

Editorial

Liebe Leserinnen und Leser,

vor Ihnen liegt nunmehr die bereits zehnte Ausgabe des E-Journals **Anwendungen und Konzepte in der Wirtschaftsinformatik (AKWI)**.

Wie in den vergangenen Ausgaben auch behandeln die Beiträge die Kernthemen der Wirtschaftsinformatik. Auch diese Ausgabe besticht durch ihre inhaltliche Breite.

Besonders freut uns Herausgeber, Ihnen verschiedene Beiträge aus kooperativen Promotionen zwischen Hochschulen und Universitäten im In- und Ausland präsentieren zu können. Für die Promovenden sind Publikationen wichtig und so erlauben wir eine gewisse Zweitverwertung erfolgreicher Publikationen, die inhaltlich zu den Zielsetzungen dieser Zeitschrift passen. Konkret wurden sehr ähnliche Beiträge in dem Tagungsband zur 33th European Conference on Modeling and Simulation (11. – 14. Juni 2019 in Caserta in Italien) publiziert. Inhaltlich passende weitere Beiträge aus diesem Tagungsband wurden zur Abrundung in diese Ausgabe aufgenommen. Weitere Promovenden mögen sich durch diese Beiträge zur Einreichung ermuntert fühlen. Den geneigten Lesern möge es Einblicke in die Forschungskraft an Hochschulen geben.

Eine weitere Neuerung betrifft die Aufnahme von Zusammenfassungen von hervorragenden Abschlussarbeiten. Damit soll ein generelles Ziel unserer Zeitschrift, nämlich das Aufzeigen der gesamten Breite der Wirtschaftsinformatik an Hochschulen, unterstützt werden. Diese Neuerung wurde erst kurzfristig ermöglicht, weswegen nur wenige Beiträge vorliegen. Mögen diese Beispiele zur Einreichung zahlreicher Zusammenfassungen für die nächsten Ausgaben ermuntern.

Über Ihr Interesse an der Zeitschrift freuen wir uns und wünschen Ihnen Freude bei der Lektüre.

Regensburg, Fulda, Wildau und Luzern, im Dezember 2019

Frank Herrmann, Norbert Ketterer, Konrad Marfurt und Christian Müller

In eigener Sache:

Alle Beiträge wurden von zwei unabhängigen Gutachtern begutachtet und von den Autoren anschließend überarbeitet. Dieser Prozess nimmt naturgemäß viel Zeit in Anspruch, da sämtliche Redakteure, Gutachterinnen und Gutachter ihre Arbeit in der immer spärlicher werdenden Freizeit leisten. Dafür gebührt ihnen unser besonderer Dank!

Unsere Zeitschrift steht kostenfrei unter <http://akwi.hswlu.ch> als E-Journal an der Hochschule Luzern unter Federführung von Konrad Marfurt zur Verfügung. Für unsere Autoren entstehen durch die Publikation keine Kosten, allerdings erhalten sie auch keine Honorare. Damit wir unter diesen Bedingungen erfolgreich arbeiten können, reichen die Autoren druckfertige Manuskripte, die unserer Formatvorlage genügen, in deutscher oder englischer Sprache ein, ordnen diesen einer der Rubriken Grundlagen, Trends, Praxis, Kurz erklärt, Buchbesprechung oder Abschlussarbeit zu. Ferner bitten wir unsere Autoren um eine Einverständniserklärung zur Publikation und eine Selbsteinschätzung. Nach der Einreichung beginnt sofort unser Begutachtungsprozess.



Christian Müller



Konrad Marfurt



Norbert Ketterer



Frank Herrmann

DESIGN AND SIMULATION OF BIOTECHNICAL MULTIDIMENSIONAL MOTION CONTROL SYSTEMS OF A ROBOT MANIPULATOR

Ekaterina Rostova
St.Petersburg Institute for
Informatics and Automation of
the Russian Academy of Sciences
14th line 39, St.Petersburg, 199178,
Russia
rostovae@mail.ru

Vladislav Sobolevsky
St.Petersburg Institute for
Informatics and Automation of
the Russian Academy of Sciences
14th line 39, St.Petersburg, 199178,
Russia
Arguzd@yandex.ru

Nikolay Rostov
Peter the Great St. Petersburg
Polytechnic University
Polytechnicheskaya Str. 29,
St.Petersburg, 195251,
Russia
rostovnv@mail.ru

Valerii Zakharov
St.Petersburg Institute for
Informatics and Automation of
the Russian Academy of Sciences
14th line 39, St.Petersburg, 199178,
Russia
Valeriov@yandex.ru

KEYWORDS

Biotechnical Systems, Human Operator, Robot-Manipulator, Motion Control, Design of Control Algorithms, Robot Velocity Control, Robot Position Control, Robot Force Control, Computer Simulation.

ABSTRACT

The design of multidimensional velocity, position and force control algorithms for a semi-automatic motion control system of a robot-manipulator has been performed. Mathematical and computer models of a motion control system with reconfigurable structure have been developed for a 3-link robot performing standard operations set by a human operator with a 3-degree-of-freedom handle.

Simulation of dynamic processes in a system with different control algorithms has been performed and their precision characteristics have been evaluated. The achieved results may be of interest for designers who develop training simulators for human supervisors of robot motion control systems.

INTRODUCTION

A human operator of biotechnical semi-automatic control systems performs remote motion control of a robot using handles in the form of joysticks with multi-degree of freedom and special calculator units. Control systems of this kind, in which an operator is considered to be a part of a control system, are widely used in high-risk work environments - in underwater vehicles, at space stations, etc. (Golovin 2011; Petukhov 2012; Soares 2008). These systems can be applied for real-time robot motion control and installed in training simulators for human operators (for example, astronauts, surgeons, etc.). The development of effective training

simulators for human operators is an important practical issue (Strakhnov et al. 2016; Denisov 2016; Alferov et al. 2006).

In semi-automatic systems different algorithms of motion control are implemented depending on a type of technological operations performed by robots (Filaretov 2011, 2013). For transportation operations that do not require high accuracy, the gripper velocity control is usually used, while for precise point-to-point and continuous-path control it is necessary to use the gripper position control. When contact operations are performed (e.g. assembling) the control of force and torque in a robot gripper is required (Arkhipov et al. 2017).

In this paper we consider a semi-automatic motion control system for a robot-manipulator in which different control algorithms are combined. The actuators of robot links are local servosystems with digital PID regulators of motor speeds, positions and torques, which depends on an operation mode.

The purpose of the work is the structural design of motion control systems with different types of algorithms and the investigation of its dynamics in typical operation modes of a robot. The main problems of the research are the following:

1. The structural design of algorithms for open-loop and closed-loop velocity, position and force control.
2. Building computer models for the analysis of dynamic processes in control systems under consideration in standard operation modes.
3. Simulation of the control systems for a robot performing the following standard operations:
 - (1) - Transportation of the gripper to a specified area in the velocity control mode;
 - (2) - Point-to-point positioning of the gripper into a given point of a trajectory;
 - (3) - Precise movement of the gripper along a trajectory in continuous-path control mode;
 - (4) - Applying specified force to a fixed contact point.

4. The evaluation of dynamic errors that occur when a robot performs transportation, positioning, continuous-path motion and force operation.

It is believed to be acceptable to use simplified computer models of a robot and joint actuators for the comparison of motion control systems with different open-loop and closed-loop structures. So, for the solution of the problems stated above only kinematic models of a robot were used, without the consideration of its nonlinear dynamics. The drives were simulated by linear dynamic models of the 2nd and 3rd order. The animations of links of the handle and the robot were created with the use of functions from MATLAB Robotics Toolbox (Corke 2012, 2017)).

ALGORITHMS OF OPEN-LOOP ROBOT MOTION CONTROL

Human-handle interface of semi-automatic motion control systems of a robot-manipulator includes the solution of the following problems:

1) The forward kinematics problem related to the handle position:

$$S_h = F_h(q_h) = (X_h, Y_h, Z_h)^T \quad (1)$$

where q_h – the vector of the handle joint positions, $F_h(q_h)$ – the vector-function corresponding to the handle kinematics scheme. The calculated coordinates X_h, Y_h, Z_h are interpreted as the vector of velocity, position or force coordinates set by a human operator.

2) In corresponding control mode the programmed velocities, positions and forces of the robot gripper are scaled as follows:

$$V_p = (V_{xp}, V_{yp}, V_{zp})^T = (M_{vx}X_h, M_{vy}Y_h, M_{vz}Z_h)^T \quad (2)$$

$$S_p = (X_p, Y_p, Z_p)^T = (M_xX_h, M_yY_h, M_zZ_h)^T \quad (3)$$

$$F_p = (F_{xp}, F_{yp}, F_{zp})^T = (M_{fx}X_h, M_{fy}Y_h, M_{fz}Z_h)^T \quad (4)$$

where M_* - scale gains.

The algorithm of open-loop robot velocity control is based on the solution of the inverse kinematics problem related to the programmed robot joint velocities:

$$\dot{q}_p = J_M^{-1}(q)V_p \quad (5)$$

where \dot{q}_p – the vector of the robot joint velocities, q – the vector of the robot joint positions, $J_M^{-1}(q)$ – the matrix inverted to the Jacobian matrix of the robot-manipulator.

The algorithm of open-loop robot position control is based on the solution of the inverse kinematics problem related to the programmed joint positions:

$$q_p = F_M^{-1}(S_p) \quad (6)$$

where $F_M^{-1}(S_p)$ – the vector-function inverted to the vector-function $F_M(q_p)$ of the robot.

And the algorithm of open-loop robot force control is based on the solution of the inverse kinematics problem related to the programmed joint torques:

$$Q_p = J_M^T(q)F_p \quad (7)$$

where Q_p – the vector of the robot joint torques, q – the vector of the robot joint positions, $J_M^T(q)$ – the matrix transposed to the Jacobian matrix of the robot.

ALGORITHMS OF CLOSED-LOOP ROBOT MOTION CONTROL

Closed-loop velocity control of a robot gripper

The algorithm of closed-loop velocity control includes the sequential solution of the following problems:

1) The forward kinematics problem related to the robot gripper velocity, which is solved in the real time mode:

$$V_r = J_M(q) \cdot \dot{q} = (V_x, V_y, V_z)^T \quad (8)$$

where \dot{q} - the vector of the robot joint velocities; $J_M(q)$ - the Jacobian matrix of the robot.

2) Calculation of the error vector for gripper velocity:

$$dV = V_p - V_r \quad (9)$$

3) The inverse kinematics problem related to the joint velocities:

$$G = J_M^{-1}(q) \cdot U_V \quad (10)$$

where $J_M^{-1}(q)$ – the matrix inverted to the Jacobian matrix of the manipulator, $U_V = (u_{vx}, u_{vy}, u_{vz})^T$ – the output vector of multivariable PID velocity controller, $G = (g_1, g_2, g_3)^T$ - the input vector of velocity drives.

Closed-loop position control of a robot gripper

The algorithm of closed-loop position control includes solution of the following problems:

1) The forward kinematics problem related to the robot gripper position, which is solved in the real time mode:

$$S_r = F_M(q_r) = (X_r, Y_r, Z_r)^T \quad (11)$$

where q_r – the vector of the robot joint positions, $F_M(q_r)$ – the vector-function corresponding to the robot kinematics scheme.

2) The error vector calculation for gripper position:

$$dS = S_p - S_r \quad (12)$$

3) The inverse kinematics problem related to the programmed robot joint velocities:

$$\dot{q}_p = J_M^{-1}(q)U_s \quad (13)$$

where \dot{q}_p – the programmed vector of the robot joint velocities, $J_M^{-1}(q)$ – the matrix inverted to the Jacobian matrix of the manipulator, $U_s = (u_x, u_y, u_z)^T$ – the output vector of multivariable PID position controller outputs.

Closed-loop force control of a robot gripper

The algorithm of closed-loop force control includes the sequential solution of the following problems:

1) The forward kinematics problem related to the robot gripper force:

$$F_r = (J_M^T(q))^{-1}Q \quad (14)$$

where q – the vector of the robot joint positions, $J_M^T(q)$ – the matrix transposed to the Jacobian matrix of the robot, Q – the vector of the robot joint torques.

2) Calculation of the error vector for gripper force:

$$dF = F_p - F_r \quad (15)$$

3) The inverse kinematics problem related to the programmed robot joint torques:

$$Q_p = J_M^T(q)U_f \quad (16)$$

where $J_M^T(q)$ – the matrix transposed to the Jacobian matrix of the robot, $U_f = (u_{fx}, u_{fy}, u_{fz})^T$ – the output vector of multivariable PID force controller.

COMPUTER MODELS OF CLOSED-LOOP ROBOT MOTION CONTROL SYSTEMS

Figure 1 shows the Simulink model of the closed-loop velocity control system with the vectorial feedback. The model includes the following blocks:

Block *Human-Handle Interface* calculates expression (1) and expression (2);

Block *fkine_V* in the feedback solves the forward kinematics problem related to the gripper velocity using expression (8);

Block *PID-V* is the multivariable PID controller of the gripper velocity coordinates.

Block *ikine_V* calculates the input actions for the robot joint velocity drives using expression (10).

Block *Velocity Drives* contains the models of velocity drives of the robot links.

Block *fkine_P* solves the forward kinematics problem related to the gripper position using expression (11).

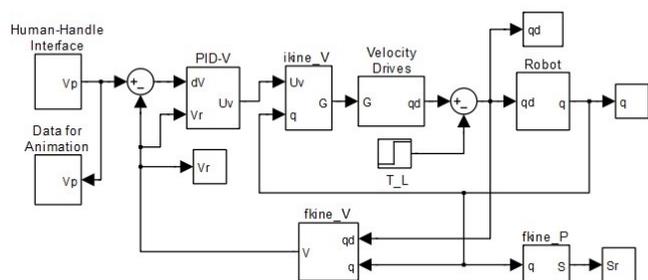


Figure 1: The model of closed-loop velocity control system

Figure 2 shows the Simulink model of the closed-loop position control system with the vectorial feedback. The model includes the following blocks:

Block *Human-Handle Interface* calculates expression (1) and expression (3).

Block *fkine_P* in the feedback solves the forward kinematics problem related to the gripper position using expression (11);

Block *PID-S* is the multivariable PID position controller.

In the closed loop block *ikine_V* calculates the input actions for the robot joint velocity drives.

The second block *ikine_V* calculates programmed gripper velocities using expression (6) in order to compensate for

velocity errors when the robot operates in continuous-path control mode.

Block *Kk* is the vector of parameters of velocity error compensators.

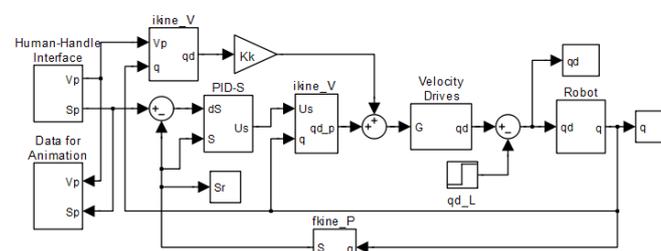


Figure 2: The model of closed-loop position control system with velocity drives

Figure 3 shows the Simulink model of the closed-loop force control system with the vectorial feedback. The model contains the following blocks:

Block *Human-Handle Interface* calculates expression (1) and expression (4).

Block *fkine_F* in the feedback solves the forward kinematics problem related to the gripper force using expression (14).

Block *PID-F* is the multivariable PID controller of the gripper force coordinates.

Block *ikine_F* calculates the input actions for the torque drives of the robot links using expression (16).

Block *Torque Drives* contains the models of the robot link torque drives.

The switch block in the model in Figure 3 allows to analyse the dynamic processes in the system for two operations: (1) – for the contact force operation and (2) – for the motion force operation.

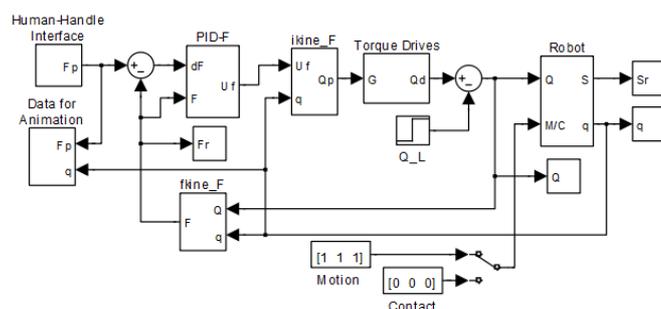


Figure 3: The model of closed-loop force control system with torque drives

Blocks *T_L*, *qd_L* and *Q_L* in the models above represent external loads (disturbances).

ANALYSIS OF DYNAMIC PROCESSES IN ROBOT MOTION CONTROL SYSTEMS

The results of simulation and dynamic analysis of open-loop position and velocity control systems for a robot which performs a transportation operation, point-to-point positioning and continuous-path motion were presented in (Rostova et al. 2018). And the results of investigation of closed-loop velocity, position and force control systems are represented below.

The processes in a control system when the robot performs a transportation operation

When the robot performs some transportation in the velocity control mode a human operator quickly turns the handle keeping it in this position and then quickly returns it to the initial position. In this mode the trajectories of handle links have trapezoidal shape.

Figure 4 represents the animations of the gripper and the robot when the gripper is moving in a required direction.

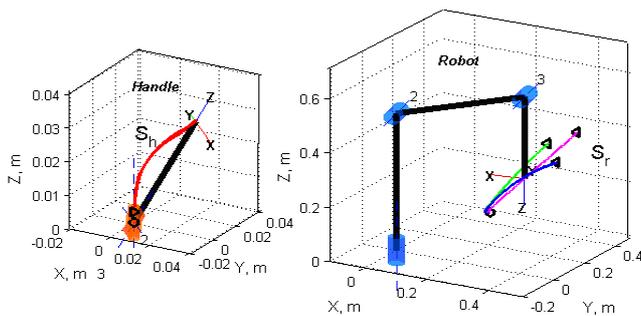


Figure 4: The handle and the robot animations for velocity control mode

The blue and green curves correspond to the open-loop velocity control when the Jacobian matrix in (5) is calculated using the programmed or real robot joint positions, correspondently. Both curves significantly differ from the required straight-line motion of the gripper and show a rather big contour error.

The pink curve that corresponds to the closed-loop velocity control is straight-line and depicts a small contour error. Figure 5 shows the gripper velocities for this case where the programmed coordinates are represented by a dash line (p) and the real coordinates - by a solid line (r).

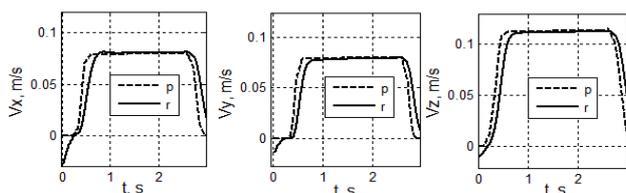


Figure 5: The gripper velocities for closed-loop velocity control

The gripper velocity coordinates have rather big errors at the beginning of the motion because of the influence of external disturbances on the load torques in the drives imitated in the block T_L .

The processes in a control system for the case of point-to-point motion of the robot gripper

In this control mode a human operator smoothly turns the handle from the initial position to the required position. The handle link trajectories are calculated by the polynomials of the 5th order.

Figure 6 shows the animations of the gripper and robot when the gripper is moving to a required position.

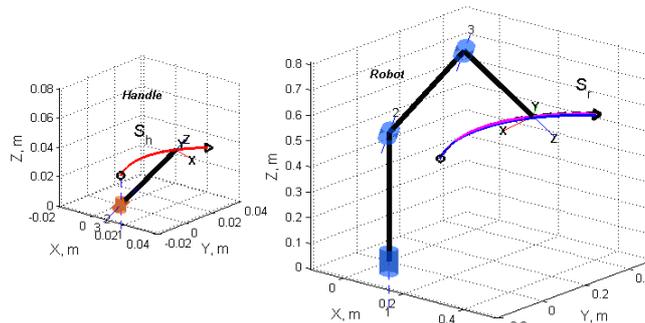


Figure 6: The handle and the robot animations for point-to-point control mode

Figure 7 represents the gripper position coordinates for a closed-loop position control system where programmed coordinates are shown with a dash line (p) and the real coordinates – with a solid line (r). As it can be seen the closed-loop position control system provides the gripper position with smaller dynamic errors. And the final position of the gripper is reached without error.

