the modern Linux graphics stack, these compositors run in user space and communicate with device drivers in kernel space to get the pixels on the screen.
To be able to forward the early graphics output of a Linux installer over ISM, it is preferable to have a solution that is independent of user space applications like the compositor.
To address this, a Linux graphics driver that acts as a virtual framebuffer device can be used. A framebuffer is a piece of memory that represents the bitmap that is going to be displayed on the screen [2]. After loading this driver, Linux acts as if a real GPU is available, albeit graphics output being written to memory, only. ISM can be used to forward this data to another LPAR.

The QEMU/KVM way

QEMU is a virtualization software for Linux that is able to emulate a multitude of hardware devices. This includes mouse, keyboard and graphics. These devices can be part of the desired solution. By adding KVM to QEMU, virtualization technologies like Intel VT or AMD-V can be used to speed up virtualization performance. QEMU/KVM can be classified as a type-2 hypervisor [1].
QEMU/KVM in combination with the Virtual Machine Manager (VMM) can be used as a full-featured virtualization solution like VirtualBox or VMWare Workstation. Users can create their own virtual machines with a GUI. The output of the guest system is displayed in a window on the screen. Also, the input devices (mouse, keyboard) of the host are forwarded to the guest system.
To realize this, QEMU can emulate paravirtualized and non-paravirtualized I/O devices. Paravirtualization means that the guest system knows that it runs in a virtual environment. Because of this, drivers can be optimized to achieve better performance. The framework of paravirtual I/O devices in QEMU/KVM is called virtio. Virtio devices are implemented in QEMU and the corresponding guest drivers are shipped with the Linux Kernel. Benchmarks of the virtio-net network device show that this approach delivers a up to 22x higher throughput than a conventional virtualized e1000 network adapter [1].
To transfer data between guest and host, virtio devices use a standardized mechanism. With a virtualized bus that supports interrupts, the other side can be notified, that new data is ready to be consumed. After that, the actual data is transmitted over a ring buffer in shared memory which is called virtqueue [3]. This virtqueue is used by the guest system in kernel space (virtio device driver) and the host system in user space (QEMU).
If it is possible to port this mechanism from using shared memory to use the ISM connection, a scenario can be, that the Support-LPAR runs QEMU and emulates all required devices. The Linux installer on the LPAR then can use the existing virtio drivers to communicate with QEMU over ISM.
By following this approach, all existing and future virtio devices can be used. No modification is required per device that should be forwarded to the Linux installer on the LPAR. This feasibility study is part of this thesis.

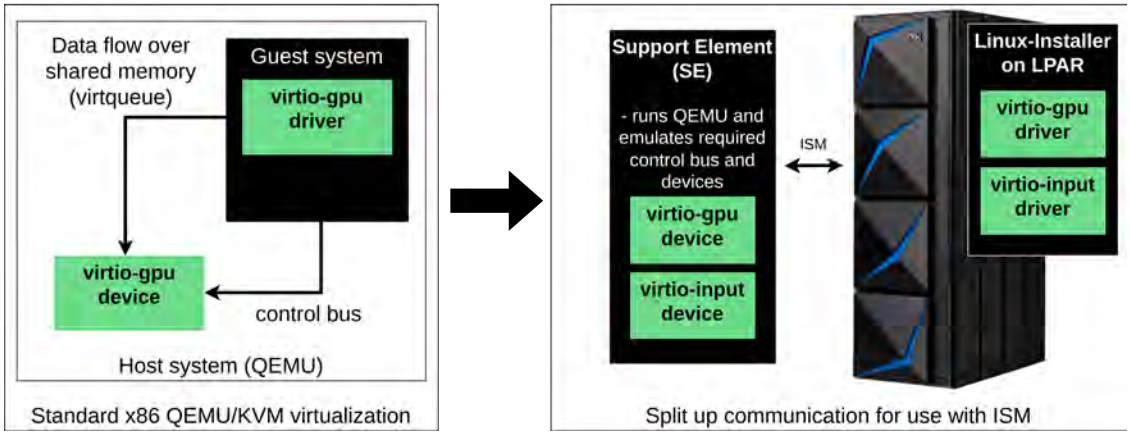

Fig. 2: Standard virtio device communication (left) needs to be split up for use with ISM. [4]

References and figures

[1] Edouard Bugnion, Jason Nieh, and Dan Tsafrir. *Hardware and Software Support for Virtualization*. Morgan & Claypool Publishers, 2017.

[2] Nicolas Caramelli. Back to the Linux Framebuffer! https://archive.fosdem.org/2020/schedule/event/fbdev/attachments/slides/3595/export/events/attachments/fbdev/slides/3595/fosdem_2020_nicolas_caramelli_linux_framebuffer.pdf, 2020.

[3] Eugenio Perez Martin. Virtqueues and virtio ring: How the data travels. https://www.redhat.com/en/blog/virtqueues-and-virtio-ring-how-data-travels, 2020.

[4] Own representation.

Neue Trends im Master Data Management

Belal Sarwar **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Daten sind der ausschlaggebende Unterschied, ob ein Geschäftsablauf einfach und schnell oder kompliziert und umständlich vollzogen werden kann. In der heutigen globalisierten Wirtschaft ist es für ein Unternehmen notwendig, eine verlässliche Quelle zu haben, die einem Informationen schafft, verwaltet und bereitstellt. Diese Notwendigkeit treibt sie zur digitalen Transformation an. Die Bereitstellung von einzigartigen Produkten oder Kundenerlebnissen wird dadurch unterstützt, dass die Unternehmen Daten verwenden, die mit einer klaren Strategie zur Verfügung stehen und nutzbar sind [4].

Was ist Master Data Management ?

Master Data Management ist eine Methode, die einem Unternehmen ermöglicht, alle geschäftskritischen Daten mit einem gemeinsamen Referenzpunkt zu verknüpfen. Wenn das richtig erfolgt, verbessert Master Data Management sowohl die Datenqualität als auch den Datenaustausch zwischen den Mitarbeitern und verschiedenen Abteilungen. Darüber kann mithilfe von Master Data Management die Datenverarbeitung in verschiedenen Systemarchitekturen, Plattformen und Anwendungen erleichtert werden.
Durch eine größer werdende Anzahl und Vielfalt an Organisationabteilungen, Mitarbeiterrollen und EDV-Anwendungen steigen die Vorteile des Master Data Management. Deshalb ist der Einsatz von Stammdatenmanagement für große und komplexe Unternehmen vorteilhafter, als für kleine und mittlere Unternehmen. Die Implementierung des Stammdatenmanagements stellt Herausforderungen dar, wenn Unternehmen fusionieren, da verschiedene Geschäftsbereiche die Bedeutung von Begriffen und Einheiten in ihrem eigenen Kontext sehen. Bei einer Fusion hat das Master Data Management jedoch Vorteile, da es helfen kann, Chaos zu vermeiden und die Effizienz der neuen, größeren Organisation zu optimieren.
Es ist essentiell, dass das Personal und die Abteilungen wissen, wie Daten beschrieben, formatiert, gespeichert und abgerufen werden können. Nur dann kann man die Vorteile des Master Data Management in Erfahrung bringen. Die Stammdatensätze sollten ebenfalls regelmäßig aktualisiert werden [3].

Bedürfnis nach verlässlichen Stammdaten

Master Data Management umfasst besonders die Themen der Datenintegrität und Datenqualität. Es soll gewährleisten, dass Daten, die das Unternehmen hat, wiederverwendet werden können.
Daten liegen in unterschiedlichsten operativen und und analytischen Systemen vor und sind Steuerungsrundlage von Geschäftsprozessen sowie die Basis von Entscheidungen, die das Unternehmen trifft. Ohne Nachhaltigkeit im Umgang mit Stammdaten ist der wirtschaftliche Vorteil von IT-Unterstützung begrenzt. Ist ein Unternehmen beispielsweise nicht mehr in der Lage, eine konsistente und konsolidierte Sicht auf die Geschäftsobjekte zu liefern, ist es essentiell, einen Ansatz für unternehmensweites Master Data Management zu implementieren.
Mit zunehmenden Wettbewerbsdruck müssen die Geschäftsmodelle und die unterstützenden Geschäftsprozesse in kleiner werdenden Zyklen angepasst werden. Durch die Globalisierung und digitalen Vernetzung der Unternehmen wird die Interaktion mit externen Geschäftspartnern immer komplexer. Deshalb ist der verlässliche Informationsaustausch mit qualitativ hochwertigen Daten von großer Bedeutung, um eine Steigerung der Effizienz in den Prozessen zu erlangen. Die Qualität und Verlässlichkeit von Daten sind besonders wichtig, wenn es darum geht, eine Verbesserung zu erlangen. Denn eine Entscheidung des Unternehmens kann nur dann gezielt und erfolgreich umgesetzt werden, wenn die hierfür als Basis dienenden Daten die folgenden Merkmale aufweisen:

- Aktualität
- Bedeutung
- inhaltliche Korrektheit

[5]

Neue Trends im Master Data Management

Noch ist es so, dass beispielsweise Produktinformationen von den Unternehmen einzeln aufbereitet werden müssen, wodurch volks- und betriebswirtschaftlich enorme Ressourcen verschleudert werden. In Zukunft soll effizienter vorgegangen werden. Es wird wenige Daten-Silos geben und anstatt dessen sollen gemeinschaftliche Stammdatenpools aufgebaut werden, die den jeweiligen Unternehmen die passgenauen Informationen zur Verfügung stellen.

Raum- und zeitunabhängige Cloudlösung

In einer Lieferkette braucht man Datenaufbereitungsanwendungen, die schnell und präzise sind. Das gilt für alle Unternehmen unabhängig von deren Größe. Ein Beispiel hierfür ist der IFCC.DataManager. Dabei handelt es sich um eine cloudbasierte Master Data Management Plattform, die unabhängig von Raum und Zeit bedient werden kann. In dieser Anwendung stehen alle Funktionen zur Verfügung, die für ein modernes Master Data Management notwendig sind. Besonderheiten von solchen flexiblen Plattformen:

- Dateneigner entscheidet selbst, welche Daten freigegeben werden sollen. Dem Dateneigner ist selbst überlassen, welche Funktionen für ihn wichtig sind
- Nutzer haben kein Lock-in-Effekt. SIe können die Daten mit anderen Unternehmen tauschen, in nach- oder vorgelagerte Systeme ausleiten

[2]

Ausblick

Im Rahmen der Arbeit werden die neuen Trends im Master Data Management ausgearbeitet. Es soll aufgezeigt werden, welche Vorteile sich für ein Unternehmen ergeben, wenn sie Master Data Management implementieren. Wenn einem Unternehmen qualitativ hochwertig aufbereitete Daten zur Verfügung stehen, kann es auf dieser Grundlage Entscheidungen treffen und ein Wettbewerbsvorteil für sich erschließen.

Abb. 1: Stammdaten [1]

Literatur und Abbildungen

[1] Christiane Klingenberg and Guido Göbel. Alles schon geregelt? Wie Sie mit Data Governance Ihrem MDM Projekt den richtigen Drive geben können. In *Alles schon geregelt? Wie Sie mit Data Governance Ihrem MDM Projekt den richtigen Drive geben können*. .msg, 2018.

[2] Ulrich Prof Dr Manz. TRENDS FÜR DAS STAMMDATENMANAGEMENT. https://allaboutsourcing.de/de/trends-fuer-das-stammdatenmanagement/, 2019.

[3] ComputerWeekly.de Redaktion. Stammdatenmanagement (Master Data Management, MDM). https://www.computerweekly.com/de/definition/Stammdatenmanagement-Master-Data-Management-MDM, 2010.

[4] Martin Roxlau. WAS IST MASTER DATA MANAGEMENT (MDM) UND WIE KANN ES MEIN BUSINESS NACH VORNE BRINGEN? https://www.sqli.de/blog/was-ist-master-data-management, 2020.

[5] Rolf Scheuch, Tom Gansor, and Colette Ziller. *Master Data Management Strategie, Organisation, Architektur*. dpunkt.verlag GmbH, 2012.

Vergleich von Binäranalysewerkzeugen zur Erkennung von Schwachstellen in der Continuous Integration

Stefan Schanz **Tobias Heer**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ads-tec Engineering GmbH, Nürtingen

Motivation und Problemstellung

Viele Unternehmen, die Software entwickeln oder mit dieser Arbeiten, verwenden mittlerweile Third Party Software. Das offeriert viele Vorteile, wie eine höhere Effektivität und einen schnelleren Entwicklungsprozess, der heute oft für Unternehmen, die Software entwickeln, vorausgesetzt wird. Dabei werden bestehende Softwarepakete verwendet, um neue Software zu bauen. Beim Bauen von Software spielt der Sicherheitsaspekt in den verwendeten Third Party Komponenten eine wichtige Rolle, dem häufig zu wenig Beachtung geschenkt wird. Das kann dazu führen, dass in populären Third Party Komponenten Sicherheitslücken aufgedeckt werden. Jedes Gerät, welches mit eben dieser Software betrieben wird, ist damit plötzlich verwundbar. Diese Schwachstellen werden publik dokumentiert und können mit öffentlich verfügbaren Exploits ausgenutzt werden.
Die Arbeit versucht potentiell verfügbare Schwachstellen auf generischem Wege zu reduzieren. Hierzu sollen Binäranalysewerkzeuge für die Untersuchung von Third Party Software auf Binärcodeebene angewendet werden. Dies bietet den Entscheidenden Vorteil, dass die Notwendigkeit des Source Codes nicht gegeben ist und jede verwendete Software einer Analyse unterzogen werden kann. Es wird tatsächlich der Binärcode untersucht, der folgend direkt auf der Maschine ausgeführt wird. Somit ist die Berücksichtigung anderer Fehlerquellen, wie beispielsweise nicht verwendete sicherheitsrelevanten Compiler Flags oder Patches, die beim Kompilieren auf den Source Code angewandt werden, möglich.
Verschiedene Binäranalysewerkzeuge sollen untersucht werden und in mehreren Szenarien evaluiert werden. Durch die Integration der Binäranalyse in die Continuous Integration ist die automatisierte Erkennung und das frühzeitige Beheben vorhandener Sicherheitslücken in Binärdateien möglich.
Die Arbeit wird in folgende Teilprobleme aufgeteilt, die gelöst werden sollen:

- Sammlung verschiedener Binäranalysetechniken: Für das spätere Verständnis der Werkzeuge und dessen Evaluation sollen verschiedene grundlegende Ansätze der Binäranalyse untersucht werden.
- Definition der betrachtenden Schwachstellen: Da nicht alle Schwachstellen erfasst werden können, müssen diese in einem klaren Rahmen definiert werden.
- Definition Evaluation der Binäranalysewerkzeuge: Es muss untersucht und folgend definiert werden, wie die Binäranalysewerkzeuge evaluiert werden. Damit soll die Basis einer aussagekräftigen Evaluation geschaffen werden.
- Entwicklung eines Werkzeugs für die automatisierte Evaluation: Für eine aussagekräftige Evaluation soll im Rahmen der Arbeit ein Werkzeug entwickelt werden, das sich um den automatisierten Vergleich der Werkzeuge kümmert.
- Evaluation der Binäranalysewerkzeuge in verschiedenen Szenarien: Anhand des definierten Evaluationsvorgangs sollen die Binäranalysewerkzeuge in verschiedenen Szenarien verglichen werden.

Design

Im Folgenden werden alle in dem Kapitel 'Motivation und Problemstellung' gelisteten Teilprobleme näher erläutert.
Sammlung verschiedener Binäranalysetechniken: Es gibt viele verschiedene Binäranalysetechniken und Ansätze, dessen Verständnis geschaffen werden soll. Die dabei entstehenden Kenntnisse bieten die Grundlage zum Verständnis der angewandten Binäranalysewerkzeuge und die Basis zur Begründung der Evaluation. Der Bereich der Binäranalyse ist groß und es sind viele verschiedenen Ansätze und Optimierungen von diesen vorhanden. Die Herausforderung besteht

nun darin zu entscheiden, wie tief in die Thematik eingestiegen werden muss, um ein Grundverständnis dieser zu erlangen.

Definition der betrachtenden Schwachstellen: Zur Ermöglichung einer aussagekräftigen Evaluierung, gilt es Schwachstellen herauszusuchen, die zum späteren Zeitpunkt der Arbeit von den Binäranalysewerkzeugen untersucht werden können. Hierbei wird die 'Juliet Test Suite v1.3 for C/C++' herangezogen, die verschiedene Arten an Schwachstellen enthält. Basierend darauf soll der Großteil der Evaluierung stattfinden. Es gibt viele verschiedene Arten von Schwachstellen in Binärdateien, von denen leider nur ein Bruchteil betrachtet werden kann.