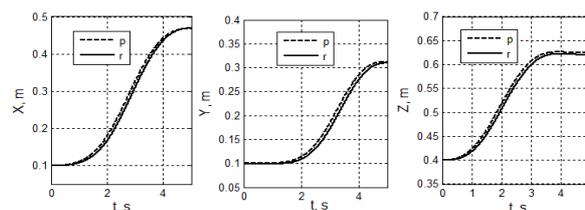


Figure 7: The gripper coordinates for point-to-point motion

The processes in a control system for the case of continuous-path motion of the robot gripper

In the continuous-path mode a skillful human operator moves the handle along the required trajectory of a complex shape, e.g. a helical line. Figure 8 illustrates the animations for the situation when the robot gripper should move along a helical line. The blue curve corresponds to the open-loop position control and the pink curve – to the closed-loop position control.

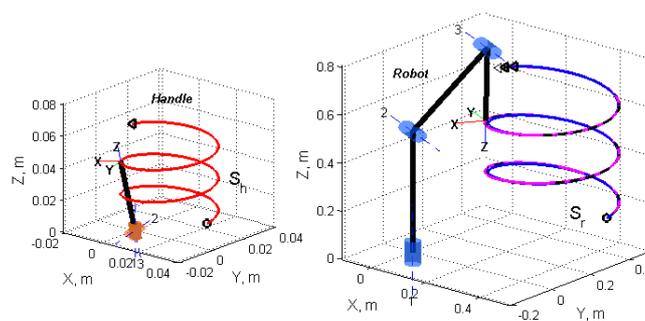


Figure 8: The handle and the robot animations for continuous-path control mode

Figure 9 represents the curves of the gripper positions for the closed-loop position control system where programmed coordinates are shown with a dash line (p) and the real coordinates – with a solid line (r).

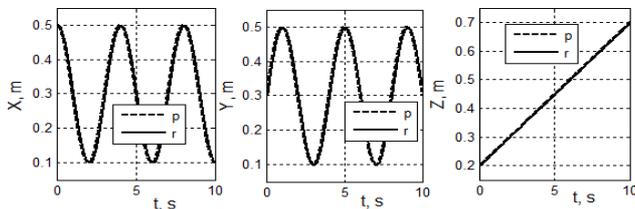


Figure 9: The gripper coordinates for continuous-path motion

The processes in a control system when the robot performs a contact operation

Figure 10 illustrates the animations when the robot performs a contact operation and applies the required force in the given gripper position without motion of the gripper. The vector of programmed force in Figure 10 is shown with red color. The green and pink color vectors correspond to real force for open-loop and closed-loop control respectively.

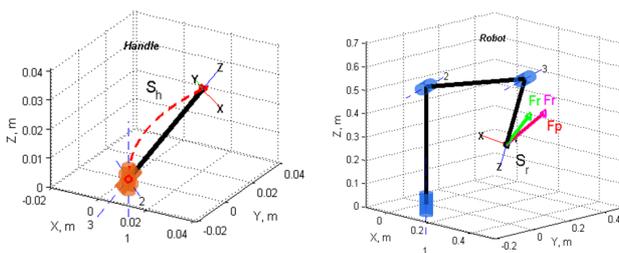


Figure 10: The handle and the robot animations for contact operation mode

Figure 11 represents the curves of programmed and real forces in the closed-loop control system when the robot applies force to a given contact point without motion of the gripper. The programmed forces are shown with red color. The real forces corresponding to the closed-loop control system are shown with pink color.

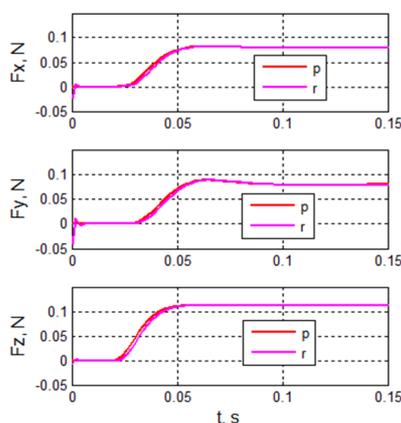


Figure 11: The gripper forces for contact operation mode

EVALUATION OF DYNAMIC ERRORS IN ROBOT MOTION CONTROL SYSTEMS

The dynamic errors of the gripper velocity, position trajectory and force for the open-loop and closed-loop control systems are calculated by the following formulas:

$$E_V = \sqrt{(V_{xp} - V_x)^2 + (V_{yp} - V_y)^2 + (V_{zp} - V_z)^2} \quad (17)$$

$$E_S = \sqrt{(X_p - X_r)^2 + (Y_p - Y_r)^2 + (Z_p - Z_r)^2} \quad (18)$$

$$E_F = \sqrt{(F_{xp} - F_x)^2 + (F_{yp} - F_y)^2 + (F_{zp} - F_z)^2} \quad (19)$$

As it can be seen from the results of simulation, the closed-loop velocity, position and force control systems provide the smaller dynamic errors.

Also, the analysis of the gripper trajectories in Figure 8 shows that the closed-loop position control system gives a smaller contour (geometrical) error compared to the open-loop position control system due to multivariable coordinated position control of the robot link drives.

The robot with the closed-loop force control system applies the required force in a given gripper position without static error.

CONCLUSION

The design of a robot motion control algorithms has shown that it is necessary to build biotechnical control systems with reconfigurable structure.

The results of simulation of dynamic processes have proved that closed-loop algorithms of velocity, position and force control give smaller dynamic errors in comparison with open-loop algorithms. The evaluation of the gripper errors for velocity, position and force control which were obtained in the process of computer simulation is approximate because the purpose of the work was to investigate robot motion control systems in general rather than assess characteristics for a particular robot.

The developed computer models of the velocity, position and force control systems allow to perform motion and contact force operations set by a human operator. In the developed models a human operator is represented as a vector of linear dynamic elements with dead time.

The developed models can be utilized in simulators for training human operators and supervisors of biotechnical robot control systems in real time mode.

From a practical point of view, the animation of both a handle and a robot on a control panel monitor with trajectories and force vectors can make the work of a human operator more effective and convenient.

ACKNOWLEDGMENTS

The research described in this paper is partially supported by the Russian Foundation for Basic Research (grants 16-29-09482-ofi-m, 17-08-00797, 17-06-00108, 17-01-00139, 17-20-01214, 17-29-07073-ofi-i, 18-07-01272, 18-08-01505, 19-08-00989), state order of the Ministry of Education and Science of the Russian Federation №2.3135.2017/4.6, state research 0073-2019-0004, and International project ERASMUS+, Capacity building in higher education, # 73751-EPP-1-2016-1-DE-EPPKA2-CBHE-JP.

REFERENCES

- Alferov G.V., Kulakov F.M., Chernakova S.A. 2006. *Informatsionnye systemy virtualnoy real'nosti v mekhatronike i robototekhnike*. SOLO, St. Petersburg, Russia.
- Arhipov M.V., Golovin V.F., Vzhesnevsky E.A. 2017. "Human-machine interface of the manipulation robot". *Extreme Robotics. Abstracts of the International Scientific and Technological Conference*, pp. 89-90, Gangut, St. Petersburg, Russia.
- Corke P.I. 2017. *Robotics, Vision and Control. Fundamental Algorithms in MATLAB*. Springer International Publishing AG.
- Corke P.I. 2012. "Robotics Toolbox 9.7 for MATLAB R4".
- Denisov A., Budkov V., Mikhalechenko D. 2016. "Designing Simulation Model of Humanoid Robot to Study Servo Control System". In *Proceedings on International Conference on Interactive Collaborative Robotics 2016*, LNCS, pp. 69-78. Springer, Switzerland.
- Filaretov V. F., Katsurin A. A. 2011. "Method of Semiautomatic Combined Control by Manipulator Using Mobile Telecamera". In *Proceedings on 11th International Conference on Control, Automation and Systems*, pp. 649-654. KINTEX. Gyeonggi-do, Korea.
- Filaretov V.F., Katsurin A. A. 2013. "Method of Semiautomatic Position Control by Manipulator Using Telecamera Which Changes Its Orientation". *Advanced Materials Research*, vol. 717, pp. 573-578.
- Golovin V.F., Arhipov M.V., Zhuravlev V.V. 2011. "Ergaticheskie i biotekhnicheskie sistemy upravleniya v meditsinskoj robototekhnike". *Mekhatronika Avtomatizatsiya Upravlenie* 5(12), pp. 54-56.
- Ignatova E.I., Lopota A.V., Rostov N.V. 2014. *Sistemy upravleniya dvizheniem robotov. Komp'yuternoe pro'ektirovanie*. Polytechnic Publishing Center, St. Petersburg.
- Mick S., Cattaert D., Pacllet F., Oudeyer P.Y. and A. de Ruyg. 2017. "Performance and Usability of Various Robotic Arm Control Modes from Human Force Signals." *Frontiers in Neurorobotics*, 25 October 2017, <https://doi.org/10.3389/fnbot.2017.00055>
- Petukhov I.V. 2012. "Issledovanie sensorno-motornogo vzaimodeystviya cheloveka-operatora i tekhnicheskoy sistemy". *Mekhatronika Avtomatizatsiya Upravlenie* 2(12), pp. 33-37.
- Rostova E., Rostov N., Sokolov B. 2018. "Structural Analysis and Animated Simulation of Biotechnical Position-Velocity Control System of a Robot-Manipulator". In *Proceedings on Interactive Collaborative Robotics, Third International Conference, ICR 2018*, Leipzig, Germany, September 18–22, Proceedings, pp. 222-232.
- Soares B.F. (2008). "Master-Slave Servo-Bilateral Control of Direct Drive Electrical Manipulators". *ABCMSymposium Series in Mechatronics*, vol. 3, pp. 246-255.
- Strashnov E.V., Torgashev M.A. 2016. "Modelirovanie dinamiki elektroprivodov virtualnykh robotov v imitatsionno-trenazhernykh kompleksakh". *Mekhatronika Avtomatizatsiya Upravlenie* 11(12), pp. 762-768.
- Wall J., Chandra V., Krummel T. 2008. "Robotics in General Surgery". *Medical Robotics*. InTech, Rijeka, Croatia, pp. 491-506.
- Yurevich E.I. 2005. *Osnovy robototekhniki*. BHV, St. Petersburg, Russia.
- Zenkevich S.L., Yuschenko A.S. 2005. *Osnovy upravleniya manipulyatsionnymi robotami*. MSTU by name N.E. Bauman, Moscow, Russia.

AUTHOR BIOGRAPHIES

EKATERINA ROSTOVA is a postgraduate student, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Laboratory of Information Technology in System Analysis and Integrated Modeling. Research interests: robot control algorithms, man-machine interface, methods, models and simulation of complex biotechnical multidimensional control systems. Her e-mail address is rostovae@mail.ru

NIKOLAY ROSTOV is an associate professor, Peter the Great Polytechnic University, Institute of Computer Science and Technology, Higher School of Cyber-Physical Systems and Control. Research interests: advanced control theory, design of optimal, adaptive and intelligent fuzzy-logic and neural network control systems, computer-aided design, simulation and optimization of robotic control systems. He is the author and co-author of 7 textbooks, 2 monographs on computer design and simulation of control systems and about 100 scientific papers. His e-mail address is rostovnv@mail.ru

VLADISLAV SOBOLEVSKY was born in Vitebsk, Republic of Belarus and went to the St. Petersburg State Technological Institute, where he studied systems analysis, control and data processing and obtained his degree in 2017. He worked for a couple of years for the JSC Concern VKO "Almaz-Antey" before moving in 2017 to the SPIIRAS where he is now working in the field of deep learning. His e-mail address is Arguzd@yandex.ru

VALERII ZAKHAROV was born in Saint-Petersburg, Russia. He studied organization of construction operations at the Saint-Petersburg State University of Architecture and Civil Engineering and obtained his master's degree in 2016. He worked for a couple of years for the building and construction company before moving in 2017 to the SPIIRAS where he is now researcher and PhD student in the Laboratory of Information Technologies in System Analysis and Modeling. His e-mail address is Valeriov@yandex.ru

SOLVING NON-QUADRATIC MATRICES IN ASSIGNMENT PROBLEMS WITH AN IMPROVED VERSION OF VOGEL'S APPROXIMATION METHOD

Maximilian Selmair
Alexander Swinarew
BMW Group
80788 Munich, Germany
maximilian.selmair@bmw.de

Klaus-Jürgen Meier
University of Applied Sciences Munich
80335 Munich, Germany
klaus-juergen.meier@hm.edu

Yi Wang
University of Plymouth
PL4 8AA Plymouth, United Kingdom
yi.wang@plymouth.ac.uk

Keywords—Vogel's Approximation Method; Hungarian Method; Generalised Assignment Problem

Abstract—The efficient allocation of tasks to vehicles in a fleet of self-driving vehicles (SDV) becomes challenging for large-scale systems (e.g. more than hundred vehicles). Operations research provides different methods that can be applied to solve such assignment problems. Integer Linear Programming (ILP), the Hungarian Method (HM) or Vogel's Approximation Method (VAM) are frequently used in related literature (Paul 2018; Dinagar and Keerthivasan 2018; Nahar et al. 2018; Ahmed et al. 2016; Korukoğlu and Balı 2011; Balakrishnan 1990). The underlying paper proposes an adapted version of VAM which reaches better solutions for non-quadratic matrices, namely Vogel's Approximation Method for non-quadratic Matrices (VAM-nq). Subsequently, VAM-nq is compared with ILP, HM and VAM by solving matrices of different sizes in computational experiments in order to determine the proximity to the optimal solution and the computation time. The experimental results demonstrated that both VAM and VAM-nq are five to ten times faster in computing results than HM and ILP across all tested matrix sizes. However, we proved that VAM is not able to generate optimal solutions in large quadratic matrices constantly (starting at approx. 15×15) or small non-quadratic matrices (starting at approx. 5×6). In fact, we show that VAM produces insufficient results especially for non-quadratic matrices. The result deviate further from the optimum if the matrix size increases. Our proposed VAM-nq is able to provide similar results as the original VAM for quadratic matrices, but delivers much better results in non-quadratic instances often reaching an optimum solution. This is especially important for practical use cases since quadratic matrices are rather rare.

LIST OF ABBREVIATIONS

GAP Generalised Assignment Problem
HM Hungarian Method
ILP Integer Linear Programming
KPI Key Performance Indicator
SDV Self-driving Vehicle
VAM Vogel's Approximation Method
VAM-nq Vogel's Approximation Method for non-quadratic Matrices

I. INTRODUCTION

The transportation problem is an extensively studied topic in operational research (Díaz-Parra et al. 2014). The methods for solving the mentioned problem aim to minimise the total transportation cost while bringing goods from several supply points (e.g. warehouses) to demand locations (e.g. customers). In general, each transport origin features a fixed amount of goods that can be distributed. Correspondingly, every point of transport destination requires a certain amount of units (Shore 1970). The underlying use case, where tasks have to be assigned to self-driving vehicles (SDVs), differs in some regards from the classical transportation problem. In our case, each vehicle has a capacity restriction of one, i.e. a maximum of one load carrier can be transported at a time. Furthermore, each task corresponds to a demand of one. This basically means that every task can only be allocated to one single vehicle. Additionally, the amount of available vehicles does rarely match the number of unassigned tasks in practice. Since the size of the matrices depends on those two factors, non-quadratic matrices (e.g. 10×50) are common. There are different approaches that can be applied to solve this kind of problem, e.g. ILP, HM and VAM. While ILP and HM manage to always generate an optimal solution, VAM often fails to do so. Furthermore, those methods vary greatly in the computational demand necessary to solve assignment problems. There are two major reasons involved in the motivation for improving the original VAM. For one, the authors wanted to keep the great performance (computational time) of the original VAM. Secondly, the insufficient results for non-quadratic matrices should be improved significantly, i.e. reaching the optimum. Following these considerations, an improved VAM version, as proposed in this paper, was developed and compared with the three established methods. The goal was to find a solution that provides optimal or near-optimal results while at the same requiring a small amount of resources (computing power).

Despite its age, the approximation method proposed by those authors is still in use nowadays and is subject to recent operations research as the contributions by Banik and Hasan (2018), Ezekiel and Edeki (2018), Hlatká et al. (2017), Ahmed

et al. (2016), and Gani and Baskaran (2014) show.

Already Shimshak et al. (1981) extended the original VAM with rules that apply in case of ties, e. g. the same maximum cost differences occur. Out of the four cases using either individual rules or a combination of them, only one case manages to generate slightly better solutions than the original VAM in regards to costs. Furthermore, they used only small matrices (5×5 , 5×10 and 10×10) which does not provide any information on the results achieved in large-scale applications. Goyal (1984) further tried to improve Shimshak's approach in case of unbalanced transportation problems, i. e. the total supply does not correspond to the total demand. Again, only a small 3×3 matrix was used, thus, lacking any real informative value. Balakrishnan (1990) realised this drawback and tested Goyal's approach with other not specified examples concluding that it is not always better than Shimshak's approach. He in turn proposed an extended approach which was tested in different scenarios and compared with those of the other authors mentioned above leading to even better solutions. In a more recent contribution, Korukoğlu and Ballı (2011) improved the original VAM by constructing a total opportunity cost matrix which they obtained through the addition of the row and column opportunity cost matrix. A row opportunity matrix is for example generated by subtracting the smallest cost value in each row from all other values in the same row. The column opportunity matrix is obtained in the same way. Korukoğlu and Ballı (2011) further deviate from the classical approach by selecting the three rows or columns with the highest penalty costs instead of choosing only the highest one. Out of those three, the cell with the lowest transportation cost is consequently selected and used for resource allocation.

This paper is structured as follows: in the second chapter a detailed description of VAM as well as a brief explanation of the HM and the ILP are given. The third chapter features the description of the proposed VAM-nq. Chapter four will provide an overview of the experiments as well as the discussion of the corresponding results. The last chapter contains a brief conclusion to this paper.

II. ESTABLISHED SOLUTION METHODS FOR THE GENERALISED ASSIGNMENT PROBLEM

This chapter is intended to provide a description on established solution methods for the Generalised Assignment Problem. These are in particular the basic VAM as well as the HM and the ILP. Since the use case at hand differs in some areas from conventional examples (e. g. the vehicles can only transport one load carrier at a time and have thus a supply of one), these variations will be considered in the description of VAM and the HM. The ILP approach will be adapted to the underlying use case as well, i. e. an appropriate objective function as well as necessary constraints will be formulated.

A. Vogel's Approximation Method

The following description of VAM is based on the original proposal by Reinfeld and Vogel (1958). VAM solves transport matrices by repeating the steps as seen below until a feasible solution is found. The cells of the matrices are filled with costs c_{ij} associated with allocating a task to a vehicle. Those costs occur when a vehicle brings goods from a point of origin i to

a destination j . Each source (origin) features a specific amount of goods that can be allocated (supply). Correspondingly, each sink (destination) usually requires a certain number of units (demand). In order to carry out the allocation under these circumstances, the following steps are necessary:

- 1) Calculate the difference between the smallest and the second-smallest cell value for each row and each column.
- 2) Select the row or column which features the biggest difference. If there is a tie, choose the row or column containing the smallest cell value.
- 3) Choose the smallest cell value of the selected row or column and allocate the corresponding task to a vehicle.
- 4) Eliminate the row and column that has been used for the allocation.
- 5) Check if there are still vehicles and tasks left to allocate, and repeat steps 1 - 4 in case that there are.

Apart from the later proposed adoption of VAM in this paper, there are different authors that tried to improve or change the classic VAM in order to achieve better results and move closer to an optimal solution which can be achieved for example by ILP or HM (Paul 2018; Dinagar and Keerthivasan 2018; Nahar et al. 2018; Ahmed et al. 2016; Korukoğlu and Ballı 2011; Balakrishnan 1990; Goyal 1984; Shimshak et al. 1981).

An example for VAM can be found in Table I through Table III. Here, the rows are represented by vehicles (V_i) and the columns by tasks (T_j). The costs (c_{ij}) are the corresponding cell values. The row differences can be found in Δ_i while the column differences are saved in Δ_j . Starting with the initial matrix (Table I), it is evident that the biggest difference can be found in the third row featuring the lowest value in the third column (Table II). Accordingly, task 3 is assigned to vehicle 3. After the allocation, the third row and column are eliminated (Table III).

TABLE I. INITIAL MATRIX TO BE SOLVED BY VAM

c_{ij}	T1	T2	T3	T4	Δ_i
V1	200	100	400	50	50
V2	60	80	30	350	30
V3	210	300	70	150	80
V4	120	510	340	80	40
V5	70	80	40	400	30
Δ_j	10	0	10	30	

TABLE II. MATRIX FEATURING THE IDENTIFIED BIGGEST DIFFERENCE (80)

c_{ij}	T1	T2	T3	T4	Δ_i
V1	200	100	400	50	50
V2	60	80	30	350	30
V3	210	300	70	150	80
V4	120	510	340	80	40
V5	70	80	40	400	30
Δ_j	10	0	10	30	

TABLE III. MATRIX AFTER ELIMINATING ASSIGNED ROW AND COLUMN

c_{ij}	T1	T2	T3	T4	Δ_i
V1	200	100	400	50	50
V2	60	80	30	350	30
V3	210	300	70	150	80
V4	120	510	340	80	40
V5	70	80	40	400	10
Δ_j	10	0	10	30	

B. Integer Linear Programming

As already stated previously, ILP is able to find an optimal solution for different scenarios, even large-scale problems. Initially, one has to formulate an objective function as well as applicable restrictions in order to receive correct results. According to Osman (1995) and following the adoption of the ILP to fit the use case at hand, the objective function reads as follows:

$$\min \sum_{j \in J} \sum_{v \in V} d_{jv} \cdot c_{jv} \quad (1)$$

$$\sum_{v \in V} d_{jv} = 1 \quad \forall j \in J \quad (2)$$

$$\sum_{j \in J} d_{jv} \leq 1 \quad \forall v \in V \quad (3)$$

$$d_{jv} \in \{0, 1\} \quad \forall j \in J, \forall v \in V \quad (4)$$

The goal of the objective function (1) is to minimise the sum of all costs (c_{jv}) for all jobs $J = 1, \dots, m$ and for all vehicles $V = 1, \dots, n$ which is the result of multiplying the decision variable (d_{jv}) with the corresponding costs which arise when a job j is assigned to a vehicle v . The first constraint (2) ensures that every job is assigned to a vehicle while the second constraint (3) makes sure that each vehicle's capacity of 1 is not exceeded, i. e. each vehicle can execute a maximum of one job at a time. The last constraint (4), which applies for both jobs and vehicles, restricts the decision variable d_{jv} to binary values.

C. Hungarian Method

The Hungarian Method was initially proposed by Kuhn (1955) to solve the Generalised Assignment Problem (GAP). Similar to the ILP, the HM is able to find an optimal solution to said problem. The algorithm solves $n \times n$ matrices (e. g. 10×10) by carrying out the following steps until an optimum solution is found:

- 1) Find the minimum value in each column and subtract this value from all other values in the corresponding column.
- 2) Find the minimum value in each row and subtract this value from all other values in the corresponding row.
- 3) Draw lines through the columns and rows so that all zero values of the matrix are covered by as few lines as possible.

- 4) Check if the number of lines equals n . If it does, an optimal allocation of the zero values is possible. If the number of lines is smaller than n , an optimal allocation is not yet feasible and step 5 has to be carried out.
- 5) Find the smallest value which is not covered by a line and a) subtract this value from each not covered row and b) add it to each covered column.
- 6) Return to step 3.

It has to be noted that in case of $n \times m$ matrices (e. g. 10×40), an extension takes place to generate $n \times n$ matrices (e. g. 40×40) since the method only works with quadratic matrices. The additional cells are filled with values that are of the same size as the highest value of the original matrix. This extension requires additional computing power since instead of 400 cells (10×40), the algorithm has to consider 1600 cells (40×40). It is evident that this is a drawback when non-quadratic matrices are to be solved. This is always the case when more tasks than vehicles have to be considered or vice versa.

D. Comparison ILP / VAM / HM

In order to compare the three methods, experiments have been carried out with different quadratic and non-quadratic matrices using an Intel Core i7-6820HQ 2.70 GHz featuring 32 GB RAM. Figure 1 shows clearly that ILP requires the most computational time for quadratic matrices. Especially, in large matrices the time it takes to finish the calculations rises significantly. HM and VAM on the other hand do not require a lot of time to finish calculating the matrices. In fact, there is almost no difference between them up until 80×80 matrices where the HM starts to take longer than VAM. From this point forward, the difference between HM and VAM grows continuously with increasing matrix size. This might lead to the conclusion that it is more sensible to use the HM since it is able to produce optimal solutions while maintaining a relatively low computation time. However, looking at Figure 2 shows that the computation time for HM increases significantly if non-quadratic matrices are involved. This is due to the fact that HM has to generate additional rows or columns to produce quadratic matrices since it is not able to deal with non-quadratic problem instances (see section II). VAM on the other hand can deal with quadratic and non-quadratic matrices regardless of their size in a relatively small amount of time which shows VAMs great scalability.

III. IMPROVED VAM FOR NON-QUADRATIC MATRICES (VAM-NQ)

Prior experiments have shown, that the original VAM is not able to produce optimal or at least near-optimal results for non-quadratic matrices (see Figure 3). In fact, the results are in some cases more than 100% worse than the optimum. It was determined that choosing the row or column featuring the maximum difference from the smaller dimension leads to those insufficient results. This means for example that if the matrix contains more columns than rows, choosing a column with the maximum difference (which is achieved by subtracting cell values in the smaller dimension) might result in worse outcomes. The same obviously applies vice versa if there are more rows than columns. This can be explained with the fact that the bigger dimension obviously features more values and

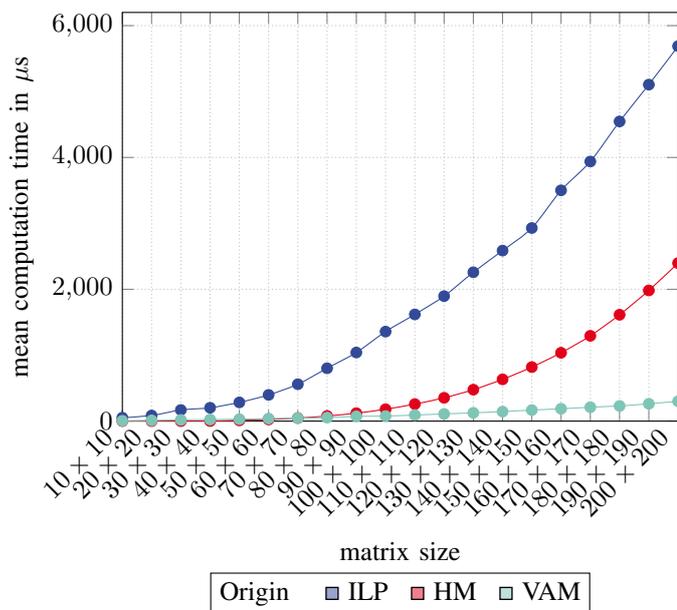


Fig. 1. Mean computational time for ILP (CPLEX-solver), HM and VAM for quadratic matrices in microseconds (5.000 samples each)

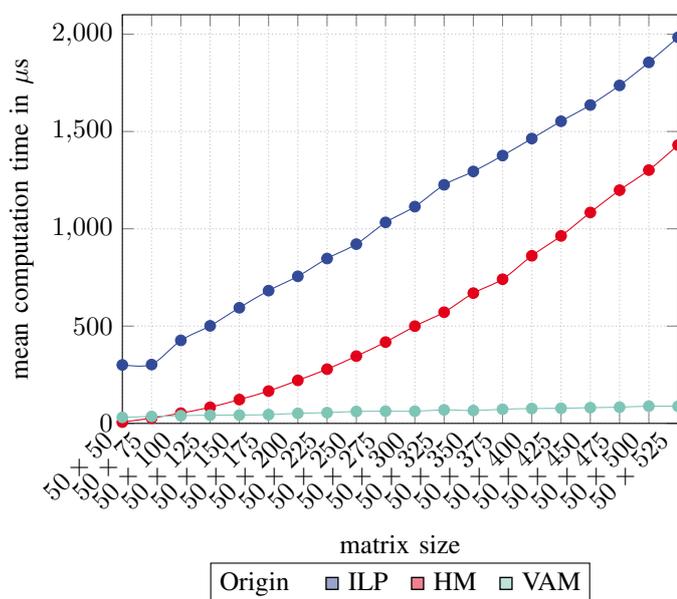


Fig. 2. Mean computational time for ILP (CPLEX-solver), HM and VAM for non-quadratic matrices in microseconds (5.000 samples each)

the chance is therefore higher to find a smaller cell value within those. In order to mitigate the above stated disadvantage of VAM, an improved version of VAM was developed.

Figure 3 shows that the results produced by VAM start to deteriorate immediately if the matrix size is increased in only one dimension, i.e. a non-quadratic matrix is created. It is evident that while VAM is able to generate optimal solutions in some cases, the cases where it fails are up to 200% worse than the optimum (see Figure 3). The deviations increase continuously with increasing matrix size, even in rather small instances. In case of 5×10 matrices for example, the results can be twice as bad as the optimum value.

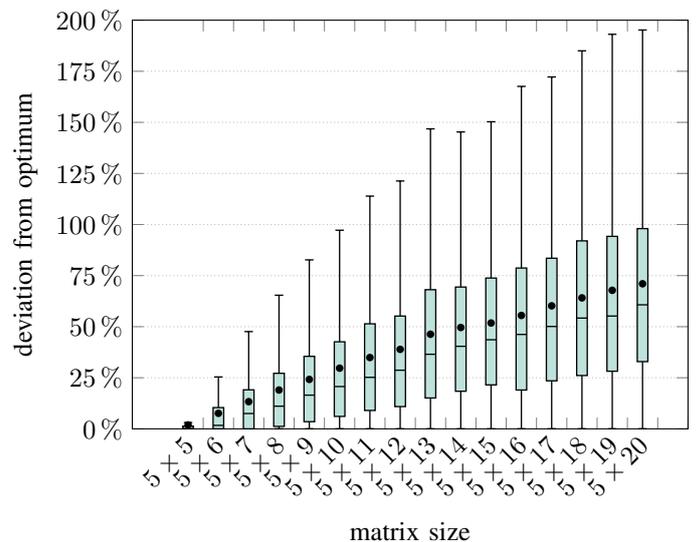


Fig. 3. Deviation of VAM from the optimal solution with increasing matrix size (5.000 samples each)

In general, there are two possible versions of non-quadratic matrices. Either there are more columns than rows or more rows than columns. The description below is based on the first case when a matrix contains more columns than rows. Accordingly, the rows and columns in the description have to be switched when the second case occurs. VAM-nq solves allocation matrices featuring more columns than rows by carrying out the following steps:

- 1) Calculate the difference between the smallest and the second-smallest cell value for each row.
- 2) Select the row featuring the biggest difference. If there is a tie among rows, choose the row containing the smallest cell value.
- 3) Determine the smallest cell value for the selected row and allocate the corresponding task to a vehicle.
- 4) Eliminate the corresponding row and column that have been used for the allocation.
- 5) Check if there are still vehicles and tasks left to allocate, and repeat steps 1-4 in case that there are.

Upon comparison of the original and the adapted VAM, it becomes evident that there are some variations and simplifications. For one, VAM-nq considers only the rows in case that there are more columns than rows (step 1). Accordingly, only the biggest differences in the rows and the corresponding smallest cell values are considered (step 2 and 3). Applying those variations to the second case (more rows than columns) would mean that only columns, their biggest differences and smallest cell values are considered in steps 1 through 3. With Table IV and Table V, the example of subsection II-A is solved with both versions showing that the proposed VAM-nq provides significant better results even in small non-quadratic cases.

TABLE IV. SOLUTION OF THE ORIGINAL METHOD (VAM) WITH OBJECTIVE OF 320

c_{ij}	T1	T2	T3	T4
V1	200	100	400	50
V2	60	80	30	350
V3	210	300	70	150
V4	120	510	340	80
V5	70	80	40	400

TABLE V. SOLUTION OF VAM-NQ WITH OBJECTIVE OF 290

c_{ij}	T1	T2	T3	T4
V1	200	100	400	50
V2	60	80	30	350
V3	210	300	70	150
V4	120	510	340	80
V5	70	80	40	400

IV. EXPERIMENTS

In order to evaluate the performance of VAM-nq as well as its ability to reach optimal solutions, experiments have been carried out by using *AnyLogic* to generate matrices of different sizes.

A. Design

The matrices have been randomly generated and randomly filled with uniformly distributed costs ranging from 0 to 1.400. Each matrix has been solved 5.000 times to provide meaningful results. The following overview shows which matrices have been used to generate and evaluate the corresponding key performance indicators (KPIs):

- Mean Deviation of VAM from the optimal solution as seen in Figure 3: 15 non-quadratic $5 \times n$ matrices with $n = \{6, \dots, 20\}$
- Mean Computation Times for VAM, HM and ILP as seen in Figure 1 and Figure 2:
 - a) 20 quadratic matrices starting with 10×10 and rising to 200×200 in steps of 10
 - b) 20 non-quadratic matrices starting with 50×50 and rising to 50×525 in steps of 25
- Mean Deviation of VAM and VAM-nq from the optimal solution:
 - a) 50 non-quadratic $50 \times n$ matrices with $n = \{51, \dots, 100\}$
 - b) 17 different mixed matrices (5×5 , 5×50 , 10×10 , 10×20 , 10×30 , 10×40 , 20×20 , 10×60 , 20×60 , 30×30 , 10×100 , 40×40 , 50×50 , 50×100 , 100×100 , 100×200 , 100×300)

B. Results of the Experiments

As can be seen from Figure 4, both the original VAM and VAM-nq are not always able to produce an optimal solution, but are instead on average deviating from it. It is also evident

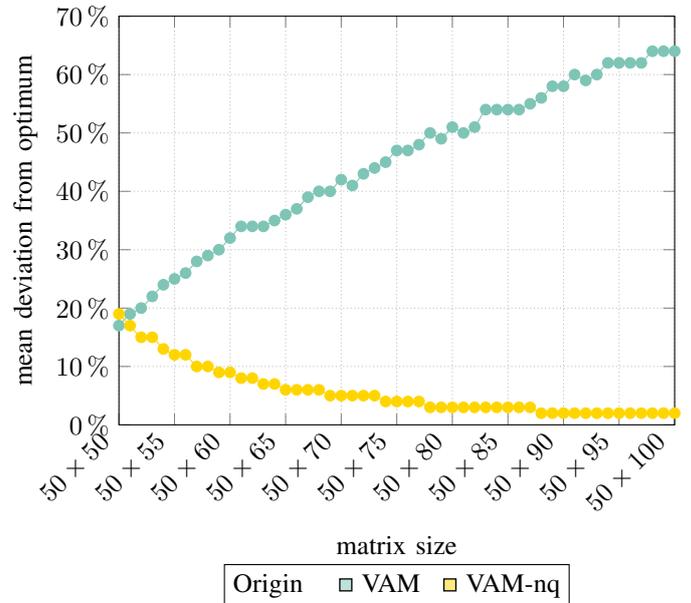


Fig. 4. Mean deviation of the original VAM and VAM-nq from the optimal solution for non-quadratic matrices in percent (5.000 samples each)

that in non-quadratic instances (as seen in Figure 4) the deviation gap between the original VAM and VAM-nq rises continuously when the size of the non-quadratic matrix is increased. While the deviation of the original VAM continues to grow, the deviation of VAM-nq approaches 0%, i.e. an optimal solution is generated more often. This shows clearly that the proposed method is more suitable to deal with non-quadratic instances than the original method. Figure 5 shows the results of experiments performed by using the original VAM and VAM-nq for different problem instances. In this case, it is also evident that in non-quadratic instances the original VAM produces results that are up to 300% worse than the corresponding optimal solution. VAM-nq on the other hand displays almost no deviation and manages on average to generate an optimal solution in all non-quadratic cases. However, it is also recognisable that in quadratic matrices the original VAM is always slightly better than VAM-nq, but the differences in those cases are negligible.

V. CONCLUSION

Experiments have shown that VAM is substantially faster in calculating results than HM and CPLEX-solver (ILP) across all matrix sizes. However, VAM is not able to generate optimum solutions in large quadratic matrices (starting with approx. 15×15) or small non-quadratic matrices (starting with approx. 5×6). In fact, VAM produces insufficient results in those cases and deviates greatly from the optimum. The proposed adapted version of VAM, introduced as VAM-nq, is able to provide slightly worse results than the original VAM for quadratic instances, but delivers much better results in non-quadratic instances reaching an optimum solution in most of the cases. Based on those findings, the authors propose to use an algorithm that includes both the original VAM and the improved VAM-nq and which is able to switch between those two according to the underlying situation. In case that the underlying matrix is quadratic, the original VAM method

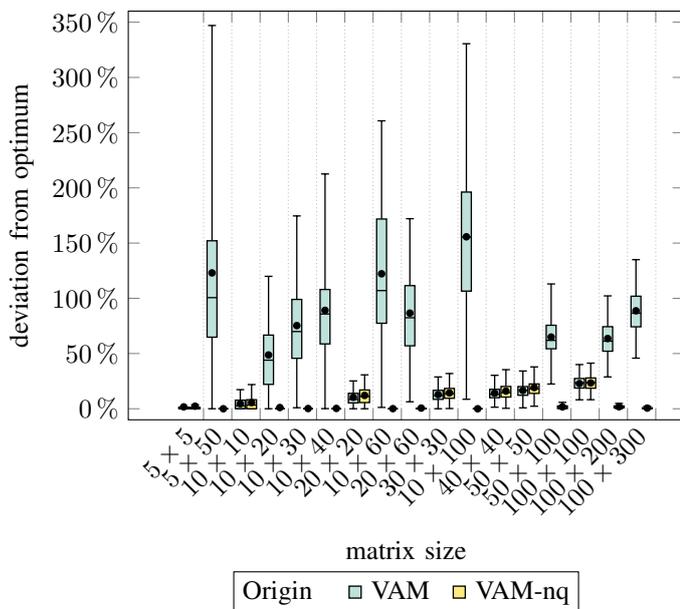


Fig. 5. Deviation of the original VAM and VAM-nq from the optimal solution for different matrix sizes (5.000 samples each)

should be used by the algorithm. For non-quadratic problem instances however, the algorithm should switch to the improved VAM-nq.

REFERENCES

- Ahmed, Mesbahuddin, Aminur Khan, Faruque Ahmed, and Md Uddin (2016). "Customized Vogel's Approximation Method (CVAM) for Solving Transportation Problems". In: *Buletinul Institutulu Politehnic* 62 (66), pp. 31–44.
- Balakrishnan, Nagraj (1990). "Modified Vogel's approximation method for the unbalanced transportation problem". In: *Applied Mathematics Letters* 3.2, pp. 9–11.
- Banik, Debapriya and Md. Zahid Hasan (2018). "Transportation Cost Optimization of an Online Business Applying Vogel's Approximation Method". In: *World Scientific News* 96.
- Díaz-Parra, Ocotlán, Jorge A. Ruiz-Vanoye, Beatriz Bernábe Loranca, Alejandro Fuentes-Penna, and Ricardo A. Barrera-Cámara (2014). "A Survey of Transportation Problems". In: *Journal of Applied Mathematics* 2014.3, pp. 1–17.
- Dinagar, D. Stephen and R. Keerthivasan (2018). "Solving Fuzzy Transportation Problem Using Modified Best Candidate Method". In: *Journal of Computer and Mathematical Sciences* 9.9, pp. 1179–1186.
- Ezekiel, I. D. and S. O. Edeki (2018). "Modified Vogel approximation method for balanced transportation models towards optimal option settings". In: *International Journal of Civil Engineering and Technology* 9, pp. 358–366.
- Gani, Nagoor and Baskaran (2014). "Improved Vogel's Approximation method to Solve Fuzzy Transshipment Problem". In: *Fuzzy Mathematical Archive* 4, pp. 80–87.
- Goyal, S. K. (1984). "Improving VAM for Unbalanced Transportation Problems". In: *Journal of the Operational Research Society* 35.12, pp. 1113–1114.

- Hlatká, Martina, Ladislav Bartuška, and Ján Ližbetin (2017). "Application of the Vogel Approximation Method to Reduce Transport-logistics Processes". In: *MATEC Web of Conferences* 134.4, p. 00019.
- Korukoğlu, Serdar and Serkan Ballı (2011). "A Improved Vogel's Approximation Method for the Transportation Problem". In: *Mathematical and Computational Applications* 16.2, pp. 370–381.
- Kuhn, H. W. (1955). "The Hungarian method for the assignment problem". In: *Naval Research Logistics Quarterly* 2.1-2, pp. 83–97.
- Nahar, J., E. Rusyaman, and S. D. V. E. Putri (2018). "Application of improved Vogel's approximation method in minimization of rice distribution costs of Perum BULOG". In: *IOP Conference Series: Materials Science and Engineering* 332.
- Osman, Ibrahim H. (1995). "Heuristics for the generalised assignment problem: simulated annealing and tabu search approaches". In: *OR Spectrum* 17.4, pp. 211–225.
- Paul, Surjit (2018). "A novel initial basic feasible solution method for transportation problem". In: *International Journal of Advanced Research in Computer Science* 9.1, pp. 472–474.
- Reinfeld, N. V. and W. R. Vogel (1958). *Mathematical Programming*. Prentice-Hall, Englewood Cliffs.
- Shimshak, Daniel, Alan James Kaslik, and Thomas Barclay (1981). "A Modification Of Vogel'S Approximation Method Through The Use Of Heuristics". In: *INFOR: Information Systems and Operational Research* 19.3, pp. 259–263.
- Shore, Harvey H. (1970). "The Transportation problem and the Vogel Approximation Method". In: *Decision Sciences* 1.3-4, pp. 441–457.