Definition der Evaluation der Binäranalysewerkzeuge: Die Evaluation der Binäranalysewerkzeuge soll definiert werden. Binäranalysewerkzeuge können in verschiedenen Kategorien verglichen werden, wie beispielsweise die Laufzeit oder der Informationsgehalt des Werkzeugs. Dies sind nur zwei von vielen Kategorien, die eine aussagekräftige Evaluation des jeweiligen Werkzeugs ermöglichen soll.

Entwicklung eines Werkzeugs für die automatisierte Evaluation: Während der Ausarbeitung der Arbeit kam das Problem auf, dass in dem für den Vergleich verwendeten Datensatz viele Binärdateien zum Einsatz kommen. Daher ist ein manueller Vergleich nicht in absehbarer Zeit möglich. Aus diesem Grund wurde die Entwicklung eines Werkzeugs umgesetzt, das den Vorgang des Vergleichs von der Ausführung bis zur Generierung von Tabellen und Diagrammen automatisiert.

Für eine adäquate Erklärung des Designs des Werkzeugs findet sich in Abb. 1 eine Übersicht über das Gesamtsystem.

Der Workflow beginnt beim 'Runner' (siehe Abb. 1), der sich um die Ausführung der verschiedenen Binäranalysewerkzeuge auf Binärdateien kümmert. Dieser erhält die zu analysierenden Binärdateien und verschiedene Docker Images. Jedes dieser Docker Images beinhaltet ein präpariertes Binäranalysewerkzeug, das in einem daraus erzeugten Container direkt auf die Binärdateien ausgeführt werden kann. Die Ergebnisse des 'Runner' werden zur Weiterverarbeitung automatisch an den 'Parser' weitergegeben.

Der 'Parser' verwendet die Rückgabewerte des 'Runner', wie in Abb. 1 zu sehen ist. Da die Rückgabewerte der verschiedenen Binäranalysewerkzeuge stark variieren, kümmert sich der 'Parser' um die Konvertierung dieser in generische Fehlerstrukturen. Diese können im weiteren Verlauf als Basis verwendet werden.

Hierfür kommen verschiedene Plugins zum Einsatz, die Implementierungen für die Verarbeitung der jeweiligen Rückgabewerte enthalten. Wie in Abb. 1 ersichtlich, gibt es Plugins für die Binäranalysewerkzeuge 'Dr. Memory', 'cwe_checker', 'Valgrind Memcheck' und 'BAP', welche die unterschiedlichen Rückgaben dieser verarbeiten können. Zudem ist das 'general' Plugin vorhanden, das Implementierungen enthält, die von mehreren Plugins verwendet werden. Die Resultate des 'Parser' werden folgend an den 'Evaluator' übergeben.

Der 'Evaluator' (siehe Abb. 1) nutzt die Daten in Form der generischen Fehlerstrukturen, die vom 'Parser' bereitgestellt werden. Die Aufgabe des 'Evaluator' ist die Extraktion mehrerer messbarer Metriken der Binäranalysewerkzeuge auf Basis der ermittelten Daten. Hierbei handelt es sich um Plugins, die jeweils eine spezifische Messung auswerten. Diese Messergebnisse werden wiederum an den 'Generator' übergeben.

Da verschiedene Informationen ermittelt werden sollen, kümmert sich das Modul 'Generator' um die Generierung von Diagrammen und Tabellen. Aus diesen können schlussendlich die notwendigen Informationen der jeweiligen Binäranalysewerkzeuge abgelesen werden.

Evaluation der Binäranalysewerkzeuge in verschiedenen Szenarien: Das erste Szenario fokussiert sich hierbei auf die Ermittlung der CWE (Common Weakness Enumeration), die von den jeweiligen Binäranalysewerkzeugen erkannt werden können. Dazu werden Binärdateien aller im 'Juliet Test Suite v1.3 for C/C++' vorhandenen CWE der Gruppe 'Simple' verwendet.

Das zweite Szenario beschreibt die Analyse von Binärdateien auf Basis der Komplexität der Binärdatei. Hierzu werden die Ergebnisse der erkannbaren Schwachstellen aus dem vorherigen Szenario verwendet und zugehörige Binärdateien aller Gruppen ('Simple', 'Kontrollfluss', 'Datenfluss') herausgesucht.

Das dritte Szenario handelt von der Untersuchung von verschiedenen Prozessorarchitekturen. Für diese wird ein simples C Programm erstellt und in verschiedene Prozessorarchitekturen cross-kompiliert.

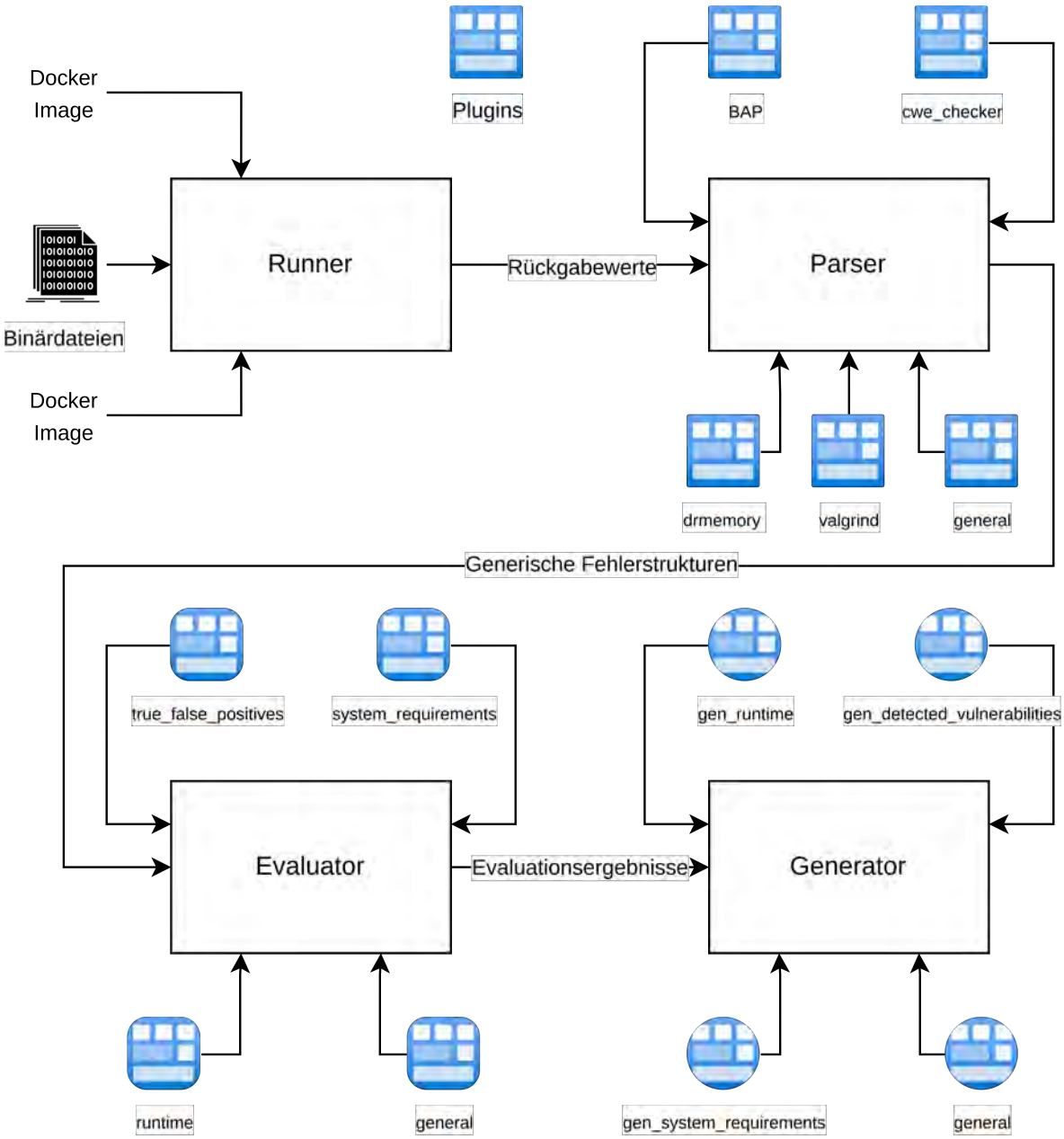

Abb. 1: Evaluation Tool: Übersicht [2]

Evaluation

Da die Arbeit zum Großteil eine Evaluationsarbeit darstellt, werden in diesem Kapitel die wichtigsten Ergebnisse der Arbeit vorgestellt.
Für den Vergleich der Laufzeit werden die Messungen des Testsystems im Bezug auf die Komplexität der Binärdateien vorgestellt. Das hier vorhandene Diagramm stammt vom 'Evaluation Tool' und basiert auf dem zweiten Szenario (alle Gruppen).
Innerhalb der folgenden Abb. 2 lassen sich verschiedene Eigenschaften feststellen. Die zwei Werkzeuge 'Dr. Memory' (blau) und 'Valgrind Memcheck' (orange) sind in der Laufzeit größtenteils niedrig. Aus diesem Grund wird die Y-Achse, die die Zeit in Sekunden zeigt, in der logarithmischen Darstellung abgebildet. In der X-Achse werden die Binärdateien gezeigt und anhand der Komplexität geordnet. Zusätzlich werden aufgrund der Übersichtlichkeit und Aussagekraft die Ausreißer, die durch das Hängenbleiben der Binärdateien entstehen nicht im Diagramm berücksichtigt.
Innerhalb der Abb. 2 lässt sich erkennen, dass sich 'Dr. Memory' (blau) und 'Valgrind Memcheck' (orange)

ähnlich verhalten. Jedoch variiert das Werkzeug 'Dr. Memory' (blau) bei der Ausführungszeit auf die Binärdateien stärker. Es lässt sich ebenfalls erkennen, dass 'Dr. Memory' (blau) bei einem niedrigerem Wert, als 'Valgrind Memcheck' (orange) startet. Das Werkzeug 'cwe_checker' (grün) befindet sich im Bereich der größeren Ausführungszeit und variiert verhältnismäßig wenig. Im Bezug auf die Komplexität der Binärdateien lassen sich in der Abb. 2 wenige Zusammenhänge identifizieren.

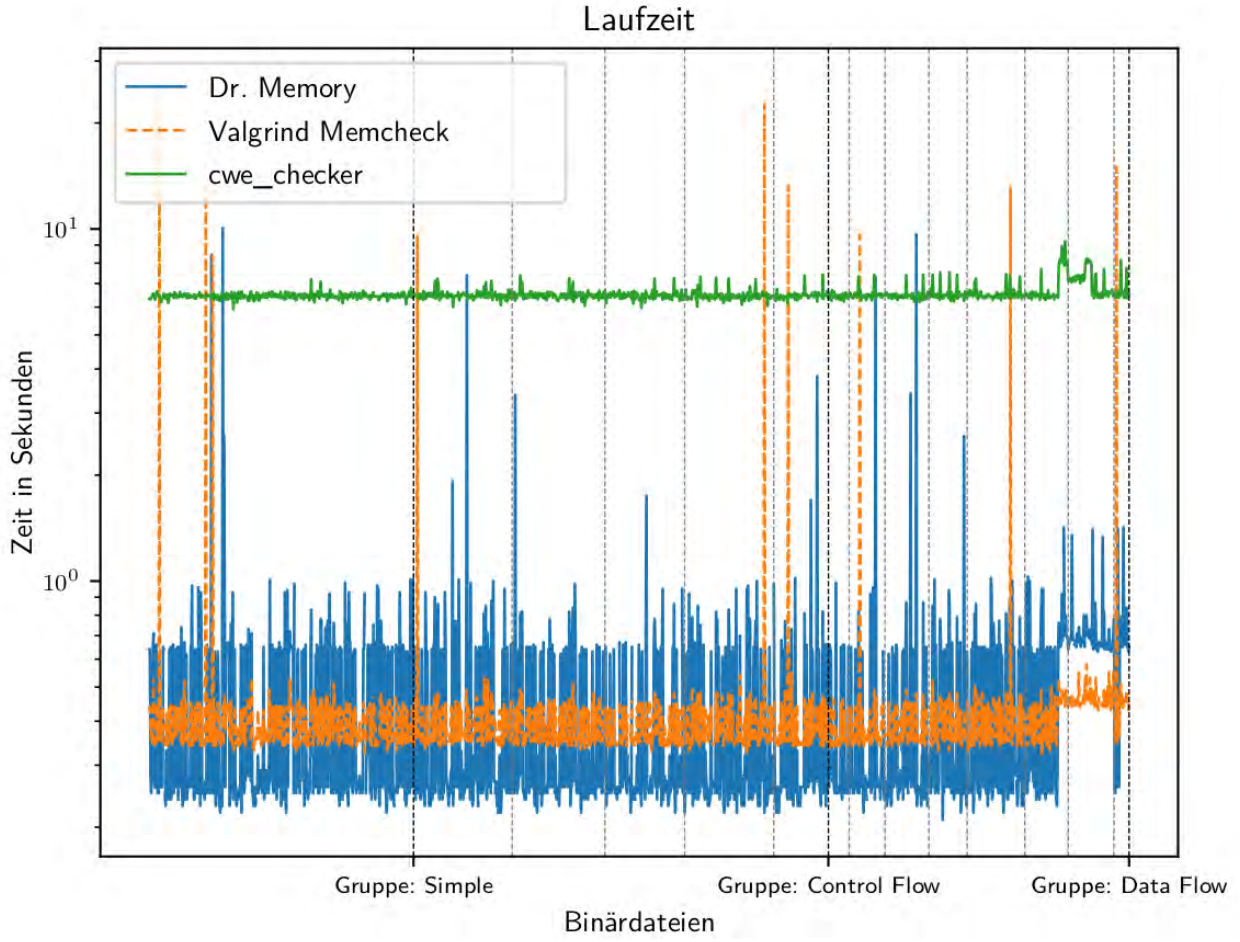

Abb. 2: Laufzeit: Übersicht [2]

Damit die Resultate für den Bereich der erkannten Arten von Schwachstellen und Genauigkeit generiert werden können, werden die Binäranalysewerkzeuge mit Hilfe des 'Evaluation Tool' auf den Datensatz mit den Binärdateien der Gruppe 'Simple' ausgeführt.
Innerhalb der Abb. 3 lässt sich für das jeweilige Werkzeug eine Zusammenfassung von CWE erkennen, die aufgrund der ermittelten Daten eine hohe Erkennungsgenauigkeit aufweisen. Hierbei wird die sogenannte 'likelihood of exploit' zugeordnet, die die Ausnutzbarkeit einer Schwachstelle beschreibt.

Dr. Memory	CWE195 (ND)(High), CWE401 (Medium), CWE415 (High), CWE476 (Medium), CWE590 (ND)(Low), CWE680 (ND)(High), CWE762 (Low)
Valgrind Memcheck	CWE122 (High), CWE124 (Medium), CWE126 (ND)(ND)(High), CWE127 (ND)(ND)(High), CWE195 (ND)(High), CWE415 (High), CWE416 (High), CWE457 (High), CWE476 (Medium), CWE590 (ND)(Low), CWE675 (ND)(ND)(ND), CWE758 (ND)(ND), CWE762 (Low)
cwe_checker	CWE23 (ND)(High), CWE36 (ND)(High), CWE78 (High), CWE427 (ND)(ND)(ND), CWE476 (Medium), CWE690 (ND)(Medium)

Abb. 3: Daten: Erkannte Arten von Schwachstellen pro Werkzeug [2]

Aus Platzgründen wird für die restlichen Resultate der Evaluation auf die Bachelorarbeit verwiesen.