AUTHOR BIOGRAPHIES

Maximilian Selmair is doctoral student at the University of Plymouth. Recently employed at the SimPlan AG, he was in charge of projects in the area of material flow simulation. Currently he is working on his doctoral thesis with a fellowship of the BMW Group. His email address is: maximilian.selmair@bmw.de and his website can be found at maximilian.selmair.de.

Alexander Swinarew is a masters graduate in logistics at OTH Regensburg who worked at BMW on his masters thesis during which the proposed method was developed. His email address is: alexander.swinarew@gmail.com.

Prof. Dr. Klaus-Jürgen Meier holds the professorship for production planning and logistic systems in the Department of Engineering and Management at the University of Applied Sciences Munich and he is the head of the Institute for Production Management and Logistics (IPL). His e-mail address is: klaus-juergen.meier@hm.edu.

Dr. Yi Wang is a lecturer in business decision making in the Faculty of Business, University of Plymouth, UK. He has special research interests in supply chain management, logistics, operation management, culture management, information systems, game theory, data analysis, semantics and ontology analysis, and neuromarketing. His e-mail address is: yi.wang@plymouth.ac.uk.

SIMULATION OF HORIZONTAL AND VERTICAL INTEGRATION IN DIGITAL TWINS

Stefan Haag and Carlo Simon
Fachbereich Informatik
Hochschule Worms
Erenburgerstr. 19, 67549 Worms, Germany
E-Mail: {haag,simon}@hs-worms.de

KEYWORDS

Horizontal and Vertical Integration, IoT, Cyber-physical System, Digital Twin, Simulation, Petri nets

ABSTRACT

Horizontal integration describes the linkage of processes to simplify the flow of materials and information between different corporations. Vertical integration describes the connection or merging of processes between top and shop floors inside the very same organization to enable comprehensive possibilities for optimization, allowing – virtually as side effect – for the customized production of batch size one orders.

Both fields offer interesting issues, for example how to gain insight into what kind of effects a possible transformation could have prior to conducting said transformation – maybe assembly of a new production line, reengineering of a supply chain or managing the change of a business process due to new market constraints. Alterations naturally come at a price, thus, testing the consequences by implementation is nearly always infeasible.

To lessen the cost of change, transformations can be analyzed using digital twins – by modelling real world machinery, work pieces and processes, then simulating them and their interaction.

The authors present a possibility to implement digital twins in arbitrary magnitude and level of detail. A web-based specification and simulation environment allows for modelling both high-level business processes and field level as well as activities across companies' borders, comprising everything between.

INTRODUCTION

The total or partial replacement of analogue processes by digital, computer-manageable processes are a main cause of the industrial revolution that can be observed today (Wolf and Strohschen 2018). It leads to disruptive changes in several industries where more and more decisions are made by algorithms instead of employees. As a precondition, real world information must be available in an integrated model in order to simulate possible future paths before conducting the best one.

Two integration scenarios are discussed as shown in figure 1: while horizontal integration means the integration of business processes along supply chains,

vertical integration means to connect the information systems from the top floor (i.e. the business processes automated by an ERP system) to those of the shop floor (i.e. the IT systems for controlling the machines in production).

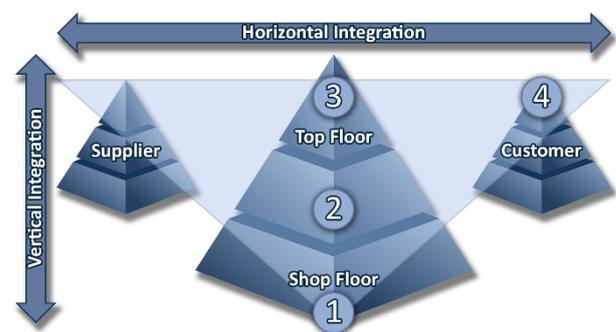


Figure 1: Integration of Processes across Companies

In this paper, possible implementations on the numbered items are introduced. For item 1 a PLC-like approach to simulate the field level is used, while an MES-like model at the operating level is shown for item 2. The company level as located at item 3 is covered by an ERP-like process. Lastly, a workflow interface can be modelled at inter-company level depicted as item 4.

In the past, a major issue has been the definition of data interfaces. Through use of standards like EDIFACT (Electronic Data Interchange for Administration, Commerce and Transport, ISO 9735:1988) a portion of these issues can be attenuated. However, this only solves the software part of the integration problem, i.e. the coupling of the information systems used in and between companies. Two further aspects complete the view:

- IoT (Internet of Things) technologies enhance the picture of a company that exists in its information systems by real time information.
- Process models extend the static view on a company by a dynamic one which allows to test future scenarios before they occur.

A model of a company which integrates data, processes, information systems and real time information concerning this company's resources is called its digital twin. It is a requirement to be able to develop cyber-physical systems for the industry of the future where the technical systems respond smartly to their environment.

The following section addresses the problems observed by the authors which limit the development of digital twins today. Research questions and methods derived from these observations are discussed in the third section. The main part of this paper elucidates the concepts of a Petri net-based environment to overcome these limitations. Further developments planned for the future are discussed at the very end.

CHALLENGES CONCERNING DIGITAL TWINS

A digital twin is a piece of software that models a part of the real – its real-world counterpart. For this, the digital twin stores all relevant status information of its real-world counterpart and is able to observe whether the real world behaves as predicted. Hence, the ability to predict the future is a central contribution of digital twins.

As a consequence, it is irrelevant whether this real-world counterpart actually exists or only its idea. For example, a production plant can already have a digital twin during its planning phase. Simulation can then be used to find out whether the planned plant will behave as assumed. All fabrication steps can be tested and optimized on the production line virtually (Kannwischer 2015). For this, the digital twin, when executed with real data, should behave like its real counterpart (Tao et al. 2017), or more precisely, it should calculate the same state as the one that can be observed in the real world.

Connectors from real world objects to their digital models are needed to compare the current and the predicted state and to deliver the relevant real-world information to the digital twins. Traditionally, sensors and user input have been these connectors. Hereby, sensors are typically restricted to local machines. User input, on the other hand, has often deficits concerning the correctness and currentness of data. In the past, these limitations have been accepted since they were the only available sources of information. In a business context a change can be observed nowadays as follows:

- Several information concerning business processes (like orders, bills, etc.) already primarily exist in the virtual world.
- By connecting machines and products to the internet, which is described by the term internet of things, also widely spread infrastructure information can be consolidated in a single place.

The other way around, actors controlled by digital twins might influence the reality. Again, the internet of things offers the ability to spread this information in a wide area.

Now, some of the major challenges are the differences concerning description languages for a business reality. They limit the ability to develop an integrated model.

Two possible solutions exist for this:

- Like connectors in the real world it might be possible to link the different types of models with interfaces to each other.
- The different views on a business from the top floor to the shop floor are described in a

monoparadigmatic way by usage of a small, generic language.

On first view, a lot seems to speak for the first solution since there exist several (business) process modeling languages that are widely used such as BPMN (Silver 2017), flowchart diagrams (information processing, ISO 5807:1985) or value stream diagrams (Klevers 2015). However, since no formal definition and semantics exists for these languages, the usefulness of interface definitions is hardly proofed. At field and production level, ladder diagrams are used as programming language when utilizing PLCs. Microcomputers (like a Raspberry Pi) are often coded on directly in a programming language such as C or C++. In each of these cases, the execution semantics are well defined, however no interface is defined concerning the above-mentioned diagrams yet.

As is illustrated in figure 2, this results in the goal of overcoming heterogeneous concepts between different business levels, while at the same time maintaining or – as a secondary effect in the case of data format mismatches – enabling the possibility to model systems integration beyond the company's borders. This can be achieved by the development of holistic digital twins.



Figure 2: Transition from Heterogeneous Specification Languages to a Monoparadigmatic Description

In this paper, a monoparadigmatic approach is pursued based on Petri nets, which have been researched for more than 50 years and are clearly defined semantically (Reisig 2010). Petri nets allow for modelling on each of the levels mentioned in the later concept sections of this paper, from workflow management down to the controller – an advantage over other simulative languages that focus on specific applications like discrete event systems.

However, the existence of this method is not enough: in terms of applied sciences, also a web-based modeling and execution software is presented which has been developed over the past years and is currently applied to integration projects to overcome the limitations described above.

RESEARCH METHOD

According to (Hevner et al. 2004), there are seven guidelines for Design Science Research. These and their implementation are briefly explained as follows:

Design as an Artifact: A viable artifact must be created. In the present case, this is a web-based specification, simulation and control environment.

Problem Relevance: The goal is to develop a technology and/or methodology-based solution to a significant and pertinent problem. The presented artifact not only simulates business processes based on real data, but also optimizes and controls them.

Design Evaluation: The presented tool is used in companies already, thus feedback from the operational practice can be evaluated. The control part is used in teaching and prepared for operational use due to feedback from student users.

Research Contribution: The contribution consists of the creation of a practical application on the theoretical basis of the Petri net theory. The theory is enriched by time concepts and the possibility to connect Petri net models with the modelled reality.

Research Rigor: In theory, the simulation options developed are compared with other simulation environments. In practice the focus is placed on the evaluation of how to use this approach in a company environment.

Design as a Search Process: The present prototype is the latest in a series that starts from the initial implementation of the underlying principles and ends in a productive system. Each implementation step has been evaluated.

Communication of Research: The results achieved so far are relevant for both research and practice. They are presented on conferences but also, more illustrative, for practitioners.

CONCEPT AT FIELD LEVEL

Possibilities for using Petri nets for machine control and plant control have a long tradition (see (Zuse 1980), (König und Quäck 1988), (Hanisch 1992)). In addition to the specification and subsequent analysis of the possible behavior of a modeled system, Petri nets can also be used to generate code for PLCs (see (Brand 2000), (Raue 2001)).

Disadvantageous to such an approach is the need for a new code generation in case of adjustments to the machinery. This is not always useful in view of the desired flexible adaptability of machines due to altering customer demands.

An interesting and cost-effective variant is the use of microcomputers such as the Raspberry Pi to handle control tasks. If a web server is installed on the computer, then the web-based modeling and simulation environment presented here can also be installed. With its own data types and functions, complex process steps can be expressed compactly as a Petri net. By use of an extension allowing access to the GPIO interface of a Raspberry Pi, a production plant can be controlled directly by the simulator.

Starting point for such a control is the question as to how the actuators (for example motor or light) and sensors (for example button or light sensor) of the production line to be controlled are mapped in the tool and connected to it.

For actuators, it makes sense to represent them with a place that is linked to the appropriate GPIO interface by

means of an attribute. If the position is marked, then the actuator is switched on, otherwise switched off. If it is a continuously controllable actuator (for example a motor whose speed can be regulated), then the continuous variable is specified by the token value.

Sensors can also be represented by means of a place and in this case a token also indicates the condition of the sensor on that place, i.e. whether the sensor supplies a signal and, in the case of a continuous sensor, its value. However, in the latter case a sampling interval in which the sensor value is updated is also to be specified in expansion of the output interface. Furthermore, such places may only be read by transitions without deleting or modifying the token.

The example in figure 3 illustrates this principle as a Petri net using a production line with a processing station.

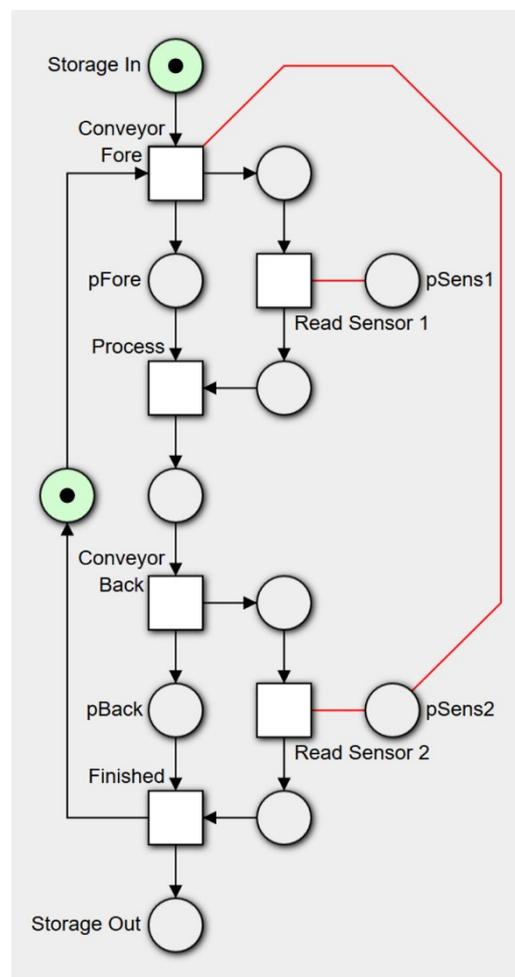


Figure 3: Process for Continuous Operation of a Two-Way Conveyor Belt with Attached Processing Station

Figure 4 shows the associated code specification except the part describing the layout has been left out to facilitate readability. In the initial state, the model waits for the signal from the light sensor at GPIO pin 24, which indicates that a work piece has arrived on the assembly line. As this is a discrete sensor, there is no need for interval sampling. When the signal arrives, the conveyor belt is set in motion in the direction of the workstation

until a signal at input port 23 is detected, thus indicating the correct positioning of the work piece. Now the processing will take place (which for reasons of space is not shown in detail here). Finally, the work piece is moved in the opposite direction back to the starting point.

```

N levelPLC {
  P pSem (label='');
  P pSI (label='Storage In');
  P pSO (label='Storage Out');
  P pFore (gpioOut=5);
  P pBack (gpioOut=6);
  P pSens1 (gpioIn=23);
  P pSens2 (gpioIn=24);
  P p0, p1s1, p2s1, p1s2, p2s2 (label='');

  M (pSem=1);
  M (pSI=1);

  T tConvFore (label='Conveyor$Fore');
  T tReadSen1 (label='Read Sensor 1');
  T tPro (label='Process');
  T tConvBack (label='Conveyor$Back');
  T tReadSen2 (label='Read Sensor 2');
  T tFin (label='Finished');

  A (pSI, tConvFore, pFore, tPro, p0,
    tConvBack, pBack, tFin, pSO);
  A (tConvFore, p1s1, tReadSen1, p2s1, tPro);
  A (tReadSen1, pSens1);
  A (tConvBack, p1s2, tReadSen2, p2s2, tFin);
  A (tReadSen2, pSens2);
  A (tConvFore, pSens2);
  A (tFin, pSem, tConvFore);
}

```

Figure 4: Specification for the Model from Figure 3 (layout specification omitted for readability)

CONCEPT AT OPERATING LEVEL

Production planning focuses on the scheduling of orders and the aggregation of individual orders into lots. One way to solve this task has already been described in (Simon 2017) and (Simon 2018). Key is the ability to process several tokens put on one place in a sequence.

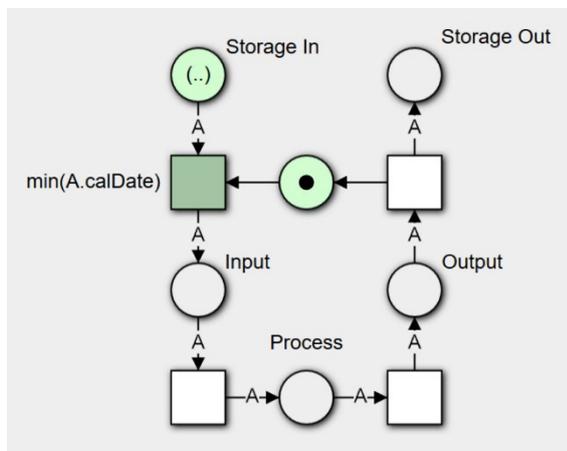


Figure 5: Process for Selection of Work Orders by Age

This corresponds to the possibility of sorting data records of a database with regard to a number of attributes in ascending and descending order. A

comparable task can be seen in figure 5, where the oldest from a set of work orders is chosen by using a minimum selector. In this example, the orders planned at the company level are inserted into the production queue one by one. During production, the work pieces must first be fed to the workstation, then processed and finally taken away again.

The associated specification is depicted in figure 6 barring the part describing the layout.

```

N levelMES {
  R recOrder {calDate: datetime,
    var: char, amt: int};

  P pSem (label='');
  P p0 (label='Storage In', type=recOrder);
  P p1 (label='Input', type=recOrder);
  P p2 (label='Process', type=recOrder);
  P p3 (label='Output', type=recOrder);
  P p4 (label='Storage Out', type=recOrder);

  M (pSem=1);
  M (p0= ("2018.04.02T08:00:00", "oxo", 75);
    ("2018.04.02T12:00:00", "oxo", 15);
    ("2018.04.10T08:00:00", "oxo", 25);
    ("2018.04.01T08:00:00", "xo", 25);
    ("2018.04.10T12:00:00", "xo", 50));

  T t0 (label='', select='min(A.calDate)');
  T t1, t2, t3 (label='');

  A (p0, t0, p1, t1, p2, t2, p3, t3, p4,
    label='A');
  A (t3, pSem, t0);
}

```

Figure 6: Specification for the Model from Figure 5 (layout specification omitted for readability)

CONCEPT AT COMPANY LEVEL

From an information technology perspective, the company and the operating levels are quite similar: both levels process orders. But while at the plant level orders are brought into production, at the company level the business prerequisites to be able to process these orders are created.

The Architecture of Integrated Information Systems (ARIS) as shown in figure 7 is a coherent frame of reference for the various concepts used to describe a company from the point of view of information systems. The central role is played by the processes by which the data of a company is linked to its operational functions. Organizational responsibilities and the specification of relevant output data complete this approach.



Figure 7: Architecture of Integrated Information Systems (ARIS) as per (Scheer 1997)

Petri nets, which are mathematically defined, also offer the possibility of modelling business processes and are therefore used for this purpose, in particular for workflow

management systems (van der Aalst and van Hees 2002). They can be interpreted unambiguously and hence may serve as the basis for an automated workflow execution.

Distinctive tokens are crucial for the use of Petri nets to model and execute business processes. The best known approaches to this are Pr/T-nets (Predicate/Transition Nets) as per (Genrich and Lautenbach 1981) and Colored Petri Nets as per (Jensen 1992). The possible usage of the latter for modeling business processes is described exemplarily in (van der Aalst and Stahl 2011).

In Pr/T-Nets, places are typed and can be marked with matching tuples. Hence, places correspond to tables of a database and their tuples to the records within a database table. In this way, the network can be linked to a database or filled with real data via a CSV interface.

A sample application is a release process for orders in case all required materials are in stock. Otherwise, an order process will be triggered automatically. Figure 8 shows the model for this example using icons for visualization purposes.

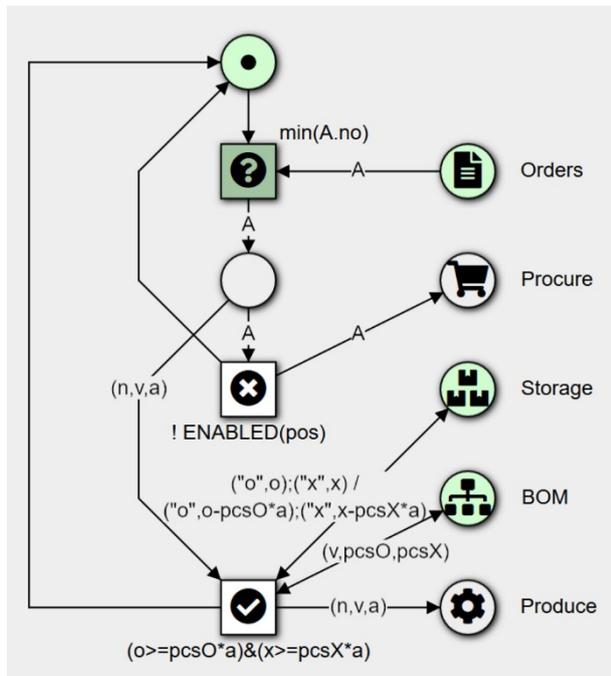


Figure 8: Process for Production Order Approval or for Repeat Order Placement, respectively

Figure 9 depicts the corresponding specification, again without mention of the layout portion.

```

N levelERP {
  R recOrder {no: int, var: char, amt: int};
  R recStorage {part: char, pcs: int};
  R recBOM {var: char, pcs0: int, pcsX: int};

  P p0 (label='');
  P p1 (label='', type=recOrder);

  P ord (label='Orders', type=recOrder);
  P sto (label='Storage', type=recStorage);
  P bom (label='BOM', type=recBOM);

```

```

P proc (label='Procure', type=recOrder);
P prod (label='Produce', type=recOrder);

M (p0=1);
M (ord='(1,"oxo",25);(2,"oxo",15);(3,"oxo",25);(4,"xo",25);(5,"xo",5)');
M (sto='("o",200);("x",80)');
M (bom='("oxo",2,1);("xo",1,1)');

T sel (label='', select='min(A.no)');
T neg (label='', condition='!ENABLED(pos)');
T pos (label='', condition='(o>=pcsO*a)&(x>=pcsX*a)');

A (ord, sel, p1, neg, proc, label='A');
A (p1, pos, prod, label='(n,v,a)');
A (sto, pos, label='("o",o);("x",x)');
A (pos, sto, label='("o",o-pcsO*a);("x",x-pcsX*a)');
A (bom, pos, bom, label='(v,pcsO,pcsX)');
A (neg, p0, sel);
A (pos, p0);
}

```

Figure 9: Specification for the Model from Figure 8 (layout specification omitted for readability)

CONCEPT AT INTER-COMPANY LEVEL

The concepts at company and inter-company levels share innately even greater similarities than those at company and operating levels. If it is possible to model and simulate business processes on an ERP level, then connecting those with external processes on the same level becomes trivial.