Verwandte Arbeiten

Meinem Kenntnisstand entsprechend gibt es bisher keine Arbeit, die den Vergleich verschiedener bestehender Binäranalysewerkzeuge thematisiert. Zudem ist die Verwendung eines Binäranalysewerkzeugs in der Continuous Integration ebenfalls ein Thema, das bisher nach bestem Wissen kaum wissenschaftlich untersucht wurde. Daher werden folgend Vergleichsweisen aus bestehenden Arbeiten beschrieben.
Das Framework angr [3] vergleicht in dessen Arbeit die Effektivität von verschiedenen Binäranalysetechniken und Evaluiert zusätzlich die verschiedenen Schwachstellenerkennungsmethoden der Binäranalyse auf Basis der CGC Binärdateien.
Eine andere Arbeit stellt das Werkzeug VYPER [1] vor, welches verschiedene Schwachstellen erkennt. Dieses Werkzeug evaluiert die Genauigkeit anhand verschiedener Szenarien. Zum einen werden selbst geschriebene Testfälle verwendet. Zum anderen wird eine Testbasis namens 'Juliet' verwendet, die insgesamt 118 verschiedene CWEs enthält. Außerdem wird das Programm anhand realen Applikationen getestet.

Ergebnis

Alles in allem lassen sich über die Binäranalyse und dem Einsatz in der Continuous Integration folgende Aussagen treffen. Der Einsatz der Binäranalysewerkzeuge innerhalb der Continuous Integration hat sowohl Vor- als auch Nachteile.
Zum Verfassungszeitpunkt ist die Binäranalyse zwar gut, kann aber weiter optimiert werden. Aufgrund der automatischen Umgebung der Continuous Integration müssen die verschiedenen Werkzeuge Resultate mit einer guten Genauigkeit liefern. Die Ergebnisse zeigen, dass nur einige CWE zuverlässig erkannt werden können. Daher können im Rahmen der betrachteten Binäranalysewerkzeuge und der Continuous Integration nur die CWE mit einer guten Genauigkeit berücksichtigt werden.
Eine Voraussetzung der Anwendung der Binäranalyse innerhalb der Continuous Integration ist die Schnelligkeit und die damit verbundene zeitnahe Rückmeldung der Analyseergebnisse. Dies ist, wie in der Evaluation gezeigt, aufgrund von komplexen und zeitaufwändigen Operationen der Binäranalyse für manche Werkzeuge eine Herausforderung.
Abschließend lässt sich folgendes über die Verwendung von Binäranalysewerkzeugen zur Erkennung von Schwachstellen in der Continuous Integration sagen. Die betrachteten Werkzeuge 'Dr. Memory', 'Valgrind Memcheck' und 'cwe_checker' können alle aufgrund der untersuchten Eigenschaften für die Verwendung innerhalb der Continuous Integration in Betracht gezogen werden.

Literatur und Abbildungen

[1] El Habib Boudjema et al. VYPER: Vulnerability detection in binary code. *Security and Privacy*, 2020.

[2] Eigene Darstellung.

[3] Yan Shoshitaishvili, Ruoyu Wang, Christopher Salls, Nick Stephens, et al. (State of) The Art of War: Offensive Techniques in Binary Analysis. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 138–157. IEEE, 2016.

Evaluation einer elastischen Control-Plane für das Edge-Cloud-Continuum

Leon Schmidt **Jörg Nitzsche**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Vector Informatik GmbH, Stuttgart

Motivation

Als zunehmend praktizierter Trend in der IT zeigt sich die Vernetzung verschiedenster Systeme mit der Cloud. Zu beobachten ist dies auch in der Automobilbranche: Bisher waren Fahrzeuge in sich geschlossene Systeme. Dieser Trend führt zur Veränderung der Architektur im Fahrzeug. Ein Teil der dadurch gewonnen neuen Anwendungsmöglichkeiten sind bereits heute sichtbar:

- In der Stadt kann die zu fahrende Geschwindigkeit ermittelt werden, beispielsweise um unnötiges Bremsen durch rote Ampeln zu verhindern. [1]
- Das Planen von Ladestopps entlang der Route wird durch Zusatzinformationen, wie die aktuelle Auslastung der Ladesäulen oder auf der Route befindliche Verzögerungen erleichtert und effizienter. [3]

Ein weiterer Trend ist die Entwicklung von hoch dynamischen Cloud-Umgebungen. Große Technologieunternehmen profitieren von der dynamischen Verfügbarkeit von Rechenressourcen. Diese ermöglicht es, die Rechenressourcen abhängig von der aktuellen Auslastung zu skalieren. Diese Skalierung gewährleistet eine bestmögliche Nutzererfahrung , ohne vorsorglich diese Ressourcen überdimensional gestalten zu müssen. Vor allem datenintensive Dienste wie beispielsweise Netflix profitieren von lokalen Rechenressourcen im Nahbereich.
Die Kombination der beiden Trends ermöglicht neue Anwendungsfälle, wie beispielsweise die Auslagerung von Fahrzeugfunktionen in die Cloud oder die Nutzung des Fahrzeugs als Rechenressouce.
Die Umsetzung dieser Anwendungsfälle erfordert die Weiterentwicklung einiger Technologien und Herangehensweisen: Im Fahrzeug müssen z.B. Microcontroller dahingehend angepasst werden, dass sie fremde Software ausführen können.
Außerdem muss auch die Infrastruktur zum Management der Knoten und der sich darauf befindlichen Applikationen und Dienste (Control-Plane) weiterentwickelt werde um den neu gestellten Anforderungen gerecht zu werden. Bereits existierende Control-Planes sind für den Betrieb im Datencenter ausgelegt. Das bedeutet, dass sie Annahmen treffen:

- Rechenknoten haben eine gute Netzanbindung
- Rechenknoten sind gesammelt in Datencentern
- Einzelne Knoten haben hohe Rechenleistung

Die Verwendung von Fahrzeugen als Rechenknoten führt zu zusätzlicher Komplexität, auf die existierende Control-Planes nicht ausgelegt sind. Eine Fahrzeugflotte besteht aus wesentlich mehr Knoten als im "klassischen" Rechenzentrum, diese haben jedoch eine vergleichsweise gering Rechenkapazität. Zudem kann eine stabile Netzwerkverbindung nicht immer gewährleistet werden. Einzelne Knoten können sehr verstreut sein und einzelne Knoten haben eine geringere Verfügbarkeit, da die Verfügbarkeit durch die Verwendung des Fahrzeugs maßgeblich eingeschränkt sein kann.

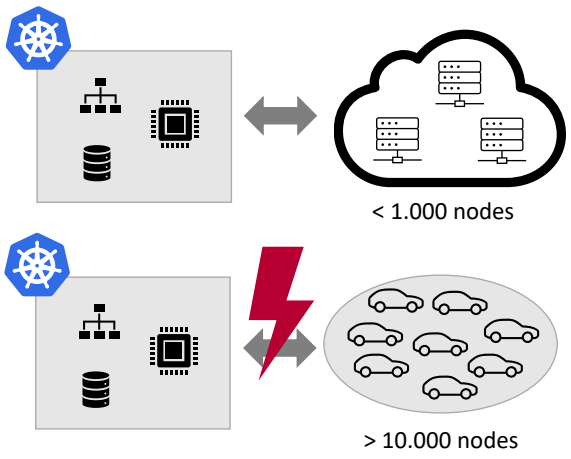

Abb. 1: Die Skallierung einer Control-Plane erfordert neue Technologien und Herangehensweisen [2]

Ziel der Arbeit

Ziel der Arbeit ist die Evaluation verschiedener Control-Plane-Architekturen. Dabei soll erreicht werden, eine hohe Anzahl von Fahrzeug-Knoten (>10000) zu unterstützen. Unter anderem sollen gängige Verfahren zur Kontrolle der Verfügbarkeit eingesetzt werden. Es soll die maximale Kapazität der Control-Plane ermittelt werden und mögliche Performance-Schwachstellen sollen identifiziert werden. Dabei sollen mögliche Handlungsoptionen erarbeitet werden und diese gegebenenfalls evaluiert werden. Die Arbeit soll einen perspektivischen Ausblick auf die notwendigen Cluster-Architekturen liefern, wobei ein Cluster aus der Ansammlung vieler Fahrzeug-Knoten (Nodes) besteht.

Vorgehen

Zur Evaluation verschiedener Control-Plane-Architekturen wird im ersten Schritt ein unmodifiziertes Kubernetes-Cluster erstellt. "Kubernetes ist ein Open-Source-System zur Automatisierung der Bereitstellung, Skalierung und Verwaltung von Container-Anwendungen"[5] . Kubernetes ist nicht darauf ausgelegt, auf mehr als 5000 Nodes zu skalieren. Ein Teil der Arbeit umfasst die Herausarbeitung der Gründe dieser Limitierung.

Zuerst wird mithilfe des Open-Source-Systems Terraform die Infrastruktur erstellt. Dieses wird verwendet, um deklarativ und automatisiert neue Infrastruktur zu planen, zu erstellen und zu verändern. Die benötigten Ressourcen werden in Dateien definiert. Terraform nutzt das Interface der Cloud-Anbieter, um diese zu realisieren.
Anschließend wird Kubespray eingesetzt, um ein Kubernetes-Cluster zu erstellen. Kubespray ist ein Tool zum automatisierten Deployment eines hochverfügbaren Kubernetes-Clusters. Es basiert auf Ansible und unterstützt die gängigen Cloud-Anbieter sowie lokale Deployments.
Da ein Test-Deployment von über 5000 Nodes sehr teuer ist und für die Umsetzung der geplanten Tests nicht notwendigerweise voll funktionsfähig sein muss, wird Virtual-Kubelet eingesetzt. Dies erlaubt es, mehrere Pseudo-Kubelets auf einer Node zu simulieren.
Da jeder virtuelle Worker-Node eine eigene IP-Adresse benötigt, werden mit Ansible zusätzliche IPs zugewiesen und die benötigten Netzwerk-Routen konfiguriert. "[Dieses] ist ein Open-Source Automatisierungs-Werkzeug zur Orchestrierung und allgemeinen Konfiguration und Administration von Computern. Es kombiniert Softwareverteilung, Ad-hoc-Kommando-Ausführung und Software-Configuration-Management." [4]

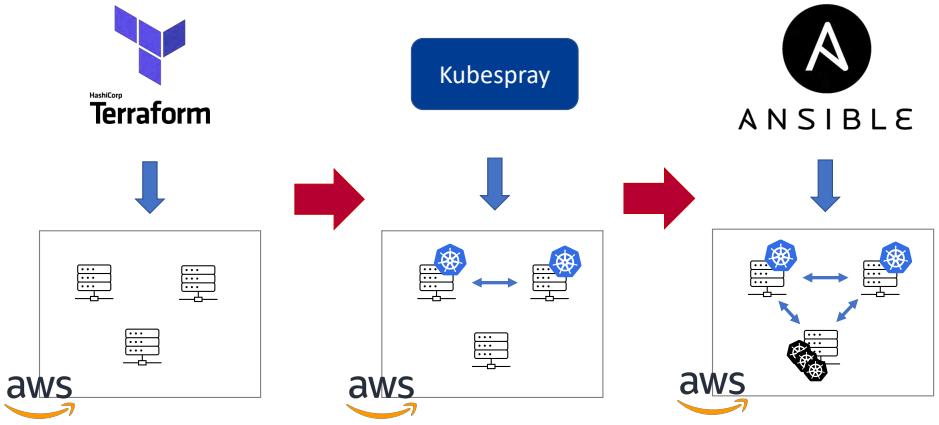

Abb. 2: Schritte zum automatisierten Deployment der Control-Plane [2]

Weiteres Vorhergehen

Das weitere Vorgehen beinhaltet die Ausarbeitung spezifischer Metriken zur Analyse der Performance. Mithilfe dieser Daten sollen die Komponenten identifiziert werden, die zu Performance-Problemen führen und Prototypen evaluiert werden, die diese Probleme lösen. Eine weitere mögliche Vorgehensweise ist die Entwicklung einer neuartigen dezentralen Control-Plane auf Basis einer skalierbaren verteilten Systemarchitektur.

Literatur und Abbildungen

[1] Audi AG. Ampelinformationen online. https://www.audi.de/de/brand/de/service-zubehoer/connect/ampelinformation-online.html, 2019.

[2] Eigene Darstellung.

[3] Tesla Inc. Supercharger. https://www.tesla.com/supercharger, 2022.

[4] Wikipedia Inc. Ansible. https://de.wikipedia.org/wiki/Ansible, 2022.

[5] Benedikt Rollik et al. Kubernetes. https://kubernetes.io/de/, 2022.

Objekterkennung anhand Bilder einer Wärmebildkamera im Kontext des vollautomatisierten Fahrens

Marc Schnaible **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Daimler Truck AG, Stuttgart

Ziel

Das Ziel dieser Arbeit ist es einen Objekterkennungsalgorithmus zu entwickeln, der in der Lage ist Objekte in Bildern einer Wärmebildkamera zu erkennen. Dabei soll dieser Algorithmus im Rahmen des vollautomatisierten Fahrens eines LKWs auf Amerikanischen Highways erstellt werden. Insbesondere wird in dieser Arbeit das Ziel verfolgt die Objekterkennung mittels einer Wärmebildkamera für die Fahrsituation Spurwechsel zu implementieren. Die Wärmebildkamera, deren Ansicht in Abbildung 1 zu sehen ist, dient zur Beobachtung des Verkehrs hinter dem Fahrzeug und auf einer Fahrspur neben dem LKW. Ein Spurwechsel wird beispielsweise bei dem Auffahren auf den Highway oder bei einem Überholmanöver vollzogen. Um einen Spurwechsel vollziehen zu können, ist es notwendig, den rückwärtigen Verkehr zu analysieren um sich anschließend in den Verkehr der Fahrspur ein sortieren zu können. Dabei dürfen andere Verkehrsteilnehmer nicht behindert oder blockiert werden. Da sich der LKW, wie auch andere Verkehsteilnehmer bewegen und sich damit die Situation ständig verändert, sollte die Lokalisierung und Klassifizierung von Objekten in Echtzeit erfolgen. Echtzeit bedeutet, dass das System für die Verarbeitung eines Bildes nicht mehr Zeit benötigen darf, wie ein festgelegter Grenzwert. Nimmt man dabei die Fahrphysikalischen Eigenschaften des LKWs und den damit einhergehenden Bremsweg des LKWs zur Hand so lässt sich auf diese Weise ein Grenzwert von 10 Hertz für die Bildwiederholrate festlegen.

Abb. 1: Identifikation eines Autos in einem Bild einer Wärmebildkamera [2]

Objekterkennung

Objekterkennung ist ein weitgefasster Begriff der mehrere Computer-Vision Aufgaben vereint. Ziel der Objekterkennung ist die Identifizierung von Objekten in digitalen Bildern. Um dies umsetzen zu können kombiniert die Objekterkennung die Computer-Vision Aufgaben Bildklassifizierung sowie die Objektlokalisierung. Die Bildklassifizierung umfasst die Vorhersage der Klasse eines Objekts in einem Bild. Bei der Objektlokalisierung werden die Position und die Größe eines oder mehrerer Objekte in einem Bild bestimmt. [1] Damit in einem Bild Objekte identifiziert werden können, muss für die Gesamtheit aller Pixel eines Eingabebildes herausgefunden werden, was diese darstellen. Dies gestaltet sich durchaus komplex, da bereits eine kleine Abweichung des Objekts im Bild eine andere Pixelrepräsentation zur Folge hat. Wird ein Objekt beispielsweise aus einem anderen Winkel aufgenommen, so variiert die Form des Objekts. Darüber hinaus kann die Größe, wie auch die Position des Objekts im Bild unterschiedlich sein. Eine weitere Herausforderung bei

der Identifizierung von Objekten ist, dass Objekte verdeckt sein könnten. Mit all diesen Herausforderungen kommen Menschen sehr gut zurecht und können trotzdem Objekte identifizieren. Dabei orientiert sich das menschliche Gehirn an bestimmten Merkmalen, wie zum Beispiel Ecken oder Kanten. Dieses Vorgehen wird zum Teil auf die Bildverarbeitung übertragen. Demzufolge wird zunächst eine Merkmalextraktion auf dem Bild durchgeführt. Dadurch werden Informationen über die Existenz, die Position, sowie die Orientierung bestimmter Merkmale im Bild gewonnen. Anschließend können diese Informationen für weitere Aufgaben der Bildverarbeitung genutzt werden. [7]
Wie bereits beschrieben vereint die Objekterkennung zwei Aufgaben der Bildverarbeitung, die Klassifizierung und die Lokalisierung von Objekten in Bildern. Für die Umsetzung dieser zwei Aufgaben gibt es grundsätzlich zwei verschiedene Herangehensweisen, die One-Stage Objekterkennung und die Two-Stage Objekterkennung.