The fundamental idea of this thought can be seen in the Workflow Management Coalition's Reference Model as depicted in figure 10. A company's workflows can be executed by workflow engines which themselves are parts of workflow enactment services. Through the use of appropriate interfaces, they may be connected to each other, even if they reside outside the company.

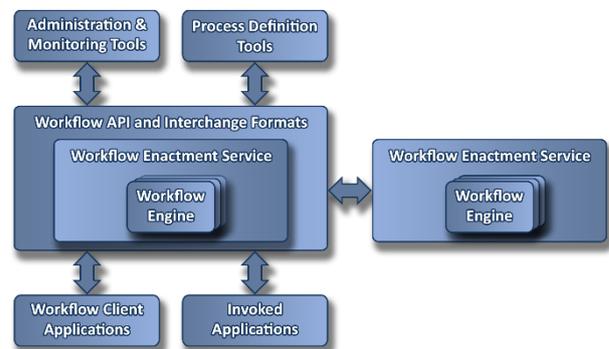


Figure 10: Components and Interfaces of the Workflow Reference Model as per (Hollingsworth 1995)

In connection with the possibility to use Petri nets as conceptual standard for the modelling and analysis of workflows (van der Aalst 1996), the claim made in the initial paragraph of this section proves true.

CONCLUSIONS

The integrated modeling and simulation environment for Petri nets presented in this article creates a monoparadigmatic and homogeneous approach for the simulation and control of processes at all levels of the automation pyramid. The system is already being used as part of consulting projects to optimize processes in retail and logistics, and currently production processes are being modeled, simulated and optimized on this basis in students' projects. Also the implementation of the system on a microcomputer is already used as part of lectures.

Overall, the results presented here make it clear that the goal of developing an integrated digital twin for both the various levels of the automation pyramid can be achieved with the help of Petri nets. A simulation of vertical integration from the top floor to the shop floor and back is already possible, as is one of horizontal integration along the supply chain.

FURTHER RESEARCH

In addition to the continuous expansion of the language base, new topics get queued due to the widespread use of the system, especially aspects of security. Different topics are relevant: The GPIOs of a Raspberry Pi control root privileges, which then have to be secured by the tool. However, before such topics can be addressed, the tool must be extended by a user management with access rights so that security-relevant information can be protected actually.

REFERENCES

- van der Aalst. 1996. "Petri-net-based Workflow Management Software". In *Proceedings of the NFS Workshop on Workflow and Process Automation in Information Systems*, A. Sheth (Ed.). UGA, Athens, GA, 114-118.
- van der Aalst, W. and K. van Hees. 2002. *Workflow Management – Models, Methods, and Systems*. MIT Press, Cambridge, MA.
- van der Aalst, W. and C. Stahl. 2011. *Modelling Business Processes: A Petri Net-Oriented Approach*. MIT Press, Cambridge, MA.
- Brand, A. 2000. *Generierung von SPS-Programmen aus Petri-Netzen*. Diplomarbeit, Universität Koblenz-Landau, Germany.
- Genrich, H.J. and K. Lautenbach. 1981. "System Modelling with High-Level Petri Nets". *Theoretical Computer Science* 13, 1, 109-135.
- Hanisch, H.-M. 1992. *Petri-Netze in der Verfahrenstechnik*. Oldenbourg, München, Germany.
- Hevner, A.R.; S.T. March; J. Park; and S. Ram. 2004. "Design Science in Information Systems Research". *MIS Quarterly* 28, 1 (Mar), 75-105.
- Hollingsworth, D. 1995. *Workflow Management Coalition: The Workflow Reference Model*. Workflow Management Coalition, Hampshire, United Kingdom.
- International Organization for Standardization. 1985. *ISO 5807:1985 Information processing*. ISO, Genf, Switzerland.
- International Organization for Standardization. 1988. *ISO 9735:1988 EDIFACT*. ISO, Genf, Switzerland.
- Jensen, K. and L.M. Kristensen. 2009. *Coloured Petri Nets*. Springer, Berlin, Heidelberg, Germany.
- Kannwischer, M. 2015. *Interaktive Präzisionswerkzeuge für die effizientere Bearbeitung. Produktivitätsfortschritte durch Industrie 4.0*. VDMA, Frankfurt, Germany (May).
- Klevers, T. 2015. *Agile Prozesse mit Wertstrommanagement*. CETPM, Ansbach, Germany.
- König, R. and L. Quäck. 1988. *Petri-Netze in der Steuerungstechnik*. VEB Verlag Technik, Berlin, Germany.
- Raue, H. 2001. *Systematische Entwicklung einer Steuerung für eine Verpackungsmaschine*. Diplomarbeit, Universität Koblenz-Landau, Germany.
- Reisig, W. 2010. *Petrinetze – Modellierungstechnik, Analysemethoden, Fallstudien*. Springer-Vieweg, Wiesbaden, Germany.
- Scheer, A.-W. 1997. *Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse*. Springer, Berlin, Heidelberg, Germany.
- Silver, B. 2017. *BPMN Quick and Easy Using Method and Style: Process Mapping Guidelines and Examples Using the Business Process Modeling Standard*. Cody-Cassidy, Los Angeles, CA.
- Simon, C. 2017. "Eine Petri-Netz-Programmiersprache und Anwendungen in der Produktion". In *AKWI Tagungsband 2017*, T. Barton; F. Herrmann; V.G. Meister; C. Müller; and C. Seel (Eds.). mana-Buch, Heide, Germany, 61-70.
- Simon, C. 2018. "Web-based Simulation of Production Schedules with High-level Petri Nets". In *Proceedings of the 32nd European Conference on Modeling and Simulation* (Wilhelmshaven, Germany, May 22-25), 275-281.
- Tao, F.; J. Cheng; Q. Qi; M. Zhang; H. Zhang; and F. Sui. 2018. "Digital Twin-Driven Product Design, Manufacturing and Service with Big Data". *The International Journal of Advanced Manufacturing Technology* 94, 9-12 (Feb), 3563-3576.
- Wolf, T. and J.-H. Strohschen. 2018. "Digitalisierung: Definition und Reife". In *Informatik-Spektrum* 41, 1 (Feb), 56-64.
- Zuse, K. 1980. *Petri-Netze aus der Sicht des Ingenieurs*. Springer-Vieweg, Wiesbaden, Germany.

AUTHOR BIOGRAPHIES



STEFAN HAAG holds degrees in Business Administration and Engineering as well as Economics with his main interests being related to modelling and simulation in graphs. After working at the Fraunhofer Institute for Systems and Innovation Research ISI Karlsruhe for several years, he is now a Research Fellow at the Hochschule Worms. His e-mail address is: haag@hs-worms.de.



CARLO SIMON studied Informatics and Information Systems at the University of Koblenz-Landau. For his PhD-Thesis, he applied process thinking to automation technology in the chemical industry. For his state doctorate, he considered electronic negotiations from a process perspective. Since 2007, he is a Professor for Information Systems, first at the Provdia School of Technology and Management Frankfurt and since 2015 at the Hochschule Worms. His e-mail address is: simon@hs-worms.de.

OPTIMAL PLANNING FOR PURCHASE AND STORAGE WITH MULTIPLE TRANSPORTATION TYPES FOR CONCENTRATED LATEX UNDER AGE-DEPENDENT CONSTRAINT

Tuanjai Somboonwivat¹

Sutthinee Klomsae²

Walailak Atthirawong³

¹ King Mongkut's University of Technology Thonburi, Thailand,

² Uttaradit Rajabhat University, Thailand,

³ King Mongkut's Institute of Technology Ladkrabang, Thailand,

E-mail: tuanjai.som@kmutt.ac.th, klomsae_ay@hotmail.com, walailaknoi@gmail.com

KEYWORDS

Age-dependent, Optimal Planning, Perishable Product, Rubber

ABSTRACT

This paper presents the mathematical model to support decision planning over the multi-period for purchasing storage and transportation of concentrated latex for the rubber glove production under age-dependent constraint. The model considers multiple suppliers with different purchasing costs, varying product ages, multi-truck load capacity and costs, and storage times with respect to perishable product. The model is then applied to a rubber glove production case study and the sensitivity analysis is performed. The result reveals that the price of concentrated latex in each period and truck load capacity decision are critical factors for total cost reduction.

INTRODUCTION

Thailand has achieved particularly remarkable in the upstream sector of rubber industry over an extended period. In 2015, the country was the world leader in rubber producer and exporter which produced about 4.5 million tons of output, followed by Indonesia, China, India, and Malaysia, respectively (Thailand Industry Outlook 2016). Nowadays, Thailand has the advantage of expanding production area, especially in the Northeast and North part of the country, which can be divided into 10 regions of plantation areas (<http://www.oae.go.th>). Rubber is a key input into a number of significant manufacturing sectors, especially rubber gloves and tires. The most suitable raw material for the manufacture of gloves is natural concentrated latex. It was found that rubber gloves had contributed an enormous export values to Thai economy. Figure 1 illustrates the flow of production materials in gloves industry from upstream to downstream.

Nevertheless, currently Malaysia becomes the world leader in developing the downstream part of the rubber

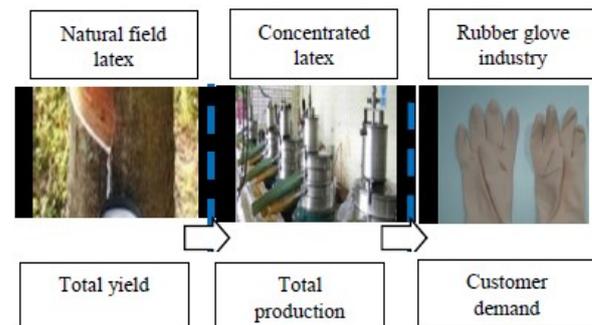


Figure 1. Flow of Production Materials in Rubber Glove Industry from Upstream to Downstream

value chain, particularly medical glove production. As such, the challenge for Thailand to gain competitive, decision planning of this industry is extensively required for purchasing and storage of the concentrated latex throughout the seasonal production cycle for the industry. It was found that the supply volume of field latex varies according to regions and seasons, which is a perishable in nature. As such, it is necessary to pay attention on production and transportation which is a generally complicated task.

According to the statistics report, most of rubber latex manufactures are located in the Middle, Southern and Eastern regions of Thailand account for 40%, 36.7% and 20% respectively (Plastics Institute of Thailand 2014). Due to the proximity of raw materials, the large size of the manufacturers is in the Southern (46%), and Eastern (38%) regions of the country. Generally, the delivery of field latex is made from the plantation to a latex concentration industry within the same region, as it is a perishable product latexes. However, the glove manufacturing factories will require concentrated latex according to their real demand which any shortages of raw material to the factories will result in higher costs. As a result, some concentrated latexes are supplied to the rubber glove companies in the same region when possible, and some of them are shipped across regions according to real demand. After that, the finish products

will be exported to customers through the closest seaports from that factory.

Along with rapid change in the marketplace and the expansion of supply chains, highly coordination on purchasing, inventory, and transportation over a multi-echelon supply chain network or a multi-period time frame is vigorous, which has very much impact on customer service and profit margins of the supply chain members. Nonetheless, research related to simultaneous purchase and storage decisions of rubber latex with regard to aging limitations as well as considering a transportation cost at the same time is still quite rare. Most of the literature in the area of the rubber supply chain is focused on the structure of the rubber industry.

The literature of the rubber supply chain is related to the structure of the rubber industry. Haan *et al.* (2003) present the production flow in the rubber supply chain of India, using semi-structured interviews and analysis of the relationships of each part. For research relevant to marketing analysis of the rubber industry, Arifin (2005) studies the supply chain of rubber production in Indonesia and assess the transmission of the price to rubber growers. Fathoni (2009) analyzes rubber market systems and analyze marketing margins and prices for rubber in Jambi Province in Indonesia. Auckara-aree and Boondiskulchok (2010) design a raw material collection system by formulating Mixed Integer Programming for location routing with a step-price policy model to find the optimal solution. Somboonwivat and Chanclay (2009) create value-added products by coherence of targets in different parts of the supply chain in Thailand. For research related to transportation, Kritchanhai and Chanpuypetch (2009) investigate gateway selections for Thailand rubber exports, using fuzzy analytic hierarchy processes. Klomsae *et al.* (2012) develop a mathematical model to determine the concentrated latex volume to purchase and storage. However, recently the perishable product management has been widely concerned in the logistics planning area, for example in the studies of Wu *et al.* (2015) and Liu *et al.* (2017). Hence, this paper presents the mathematical model for decision planning which simultaneously solve the volume of concentrated latex purchased and stored in each age including transportation type in multi-period time frame taking into consideration of product aging and deterioration through each time period. The parameters in the model include purchasing costs of product with varying price and age, transportation costs that vary by truck load capacity and distance between location of suppliers and manufacturers, storage costs, and expiration costs of outdated concentrated latex.

The organization of this paper is as the following. The next section describes the multi-period planning problem of purchasing and storage of concentrated latex from multiple suppliers and transportation types. Then the mathematical model is developed. A numerical example is presented to illustrate an application of the

formulation. Finally, summary and concluding remarks are addressed.

PROBLEM DESCRIPTION

The rubber glove industry uses rubber concentrated latex to produce the glove products. The latex requirement is based on dependent demand which the latex can be purchased from different suppliers in different regions. The decision planning for purchase, storage and transportation of concentrated latex for rubber glove production begins with the determination of the volume of concentrated latex from different regions and multiple suppliers with number of round trips of different transportation types. The cost of transportation will depend on distance between location of suppliers and manufacturers, as well as size of truck and truck load capacity as well. Since age is significant factor in perishable latex requirement operations, the latex planning includes age of latex for each supplier and transportation type.

Figure 2 delineates the system of purchase, storage and transportation of perishable concentrated latex. There are multiple prices and ages of latex for each supplier. The number of round trips (X_{ijkt}) is decided for each supplier for each age of latex. The volume of purchased concentrated latex for supplier i at age j deliver by truck type k in period t (B_{ijkt}) and the volume of stored concentrated latex at the beginning of period t (I_{jt}^B) must be high enough for the demand of concentrated latex to glove producers (D_t). At the beginning of period $t+1$ the age of the concentrated latex is increased to $j+1$ ($I_{(j+1)(t+1)}^B$). The age of concentrated latex continues to increase until it reaches the expiration date, then that concentrated latex will be eliminated.

The primary objective of this paper is to develop a mathematical model for solving the concentrated latex requirement planning under age-dependent and multiple transportation types constrains. The decision criteria used is to minimize the total costs of purchasing, storing, transportation and expiration.

At period T, volume of expired concentrated latex equal to volume of concentrated latex inventory for age j at the ending of period t .

$$F_t = I_{jt}^E \quad \forall t \quad (6)$$

In order to determine the volume of concentrated latex of age j used in production in period t : if demand for concentrated latex for production in period t is greater than or equal to the sum of the volume of concentrated latex inventory for age $j-1$ at the ending of period $t-1$ and the volume of concentrated latex purchased from supplier i of age j in period t then: the volume of concentrated latex for age j used in production in period t will equal the sum of the volume of concentrated latex for age $j-1$ used in production in period $t-1$ and the volume of concentrated latex purchased from supplier i of age j in period t (Equation 7). But if this condition is not true, then: the volume of concentrated latex age j used in production in period t will equal to the demand for concentrated latex for the production in period t (Equation 8).

$$\text{If } D_t - (I_{(j-1)(t-1)}^E + \sum_{i=1}^I B_{ijt}) \geq 0$$

$$\text{True: } U_{jt} = I_{(j-1)(t-1)}^E + \sum_{i=1}^I B_{ijt} ; \quad \forall t \quad (7)$$

$$\text{False: } U_{jt} = D_t ; \quad \forall t \quad (8)$$

The constraints address to determine the volume of concentrated latex for age j used in production in period t . For example, concentrated latex of the oldest age available for production is $j=4$ and determine the volume of concentrated latex available for production of age $j=3$, and determine the volume of concentrated latex available for production in period t as demonstrated in Equation 9 and Equation 10.

$$\text{If } (I_{(j-2)(t-1)}^E + \sum_{i=1}^I B_{i(j-1)t} - (D_t - U_{jt})) > 0$$

$$\text{True: } U_{(j-1)t} = D_t - U_{jt} ; \quad \forall t \quad (9)$$

$$\text{False: } U_{(j-1)t} = I_{(j-2)(t-1)}^E + \sum_{i=1}^I B_{i(j-1)t} ; \quad \forall t \quad (10)$$

Non-negativity constraint:

$$B_{ijt}, X_{ijkt}, I_{jt}^E, I_{jt}^B, F_t, U_{jt} \geq 0 ; \quad \forall j, k, t \quad (11)$$

NUMERICAL EXAMPLE

In this example, the concentrated latex can be purchased from 2 suppliers which are transported from different regions to the production facilities by 2 transportation types. The age of the latex is tracked over varying periods, allowing for a maximum aging limit of 4 months. The problem is to determine the volume of concentrated latex at each age purchased from each supplier, the number of round trips for concentrated latex transported from suppliers to the production facilities, the volume of concentrated latex used in production, the volume of concentrated latex inventory stored in each period, and the volume of expired concentrated latex. The model is applied to an example data set in order to generate planning decisions for the purchase and storage of concentrated latex with multiple prices and ages. The latex volumes must vary in response to production demand for rubber glove production over a 3 period time interval. The prices and volumes of supplied latex are shown in Table 1. Production demand, holding cost and expiration cost are presented in Table 2. Transportation costs and weight per round of concentrated latex for each transportation type are shown in Table 3.

Table 1. Prices and Volumes of Supplied Latex

Supplier i	Age j	Period 1				Period 2				Period 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	P _{ijt} (baht)	36.99	36.43	35.7	0	35.16	34.81	34.46	0	35.75	35.39	35.04	34.86
	C _{ijt} (Ton)	1,350	1,350	350	0	1,350	1,350	1,350	0	1,350	1,350	1,350	1,350
2	P _{ijt} (baht)	36.74	36.19	35.45	0	34.91	34.56	34.21	33.86	35.5	35.15	34.79	34.44
	C _{ijt} (Ton)	1,420	1,420	1,420	0	1,420	1,420	1,420	1,420	1,420	1,420	1,420	1,420

Table 2. Demand, Holding Cost and Expiration Cost

Period	Period 1				Period 2				Period 3			
Age j	1	2	3	4	1	2	3	4	1	2	3	4
I (kg.)		13,000										
H _{ikt} (Baht)	5	5	5	5	5	5	5	5	5	5	5	5
E _t (Baht)	11	11	11	11	11	11	11	11	11	11	11	11
D _t (kg.)	457,000				426,000				442,000			

Table 3. Transportation Costs and Weight per Round of Concentrated Latex for Each Transportation Type

Supplier i	Transportation type	T _{ikt} (Bath)	W _k (kg./Round)
1	10-wheel truck (k=1)	4,694	12,000
	Trailer (k=2)	8,605	25,000
2	10-wheel truck (k=1)	15,679	12,000
	Trailer (k=2)	21,877	25,000

Results

LINGO software is employed to solve this problem and the solution is presented in Table 4. The results show the optimal planning for purchasing, storage and transporting of perishable concentrated latex used in glove production over a 3-period time interval. The concentrated latex is purchased from both suppliers which purchasing cost is more sensitive to decision making than transportation and holding costs. Since the price of concentrated latex decreases when the age of concentrated latex goes up, so the purchasing decision is prioritized from the oldest to the youngest ones. For example, it can be seen from the planned purchase of concentrated latex of age 3 from supplier 1 and 2 in period 1, concentrated latex of age 4 from supplier 2 and of age 3 from supplier 1, and concentrated latex of age 4 from both suppliers in period 3. Then, the

sensitivity analysis is performed by changing the purchase price of concentrated latex from the original price ranging from -30% to 30%. The computational results for a 3-period time interval each purchase price in details are presented in Table 5. The results indicate that the volume of concentrated latex purchased from supplier 1 increases and supplier 2 decreases if the purchase price of latex changes by -30%, as shown in Table 6. In the opposite way, if the purchase price of latex changes by +30%, the volume of concentrated latex purchase from supplier 1 decreases and supplier 2 increases, as shown in Table 7. Moreover, the result from sensitivity analysis also shows that the 25-ton trailers are normally selected as the transportation type. Due to the aging of latex and its holding cost, the results also suggest that the rubber glove industry should purchase the concentrated latex with the only required amount for each period.