Two-Stage Objekterkennung

Die Two-Stage Methode verwendet die zuvor erläuterte Informationen aus der Merkmalextraktion als Ausgangspunkt für die weiteren Schritte zur Identifizierung von Objekten im Bild. Basierend auf den Merkmaldaten werden bei der Two-Stage Methode Vorschläge für Regionen, in denen sich jeweils ein Objekt befinden kann, erzeugt. Nachdem mehrere Bildregionen mit besonderem Interesse vorgeschlagen wurden, wird für jede Region eine Bildklassifizierung und eine Lokalisierung zur Bestimmung der Bounding Box Koordinaten durchgeführt. Dieses Vorgehen ist zu einem gewissen Grad aus den Erkenntnissen des Aufmerksamkeitsmechanismus aus der Erforschung des menschlichen Gehirns übertragen worden. Auch das menschliche Gehirn analysiert zunächst das gesamte Szenario bevor es sich anschließend auf Regionen von Interesse konzentriert. [4]

One-Stage Objekterkennung

Im Gegensatz zu der Two-Stage Objekterkennung wird bei der One-Stage Objekterkennung direkt auf den Merkmalinformationen des Bildes eine Klassifizierung und Lokalisierung von Objekten durchgeführt. Dadurch können im Vergleich zu der Two-Stage Methode deutlich kürzere Laufzeiten erreicht werden, was diese Methode für Echtzeitanwendungen interessant macht. [4]
In den letzten Jahren hat sich herausgestellt, dass künstliche neuronale Netze sehr schnell mit einer hohen Genauigkeit Objekte in Bilder klassifizieren und lokalisieren können. So werden auch in dieser Arbeit künstliche neuronale Netze beziehungsweiße konvolutionale Neuronale Netze (kurz: CNN) verwendet. CNNs eignen sich besonders für Bilder, da bei diesen die geometrischen Informationen eines Bildes erhalten bleiben. [4]

FLIR Thermal Datensatz

Für die Entwicklung des Objekterkennungsalgorithmus anhand von Bildern einer Wärmebildkamera wurde der datengesteuerte Ansatz gewählt. Dieser wurde gewählt, da er für die Problemstellung dieser Arbeit die gängige Praxis widerspiegelt und eine Alternative hohe Komplexität mit sich bringen würde.
Für den datengesteuerten Ansatz wird allerdings ein Datensatz an Bildern einer Wärmebildkamera mit entsprechenden Annotationen benötigt. Im Rahmen dieser Arbeit wurde der FLIR Thermal Datensatz gewählt. Dieser Datensatz beinhaltet 26442 Bilder mit 520000 Annotation. [5] Die Bilder des Datensatzes zeigen gänige Situationen aus dem Straßenverkehr in Europa sowie den USA. Zur Aufnahme dieser Bilder wurde die Wärmebildkamera auf Höhe des Fahrers am Auto in Fahrtrichtung befestigt. Des Weiteren werden alle für diese Arbeit relevanten Objektklassen durch diesen Datensatz repräsentiert. Zusätzlich wurde dieser Datensatz gewählt, da er für Wissenschaftliche Zwecke öffentlich ist und damit einen Vergleich ermöglicht. Ein weiterer Vorteil ist, dass im Rahmen dieser Arbeit kein eigener Datensatz erstellt werden muss, was einen erheblichen Zeitaufwand mit sich bringen würde, was jedoch bei einer Produktanwendung aus Lizenzierungsgründen notwendig wäre.

Umsetzung

Für die Umsetzung wurden zunächst verschiedene Modellarchitekturen ausgewählt, die den aktuellen Stand der Technik repräsentieren. Hierbei wurden in erster Linie zwei Metriken als Auswahlkriterien herangezogen. Die Inferenzgeschwindigkeit und die Mean Average Precision. Die erste Metrik beschreibt die Zeit wie lange das Modell zur Analyse eines Bildes benötigt. Der Grenzwert für die Inferenzzeit eines Modells wurde auf Grund der Echtzeitfähigkeit auf maximal 100 Millisekunden festgelegt. Die Mean Average Precision beschreibt mit welcher Genauigkeit ein Modell Objekte in Bildern identifiziert. Um so höher der Wert der Mean-Average Precision, desto genauer ist das Modell. [3] Für die Realisierung der Objekterkennung mittels CNNs wurde die Methode des Transferlernens herangezogen. Der Grundgedanke beim Transferlernen ist der, dass ein Modell welches mit einem großen und ausreichend allgemeinen Datensatz trainiert wurde als generisches Modell der visuellen Welt dienen kann. So können die von diesem Modell erlernten Feature-Maps genutzt werden, um andere Problemstellungen schneller und mit weniger Beispielen zu lösen. Ein Vortrainiertes

Modell ist also ein gespeichertes Netzwerk, welches zuvor an einem großen Datensatz trainiert wurde. Das Vortrainierte Modell kann in einem weiteren Schritt mit einem Problemspezifischen Datensatz weiter trainiert werden. [6] Für Objekterkennungsaufgaben wird in der Regel für das Vortrainieren eines Netzes der Datensatz Common Objects in Context kurz COCO verwendet. Dieser ausreichend große Datensatz kann als generisches Modell der visuellen Welt dienen. Die zuvor beschriebenen Metriken werden für die Auswahl geeigneter Modelle anhand des Testdatensatzes von COCO ermittelt. Anschließend wurden die ausgewählten Modelle unter möglichst gleichen Bedingungen auf dem zuvor beschriebenen FLIR Thermal Datensatz trainiert. Unter Berücksichtigung des Trainingsverlusts und des Validierungsverlusts wurde das Optimum der einzelnen Modelle im Trainingsverfahren ausgewählt. Das Beste Ergebnis des Optimierungsverfahrens wird hauptsächlich anhand des Validierungsverlusts bewertet, da dieser die Evaluierung des Netzes auf Daten widerspiegelt, die nicht für das Training herangezogen wurden, jedoch der Applikation entsprechen. Zudem verdeutlicht der Verlauf des Trainingsverlusts wie sehr das Modell noch optimiert wird. Nachdem das Optimum des Trainingsverfahrens der einzelnen Modelle ermittelt wurde, konnten die Modelle evaluiert und gegenübergestellt werden. Für die Evaluierung wurde zum einen der Testdatensatz von FLIR und zum anderen ein eigener Testdatensatz, der mit dem LKW aufgenommen wurde, herangezogen. Da der eigene Testdatensatz genau die Situationen zeigt, die auch später in der Anwendung wieder zu finden sind, wird in dieser Arbeit die Evaluierung mit dem eigenen Testdatensatz näher erläutert. Für die Gegenüberstellung der verschiedenen Modellarchitekturen wurde die Precision-Recall-Kurve gewählt. Die Precision-Recall-Kurve zeigt den Kompromiss zwischen Präzision und Recall für verschiedene Schwellenwerte. Eine große Fläche unter der Kurve steht sowohl für einen hohen Recall als auch für eine hohe Präzision, wobei sich eine hohe Präzision auf eine niedrige Falsch-Positiv-Rate und ein hoher Recall auf eine niedrige Falsch-Negativ-Rate bezieht. Die Formeln 1 bis 2 beschreiben wie sich der Recall und die Präzision zusammen setzen. Eine Vorhersage für ein Objekt gilt als True Positive kurz TP, wenn das vorhergesagte Objekt mit dem Ground Truth Objekt eine größere Intersection over Union als 0,5 hat. Intersection over Union ist ein Begriff, der das Ausmaß der Überlappung von zwei Feldern beschreibt. Je größer der Überschneidungsbereich, desto größer die Intersection over Union. Ist die Intersection over Union kleiner als 0,5, so gilt die Vorhersage als False Positive kurz FP. Wurde hingegen bei einem exisitierenden Ground Truth Objekt keine Vorhersage durch das Modell getroffen, so gilt dies als False Negative. [3]

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Um die verschiedenen Modellarchitekturen besser bewerten zu können wurden Grenzwerte für die Präzision und den Recall festgelegt. Diese Grenzwerte wurden durch den Austausch mit Kollegen und den dabei ermittelten Rahmenbedingungen der Applikation näher definiert. Da sich die Modelle allerdings für Objekte mit unterschiedlichen Entfernungen zur Kamera auch unterschiedlich verhalten, wurden die Objekte des Test Datensatzes entsprechend ihrer Entfernung zur Kamera in drei Bereiche unterteilt. Die Einteilung dieser drei Bereiche wurde ebenfalls anhand der ermittelten Informationen über die Szenarienparametrisierung getroffen. Die Entfernung der Objekte wurde mittels der Referenzmesstechniken INS und GNSS beim Aufnehmen der Bilder ermittelt.
Abbildung 2 zeigt im ersten Diagramm die Precision-Recall-Kurve für den gesamten Bereich und in den anderen Diagrammen die Kurven für die einzelnen Entfernungsbereiche. Zudem sind in den Diagrammen die entsprechenden Grenzwerte mit eingezeichnet. Die grauen Flächen zeigen Bereiche die unterhalb der Grenzwerte liegen und damit nicht in Betracht gezogen werden sollten. In allen Bereichen sticht das Modell YOLOR hervor. Darüber hinaus hat dieses Modell eine Laufzeit von 0,021 Sekunden und erfüllt damit die Anforderungen für die Echtzeitfähigkeit. Das Modell Faster R-CNN ist das einizige Modell nach der beschriebenen Two-Stage Methode. Anhand des Vorgehens dieser Methode wäre zu erwarten, dass Modelle eine höhere Genauigkeit erzielen können. Diese These lässt sich nur bedingt bestätigen. Zwar wird eine hohe Genauigkeit erreicht allerdings ist die Erkennungsrate welche durch den Recall wiedergespiegelt wird im Vergleich zu anderen Modellen eher gering.

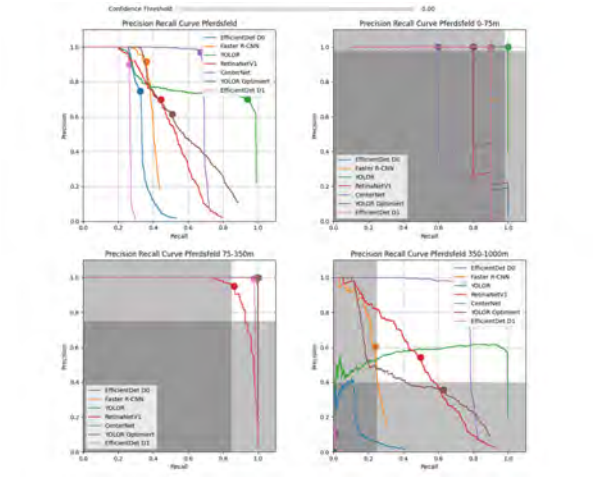

Abb. 2: Precision-Recall Kurve [2]

Im weiteren Verlauf dieser Arbeit wird versucht das anhand dieser Evaluierung hervorstechende Modell YOLOR weiter zu optimieren und im Anschluss für die Verwendung in der Anwendung zu verarbeiten.

Literatur und Abbildungen

[1] Jason Brownlee. A Gentle Introduction to Object Recognition With Deep Learning. https://machinelearningmastery.com/object-recognition-with-deep-learning/, 05 2019.

[2] Eigene Darstellung.

[3] Ahmed Fawzy Gad. Evaluating Object Detection Models Using Mean Average Precision (mAP). https://blog.paperspace.com/mean-average-precision/, 2020.

[4] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*, 2019.

[5] Teledyne FLIR LLC. KOSTENLOSER Teledyne FLIR-Wärmebilddatensatz für das Algorithmustraining. https://www.flir.de/oem/adas/adas-dataset-form/, 2021.

[6] Hannah Mitera, Chanjong Im, Thomas Mandl, and Christa Womser-Hacker. *Objekterkennung in historischen Bilderbüchern: Eine Evaluierung des Potenzials von Computer-Vision-Algorithmen*. Springer-Verlag GmbH, 2021.

[7] Zhao Zhong-Qiu, Zheng Peng, Xu Shou-tao, and Wu Xindong. Object Detection with Deep Learning: A Review. *IEEE*, 2018.

Entwicklung eines generischen Business-Intelligence-Ansatzes

Marc Schnalke **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung und Zielsetzung

Ein im März 2022 veröffentlichter Bericht des Forschungs- und Beratungsunternehmens Ëmergen Research"bewertet das Marktvolumen von Business-Intelligence-Plattformen auf 33,91 Billionen US-Dollar und prognostiziert bis in das Jahr 2028 ein durchschnittliches Wachstum von 11,9% auf einen Wert von 84,25 Billionen US-Dollar [3]. Ein Blick auf diese Beträge zeigt deutlich, dass Business Intelligence in den Unternehmen nicht mehr wegzudenken ist. Der Prozess zur Implementierung eines Business-Intelligence-Gesamtsystems kann dabei sehr komplex werden, um das volle Potenzial einer solchen Lösung auszuschöpfen. Diese Arbeit soll einen generischen Entwicklungsprozess für die Planung und Umsetzung eines solchen Projektes in einem Unternehmen mit den verschiedenen Architektur- und Technologievarianten beschreiben.

Rahmenkonzept

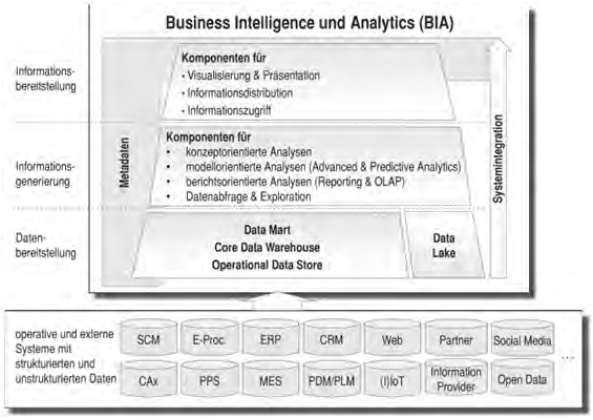

Abb. 1: Rahmenkonzept eines Business-Intelligence-Gesamtsystems [1]

Der Aufbau eines Business-Intelligence-Gesamtsystems lässt sich auf das in Abbildung 1 dargestellte Rahmenkonzept zurückführen, dessen Teilsysteme immer im unternehmensspezifischen Umfeld definiert werden müssen 1. Zur Konzipierung eines Business-Intelligence-Gesamtsystems bilden die operativen Quellsysteme eines Unternehmens die Datenbasis. Diese werden auch als „online transactional processing“ (OLTP)-Systeme bezeichnet [4]. Zusätzlich können noch externe Quellen die Datenbasis des Business-Intelligence-Systems ergänzen. Neben den Datenquellen kristallisieren sich drei aufeinander aufbauende Schichten des zu entwickelnden Business-Intelligence-Gesamtsystems heraus [1]:

- **Datenbereitstellung**: Die unterste Schicht des Gesamtsystems bildet die Datenbereitstellung, die in der Praxis sehr unterschiedlich gestaltet wird. Unstrukturierte Daten, bspw. aus Big-Data-Anwendungen können in Data Lakes gespeichert werden. Strukturierte Daten werden in der Regel über Data-Warehouse-Architekturen verarbeitet, die wiederum individuelle Gestaltungsmöglichkeiten, wie bspw. einen Zwischenschritt über einen Operational Data Storage (ODS), bieten. [1]
- **Informationsgenerierung**: In der zweiten Schicht wird die Informationsgenerierung in Form von verschiedenen Analysen praktiziert. Die am häufigsten verwendete Analysemethode wird als „online analytical processing“ (OLAP) bezeichnet. Im Vergleich zu den ebenfalls generierten Standard-Reports können nach Dimensionen definierte individuelle ad-hoc Reports erstellt werden [4]. In den letzten Jahren ist zusätzlich noch der Bereich von Analyselösungen, basierend auf mathematisch-statistischen oder algorithmischen Modellen, aufgekommen [1].
- **Informationsbereitstellung**: Die finale Schicht dient zur Visualisierung und Gewährleistung des Zugangs auf die in der vorherigen Phase gewonnen Informationen. Typischerweise erfolgt die Darstellung in Form von Dashboards durch Front-End-Lösungen. [1]

Vorgehensweise

Zunächst soll der Grundlagenteil die unterschiedlichen Gestaltungsvarianten der drei Schichten thematisieren, um einen Überblick zu geben, welche Möglichkeiten für einen generischen Ansatz in Frage kommen. Außerdem sollen im Grundlagenteil aktuelle Trends im Bereich der Business-Intelligence herausgearbeitet werden, um die Aktualität des generischen Ansatzes zu garantieren. Bevor die einzelnen Schichten des Business-Intelligence-Gesamstsystems für den generischen Ansatz entwickelt werden können, muss eine individuelle Business-Intelligence-Strategie in einem Unternehmen entwickelt werden. Mit einer Business-Intelligence-Strategie wird die Ausrichtung eines Business-Intelligence-Gesamtsystems in einem Unternehmen bestimmt. Sie bildet dadurch den Grundstein für einen langfristigen Unternehmenserfolg, indem sie an den Unternehmenszielen ausgerichtet für eine bestmögliche Versorgung mit aus Daten gewonnenen Informationen sorgt [2]. Dabei sollen die unterschiedlichen Aufgaben im Prozess der Entwicklung einer Business-Intelligence-Strategie herausgearbeitet und erläutert werden. Ziel der Strategieentwicklung ist das Definieren einer an den Zielen eines Unternehmens ausgerichteten und dadurch individuellen Business-Intelligence-Architektur, die als Basis für die Entwicklung der weiteren Schichten dient.