Table 4. Results

Supplier i	Age j	Period 1				Period 2				Period 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Purchase B_{ijt} (Ton)			344				1					442
	Transport X_{ijkt} (Round)												
	- 10-wheel truck (k=1)							1					
	- Trailer (k=2)			14									18
2	Purchase B_{ijt} (Ton)			100					425				
	Transport X_{ijkt} (Round)												
	- 10-wheel truck (k=1)												
	- Trailer (k=2)			4					17				
Use U_{jt} (Ton)		13	444				1	425					442
Ending Inventory I_{jt}^E (Ton)													
Eliminate F_t (Ton)													
Purchase Cost (Bath)		15,825,800				14,424,960				15,408,120			
Transportation Cost (Bath)		207,978				376,603				154,890			
Inventory Cost (Bath)													
Eliminate Cost (Bath)													
Total Cost (Bath)		46,398,351											

Table 5. Sensitivity Analysis for Changing Purchase Price

Supplier i	% Change	-30%	-20%	-10%	0%	10%	20%	30%
1	Purchase B_{ijt} (Ton)	1,212	1,212	787	787	787	787	362
	Transport X_{ijkt} (Round)							
	- 10-wheel truck (k=1)	1	1	1	1	1	1	1
	- Trailer (k=2)	49	49	32	32	32	32	15
2	Purchase B_{ijt} (Ton)	100	100	525	525	525	525	950
	Transport X_{ijkt} (Round)							
	- 10-wheel truck (k=1)							
	- Trailer (k=2)	4	4	21	21	21	21	38
Use U_{jt} (Ton)	1,325	1,325	1,325	1,325	1,325	1,325	1,325	1,325
Ending Inventory I_{jt}^E (Ton)								
Eliminate F_t (Ton)								
Purchase Cost (Bath)	32,138,480	36,732,840	41,090,020	45,658,880	50,228,740	54,788,920	59,125,030	
Transportation Cost (Bath)	513,847	513,847	739,471	739,471	739,471	739,471	965,095	
Inventory Cost (Bath)								
Eliminate Cost (Bath)								
Total Cost (Bath)	32,652,327	37,246,687	41,829,491	46,398,351	50,968,211	55,528,391	60,090,125	

Table 6. Volume of Concentrated Latex Purchase for Changing Purchase Price -30%

Supplier i	Age j	Period 1				Period 2				Period 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	- 10-wheel truck (k=1)							1					
	- Trailer (k=2)			344				425					442
2	- 10-wheel truck (k=1)												
	- Trailer (k=2)			100									

Table 7. Volume of Concentrated Latex Purchase for Changing Purchase Price +30%

Supplier i	Age j	Period 1				Period 2				Period 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	- 10-wheel truck (k=1)							1					
	- Trailer (k=2)			344									17
2	- 10-wheel truck (k=1)												
	- Trailer (k=2)			100					425				425

CONCLUSION

The main contribution of this paper is the proposed mathematical model to support decision planning over the multi-period interval, as well as planning for purchasing of concentrated latex from different regions, from multiple suppliers with multiple transportation types under age-dependent constraints. Purchase and storage of concentrated latex with constraint of perishability through a multi-period time interval is a complex problem. The supply volume differs in different regions according to seasonal fluctuations, and the concentrated latex is perishable over time if stored for too long. This model can also be extended by including the quality of concentrated latex in each age. In addition, the matching between glove product and concentrated latex prices under uncertainty could be interesting to explore.

REFERENCES

- Auckara-aree, K. and Boondiskulchok, R. 2010. "Designing the raw material collection system for profit maximization under a step-price policy", *Songklanakar Journal of Science Technology*, Vol. 32, 581-588.
- Fathoni Z. 2009. "Evaluation of market system and market integration for rubber cultivation in jambi province Indonesia", Wageningen University and Research.
- Haan J. de., Groot G. de., Loo E. and Ypenburg M. 2003. "Flows of goods or supply chains; lessons from the natural rubber industry in Kerala, India", *International Journal of Production Economics*, Vol. 81-82, 185-194.
- Klomsae, S., Somboonwivat, T. and Atthirawong, W. 2012. "Optimal multi-period planning for purchase and storage of rubber latex with perishability constraints", *Proceedings of the 7th International Engineering and Management System*, Kitakyushu, Japan, 582-590.
- Kritchanchai, D. and Chanpuyetch W. 2009. "Gateway selections for Thailand rubber export", *Proceedings of the Asia Pacific Industrial Congress on Logistics and SCM Systems*, 244-250.
- Liu, H., Zhang, J., Zhou, C. and Ru, Y. 2017. "Optimal purchase and inventory retrieval policies for perishable seasonal agricultural products", *Omega*, In press, 1-13.
- Wu, T., Shen, H. and Zhu, C. 2015, "A multi-period location model with transportation economies-of-scale and perishable inventory", *International*

Journal of Production Economics, Vol. 169, 343-349.

AUTHOR BIOGRAPHIES



TUANJAI SOMBOONWIWAT is an Associate Professor in the Industrial Management section, Department of Production Engineering Faculty of Engineering, King Mongkut's University of Technology Thonburi, Thailand. She received her M.Eng. in Industrial Engineering from Chulalongkorn University, Thailand and Ph.D. in Industrial Engineering from Corvallis, Oregon State University, USA. Her research interests include green supply chain and logistics, business process and applications of operations research. Her e-mail address is: tuanjai.som@kmutt.ac.th



SUTTHINEE KLOMSAE received M.Eng. in Industrial and Manufacturing Systems Engineering from King Mongkut's University of Technology Thonburi in 2006. She works at Uttaradit Rajabhat University. Her research interests are in area of logistics and supply chain management. Her e-mail address is: klomsae_aey@hotmail.com.



WALAILAK ATTHIRAWONG is an Associate Professor in Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Thailand. She received a Ph.D. in Manufacturing Engineering and Operations Management at the University of Nottingham, England in 2002. Her teaching and research interests include applied operations research, supply chain management, forecasting and location decision. Her e-mail address is: walailaknoi@gmail.com.

Geschäftsprozesse in Low-Code Plattformen mit einem SAP-ECC Backend

Vergleich zweier Plattformen auf Basis eines DMS-Prozesses

Daniel Degenhardt
Hochschule Fulda
Leipziger Str. 123
36037 Fulda

Daniel.Degenhardt@cs.hs-fulda.de

Norbert Ketterer
Hochschule Fulda
Leipziger Str. 123
36037 Fulda

Norbert.Ketterer@cs.hs-fulda.de

ABSTRACT

In dieser Arbeit wird untersucht, in wieweit sich Prozesse des Dokumentenmanagements in zwei Low-Code Plattformen abbilden lassen. Die eigentlichen Funktionen des Dokumentenmanagements werden dabei über einen Backend in Form eines SAP-ERP Systems mittels SOAP- und REST-Services implementiert. Insbesondere soll die Art der technischen Anbindung zwischen den jeweiligen Plattformen und dem Backend untersucht werden; speziell das Zusammenspiel zwischen Funktionen in Front- und Backend und die Verwendbarkeit der verschiedenen Services. Ein weiterer Aspekt der Fragestellung ist auch, wie die Low-Code Plattformen die Entwicklung der Funktionen des Dokumentenmanagements und deren Einbettung in den Geschäftsprozess unterstützen. Diese Frage soll durch einen Prototypen validiert werden. Als Untersuchungsgrundlage dienen die beiden Low-Code Plattformen „Bizagi“ und „Mendix“. Es wird in der Untersuchung ein Prozess des Dokumentenmanagements betrachtet, da hier die Parametrisierung der Aufrufe zum Backend besonders komplex ist, denn es sind statt der sonst üblichen, durch einfache Variablen oder Listen darstellbaren Parameter (z.B. Werke, Materialnummern und Kunden), die Identifikation der Dokumente aber auch die Originale der Dokumente zu übertragen.

Keywords

Dokumentenmanagement, SAP, Low-Code, Bizagi, Mendix, Digitalisierung Geschäftsprozesse, REST, SOAP

EINLEITUNG

Für einen kontinuierlichen unternehmerischen Erfolg ist die Nachvollziehbarkeit aller Information, die im Zuge der Durchführung der Geschäftsprozesse anfällt oder von diesen als Referenz benötigt wird, unabdingbar. In diesem Zusammenhang definiert sich ein Dokument als Trägereinheit von inhaltlich zusammengehörender Information, die sich in unterschiedlichen Strukturierungsgraden ausdrückt und als Nach-

weis oder Beleg für betriebliche Vorgänge dient [1] (S. 27f). Angesichts der stetig wachsenden Komplexität von betriebswirtschaftlichen Prozessen sind Unternehmen dazu veranlasst, sich mit dem effizienten Management ihrer Dokumente zu beschäftigen. Beispiele von Prozessen sind etwa PLM-Prozesse, in denen eine Dokumentation zum Zweck einer stetigen Produktqualität oder aufgrund von gesetzlichen Vorschriften wie Aufbewahrungsfristen für bestimmte Geschäftsdokumente existiert.

Gemäß Götzer behandelt das Dokumentenmanagement primär Verwaltungsfunktionen der Dokumente [1] (S. 15 ff), nicht die Inhalte der Dokumente selbst. Als Teilgebiet des Informationsmanagements zeichnet sich das Dokumentenmanagement durch eine starke Wechselwirkung mit anderen unternehmerischen Prozessen aus und erstreckt sich damit in Form von unterstützenden Prozessen über die gesamte Wertschöpfungskette des Unternehmens. Im Zuge der Einführung von Dokumenten-Management-Organisationen gilt es von den Fachabteilungen zu ermitteln, was für Dokumente auf welche Art und Weise verwaltet werden müssen und in welchem Zusammenhang diese mit den einzelnen Geschäftsprozessen stehen; danach kann festgelegt werden, wie dies durch den Einsatz eines Informationssystems unterstützt werden soll.

In dieser Arbeit werden die Dienste des Dokumentenmanagementsystems (genauer CA-DMS/ Cross - Application Components - DMS) des ERP-Systems in Services gebündelt, die von parallel zum ERP-System laufenden Applikationen genutzt werden können, die mit Hilfe der Low-Code Plattformen implementiert werden. Dieses Szenario betrifft also die Digitalisierung eines speziellen, einfachen Geschäftsprozesses mit Hilfe einer Low-Code Plattform, wobei dieser Prozess trotzdem Services des ERP-Backends nutzen soll, um auf relevante Information Zugriff zu haben. Innerhalb des Dokumentenmanagements stellen sich hierbei besondere Herausforderungen - speziell die Verwaltung der Originale.

PROBLEMDEFINITION

Prozesse im Dokumentenmanagement

Die betriebswirtschaftlichen Prozesse des Dokumentenmanagements lassen sich in zwei Ebenen, nämlich **Metaprozesse** und **operative Prozesse** untergliedern. Die Metaprozesse definieren hierbei den Handlungsrahmen, in dem sich die operativen Prozesse ausprägen können. Dieser ist

stark an die Verfügbarkeit von Funktionen des Dokumentenmanagements gekoppelt, etwa

- der Festlegung der Dokumentenarten, der
- Definition der Verknüpfungslogik mit Objekten des Anwendungssystems, den verwendeten
- Statusnetzen und dem Bezug zum Klassensystem.

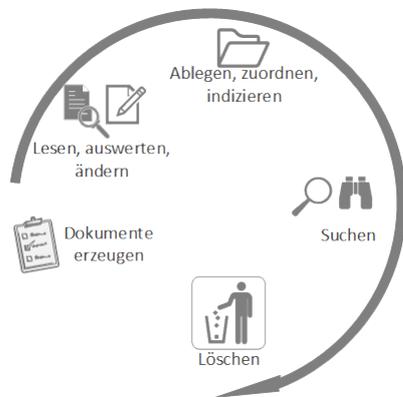


Abbildung 1: Dokumentenlebenszyklus nach Götzer [1], Kapitel 1

Dokumentenverwaltung im ERP-System

Die Architektur eines Dokumenten-Management-Systems (kurz „DMS“) besteht aus einer Anzahl von zueinander abhängigen Komponenten (Abbildung 2), die sich wiederum in eine Menge von Subkomponenten und -funktionen untergliedern lassen. Die Komponente „Eingang“ und „Klassifizierung“ muss gewährleisten, dass einerseits nur Dokumente dem System zugeführt werden können, die durch die Metaprozesse definiert sind, und dass diese darüber hinaus nach den Vorgaben der Dokumentenart ausreichend indiziert werden. Die Komponente „Ablage“ und „Archivierung“ stellt die Schnittstelle zum Transport von Nutzinhalten zwischen den Massenspeichersystemen (wie Dateisystemen, in der Abbildung als RAID- und File-System dargestellt) und dem DMS. Zur Eingrenzung des Ergebnisraums von Suchanfragen und zur Darstellung von einzelnen Dokumenten werden Funktionalitäten der „Recherche“ und der „Anzeige“ genutzt. Werden die Dokumenten im DMS auch physisch abgelegt und nicht nur verwaltet (also nur „Metainformation“ abgelegt), werden sie in besonderen Speicherstrukturen gespeichert. Dies sind speziell strukturierte Datenbanktabellen oder auch Dateisysteme.

Ein Beispiel eines Flusses der Dokumenteninformation durch das DMS zeigt Abbildung 3; das Dokument wird in Nutz- und Verwaltungsinformation aufgeteilt und in den verschiedenen Komponenten des DMS abgelegt. Dieser Fluss setzt den Dokumentenlebenszyklus nach Abbildung 1 um. Es ist im Rahmen der Fragestellung zu untersuchen, welche Funktionen im DMS des ERP-Systems genutzt werden können, um diesen Fluss zu implementieren und wie diese durch Services in den Low-Code Plattformen nutzbar sind.

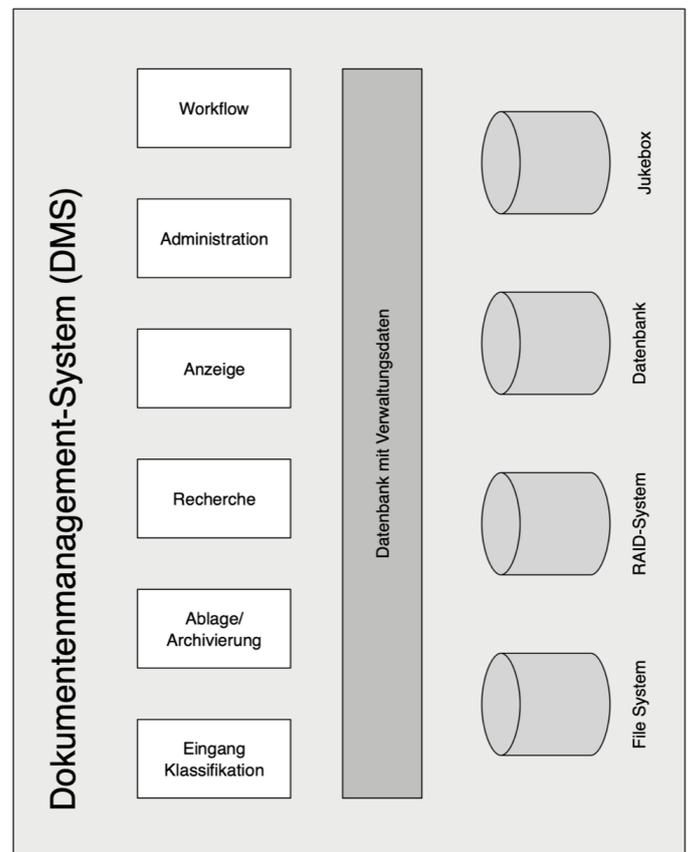


Abbildung 2: Architektur eines DMS - nach [2](S. 29)

Implementierung von Backend-Services im ERP-System

Um die Funktionen des Dokumentenmanagements durch die Low-Code Plattform ansprechen zu können, sind diese im ERP-System zu identifizieren, zu gruppieren und als Service zu veröffentlichen, dass sie sinnvoll durch die Low-Code Plattform angesprochen werden können. Eine wesentliche Frage wird hierbei sein, wie die Verwendung von aktuellen Kommunikationsmitteln, wie SOAP und REST die Implementierung im Backend und die Einbindung in die Low-Code Plattform beeinflussen.

Softwareentwicklung in den Low-Code Plattformen

Das Dokumentenmanagement ist eine Teilfunktion, die innerhalb eines digitalisierten Gesamtprozesses notwendig sein kann und im Falle dieser Untersuchung durch ein zentrales Backend bereitgestellt werden soll. Hierzu sind in der jeweiligen Low-Code Plattform die Geschäftsprozesse gemäß der individuellen Methodik der Plattform abzubilden, um dort dann die Funktionen des Dokumentenmanagements, die durch das Backend bereitgestellt werden, aufrufen zu können.

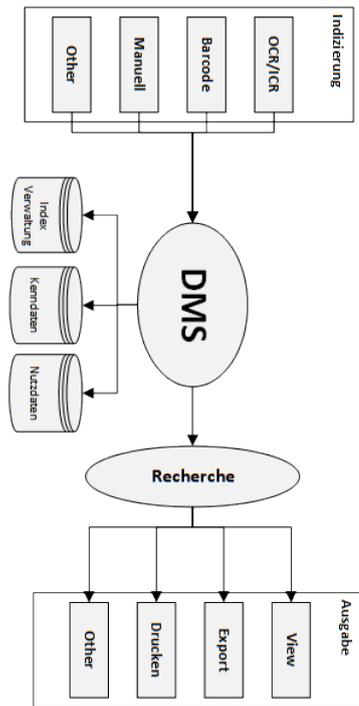


Abbildung 3: Dokumentenfluss innerhalb eines DMS - nach [1] (S. 164)

STAND DER TECHNIK

Architektur von SAP-DMS

Das Backend, welche als ERP-System in dieser Untersuchung die Funktionen des CA-DMS zur Verfügung stellt, wird zusammen mit der NetWeaver-Plattform ausgeliefert und bietet aufgrund der vollständigen Integration des CA-DMS in das ERP-System einen direkter Bezug zwischen den Dokumenten und den operativen Geschäftsprozessen; i.d.R. ist eine direkte Verknüpfung von Dokumenteninfosätzen und den durch die Geschäftsprozesse erzeugten Objekte möglich (etwa Kunden, Materialien, Aufträge). In diesem Kontext kann dann ein Dokument jegliche Form von speicherbarer Information darstellen, wie technische Zeichnungen, Grafiken, Programme oder einfach Textdokumente [3], die als Originale abgelegt werden können, die dann mit Dokumenteninfosätzen verknüpft sind. Diese Infosätze können selber wieder miteinander verknüpft oder auch klassifiziert werden (neben der Originalverknüpfung und Geschäftsobjektverknüpfung). Die Definition eines Dokuments beschränkt sich also nicht nur auf Dokumente in Papierform, sondern erweitert den Begriff auf beliebige Informationskonstellationen.

Bei Dokumenteninfosätze handelt es sich um Objekte, welche die Ausprägungen von Metadaten eines einzelnen Dokuments innerhalb der DMS-Datenbank abspeichern und somit eine Repräsentation des Nutzdokuments darstellen. Weiterhin definieren Dokumenteninfosätze diverse Schnittstellen zu internen Geschäftsobjekten (Materialien, Kunden, Lieferanten) und externen Diensten, über welche die Nutzdokumente in das Dateisystem abgelegt werden.

Zur Verwaltung von Dokumenten, die primär außerhalb des ERP-Systems erzeugt werden, eignet sich insbesondere der

sogenannte **Knowledge Provider (KPRO)** als Vermittler zwischen dem DMS- und dem Dateisystem. Dieser greift auf Funktionen zurück, durch die eine bidirektionale Kommunikation mit dem ERP-System gewährleistet werden kann. Wie sich der Abbildung 4 entnehmen lässt, kommunizieren die einzelnen Applikationen mit der für sie zuständigen Komponente innerhalb des KPROs, der wiederum die Kommunikation mit dem Content-Server übernimmt.

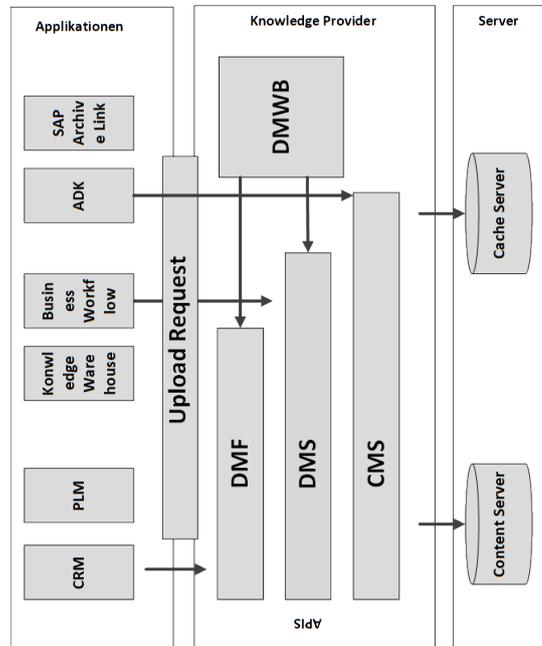


Abbildung 4: Architekturkomponenten SAP-DMS - nach [4]

Die Anlage eines Dokumenteninfosatzes erfordert intern im DMS als ersten Schritt das Definieren eines Primärschlüssels. Das System validiert die Einstellung der Dokumentenart und sichert damit zu, dass nur Dokumenteninfosätze mit einer gültigen Dokumentenart unter den speziellen Regeln der Dokumentenart angelegt werden können; etwa einem speziellem Statusnetz. Anschließend kann der Dokumenteninfosatz nach den zusätzlichen Kriterien der Dokumentenart genauer spezifiziert werden; etwa dem Einchecken der Originale, der Verknüpfung weiterer Objekte der Geschäftsprozesse, Klassifikationen, Hierarchisierungen, der Verwaltung der Status und von Signaturen - ein Beispiel der Anlage zeigt Abbildung 5.

Das Screenshot zeigt die SAP-GUI für die Anlage eines Dokumenteninfosatzes. Die Überschrift lautet 'Dokument anlegen: Einstieg'. Es gibt eine Tabelle mit den folgenden Daten:

Dokument	123456
Dokumentart	DRW
Teildokument	000
Dokumentversion	00

Abbildung 5: Anlage eines Dokumenteninfosatzes in der SAP-GUI

Es existiert ein OASIS-Standard auf Basis des SOAP-Protokolls (CMIS), der die Interoperabilität zwischen Enterprise-Content-Management-Systemen vereinheitlicht. Er wird zwar von einer der verwendeten Low-Code Plattformen unterstützt, aber nicht von dem ERP-System, weshalb er nicht weiter betrachtet werden soll.

Service-orientierte Architekturen

Die Geschäftsprozesse der Fachabteilungen lassen sich in einzelne Aktivitäten, die elementare Aufgabenblöcke repräsentieren, zerlegen. Es sollen deshalb Services erzeugt werden, die die Aktivitäten implementieren, soweit sie das ERP-System als Backend betreffen. Auf Basis der bestehenden Systemlandschaft lassen sich somit Services generieren, die ein Abbild der Aufgabenblöcke repräsentieren; einzelne Services können dann über die Low-Code Plattformen zu individuellen Applikationen innerhalb oder über Unternehmensgrenzen hinweg kombiniert werden.

SOAP und REST Web-Services

Ein Ansatz zur Realisierung einer Service Orientierten Architektur ist die Bereitstellung der zuvor beschriebenen „Aufgabenblöcke“ in Form von Web-Services. Diese bilden hierbei eine Abstraktionsschicht, die es ermöglicht, Funktionalitäten der eigenen Systemlandschaft unabhängig von der zugrundeliegenden Technologie plattformübergreifend durch den Einsatz des SOAP-Protokolls bereitzustellen, wie dies in Abbildung 6 skizziert ist; die Beschreibung der Schnittstelle erfolgt dabei durch die WSDL [5] (S.61).



Abbildung 6: Abstraktionsschicht von Web-Services

Eine WSDL-Datei stellt einen in XML spezifizierten Vertrag zwischen den Web-Service-Akteuren dar und beschreibt Anforderungen, die ein Dienstnutzer erfüllen muss, damit er eine bestimmte Funktion des Diensteanbieters in Anspruch nehmen kann.

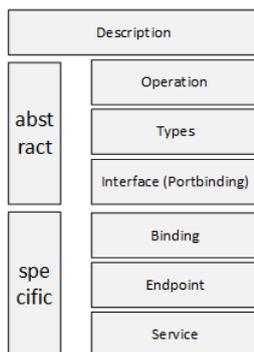


Abbildung 7: Aufbau einer WSDL-Datei nach Melzer, plus Types-Komponente [6] (S. 116)

Die WSDL-Datei kann somit als Metadaten-Dokument angesehen werden. Innerhalb der WSDL-Datei wird der Web-Service aus zwei Ebenen beschrieben, die sich in die Schnittstellenbeschreibung und die Implementierungsbeschreibung

aufgliedern (Abbildung 7)). Die **abstrakte Ebene** beschreibt, was der Dienst anbietet und wie dieser implementiert ist; die **konkrete Ebene** beschreibt, wie der Dienst aufgerufen werden kann.

Ein weiterer relevanter Ansatz zur Realisierung von Web-Services, setzt auf dem REST- Architekturstil nach Roy Fielding auf. Dieser ermöglicht eine Kommunikation zwischen Diensten über das Internet hinweg, ohne die Notwendigkeit der SOAP- Abstraktionsschicht.

REST zielt darauf ab, lose gekoppelte, skalierbare und zugleich verteilte Systeme zu erstellen [7] (S. 16 f). Hierbei kommt kein spezielles Nachrichten-Austausch-Protokoll zum Übermitteln von Nutzdaten wie etwa SOAP zum Einsatz. Der Austausch von Nutzdaten beschränkt sich auf das HTTP/S Protokoll. Der Dienstanbieter übermittelt die Nutzdaten eingebettet in der Zeichenkette des URL an den Dienst. Die Idee des REST-Ansatzes besagt, dass jede Art von beschreibbaren Daten eine Ressource darstellt, die durch eine URI eindeutig identifizierbar ist [7] (S. 7 ff). Ressourcen können durch die HTTP-Methoden (GET,POST,PUT,DELETE) im Sinne der CRUD-Operationen modifiziert werden. Im Gegensatz zum SOAP-Ansatz, der sich ausschließlich auf XML festlegt, kann eine Ressource im REST-Ansatz neben XML weitere Formate wie JSON, TXT, HTML usw. annehmen. Um mit einem REST basierten Service kommunizieren zu können, bedarf es keinem Service-Vertrag, wie es bei den SOAP-Services durch die WSDL verlangt wird. Die einfache Kenntnis des Service Endpunkts in Verbindung mit Authentifizierungsaspekten reicht aus, um eine Kommunikation zu ermöglichen.

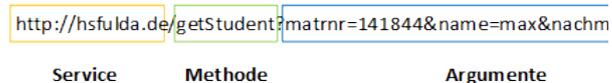


Abbildung 8: Aufruf eines fiktiven REST-Service

Der REST-Ansatz unterliegt keinerlei Standardisierung, wie es bei SOAP-Web-Services der Fall ist. Dies lässt sich darauf zurückführen, da es sich bei REST lediglich um ein Architekturparadigma handelt. Der Umstand bietet somit Interpretationsspielraum für die Ausgestaltung und Implementierung von Web-Services auf Basis der REST-Idee. Um diese Services dennoch nach qualitativen Kriterien beurteilen zu können, hat Richardson ein Reifegradmodell entwickelt, das eine Orientierungshilfe für die Beurteilung von REST-Services darstellt (Abbildung 9).

Auf Level 2 werden die HTTP-Operationen im Sinne der CRUD-Operationen genutzt, über welche die URI modifiziert wird. Eine Ressourcen-Repräsentation findet ausschließlich durch GET-Befehle und eine Ressourcen-Änderungen ausschließlich durch POST- Befehle statt - die CRUD Operationen werden auf Level 1 nicht verwendet. Auf Level 3 stellt der Dienstanbieter dem Dienstnutzer einzelne Funktionen und Information als „Hyper-Media-Control“ bereit. Der Client kann dadurch mit allen weiteren dem Dienst zugeordneten Ressourcen kommunizieren, ohne initial deren Lokalisation zu kennen. Dies lässt die Möglichkeit offen, einzelne andere Dienste/Ressourcen abzuändern ohne dass dadurch ein Problem beim Client auftritt. Der Client ist somit stets

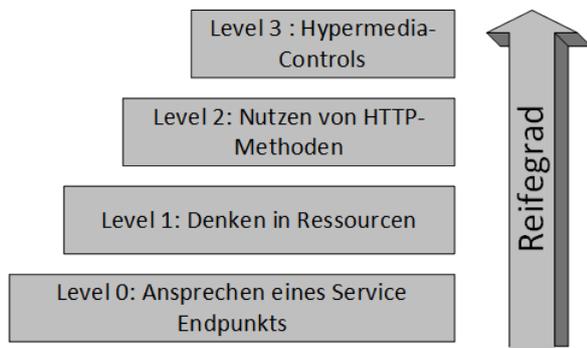


Abbildung 9: Richardson Maturity Model [8]

über den aktuellsten Ressourcenzustand informiert. Rest- und SOAP-basierte Services werden in Tabelle 1 kurz gegenübergestellt.

SOAP-Web-Services	REST-Web-Services
Erfordert eine client-seitige SOAP-Library	Keine Library erforderlich
Alle Aktionen via HTTP-POST	Nutz CRUD-Operationen des HTTP-Protokolls
Kann zustandsbehaftet oder zustandslos sein	Ist zustandslos
N Ressourcen für N Aktionen	Eine Ressource für N Aktionen
Unterstützt nur XML	Unterstützt XML, JSON, TXT usw.
WSDL beschreibt sehr genau wie der Dienst aufgebaut ist	Keine Vertragsvereinbarungen notwendig

Table 1: SOAP vs. REST

Web-Services innerhalb von NetWeaver

In der NetWeaver-Plattform des Prototypen existiert ein Modul, welches das Empfangen und Versenden von HTTP-Nachrichten gewährleistet. Die technische Grundlage zur internetbasierten Kommunikation bildet der Internet-Communication-Manager und das Internet-Communication-Framework. Die beiden Dienste sind in der Lage, unterschiedliche Protokolle wie SMTP, HTTP/S zu verarbeiten. Unabhängig von dem gewählten Web-Service-Ansatz (SOAP/REST) ist stets die Kombination der beiden Dienste für die Nachrichtenverarbeitung zuständig. Eingehende HTTP-Anfragen können durch den ICM modifiziert und weitergeleitet werden. Nach der initialen Annahme der Nachricht wird diese an das ICF weitergereicht. Dieses entpackt die HTTP-Nachricht und reicht den Aufruf an die zuständigen Programme also die eigentlichen Dienste weiter ([9], S.958).

SOAP-Services können direkt erstellt werden; insbesondere mit Hilfe des eingebauten Web-Service-Wizards, der es ermöglicht, innerhalb von neun Schritten einen Funktionsbaustein web-fähig zu machen. Es muss hierzu im Wesentlichen der Service-Name und der Endpunkt, der die Geschäftslogik beinhaltet, angewählt sowie der Service konfiguriert werden (Zertifikate, Autorisierung). Der erstellte Service kann anschließend im Repository-Browser unter dem Punkt der Enterprise-Services gefunden werden. Damit der

Service von außerhalb des NetWeaver-basierten Systems angesprochen werden kann, muss für diesen noch ein konkreter Eintrag im ICF hinterlegt werden. Die Kommunikation zwischen einem Dienstaufrufer und Dienstanbieter zeigt Abbildung 10: Eine eingehende Anfrage wird durch das Internet-Communication-Framework an die SOAP-Runtime weitergeleitet. Diese wird die SOAP-Nachricht interpretieren und den Nutzinhalt an den Funktionsbaustein weiterreichen, der daraufhin die Dienstleistung durchführt. Das Ergebnis des Aufrufs wird wiederum durch die Runtime in eine SOAP-Nachricht eingebettet und durch die Dienste des ICF und ICM in Form einer HTTP-Nachricht an den Dienstaufrufer zurückgesendet.

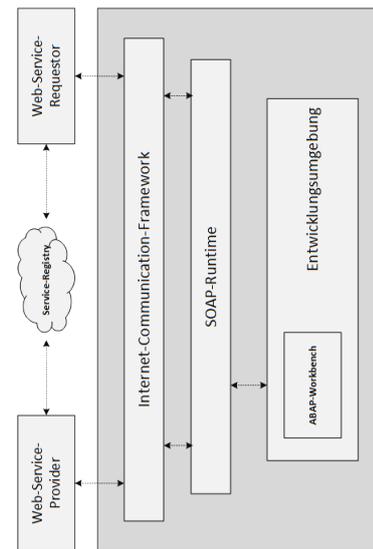


Abbildung 10: Schema der SOAP-basierten Nachrichtenverarbeitung [9] (S. 978)

Ein REST-basierter Service ist dagegen nicht so direkt in der verwendeten NetWeaver-Version aus einer Funktion erzeugbar, wie im Falle eines SOAP-Service. Deshalb wurde für die Untersuchung prototypisch eine ABAP-Klasse programmiert, die das Interface „IF_HTTP_EXTENSION“ implementiert. Das Interface verfügt ausschließlich über eine Methode „Handle_Request“, die gewährleistet, dass eine vom ICF übertragene HTTP-Nachricht durch die Klasse behandelt wird. Durch dieses Interface wird das ABAP-Programm in die Lage versetzt, HTTP-Nachrichten zu verarbeiten. Im ICF-Baum kann dann ein Service angelegt werden, der über eine URL ansprechbar ist. Dem Service wird dabei die ABAP-Klasse als Handler zugewiesen. Zur Laufzeit wird der ICF-Baum per Top-Down-Ansatz nach dem gewünschten Dienst abgesucht und es erfolgt eine dynamische Zuweisung der Handler-Klasse und die Übertragung der Nachricht an diese.

Low-Code Plattformen und Software Engineering

Durch die Zuhilfenahme von Low-Code Plattformen sollen Fachabteilungen befähigt werden, insbesondere für einfache Prozesse eigene Applikationen zur Unterstützung bei den anfallenden Geschäftsprozessinstanzen zu implementieren. Die hierfür notwendige Komplexitätsreduktion wird durch die Bereitstellung von Modellierungswerkzeugen gewährleistet.

Die Plattformen ermöglichen es den Unternehmen, schnell durch Modellierung Software zu erstellen, können aber auch angebundene Geschäftsanwendungen (wie eine Landschaft von OLTP-Systemen) mit einem einheitlichen User-Interface belegen und diese zusätzlich um Spezialfunktionen ergänzen.

Ausgehend von einem fachlichen Gegenstandsbereich werden Elemente und Beziehungen beschrieben, die durch eine formale Notation (z.B. UML- oder BPMN) spezifiziert werden. In mehreren Schritten werden die erzeugten Modelle anschließend auf eine darunter liegende Ebene transformiert und letztlich von den Plattformen in Quellcode übersetzt (in Anlehnung an [10]).

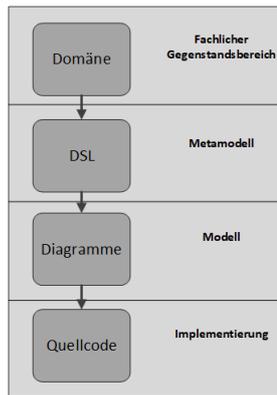


Abbildung 11: Vorgehensweise der modellbasierten Software-Entwicklung [10] (S.19)

Typisch gegenüber der herkömmlichen Software-Entwicklung ist, dass die Plattformen bereits Lösungen für Anforderungen mitliefern, die in vielen Software-Projekten eine Rolle spielen. So gibt es oftmals eigene Module für die Benutzerverwaltung, welche durch die Anbieter stetig gewartet und verbessert werden. Darüber hinaus müssen sich Anwender nicht mit der Installation der mitunter komplexen Infrastrukturen für Entwicklung, Deployment und Wartung befassen. Die Verwendung dieser Angebote resultiert in schnelleren Entwicklungszyklen und bringt Kostenvorteile für Nutzer der Plattformen. Die Plattformen bestehen aus vorgefertigten Komponenten, die eine klar definierte Schnittstelle besitzen und aus denen die Applikationen zusammengesetzt werden. Soll eine Komponente in die eigene Applikation eingebunden werden, ist es nicht erforderlich, ein Verständnis für die zugrundeliegende Logik der Komponente zu besitzen. Es reicht aus, wenn die Schnittstellenanforderungen zur Erbringung der Dienstleistung erfüllt werden. Die Komponenten übernehmen beispielsweise Aufgaben zur Integration der Bestandteile der Applikation im Sinne eines Service-Busses, der Datenbankanbindung, der Benutzerschnittstelle, der Integration mit Um-Systemen aber auch Spezialaufgaben, wie das Bereitstellen eines Mail-Clients.

Low-Code Plattform Bizagi

Bizagi ist ein Business-Process-Management-System, das es Anwendern ermöglicht, direkt auf Basis von **Geschäftsprozessen** Applikationen zu erstellen. Hierbei repräsentieren diese digitale Abbilder der Geschäftsprozesse in BPMN und zielen darauf ab, die Zusammenarbeit von menschlichen und maschinellen Akteuren zu organisieren. Innerhalb des

Bizagi-Modelers besteht die Möglichkeit, Geschäftsprozesse zu modellieren. Das Bizagi-Studio ermöglicht ebenfalls eine Prozessmodellierung (mit Einschränkungen, z.B. bei der Verwendung von Pools/ Messages), aber dafür dann die Definition der Datenmodellebene und der Front-End-Ebene sowie die Anreicherung der Geschäftsprozesse um Geschäftsregeln und technische Komponenten. Der Aufbau der Plattform ist analog zur Architektur einer Geschäftsanwendung - besteht also aus Frontend, Geschäftsregeln und Datenbank-anbindung (Abbildung 12).

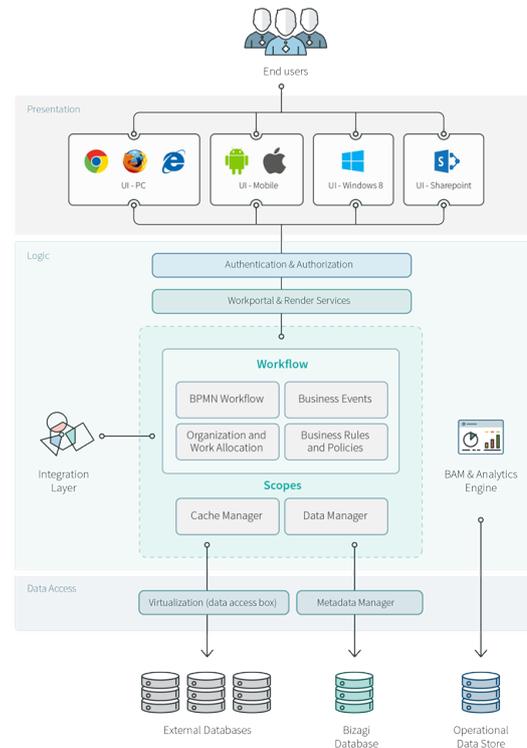


Abbildung 12: Architekturmodell von Bizagi [11]

Das Vorgehensmodell von Bizagi besteht aus der iterativen Anwendung der folgenden Schritte:

1. Erstellung eines **Geschäftsprozessmodells**,
2. Erstellung eines **Datenmodells**,
3. Erstellung des **Front-End-Modells**,
4. Erstellung von **Geschäftsregeln**.

Das Geschäftsprozessmodell wird mit Hilfe von BPMN und das Datenmodell wird mit Hilfe eines E-R-Diagramms dargestellt. Somit entsprechen diese beiden Schritte dem Standardvorgehen innerhalb einer prozessbezogenen Systementwicklung. Die Gestaltung des Front-Ends und der Geschäftsregeln ist etwas toolspezifischer, weshalb hier nur auf die letzten beiden Schritte eingegangen werden soll.

Die Komponenten der FORMS-Umgebung erstellen das Modell für die Präsentationsebene und verknüpfen dieses mit der Geschäftsebene und dem Datenmodell. Dabei stellen Geschäftsprozessaktivitäten des Typs „User-Task“ das In-

terface zu den Anwendern dar und können mit Front-End-Elementen angereichert werden.