Literatur und Abbildungen

[1] Henning Baars and Hans-Georg Kemper. *Business Intelligence & Analytics - Grundlagen und praktische Anwendungen*. Springer Vieweg, 2021.

[2] Peter Gluchowski and Peter Chamoni. *Analytische Informationssysteme*. Springer Berlin Heidelberg, 2021.

[3] Emergen Research. Business Intelligence and Analytics Platforms Market By Platform Type (Business Intelligence Platforms, Advanced & Predictive Analytics Platforms, CPM Suite), By Organization Size (Large Enterprise, Small & Medium Sized Enterprises), By Mode of Deployment. https://www.emergenresearch.com/industry-report/business-intelligence-and-analytics-platforms-market, 03 2022.

[4] Mohammad Ismail Rostaminiya and Arash Hossin Zadeh Fard. Investigate the Business Intelligence Development and Improvement. *Dutch Journal of Finance and Management*, 2019.

Untersuchung und Vergleich von verschiedenen Backend-as-a-Service-Anbietern

Josua Seibold **Jürgen Koch**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma pep.digital GmbH, Esslingen

Einleitung

Jedes Jahr steigt die Zahl der Menschen, die ein Smartphone benutzen [3]. Prognostiziert ist auch, dass die Zahl der Nutzer in den nächsten Jahren noch weiter steigen wird. Ein Ende des Wachstums zeichnet sich nicht ab, da die Zahl der Smartphone-Besitzer und die Zeit, die Menschen am Smartphone verbringen, nach oben gehen. Dadurch werden Internetanwendungen und Apps immer bedeutender [2]. Über das Smartphone kann fast jeder Mensch erreicht werden, deshalb entstehen viele Internetseiten und Apps, die bestimmte und spezifische Aufgaben erfüllen. Diese Anwendungen werden dabei nicht nur von IT-Riesen, wie Google, Amazon und Meta entwickelt, sondern auch von vielen Unternehmen, die bisher keine oder wenig digitale Interaktionen mit Kunden haben. Durch das schnelle Wachstum ist es von großer Bedeutung, Produkte schnell auf den Markt zu bringen.

Cloud

Die Cloud nimmt bei der Entwicklung neuer Internetseiten und Apps eine immer wichtigere Rolle ein. Viele Services können über die Cloud schnell und einfach genutzt werden, die hauptsächlich die Entwicklung und den Betrieb von digitalen Systemen beschleunigen. Der Trend, Cloud-Services zu nutzen, wird immer größer.

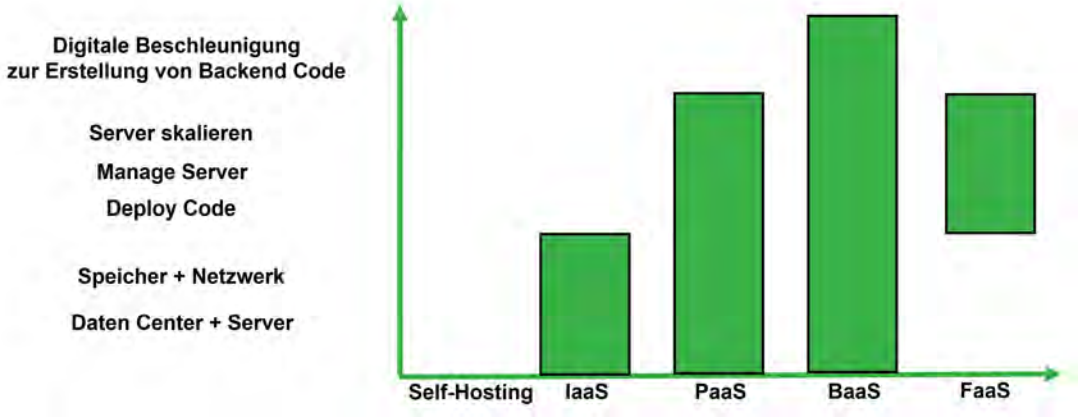

Abb. 1: ?-as-a-Service [1]

Vor der Nutzung von Cloud Services musste jedes Unternehmen ihre eigene Infrastruktur errichten, um digitale Dienstleistungen anzubieten. Da dies nicht für alle Unternehmen sinnvoll ist, entwickelten sich erste Züge der heutigen Cloud. Zum Beispiel werden virtuelle Maschinen (VM) angeboten, deren Rechenleistung für die eigenen Zwecke verwendet werden kann. Auf dieser Stufe wird die Software komplett durch den Nutzer geschrieben und betrieben. Nur die Hardware-Seite wird durch den Cloud-Anbieter übernommen. Mit jeder Erweiterung der Cloud wird mehr durch den Cloud-Anbieter übernommen. Das bedeutet, dass nicht nur die Hardware gemietet ist, sondern auch Teile des Softwarestacks. Dazu gehören zum Beispiel Betriebssysteme, Server-Software und Hosting. All diese Erweiterungen der Cloud-Services, führen dazu, dass ein Entwickler oder ein Unternehmen, sich nicht damit beschäftigen muss, sondern sich auf das Produkt konzentrieren kann. Diese Entwicklung endet aktuell bei dem Backend-as-a-Service (BaaS), mit dem ein

komplettes Backend angemietet werden kann.

Backend-as-a-Service (BaaS)

Das BaaS sorgt dafür, dass ein Entwickler, für wichtige Funktionen keinen Backend-Code schreiben muss, sondern nur über eine Web-UI sein Backend konfigurieren kann. Ein BaaS bringt den Code mit, der für wichtige Features eines Backends benötigt wird. Viele Anwendungen arbeiten mit Daten, wozu Nutzerdaten und inhaltliche Daten gehören. Diese Daten werden in den meisten Fällen in Datenbanken gespeichert. Das BaaS hostet für den Kunden eine Datenbank und stellt vorkonfiguriert eine Schnittstelle zur Verfügung. Der Kunde muss nur noch aus seinem Frontend die Informationen aus dem Backend abfragen. Weitere Funktionen von BaaS sind die Authentifizierung von Nutzern, das Speichern von großen Dateien und das Nutzen von Cloud Funktionen.

Anbieter von Backend-as-a-Service

Der Markt für BaaS-Systeme ist groß und deshalb gibt es auch viele Anbieter solcher Systeme. Für einen Vergleich können nicht alle Systeme bis ins Detail untersucht werden, aber so, dass eine allgemeine Aussage über verschiedene BaaS möglich ist.

1. Google Firebase

   Googles Firebase ist das bekannteste BaaS, es wurde 2011 gegründet und gehört damit zu den ersten BaaS [3]. Firebase wird von vielen Entwicklern genutzt und bietet viele Features und Integrationen.

2. Nhost

   Nhost ist ein neuer Player. Mit der aktuellen Version können alle wichtigen Funktionen umgesetzt werden. Nhost hat sich zum Ziel gesetzt ein Konkurrent von Firebase zu sein.

3. Back4App

   Back4App ist ein Open-Source basiertes BaaS. Es basiert auf der Open-Source Lösung Parse. Parse ist schon länger in der Entwicklung und wird unter anderem durch Back4App gehostet.

4. Backendless

   Backendless ist ein All-In-One BaaS. Anders als die bisher genannten Anbieter, ist es mit Backendless auch möglich, ein passendes Frontend zu bauen. Großen Wert legt Backendless darauf, dass alles ohne Code funktioniert.

5. AWS Amplify

   AWS Amplify ist der Versuch von Amazon, als größtem Cloud-Anbieter Konkurrenz zu Firebase zu bieten. Genau wie Firebase profitiert Amplify von der großen Erfahrung im Cloud Bereich und den schon vorhandenen Produkten von AWS.

6. Mongo DB Realm

   Mongo DB Realm nutzt die bekannte Datenbank MongoDB und stellt diese als Teil eines BaaS den Kunden zur Verfügung.

7. Supabase

   Supabase ist ein sehr neues BaaS, welches auf Open-Source-Software entwickeln will. Anders als Back4App nutzen sie nicht Parse, sondern entwickeln ihre eigene Software. Hinter dem Projekt steht zum Beispiel Mozilla.

8. 8Base

   8Base ist wie Nhost ein BaaS, welches versucht Firebase zu verdrängen. 8Base entwickelt sich in Richtung von Backendless, denn aktuell ist ein Frontend-Builder in der Beta-Phase.

Ziele

Das Ziel dieser Arbeit ist es, einen Überblick und eine Vergleichsbasis zwischen BaaS-Systemen herzustellen. Auf Basis dieses Vergleichs, soll eine fundierte Aussage darüber getroffen werden können, welches BaaS allgemein eine gute Wahl ist, und welches in einem Bestimmten Anwendungsszenario am besten geeignet ist.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] L. Rabe. Umfrage zur Nutzung von Anwendungen im Internet nach Geschlecht in Deutschland 2020. https://de.statista.com/statistik/daten/studie/1189123/umfrage/nutzung-von-anwendungen-im-internet-nach-geschlecht/, 02 2022.

[3] F. Tenzer. Anteil der Smartphone-Nutzer* in Deutschland in den Jahren 2012 bis 2021. https://de.statista.com/statistik/daten/studie/585883/umfrage/anteil-der-smartphone-nutzer-in-deutschland/, 11 2021.

Autonomes und kooperatives Einfädeln in eine Fahrzeugkolonne oder in ein Platoon

Sungeeta Singh **Thao Dang**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Mercedes-Benz AG, Böblingen-Hulb

Motivation

Das Autofahren ist allgemein bekannt als das gefährlichste Transportmittel. Durch menschliche Fehler beim Fahren oder auch durch Einschränkungen in der Sicht können Unfälle verstärkt auftreten. Im Rahmen des Projekts IMAGinE [4] sollen kooperative Assistenzsysteme zur Gefahrenerkennung und -reduzierung entwickelt werden. Hierfür ist es nötig, ein kollektives Verständnis über das Umfeld zu erhalten, in dem ein Informationsaustausch und eine Manöver-Abstimmungen zwischen den Fahrzeugen stattfinden. Eine besonders schwierige Verkehrssituation ist bspw. das Einfädeln auf eine Autobahn. Hierbei muss ein Fahrer mehrere Aufgaben gleichzeitig lösen: Anpassung der Geschwindigkeit an den Fahrzeugen auf der Hauptfahrbahn, Beobachtung der Fahrzeuge in der Umgebung, Entscheidung für eine optimale Fahrlücke und Vorhersagen der Entwicklung dieser Lücke. Mithilfe einer V2X-Kommunikation kann das Einfädelungsverfahren kooperativ unter den Fahrzeugen abgestimmt werden und somit das Fahren angenehmer und vor allem sicherer werden.

Zielsetzung

Das Ziel dieser Bachelorarbeit ist es ein bereits implementiertes und autonomes Einfädelungsverfahren so zu erweitern, dass sich das Egofahrzeug vor, hinter oder zwischen einer Fahrzeugkolonne oder einem Platoon einfädeln kann. Die beschriebene Verkehrssituation ist in der Abbildung 1 mit dem Egofahrzeug in dem Einfädelungsstreifen und einem Platoon aus vier Fahrzeugen in der linken Hauptfahrbahn dargestellt. Die Boxen sowie die Verbindungslinien symbolisieren das Platoon. Hierbei berechnet das Egofahrzeug die für sich optimale Möglichkeit in dieser Verkehrssituation. Hierfür prüft das Egofahrzeug die Einfädelung zwischen zwei Fahrzeugen der Hauptfahrbahn. Die möglichen Situationen werden mit einer dynamischen Prädiktion erzeugt und mit einem nachgelagerten Trajektorienplaner bewertet. Anhand dieser kann das Fahrzeug die Situationen einschätzen und potenzielle Kooperationspartner identifizieren. Nach der Auswahl der bestmöglichen Kooperationspartner wird eine Anfrage per V2X gestellt. Akzeptiert das Umgebungsfahrzeug die Manöveranfrage, kann der kooperative Einfädelungsprozess ausgeführt werden.

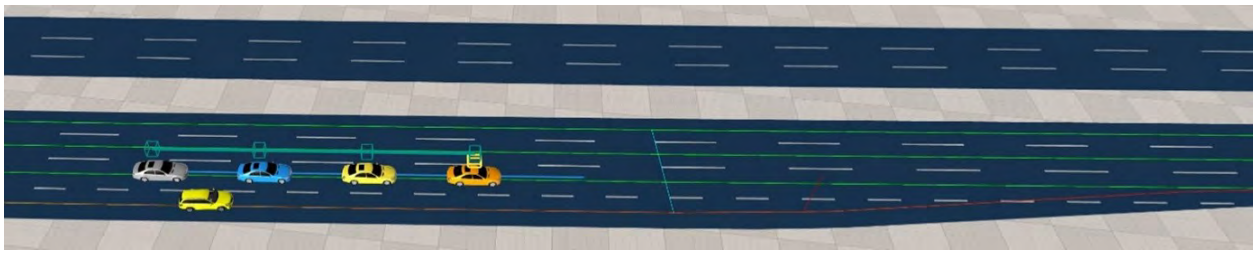

Abb. 1: Bildschirmaufnahme des Szenarios in PHABMACS [1]

PHABMACS

PHABMACS (Physics Aware Behavior Modelling Advances Car Simlator) ist ein von DCAITI (Daimler Center for Automotive IT Innovations) entwickeltes Simulationswerkzeug für das virtuelle Testen von ADAS (Advanced Driver Assistance Systems), das ein Framework in JAVA innerhalb einer physikalisch realistischen Umgebung zur Verfügung stellt [6]. Hierfür werden die vom HighD Datensatz aufgezeichneten Verkehrsszenarien in einer virtuellen Karte mit Infrastruktur (siehe Abb. 1) und simulierten Fahrzeugen dargestellt [5].

Für die vorliegende Abschlussarbeit wird PHABMACS verwendet, um Zwischenergebnisse, wie z.B. Prädiktionsergebnisse darzustellen und die Erweiterung des Einfädelungsmanövers mit der Kooperation zu testen.

V2X

Das vernetzte Fahren bringt Vorteile, wie Verkehrsflussoptimierung, Komfort, Einsparung im Spritverbrauch und Reduzierung der Verkehrsunfälle, mit sich. Hierfür ist eine Kommunikation beispielsweise per V2X zwischen den Roadusern notwendig. V2X umfasst die Kommunikation zwischen zwei Fahrzeugen sowie zwischen einem Fahrzeug und der Infrastruktur. Eine Kooperation über das V2X-Modul findet für die Erstellung eines Umfeld-Modells oder für eine Manöverabstimmung statt. Die zwei wichtigen Nachrichtentypen, um ein Modell über die Verkehrssituation zu erhalten, sind CAM (Cooperative Awareness Message) und CPM (Cooperative Perception Message). Mit CAM sendet ein Fahrzeug Daten über seinen aktuellen Zustand aus. Darunter gehören Informationen, wie Zeitstempel, Position, Geschwindigkeit und Bewegungszustand. Hierdurch werden die Fahrzeuge in das Umfeld-Modell aufgenommen und können somit ihre Manöver aufeinander abstimmen. CPM teilt Informationen über andere Roaduser und Hindernisse mit. Diese Roaduser inkludieren ebenfalls Fahrzeuge, die kein V2X-Modul zum Kommunizieren besitzen und mit lokalen Wahrnehmungssensoren, wie z.B. eine Kamera oder Radar, detektiert werden. Durch die Informationen einer CPM erweitert sich das von CAM erhaltene Umfeld-Modell, da weitere Fahrzeuge detektiert werden können. Das Konzept hinter der CPM ist, dass alle Kommunikationsteilnehmer ein gemeinsames Verständnis über die Verkehrssituation behalten. Außerdem ergibt sich eine Verbesserung der geschätzten Parameter, wie die Position und Geschwindigkeit über Fahrzeuge, die bereits durch die CAM-Nachricht detektiert wurden, da weitere Sensordaten über diese Fahrzeuge in der Schätzung mit einfließen [3]. Mit der CMM (Cooperative Maneuver Message) kann ein sitzungsbasiertes Manöverabstimmungsverfahren eingeleitet werden. Eine Sitzung wird aufgebaut, da eine zustandsbasierte Abstimmung zwischen den Kommunikationspartnern stattfindet.

Systembeschreibung

Im Folgenden wird das bisher verwendete System für das Einfädelungsmanöver beschrieben. Die Erweiterung dieses Systems ist unter *Implementierung* nachzulesen. Für die Einfädelung wird eine kooperative und Trajektorien-basierte Manöverplanung verwendet. Bisher findet eine Kooperation zwischen den Fahrzeugen durch einen Informationsaustausch mit den V2X-Nachrichten CAM und CPM statt. Das Constant-Velocity (CV) Prädiktionsmodell verwendet die aus den Nachrichten erhaltenen Informationen als Eingabedaten, um das Verhalten der Fahrzeuge vorherzusagen. Mithilfe dieser Prädiktion kann die Wahrscheinlichkeit, ob sich ein Fahrzeug an ausgewählten Positionen befindet, berechnet werden. Hiermit können die Fahrzeuglücken detektiert werden. Für die Trajektorien-Berechnung werden Punkte in diesen Fahrzeuglücken festgelegt. Zwischen diesen Punkten und der aktuellen Position des Egofahrzeugs findet jeweils eine Hermite-Interpolation statt. Diese Interpolation stellt die Trajektorie dar. Die Trajektorie mit den geringsten Kosten wird für das Manöver ausgewählt. In den Kosten ist die longitudinale und laterale Beschleunigung, der Ruck, der Sicherheitsabstand zu den Fahrzeugen in der Umgebung, TTC (Time To Collision) und die Dauer des Spurwechsels für den Einfädelungsprozess einkalkuliert [2].

Implementierung

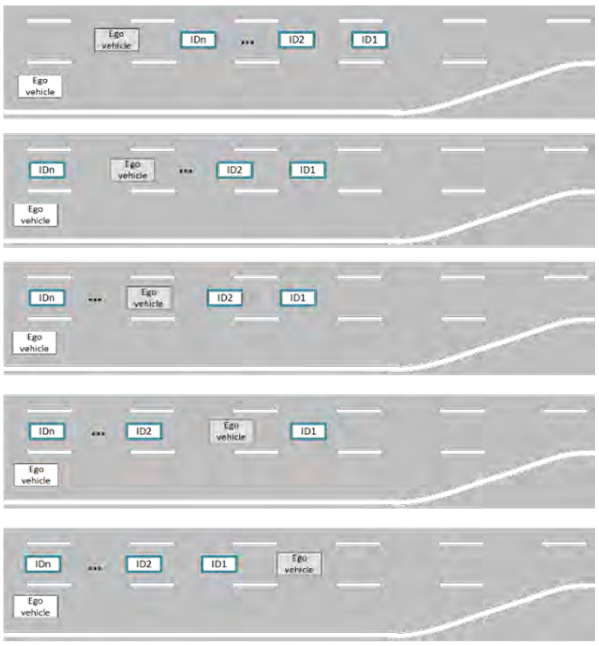

Abb. 2: Möglichkeiten zum Einfädeln für das Egofahrzeug bei n Fahrzeugen auf der Hauptfahrbahn [1]

In der Abbildung 2 ist dargestellt, welche Einfädelungsmöglichkeiten für das Egofahrzeug bestehen. Die Box mit der Beschriftung *Ego vehicle* stellt dar, in welcher Anordnung sich das Egofahrzeug einfädeln würde, wenn es sich für die jeweilige Möglichkeit entscheidet. Die ID-Nummer repräsentiert jeweils ein Fahrzeug in der Hauptfahrbahn. Eine beliebige Anzahl an Fahrzeugen kann in dieser Spur sein, weshalb das n in der Abbildung

eine beliebige natürliche Zahl annehmen kann. Sind *n* Fahrzeuge mithilfe des V2X-Umfeldmodells in der Hauptfahrbahn detektiert, werden *n+1* Möglichkeiten berechnet. Für jede Möglichkeit wird die Trajektorien-basierte Bewegung geplant. Hierbei wird die in [2] beschriebene Bewegungsplanung verwendet. Jedoch wird ein neues Prädiktionsmodell für den Lückenaufbau verwendet. Das dynamische Prädiktionsmodell besitzt einen Geschwindigkeitsregler, der den Beschleunigungs- und Verzögerungsvorgang prädiziert. Schließlich liefert jede der *n+1* Bewegungsplanungen eine optimale Trajektorie. Nun wird iterativ die Möglichkeit mit den geringsten Trajektorienkosten für die CMM-Anfrage selektiert. Der Einfädelungsvorgang wird ausgeführt, wenn eine Manöveranfrage akzeptiert wird. Bei einer Ablehnung der Anfrage wird die nächst-optimale Möglichkeit ausgewählt.

Ausblick

In dem nächsten Schritt wird das Prädiktionsmodell für eine bessere und realistische Vorhersage angepasst. Außerdem sollen die identifizierten Kooperationspartner in die Kommunikation mit eingebunden werden. Damit kann der Lückenaufbau initiiert und anschließend das Einfädelungsmanöver ausgeführt werden. Am Ende der Arbeit wird das Einfädelungsverfahren ohne und mit der neu implementierten Kooperation untersucht, indem der zeitliche Geschwindigkeits- und Beschleunigungsverlauf des Egofahrzeugs aus beiden Use Cases verglichen wird. Hiermit soll die Hypothese bestätigt werden, dass das Egofahrzeug durch die im Rahmen der Abschlussarbeit implementierte Erweiterung eine komfortablere und sichere Fahrweise annehmen kann.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Lucas Eiermann, Nick Bührer, Gabi Breuel, et al. Trajectory Based Motion Planning for On-Ramp Merging-Situations Considering Extended Evaluation Criteria. In *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 299–306. Institute of Electrical and Electronics Engineers, 2020.

[3] European Telecommunications Standards Institute ETSI. *ETSI TR 103 562 V2.1.1 Technical Report*. European Telecommunications Standards Institute, 2019.

[4] IMAGinE. IMAGinE – Lösungen für kooperatives Fahren. https://www.imagine-online.de/, 2022.

[5] Robert Krajewski, Julian Bock, Laurent Kloeker, et al. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2118–2125. IEEE, 2018.

[6] Ilja Radusch. PHABMACS. https://www.fraunhoferventure.de/content/dam/venture/de/documents/Angebote/projekte/techbridge/PHABMACS-Product-Sheet.pdf, 2022.

Implementierung von interaktiven Dashboards zur Zeitreihenanalyse mit Jupyter Notebook

Sebastian Stein **Dirk Hesse**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt an der Fakultät Informatik und Informationstechnik, Esslingen

Einleitung

Eine Vielzahl von kleinen und mittelständischen Unternehmen verfügen über große Datenbestände, wissen jedoch nicht, was sie mit diesen Daten über den Zweck der Archivierung hinaus anfangen sollen. Die Gründe für die fehlende Datenanalyse liegen darin, dass die Verantwortlichen oft nicht wissen, wo die Informationen zu finden sind, was sie aussagen und welche Anwendungsfälle sich damit verbinden lassen. Dies kann durch die Beschäftigung eines einzigen Datenwissenschaftlers in Gang gesetzt werden. In dieser Arbeit wird ein maßgeschneiderter Ansatz aufgezeigt, der von Datenwissenschaftlern und Datenwissenschaftlerinnen schrittweise und mit einfachen Mitteln umgesetzt werden kann.

Ziel der Arbeit

Das Ziel dieser Arbeit ist es ein für den Benutzer interaktives Dashboard in Python innerhalb der Jupyter Notebooks Umgebung aufzubauen. In Abbildung 1 ist ein Bestandteil des Dashboards zu sehen. Die Interaktive Komponente hierbei stellt das Drop-Down Menü zur Wahl des zu betrachtenden Produktes und das Schiebewerkzeug zum Verändern des zu betrachtenden Zeitraums.

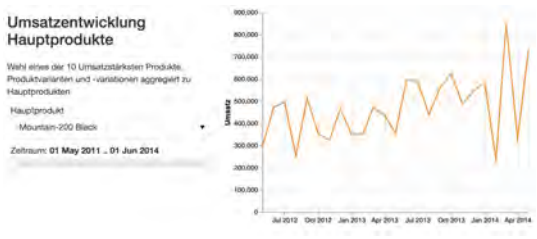

Abb. 1: Dashboardkomponente [1]

Der Fokus liegt hier auf der Erstellung des Dashboards mit einem Aufwand, der nicht weit über den der explorativen Datenanalyse hinausgeht. Zur explorativen Datenanalyse werden in der Arbeit gängige und relevante Python Bibliotheken wie pandas, Numpy, scikit-learn, statsmodels und Bibliotheken zur Visualisierung wie Bokeh und Altair beleuchtet und eingesetzt. Auf Webprogrammierung wird bei der Erstellung des Dashboards fast vollständig verzichtet. Die Anwendung ist für einen Endnutzer über den Browser und ohne installierte Python Umgebung erreichbar und verwendbar. Die darunterliegenden Daten stammen aus einer Beispieldatenbank und liegen auf einem lokalen Microsoft SQL Server. Das übergreifende Ziel dieser Arbeit ist es einen Mehrwert für das fiktive Unternehmen, mittels Datenanalyse und spezifisch Zeitreihenanalyse zu generieren.

Data Science als Prozess

Im ersten Schritt wird Data Science anhand von verschiedenen Modellen und Rahmenwerken aus der Literatur beschrieben und als ein Prozess dargestellt. Es werden Rahmenwerke wie 'Knowledge Discovery in Databases' und CRISP-DM beschrieben und angewandt, um die Einzelschritte zur Erstellung eines qualitativ hochwertigen Projektes abzuarbeiten. CRISP-DM ist ein erstmals im Jahr 1996 vorgestellter Prozess und steht für 'Cross Industry Standard Process for Data Mining'. CRISP-DM soll den komplexen Prozess des Data Minings in 6 Schritten standardisieren. Es wird begonnen mit dem Erlangen von einem Verständnis für das zugrundeliegende Geschäft und die damit einhergehenden Prozesse und Abläufe um daraus ableiten zu können welche Probleme durch das Data Mining gelöst werden können. Im zweiten Schritt wird Verständnis über die Daten angestrebt, es soll herausgefunden werden welche Daten für die weiteren Schritte von Relevanz sind und wie diese mit anderen Daten verknüpft sind, dies kann beispielsweise durch ein ERM-Modell der Datenbank unterstützt werden. Den dritten Schritt stellt die Datenaufbereitung dar. Es ist der Zeitintensivste Schritt, er beinhaltet die Datenbereinigung, Vorverarbeitung, den Umgang mit fehlenden Werten und der Zusammenführung von

Daten aus verschiedenen Quellen. Im nächsten Schritt werden die aufbereiteten Daten verwendet für die Erstellung von Modellen, beispielsweise durch den Einsatz von Algorithmen. Im vierten Schritt werden Ergebnisse des bisher durchgeführten Prozesses bewertet. Sind die Ergebnisse nicht zufriedenstellen, so kann eine neue Iteration vorgenommen werden. Sind die Ergebnisse zufriedenstellen, so kann im fünften und letzten Schritt das Modell in Betrieb genommen werden [2].

Datenaufbereitung

Im Hinblick auf die Erstellung von Datenanalysen und spezifisch einer Zeitreihenanalyse müssen die zu verwendenden Daten erst auf ihre Beschaffenheit geprüft werden. Ein Beispiel hierfür stellt die Ausreißeranalyse dar. Es handelt sich um eine Maßnahme zur Ermittlung von abweichenden, nicht erwarteten Werten oder Erscheinungen in einem Datensatz, die vom gewöhnlichen Zustandsbild abweichen.
Der partielle Datensatz 'Umsatz pro Transaktion für die Kundengruppe Geschäftskunden' wird mittels einer SQL-Abfrage in der Datenbank ausgelesen und mit Python in einen DataFrame der Bibliothek Pandas überführt. Hier kann nun die zuvor beschriebene Maßnahme umgesetzt werden.

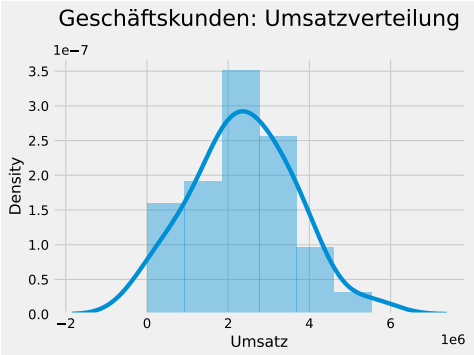

Abb. 2: Wahrscheinlichkeitsverteilung für Umsatz von Geschäftskunden [1]

Im ersten Schritt wird die Umsatzverteilung in Abbildung 2 auf Ausreißer und Unregelmäßigkeiten untersucht. Im nächsten Schritt wird der Isolation Forest Algorithmus mit dem Einsatz der scikit-learn Bibliothek herangezogen. Dieser berechnet für jede Umsatzzeile einen sogenannten Anomaliewert. Die blaue Kurve in Abbildung 3 zeigt diese Anomaliewerte. Die roten Flächen visualisieren, ab welchem Wert der Umsatz einer Transaktion als Anomalie zu werten ist. So können Einzelfälle im gekennzeichneten roten Bereich genauer untersucht werden, um zu klären, ob die Daten fehlerbehaftet sind oder ob die Daten einen realitätsgetreuen Zustand widerspiegeln.

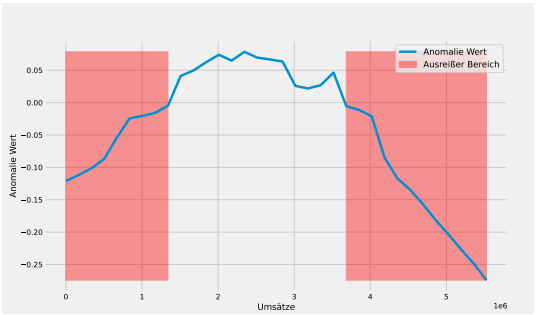

Abb. 3: Untersuchung von Anomalien mittels Isolation Forest [1]

Ausblick

In den nächsten Schritten wird eine Zeitreihenanalyse mit verschiedenen Modellen durchgeführt und ausgewertet. Weitere Bausteine werden in das Dashboard eingebunden und schließlich wird das Dashboard versuchsweise als Webanwendung in Betrieb genommen.

Literatur und Abbildungen

[1] Eigene Darstellung.

[2] Jochen Hipp and Rüdiger Wirth. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–40. Practical Application Company, 2000.

Konzeption und Implementierung eines webbasierten Editors zur Unterstützung von Crossmedia-Publishing

Simon Weber **Andreas Rößler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma mpr werbefactory Marketing- und Produktionsgesellschaft mbH, Reichenbach a. d. Fils

Einführung

Der Verkauf in lokalen Räumen wird durch die Nebenwirkungen des Digitalen Zeitalters vor neue Herausforderungen gestellt. Verschiedene Faktoren, wie z.B. Abstumpfung der Werbewirkung oder sinkende Loyalität der Konsumenten, „stellt die Verkaufsraumgestaltung vor neue Herausforderungen"[3]. Hier kommt Digital Signage (DS) ins Spiel. Und zwar ist darunter die Darstellung von Werbe- bzw. Produktinhalten auf Digitalen Systemen im Verkaufsraum gemeint, die zentral über das Internet angesteuert werden können. [3]

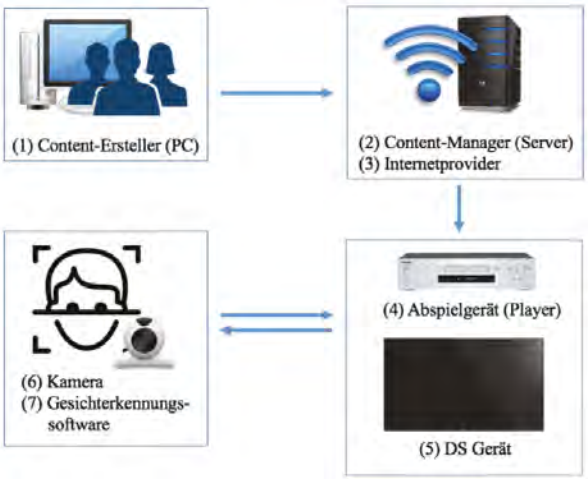

Abb. 1: Technische Funktionsweise DS mit Gesichtserkennung [3]

Abbildung 1 zeigt die beispielhafte Funktionsweise, hier mit der Erweiterung hinsichtlich Gesichtserkennung, zur Auswertung der Emotionen der Kunden. Die durch den Content-Ersteller bereitgestellten und den Content-Manager (Server) gespeicherten Daten zur Bereitstellung von DS-Inhalten, können aber nicht nur für DS-Systeme verwendet werden: Viel mehr bietet sich hier ein Ansatz, mit Hilfe von medienneutralen Daten, Crossmedia Publishing zu betreiben.

Motivation

Das Ziel der mpr werbefactory ist es aus bereitgestellten medienneutralen Daten eine Form des Single-Source-Publishings [1] zu kreieren. Zum jetzigen Zeitpunkt existieren bereits Daten aus vorhandenen Webtools diverser Kunden. Diese sollen in einer eigenständigen Webapp verwaltet werden können und in verschiedenen Formaten ausgegeben werden. Die Formate sollen PDF und Bilddateien (JPEG, PNG) umfassen, könnten aber auch auf fertige Html-Dateien, zur Einbindung in Websites, oder Filme (MPEG) ausgeweitet werden. Grund hierfür ist das zur Verfügung stellen von Daten und Inhalten, die bis jetzt für Printmedien, wie Prospekte, verwendet wurden. Das soll dem Kunden die Nutzung seiner angelegten Produkte in möglichst vielen Medien ermöglichen. Des Weiteren soll das ganze Programm möglichst kundenunabhängig (Branding, Farbwahl und Layout) entwickelt werden, um eine einfache und schnelle Anpassung an mögliche weitere Kunden zu ermöglichen.

Konzeption und Implementierung

Zur Umsetzung der benötigten Funktionalitäten werden beispielhaft folgende Technologien verwendet:

- Server: Linux mit Apache HTTP Server
- Backend: API basierend auf dem PHP Framework CodeIgniter 4
- Frontend: JS-App basierend auf dem VueJS Framework, um Interaktivität zu ermöglichen; CSS-Framework (Bootstrap)

Die Verwendung des CodeIgniter Frameworks ist eine zwingende Voraussetzung, darüber hinaus bietet das MVC-Framework aber auch viele Funktionen und

Methoden zur Verarbeitung von Daten via Datenbank und Dateien. Das einfache Arbeiten mit Dateien ist hier wichtig, da die medienneutralen Daten im XML-Format bereitgestellt werden. Dieses Format bietet im Vergleich zu anderen Formaten, wie das im Web-Umfeld gängige JSON-Format, die Möglichkeit zur einfachen Validierung durch den XML-Standard oder die Trennung von Daten, Struktur oder Layout durch z.B. DTDs oder XSLTs [4]. Alle verwendeten Technologien stehen als Open-Source-Projekte frei zur Verfügung.

Das Frontend der entwickelten App soll auch hinsichtlich der Usability analysiert werden. Hier ist vor allem zu beachten, dass der wahrscheinliche Nutzerkreis aus wenig technikaffinen Nutzern besteht. Zur Bewertung der Usability wurde zuerst ein Prototyp mit Adobe XD erstellt und analysiert. Dieser wird im folgenden näher beschrieben.

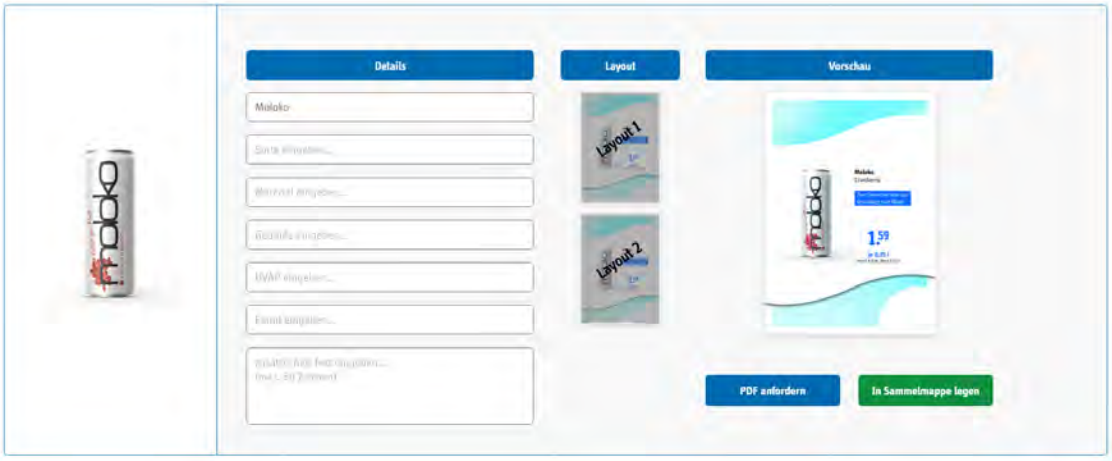

Abb. 2: Bearbeitung eines einzelnen Artikels [2]

Abbildung 1 zeigt einen einzelnen angelegten Artikel (hier am Beispiel eines Getränkehändlers) mit den verschiedenen Möglichkeiten zur Datenerfassung und Auswahl von vordefinierten Layouts. Die Artikel können, nach dem Erfassen, über ein Web-Content-Management-System verwaltet und abgerufen werden. Abbildung 2 zeigt dies beispielhaft:

Abb. 3: Verwaltung der Artikel über Webapp [2]

Die verschiedenen Artikel können durchsucht und gefiltert werden und danach vom Nutzer einem Warenkorb hinzugefügt werden, um sie schließlich in eines der unterschiedlichen Ausgabeformate zu exportieren.

Ausblick

In vielen Bereichen des Lebens hat die Digitalisierung Einzug gehalten. So auch am Point of Sale. Viele Händler nutzen keine klassischen Printmedien mehr, sondern nutzen ausschließlich digitale Formen der Werbung. An diesem Punkt kann die entwickelte Webapp einen Mehrwert bereitstellen, in dem sie dem Nutzer eine benutzerfreundliche Oberfläche mit der Möglichkeit zur simplen Erstellung von Crossmedia-Formaten dienen kann. Weitere Möglichkeiten der Weiterentwicklung wären z.B. das Erstellen von fertigen Filmchen oder aber auch das Berücksichtigen von Emotionsdaten der Gesichtserkennung der DS Systeme. Am Beispiel des Getränkehändlers könnte so dem müden Kunden passende Werbung für koffeinhaltige Getränke angezeigt werden.

Literatur und Abbildungen

[1] Peter Bühler, Patrick Schlaich, and Dominik Sinner. *Crossmedia Publishing*. Springer Vieweg, 2019.

[2] Eigene Darstellung.

[3] Ricarda Carina Rainer. *Digital Signage am Point of Sale*. Springer Gabler, 2020.

[4] Oliver Zmorek. Entwicklung eines XML-Webworkflows am Beispiel der HTWK-Zeitschrift »Streifband«, 2009.

Vehicle to Building - Technische Umsetzung von bidirektionalem Laden

Peter Wild **Mirko Sonntag**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma Hummel Systemhaus GmbH & Co. KG, Frickenhausen

Einleitung

Immer mehr Unternehmen und Privatpersonen entscheiden sich für Photovoltaikanlagen um unabhängiger vom Energiemarkt zu werden und um gleichzeitig in erneuerbare Energien zu investieren. Die von der Photovoltaikanlage erzeugte Energie soll vor allem zum Decken des eigenen Energiebedarfs genutzt werden. Es gibt zwei Möglichkeiten dies zu optimieren, zum einen kann bei hoher Energieproduktion darauf geachtet werden elektrische Geräte zu betreiben und diese Energie damit direkt zu nutzen. Die zweite Möglichkeit wäre ein Energiespeicher, in dem überschüssige Energie gespeichert wird.
Fahrzeuge stehen oft ungenutzt vor dem Haus oder Bürogebäude, außerdem haben die meisten Elektrofahrzeuge eine weitaus größere Energiekapazität als in Gebäuden verbaute Speicher. Zusätzlich muss der Speicher der Fahrzeuge für die meisten Fahrstrecken nicht vollgeladen sein. Dadurch bieten sie sich als Pufferspeicher an, während sie lokal an einen Ladepunkt angeschlossen sind. Dies soll mit bidirektionalem Laden ermöglicht werden, bei Überschuss von der Photovoltaikanlage werden die Fahrzeuge geladen und können zum Decken des Energiebedarfs wieder entladen werden.

Abb. 1: ENNAGY Ladesäule [2]

Bidirektionales Laden

Bidirektionales Laden beschreibt die Möglichkeit, Elektrofahrzeuge nicht nur aufladen zu können, sondern auch entladen zu können. Dadurch kann das Fahrzeug als lokaler Pufferspeicher genutzt werden. Dies funktioniert analog zu bereits großflächig genutzten Energiespeichern für Gebäude. Mit „Vehicle to Load“ ist dies bei Elektrofahrzeugen von einigen Herstellern bereits möglich [4]. Hiermit können elektrische Kleingeräte direkt an das Fahrzeug angeschlossen werden, die dann von dem Fahrzeug mit Wechselspannung versorgt werden. Dies ist meist auf niedrige Kilowatt Beträge begrenzt. Bei „Vehicle to Building“ soll das Fahrzeug genauso schnell entladen werden können, wie es auch geladen werden kann. Dafür wird jedoch eine Ladesäule oder ein Ladepunkt benötigt, die für bidirektionales Laden ausgelegt ist [1].
Mit dem Wandel hin zu erneuerbaren Energien spielt das Zwischenspeichern von Energie zum Ausgleich von Produktionsschwankungen eine immer größere Rolle. Durch diese Dezentralisierung der Stromerzeugung muss auch die Speicherung dezentral gelöst werden. Auf einer regionalen Ebene kann bidirektionales Laden, dann „Vehicle to Grid“ genannt, zur Stabilisierung des Stromnetzes beitragen.

Zielsetzung

Ziel der Arbeit ist es bidirektionales Laden als Produkt für Privat- und Gewerbekunden zu realisieren. Damit soll das Produktportfolio von HUMMEL Systemhaus, vor allem im Bereich der Ladeinfrastruktur, erweitert werden. Zuerst muss eine Marktanalyse durchgeführt werden und anhand der Ergebnisse ein Produkt entwickelt werden. Hierbei muss nicht nur die benötigte Hardware betrachtet werden, sondern auch die verschiedenen Kommunikationsmethoden untersucht werden, da eine softwarebasierte Regelung entworfen werden soll. Durch die Analyse der verschiedenen Kommunikationsmethoden soll die Regelung möglichst einfach erweitert werden können und eine Anpassung

auf Ladepunkte und Fahrzeuge verschiedener Hersteller ermöglichen.

Regelung

Für die Regelung muss zuerst die grundsätzliche Kommunikation implementiert werden. Die meisten regelbaren Ladepunkte nutzen OCPP („Open Chargepoint Protocol“), um mit einem Backend zu kommunizieren [3]. Dieses Backend ist zum Beispiel für die Authentifizierung und Abrechnung von Ladevorgängen zuständig. Über das Backend sind jedoch auch Einstellungen an dem Ladepunkt möglich, wodurch man unter anderem auch die Ladeleistung anpassen kann.
Zusätzlich ist es wichtig den Ladestand des angeschlossenen Fahrzeugs zu kennen, um eine gewisse Mindestreichweite zu gewährleisten, die je nach geplanter Strecke variieren kann. Um diese Informationen vom Fahrzeug zu erhalten, gibt es von manchen Herstellern bereits Dienste mittels denen man die Statusinformationen abfragen kann.
Für Privathaushalte mit Photovoltaikanlage ist der Einsatz des Fahrzeugs als Energiespeicher interessant, hierbei wird das Fahrzeug wie bei einem Gebäudespeicher über den Tag mit der überschüssig erzeugten Energie geladen und nachts entladen, um den Haushaltsbedarf zu decken. Dies setzt voraus, dass das Fahrzeug nicht zur Fahrt zur Arbeitsstelle genutzt wird, da dann das Fahrzeug nicht vor Ort ist, wenn die Photovoltaikanlage Strom produziert.

Für Gewerbekunden ist zusätzlich eine Lastspitzenvermeidung interessant. Je nach Anzahl der Elektrofahrzeuge können hier große Lastspitzen abgefedert werden, indem zu Zeiten von hohem Bedarf die Fahrzeuge am Standort entladen werden, um so den Bedarf zu decken und eine niedrigere Spitzenleistung am Netzknoten zu erzielen. Vor allem wenn diese Lastspitzen nur vereinzelt auftreten, können so Kosten gesenkt werden, die ansonsten durch diese Lastspitzen anfallen würden. Hierbei muss vor allem nach dem Netzbezug geregelt werden, der möglichst unter einem gesetzten Limit verbleiben sollte. Dieses Limit muss je nach Standort individuell ermittelt werden.

Ausblick

Da bidirektionales Laden noch im Entwicklungsstadium ist, gibt es noch sehr wenige Hersteller die hierzu Produkte anbieten. Es gibt bereits einige bidirektionale Ladelösungen mit Gleichstrom Ladetechnik und „Vehicle to Load“ Lösungen, bei denen direkt im Fahrzeug Steckdosen integriert sind, an die man elektrische Kleingeräte anschließen kann. Im Bereich „Vehicle to Grid“ sind bereits einige Forschungsprojekte gestartet, deren beinhaltete Produkte auch für „Vehicle to Building“ nutzbar sein werden [5]. Dies umfasst die Technik der Elektrofahrzeuge und die dazugehörige Ladetechnik. Da noch nicht sicher ist welche Hardware genutzt werden kann, wird eine allgemeine Regelung erarbeitet, die einfach an verschiedene Hardware angepasst werden kann.

Literatur und Abbildungen

[1] Audi AG. Electric cars as part of the energy transition: Audi is researching bidirectional charging technology. https://www.audi-mediacenter.com/en/press-releases/electric-cars-as-part-of-the-energy-transitionaudi-is-researching-bidirectional-charging-technology-12996, 07 2020.

[2] ENNAGY by HUMMEL Systemhaus. ENNAGY - The Co-Energy Community. https://www.enna.gy/, 2021.

[3] Open Charge Alliance. Participants - Open Charge Alliance. https://www.openchargealliance.org/participants/, 2022.

[4] Hyundai Deutschland. IONIQ 5 Charging | Eco - Hyundai Worldwide. https://www.hyundai.com/worldwide/en/eco/ioniq5/charging, 2022.

[5] Christiane Köllner. Wann kommt bidirektionales Laden von E-Autos? https://www.springerprofessional.de/ladeinfrastruktur/elektrofahrzeuge/wann-kommt-bidirektionales-laden-von-e-autos-/18217570, 04 2022.

Evaluierung und Implementierung eines 3D-Objektverfolgungssystems im Kontext von mobilen Anwendungen

Mick Zuelch **Markus Enzweiler**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma TeamViewer Germany GmbH, Göppingen

Einleitung

Augmented Reality ist eine Technologie welche in der Industrie einen stetigen Zuwachs verzeichnet. Der weltweite Marktumsatz ist von knapp 4 Milliarden USD im Jahr 2019 bereits auf 9,53 Milliarden USD im Jahr 2021 angestiegen [2]. Im gleichen Zeitraum hat sich die Zahl der Nutzer von Augmented Reality fast verdoppelt. Zudem wird die Anzahl der Nutzer im Jahr 2025 bereits auf 1,7 Milliarden geschätzt [1]. Die Einsatzmöglichkeiten von Augmented Reality sind dabei sehr flexibel. Egal ob dabei eine Maschine oder ein Motor virtuell angezeigt wird oder eine Remotesupport Sitzung durchgeführt wird. Durch die rasante Verbreitung der Technologie haben Unternehmen zudem ein großes Interesse die immer größer werdende Nachfrage nach Augmented Reality zu decken. In Verbindung mit Künstlicher Intelligenz und Objekterkennung sind die Einsatzmöglichkeiten von Augmented Reality in der Industrie nahe zu grenzenlos.

Problemstellung

TeamViewer hat mit TeamViewer Assist AR bereits eine Remotesupport Lösung im Produktportfolio. Dabei können zum Beispiel Instruktionen, mit Hilfe von Augmented Reality Objekten, auf dem Handy eines Nutzers angezeigt werden. Eine fachkundige Person hilft dem Nutzer dabei ein Problem zu lösen oder beispielsweise eine Maschine Instand zu halten. Bei den Augmented Reality Objekten handelt es sich um nummerierte Pfeile und Handzeichnungen, die während der Remotesitzung platziert werden können. Dabei wird jedoch immer eine fachkundige Person benötigt, welche möglicherweise nicht direkt zur Verfügung steht. Daher soll mithilfe von Maschinellem Lernen, Computer Vision und Augmented Reality eine Lösung geschaffen werden, welche die Wartung und Nutzung von Maschinen und Gegenständen unabhängig von einer fachkundigen Person zulässt.

Ziel der Arbeit

Das Ziel der Arbeit ist es eine Android Applikation zu entwickeln. Diese soll in der Lage sein vordefinierte Objekte im Raum zu finden. Hinzu kommt, dass die Applikation den Zustand, in welchem sich die vordefinierten Objekte befinden, erkennen soll. Mithilfe der Informationen wo sich ein Objekt befindet, um welches es sich handelt und in welchem Zustand es sich befindet, sollen dem Benutzer gezielte Informationen und Hinweise über das erkannte Objekt angezeigt werden. Die Informationen sollen dabei über Augmented Reality bei dem Objekt platziert werden. Diese Technologie kann zum Beispiel genutzt werden um schnell und effizient Informationen und Instruktionen für die Bedienung von Maschinen aufzuzeigen. Dabei befasst sich die Arbeit mit den Themen Objekterkennung, Objektzustandserkennung und Augmented Reality. Zur Demonstration und Evaluation soll ein komplettes Augmented Reality Tutorial erstellt werden, welches einem Nutzer Hinweise zur Kaffeezubereitung mit einer French-Press gibt.

Objekterkennung und Objektzustandserkennung

Die Objekterkennung und Objektzustandserkennung sind zwei Probleme, welche in der Thesis gelöst werden müssen. In der Arbeit sollen dabei verschiedene Methoden zur Objekterkennung und Objektzustandserkennung, mit Hilfe von Maschinellem Lernen, Computer Vision oder einer Kombination aus beidem implementiert und evaluiert werden. Das Maschinelle Lernen ist eine der Möglichkeiten, welche im Folgenden genauer erläutert wird. Hierfür wurde ein You only look once (Yolo) Objekterkennungssystem trainiert [4], um verschiedene Objekte und deren Zustand zu erkennen.

Abb. 1: Objekterkennung einer Tasse [3]

Bei einer Tasse, wie in Abbildung 1 zu sehen, wird zum Beispiel zwischen den Zuständen leer oder voll unterschieden. Dabei wurden im Datensatz verschiedene Bilder von vollen und leeren Tassen integriert und manuell gelabelt. Neben der Tasse können noch vier weitere Objekte inklusive verschiedener Zustände erkannt werden. Daraus ergeben sich insgesamt 12 Klassen auf welche das Yolo Netz trainiert wurde.

Augmented Reality

Durch die Objekterkennung ist es möglich ein Objekt in dem 2D Bild der Umgebung zu markieren. Jedoch fehlt hier die Information an welcher Stelle sich das Objekt im Raum befindet. Hier kommt vor allem Google AR Core zum Einsatz. Zunächst wird mit AR Core ein 3D Raum, durch die Kamera und die Sensoren eines Android Geräts , erzeugt. Hier beginnt nun ein enges Zusammenspiel aus den Erkenntnissen der Objekterkennung und AR Core. Durch die Objekterkennung sind die exakten 2D Koordinaten eines Objektes auf dem Bildschirm bekannt. Durch den bereits existierenden 3D Raum kann die 2D Bildschirmkoordinate in eine 3D Koordinate im Raum umgerechnet werden.

Abb. 2: Augmented Reality Objekt mit Informationen über die Tasse [3]

An dieser Koordinate kann nun, wie in Abbildung 2 zu sehen ist, ein beliebiges Objekt platziert werden. Da durch die Objekterkennung der Zustand des Objekts bekannt ist, können gezielte Informationen für den Nutzer platziert werden. Somit kann ein Nutzer mit gezielten Informationen ein Objekt bedienen, ohne von einer Remoteverbindung abhängig zu sein.

Aktueller Stand und Ausblick

Zum Zeitpunkt der Veröffentlichung dieses Artikels existiert bereits eine Demo Applikation, welche mit Hilfe von einem Yolo Netz Objekte und deren Zustand erkennen kann. Hilfestellungen und Informationen für die Nutzer können ebenfalls im 3D Raum dargestellt werden. Somit lassen sich bereits komplette Workflows umsetzen. Im weiteren Verlauf wird an verschiedene Methoden der Objekterkennung, zum Beispiel über OpenCV, geforscht. Außerdem werden die verschiedenden Varianten untereinander verglichen und evaluiert.

Literatur und Abbildungen

[1] Thomas Alsop. Global mobile augmented reality (AR) users 2019-2024. https://www.statista.com/statistics/1098630/global-mobile-augmented-reality-ar-users/, 11 2021.

[2] Thomas Alsop. Mobile augmented reality (AR) market revenue worldwide from 2019 to 2025. https://www.statista.com/statistics/282453/mobile-augmented-reality-market-size/, 03 2022.

[3] Eigene Darstellung.

[4] Jacob Solawetz, Joseph Nelson, and Samrat Sahoo. How to Train YOLOv4 on a Custom Dataset. https://blog.roboflow.com/training-yolov4-on-a-custom-dataset/, 05 2021.

Konzeptionierung eines Privileged Access Workstation Virtualisierungshosts

Friedemann Zurhorst **Tobias Heer**

Fakultät Informatik und Informationstechnik, Hochschule Esslingen

Arbeit durchgeführt bei der Firma ERNW Enno Rey Netzwerke GmbH, Heidelberg

I. Motivation und Problemstellung

In der heutigen Bedrohungslandschaft sind Sicherheitslücken als auch Angriffe fast schon allgegenwärtig geworden. Demnach ist es eine Herausforderung, eine sichere Administrationsweise des Active Directories zu entwerfen und zu implementieren. Teil der Active Directory Sicherheitsstrategie ist das Admin-Tiering-Konzept von Microsoft, in dem die vorhandenen Systeme in Tiers beziehungsweise Sicherheitszonen kategorisiert werden. Hierbei sind Systeme, die die volle Kontrolle über die Umgebung ermöglichen in Tier 0, der sichersten Zone eingeordnet. Anschließend folgen Anwendungsserver in Tier 1, und unsichere Workstations in Tier 2.
Zweck des Admin-Tierings ist es, die verschiedenen Zonen hinsichtlich der administrativen Credentials zu isolieren, so dass diese nicht Tier-übergreifend verwendet werden könnten. Dies erfolgt durch die Definition von eigenen administrativen Konten in jedem Tier. Das Admin-Tiering-Konzept wird durch das sog. Clean-Source-Prinzip [1] ergänzt. Dieses besagt, dass alle Sicherheitsabhängigkeiten genauso vertrauenswürdig sein müssen, wie das zu sichernde Objekt. Das bedeutet, dass wenn eine Workstation zur Administration eines Servers genutzt wird, diese genauso sicher sein muss wie der Server selbst.
Im Idealfall hat jede Person, die systemkritische Aufgaben durchführt, so viele separate Accounts und Workstations, wie sie Tiers administriert. Die Tier 2 Workstation hat keine besondere Anforderung hat. Sie dient der alltäglichen Arbeit, und wird auch „Office Workstation“ genannt. Die Nutzung mehrerer Geräte stellt enorme Ansprüche an Material-, und Prozesswirtschaft als auch die technische Implementierung.
Alternativ zu mehreren separaten Geräten, gibt es die Möglichkeit zur Nutzung einer oder mehrerer virtueller Maschinen auf einer speziell gehärteten Workstation. Hierbei übernimmt der Hypervisor die Trennung, sodass die virtuellen Maschinen nicht untereinander kommunizieren können. Hier bestehen verschiedene Nutzungsansätze, wie zum Beispiel die Nutzung des Hosts als Administrationssystems und einer extra Office Workstation VM. Auch hier ergeben sich verschiedene Faktoren, die die bei der Umsetzung beachtet werden müssen. Im speziellen:

- Netzwerkanbindung des Hosts beziehungsweise der Admin VM an das Firmennetz
- Sichere Konfiguration des Internetzugriffs (z.B. ausgehende Filterung) des Hosts und der VMs
- Durchreichen von Peripherie an einzelne VMs
- Sichere Konfiguration der Remote Arbeit beziehungsweise Homeoffice
- Welche Konten werden zur Anmeldung an welchem System/welcher VM verwendet, und welche Berechtigungen dürfen oder müssen diese haben?
- Spezielle Härtung der verschiedenen Systeme

II. Design

Die genannten Punkte aus I. sind nicht trivial zu lösen. Die Netzwerk- und Internetkonnektivität stellt eine Herausforderung dar. In einer sicheren Umgebung werden Administrations- und Produktionsnetzwerk getrennt. Eine Kommunikation zwischen diesen Netzen muss sehr präzise definiert sein. Hier stellt sich die konzeptionelle Frage der Anbindung. Der Host, oder auch die Admin VM, muss sich in das sichere Administrationsnetzwerk verbinden. Die Office VM jedoch benötigt diesen Zugang nicht, sondern lediglich Zugriff auf das Produktionsnetzwerk. Hier werden sowohl technische als auch organisatorische Fragen zur konkreten Umsetzung und sicheren Konfiguration der geklärt werden müssen.
Ebenfalls Teil der Netzwerkkonfiguration ist die Anbindung an das Internet. Die zunehmende Nutzung von Clouddiensten und die damit verbundene Verwendung

von hybriden Active Directory-Umgebungen erfordert ebenfalls einen sicheren administrativen Zugang zu diesen Diensten. Folglich muss die Admin-Workstation ebenfalls sicher an das Internet angebunden werden. An diesem Punkt muss etwa geklärt werden, auf welche Art und Weise dies konfiguriert wird (z.B. welche Dienste in der Cloud über welche Protokolle und mit welchen Identitäten erreicht werden dürfen). Die Nutzung von (Hardware-)Peripherie in einer virtualisierten Umgebung bringt eine ganze Reihe von Herausforderungen mit sich, da Geräte an die VMs durchgereicht und dort (z.B. bei Ton und Bild verzögerungsfrei) funktionieren müssen. Dies muss sowohl vom Hypervisor als auch teilweise von den Geräten unterstützt werden. In der modernen Arbeitsweilt ist es bei weitem nicht mehr ausreichend lediglich Maus und Tastatur zur benutzen. Durch steigenden Bedarf an unterschiedlichen Kommunikationsformen und durch die Anforderung, vom Homeoffice „nahtlos“ arbeiten zu können, ist beispielsweise auch die Nutzung von Mikrofon, kabellosen Kopfhörern, und Webcams für Online-Meetings nötig. Insbesondere die spezielle Peripherie wie Hardware-Tokens (z.B. in der Form von Smartcards und Smartcard-Lesegeräten), Druckern oder anderen Geräten müssen nicht nur der Office VM bereitgestellt werden. Der tatsächliche Implementierungsaufwand ist davon abhängig, welche konkrete Peripherie existiert und genutzt wird.
All diese Fragen müssen mit dem Kunden besprochen werden. Dabei müssen insbesondere die bisherigen Arbeits- und Administrationsprozesse berücksichtigt werden.
Was die Homeoffice-Tätigkeit betrifft, muss etwa geklärt werden, wie und von wo eine VPN-Einwahl stattfindet, ob es für die verschiedenen VMs auch verschiedene VPN-Tunnel gibt, wie jeder der Zugänge gesichert wird, welche Protokolle gesprochen werden dürfen usw.
Da der Virtualisierungshost eine zentrale Rolle hinsichtlich der Sicherheit der VMs einnimmt, muss dieser gegen Angriffe speziell abgesichert beziehungsweise gehärtet werden. Eine Infektion des Hosts hätte eine Kompromittierung der Admin-Workstation und somit der gesamten zu verwaltenden Umgebung zur Folge. Es müssen Maßnahmen definiert werden, die die Sicherheit maßgeblich erhöhen und Risiken minimieren. Dazu zählen beispielsweise die Regelung des Zugriffs auf Systemressourcen, das Abschalten von Legacy-Protokollen, Definition von Logging-Policies uvm.

III. Evaluation

Aufgrund der individuellen Konfiguration der PAWs für das jeweilige Unternehmen, ist ein allgemeingültiger Ansatz zur Messung des Erfolges der Umsetzung schwierig. Pauschale Aussagen lassen sich schwer bis gar nicht treffen. Um so wichtiger ist es, zu analysieren ob und wie gut die konzeptionierte Lösung für den Kunden ist. Hierfür muss am Anfang des Projekts eine Anforderungsanalyse in Zusammenarbeit mit dem Kunden durchgeführt werden. Diese stellt klar, was gefordert und gewünscht, aber auch was technisch möglich ist.
Auf der Basis der Anforderungsanalyse werden Maßnahmen beziehungsweise technische Konfigurationen sowie organisatorische Prozesse abgeleitet. Anhand dessen lässt sich später messen, ob die Anforderung erfüllt wurden. Dies wird der grundlegende Pfeiler der Qualitätssicherung und Evaluation sein, und ist erst beim Abschluss des Projektes möglich. Weiterhin kann eine Risikoabschätzung der einzelnen technischen und organisatorischen Maßnahmen durchgeführt werden. Dadurch kann die Erhöhung der Sicherheit durch die Einführung einer solchen PAW bewertet werden.

IV. Verwandte Arbeiten

Das Admin-Tiering-Konzept ist der Defacto-Standard für On-Premises Active Directory-Umgebungen. Andere Verzeichnisdienst- und Identity-Softwarelösungen werden hier nicht betrachtet.
Was die PAW betrifft, so gibt es mehre Möglichkeiten eine PAW zu konzeptionieren (separates Gerät, virtualisiert usw.). Zwar werden Jumphosts oder Privileged Access Management Lösungen wie CyberArk manchmal als Alternativen betrachtet, sie sind aber keine wirkliche Alternative zu den Forderungen des Clean-Source-Prinzips. Wenn es darum geht, konkrete Maßnahmen beziehungsweise Konfigurationsmöglichkeiten für PAWs in öffentlich zugänglichen Informationsquellen zu finden, sind die Ergebnisse rar. Hintergrund davon ist, dass PAW Konzepte meist durch externe Dienstleister erarbeitet werden, und jedes Konzept individuell auf die jeweilige Firma zugeschnitten ist. Sowohl der Dienstleister als auch die Firma haben kein Interesse daran, ihr Know-How kostenlos zur Verfügung zu stellen.
Seitens Microsofts gibt es aber eine High-Level Übersicht über die Beschreibung [3] und grundlegender Konfiguration einer PAW [4]. Weitere Übergeordnete Informationen zur Privileged Access Strategie lassen sich im entsprechenden Whitepaper finden [2].

V. Ergebnis

Durch den Entwurf und Implementation dieser Lösung, soll eine für den Kunden zugeschnittene, sichere PAW-basierte Administrationslösung bereitgestellt werden. Diese erfüllt die heutigen Sicherheitsanforderungen, um sowohl aktuelle als auch zukünftige Bedrohungen abzuwehren. Weiterhin soll durch diese Arbeit auch Lösungen für die typischen Probleme einer virtualisierten PAW aufgezeigt werden.

Literatur und Abbildungen

[1] John Flores. Clean Source Principle. https://docs.microsoft.com/en-us/security/compass/privileged-access-success-criteria#clean-source-principle, 2021.

[2] John Flores and J.P. Delauri. Securing Privileged Access. https://docs.microsoft.com/en-us/security/compass/overview, 2022.

[3] Jian Yan. Privileged Access Workstation (PAW). https://docs.microsoft.com/en-us/archive/blogs/datacentersecurity/privileged-access-workstationpaw, 2017.

[4] Jian Yan. PAW deployment guide. https://docs.microsoft.com/en-us/archive/blogs/datacentersecurity/paw-deployment-guide, 2018.