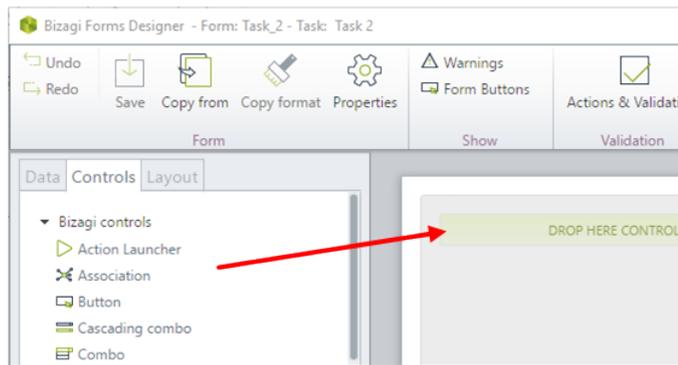


Abbildung 13: Standardkomponenten für das Bizagi-Front-End

Ein wesentlicher Teil der Frontendgestaltung ist das Abbilden der Attribute des Datenmodells auf die Front-End-Komponenten. Hier kann der Designer auf einen Pool an Front-End-Komponenten zurückgreifen, wie Buttons, Listen oder Druckausgaben (Abbildung 13) aber auch externe Komponenten, etwa Spracheingabe oder über eine vordefinierte API auf einen Amazon Cloud Connector. Auch kann die Datenvalidität geprüft werden, etwa indem bestimmte Eingabefelder als Pflichtfelder für Endanwender deklariert werden oder Validierungsregeln explizit erstellt werden (Abbildung 14).



Abbildung 14: Regeln zur Datenvalidität

Zur Steuerung der Applikation nach organisatorischen Vorgaben bietet Bizagi zwei Ansätze, um den Sequenzfluss geschäftsprozesskonform zu implementieren:

- **Einfache Businessregeln** - einfache Validierungsregeln, welche Variablen, die zur Laufzeit mit Daten befüllt sind, mittels Gateways auswerten und auf Basis der Ergebnisse den BPMN-Sequenzfluss lenken; Definition Analog zu Abbildung 14.
- **Activity-Actions** - graphisches Programmiertool, um komplexere Verarbeitungslogiken abzubilden, wobei diese auch an Ereignisse und an Elemente der Eingabemaske gehängt werden können (Abbildung 15).

Externe SOAP- und REST-Services können eingebunden werden, wobei im Falle von SOAP-Services die Spezifikation

nen der WSDL-Datei in Bizagi-Aktionen und -Funktionen umgewandelt werden kann.

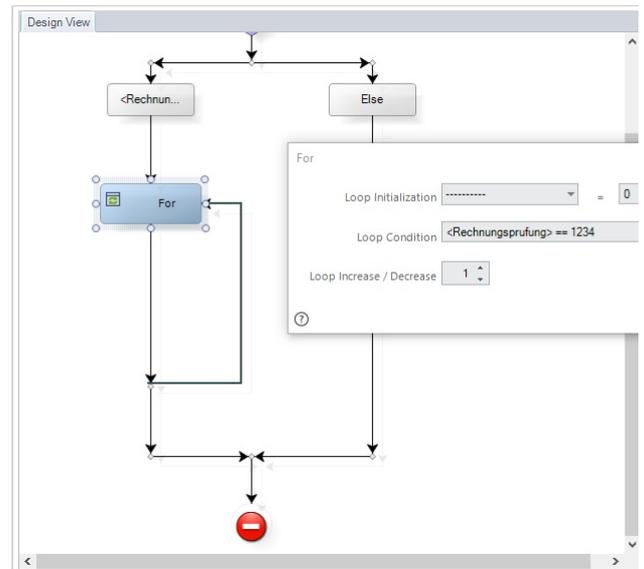


Abbildung 15: Grafische Modellierung von Programmierlogiken

Low-Code Plattform Mendix

Der Fokus von Mendix liegt auf der Erstellung individueller Applikationen, die im eigenen Unternehmen und darüber hinaus zur Unterstützung der Geschäftsprozesse eingesetzt werden können. Im Gegensatz zu Bizagi implementiert Mendix aber nicht direkt den Geschäftsprozess, sondern Ausgangspunkt ist eine User-Story, die potentiell in verschiedene Applikationen abgebildet wird.

Das Vorgehensmodell von Mendix basiert auf den folgenden Schritten:

1. Erstellung der **User-Stories**,
2. Umsetzung der User-Stories in ein **Front-End-Modell**,
3. Erzeugung eines **Datenmodells** auf Basis des Front-End-Modells,
4. Festlegen der **Applikationslogik** durch das Business-Modell,
5. **Deployment** und **kontinuierliche Anpassung** der Applikation.

Die Grundarchitektur von Mendix besteht wieder aus einem 3-Schichtenmodell mit dem User-Interface, der Applikationsebene sowie der Datenbankschnittstelle (Abbildung 16).

Die Möglichkeiten der Gestaltung der Nutzeroberfläche reichen hier von der Gestaltung von Navigationsmenüs über farbliche Hervorhebungen bis hin zu Ausgestaltung einzelner Front-End-Elemente wie Buttons, Listen, Views und Widgets.

Auf Basis des UI-Modells wird eine Datenrepräsentation generiert, welche die Front-End-Elemente in Form von Datenobjekten widerspiegelt. Entitäten werden in Mendix als Objekte im Sinne der objektorientierten Programmierung

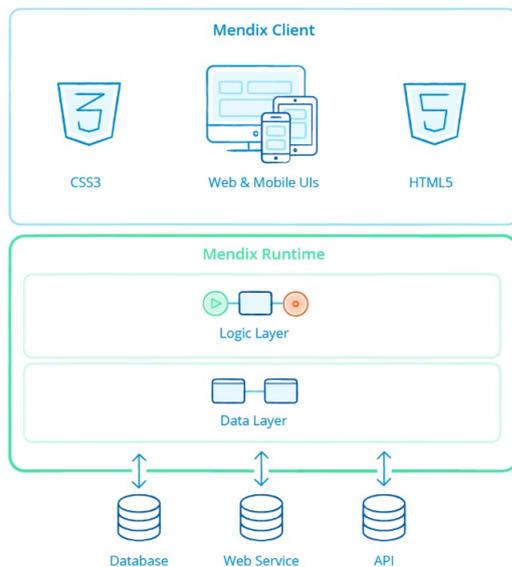


Abbildung 16: Architekturmodell von Mendix [12]

verstanden, die sich durch das Modellierungsframework der Datenmodellebene definieren lassen. Dabei können die Attribute sowie die Entitätsbeziehungen definiert werden.

Der Rahmen zur Formulierung der zeitlichen und logischen Abfolge der Schritte der Applikation wird in Mendix durch Elemente der eigenen Modellierungssprache definiert. Die Elemente ähneln stark einer BPMN-Modellierung, wobei in Mendix (im Unterschied zu Bizagi) der Ausgangspunkt nicht der Geschäftsprozess darstellt, sondern die durch die User-Stories definierten Applikationen, welche nach einem objekt-orientierten Programmiermodell umgesetzt werden.

Das Ausformulieren von Programmlogik kann entweder in einem Microflow [13] oder einem Nanoflow [14] modelliert werden. Nanoflows können nur Elemente verwenden, die sich offline ohne JAVA ausführen lassen. Daher eignen sich diese insbesondere für die Modellierung von Logiken auf dem UI-Layer. Für komplexere Geschäftslogiken empfiehlt sich der Einsatz von Microflows, da diese auf den vollen Sprachumfang zugreifen können. Innerhalb der Flows liegt der Fokus dann, gemäß dem objektorientierten Programmierparadigma von Mendix, auf dem Erzeugen und Verarbeiten von Objekten, die innerhalb des Datenmodells definiert werden. Tabelle 2 zeigt tabellarisch den Unterschied zwischen beiden Flow-Modellen.

Zur Verbindung von Mendix mit weiteren externen Diensten, existieren eine Reihe von Konnektoren; zusätzlich zu den Web-Service-Konnektoren, welche die Einbindung von eigenen SOAP- und REST-Services ermöglichen, besteht die Möglichkeit, spezielle Protokolle wie MQTT oder ODATA anzusprechen.

		Microflow	Nanoflow
Laufumgebung		Server	Client (browser/-device)
Sprache		Java / Scala	Java Script
Offline Apps	Mobile	Werden nicht unterstützt	Werden unterstützt

Table 2: Micro- vs. Nanoflow

Aufbau des Prototypen Implementierung der Backend-Services

Die Backend-Services werden mit Standardbausteinen des ERP-Systems implementiert - egal ob der Service später als SOAP oder REST mit der Low-Code Plattform verbunden wird. Hierzu werden teils gleiche Funktionsbausteine durch Wrapper-Funktionen umschlossen, um die Komplexität der Schnittstellen zu reduzieren und durchgehend RFC-fähig zu machen.

Die wesentlichen Funktionsbausteine in der verwendeten Version des ERP-Systems sind:

- **CVAPL_DOC_GETDETAIL** - liefert alle betriebswirtschaftlichen und technischen Details zu einem Dokumenteninfosatz zurück.
- **SDOK_PHIO_GET_URL_FOR_GET** - liefert eine URL, die einen externen Methodenaufruf auf dem Content-Server durchführt.
- **SDOK_PHIO_LOAD_CONTENT** - konvertiert ein Original-File in eine Binärdarstellung (zur Ablage im Contentbereich des DMS).
- **CMS_BINARY_TO_XSTRING** - konvertiert einen Binärstream in einen XString.
- **SSFC_BASE64_ENCODE** - kodiert einen XString in einen BASE64-String.
- **SSFC_BASE64_DECODE** - dekodiert einen BASE64-String in einen XString.
- **BAPL_DOCUMENT_CHANGE2** - Standard-API zum Abändern von Dokumenteninfosätzen.
- **BAPL_DOCUMENT_CREATE2** - Standard-API, der zum Erzeugen von Dokumenteninfosätzen eingesetzt wird.

SOAP Ansatz

Da eine Implementierung von SOAP-fähigen Web-Services in der verwendeten NetWeaver-Version direkt „out of the Box“ möglich ist, müssen die obigen Bausteine nur in neue Funktionen zusammengestellt werden, für die dann ein neuer SOAP-basierter Web-Service erzeugt wird, um in die entsprechende Funktion der Low-Code Plattform als Web-Service eingefügt werden zu können. Im Wesentlichen handelt es sich dann um Funktionen zum Erzeugen, Ändern und Lesen von Dokumenten, wobei hier wichtig ist, die Möglichkeit zu besitzen, auch eine Originaldatei zu verarbeiten.

REST Ansatz

Es müssen zunächst auf Basis des ICF die REST-Services erzeugt werden, da die verwendete NetWeaver-Version des ERP-Systems die Erzeugung noch nicht direkt „out of the Box“ unterstützt. Beispielhaft wird ein REST-basierter Service zum Anzeigen von Dokumenten genauer betrachtet.

Die Implementierung der Klasse Z_SHOW_Document folgt den im Stand der Technik zu den Rest-Services vorgestellten Kriterien und bietet die Möglichkeit, mittels einer URL auf Originale in Dokumenteninfosätze (Ressource) von außerhalb des SAP-Systems zuzugreifen. Einem aufrufenden REST-Client wird durch den Web-Service Z_SHOW mitgeteilt, über welche URI er auf das Originaldokument des angefragten Infosatzes, das innerhalb des Content-Servers liegt, zugreifen kann (Abbildung 17). Hierbei geht die Spezifikation der Dienstleistung innerhalb der Klasse Z_SHOW_Document dem Erzeugen des Services voraus. Die Klasse bindet das in im Stand der Technik erwähnte Interface ein und verpflichtet sich damit, die Methode Handle_Request zu implementieren.

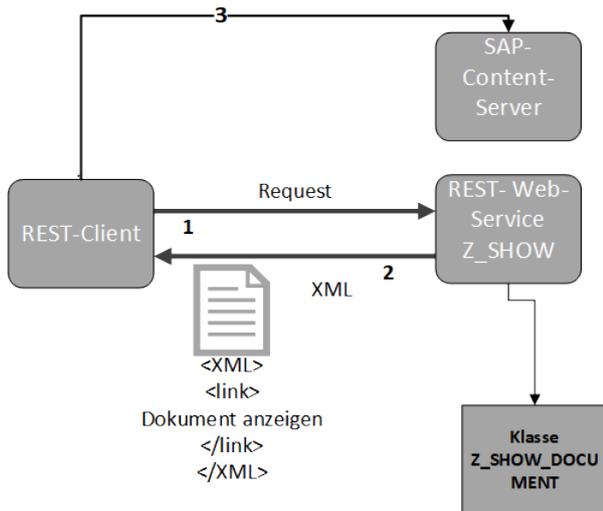


Abbildung 17: Diagramm des HATEOAS-REST-Services

Damit ein Client überhaupt auf die Dienstleistungen der Klasse zugreifen kann, ist es notwendig einen Endpunkt anzulegen, über den diese ansprechbar sind. Hierzu kann innerhalb des ICF-Service-Baums, der sich über die Transaktion SICF aufrufen lässt, ein Service-Eintrag definiert werden. Dem Service Z_SHOW wird hierbei die Klasse mit der Dienstleistung zugewiesen. Dieses Vorgehen lässt sich analog zu der im Stand der Technik beschriebenen Methodik verstehen.

Dem REST-Client steht somit ein zentraler Zugriffspunkt zur Verfügung, über den er sich Zugang zu den Funktionalitäten des SAP-Dokumenten-Managements verschaffen kann. Es ist durchaus denkbar, die Klasse mit weiteren Funktionalitäten und einer Differenzierung zwischen einzelnen Methoden zu erweitern.

Implementierung der Low-Code Anwendungen

Implementierung in Bizagi

Da die Anwendungsentwicklung unter dieser Plattform auf der Definition eines Geschäftsprozesses basiert, wird hier ein Beispielprozess in BPMN zugrunde gelegt. Der wesentliche Aspekt dieses Prozesses sind dabei nur die Aufrufe des Backends im Rahmen des Dokumentenmanagements - etwa durch eine Anlage eines Infosatzes im DMS des ERP-

Systems. Den relevanten Ausschnitt eines Prozesses aus Bizagi zeigt Abbildung 18; die Aktivität „Infosatz an SAP übermitteln“ beinhaltet den Aufruf in das Backend.

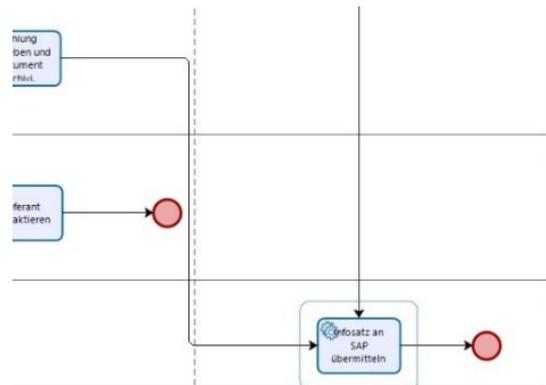


Abbildung 18: Ausschnitt Prozessmodells - Prozess „Rechnungsprüfung“

Auf die Darstellung des Datenmodells soll hier verzichtet werden - im Wesentlichen muss das Datenmodell im Rahmen des Prototypen den Prozessfluss innerhalb der Applikation nachvollziehbar machen und einen Aufruf der Web-Service-Schnittstelle gewährleisten.

Zur Sicherstellung der Datenvalidität des Dokumenteninfosatzes sind Regeln zu hinterlegen, welche die korrekte Bedienung der Schnittstelle zum Backend zur Laufzeit gewährleisten. Dies kann bereits bei der Definition der Forms des Frontend geschehen. Eine beispielhafte Regel ist, dass der Dokumententyp maximal einen Char-Wert der Länge 3 annehmen darf (Abbildung 19 - Maske 2). Sollte ein Mitarbeiter darüber hinaus einen Dokumententyp anlegen, der nicht vom Typ DRW ist, soll das Programm diesen automatisch in DRW umwandeln (Abbildung 19 - Maske 3).

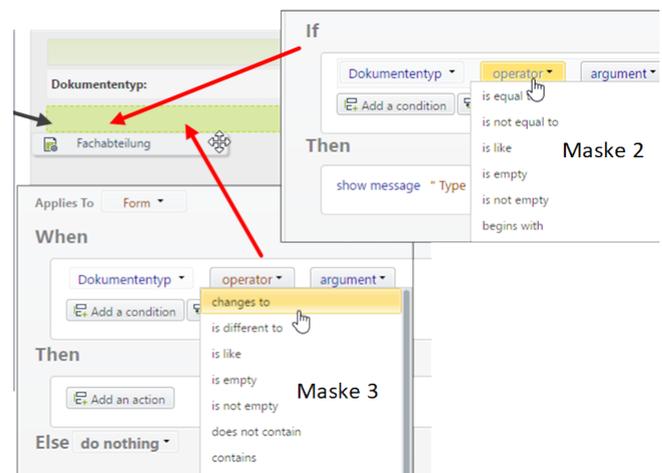


Abbildung 19: Design einer Form mit Regeln

Eine Form ist genau einer User-Task (Abbildung 20) zugeordnet; Anwender können in dem beispielhaft dargestellten Prozessschritt entscheiden, ob eine Rechnung den Status „ok“ oder „nicht ok“ erhält. Der Parameter wird nach dem

Verlassen der Form durch das Expression-Gateway aus Abbildung 21 ausgewertet. Unter diesem Prozessschritt kann hierarchisch der Prozess gehängt werden, der das DMS im Backend aufruft.

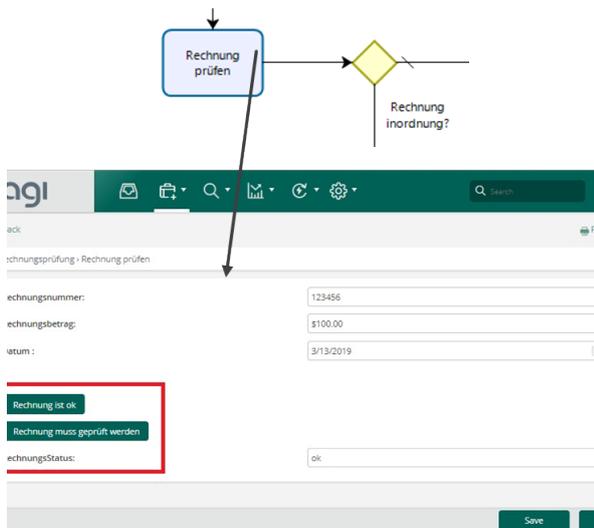


Abbildung 20: Gegenüberstellung Form und Task

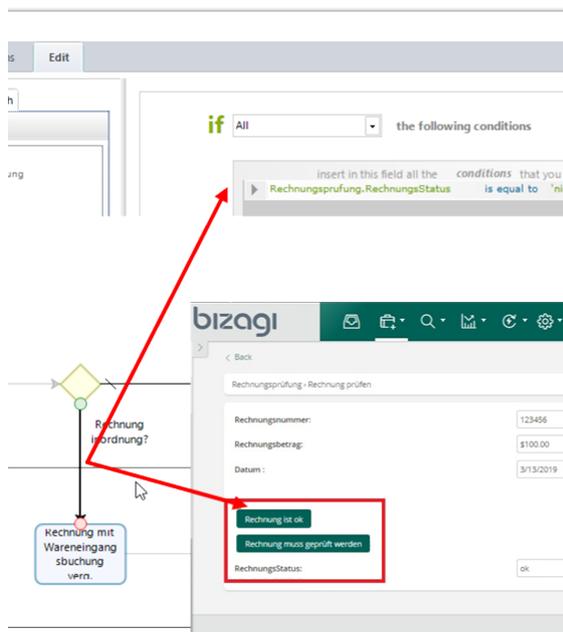


Abbildung 21: Expressions zum Steuern des Prozessflusses

Sobald der Prozess die Schritte der Datenerfassung und der Dokumentenprüfung durchlaufen hat, kann der Dokumenteninfosatz gemeinsam mit dem Originaldokument mittels des Web-Service-Aufrufs innerhalb des DMS des Backends archiviert werden. Durch den Aufruf der Service-URL interpretiert Bizagi die WSDL-Datei und visualisiert diese in Form einer Entität (Abbildung 22). Die Abbildung zwischen den Bizagi-Daten und der Schnittstelle ist intuitiv nachvollziehbar und wird durch den Web-Service-Wizard unterstützt. Innerhalb des Wizards kann neben der Abbildung

auch eine Transformation und Berechnung der Bizagi-Daten mit anschließender Weiterleitung an die Schnittstelle erfolgen.

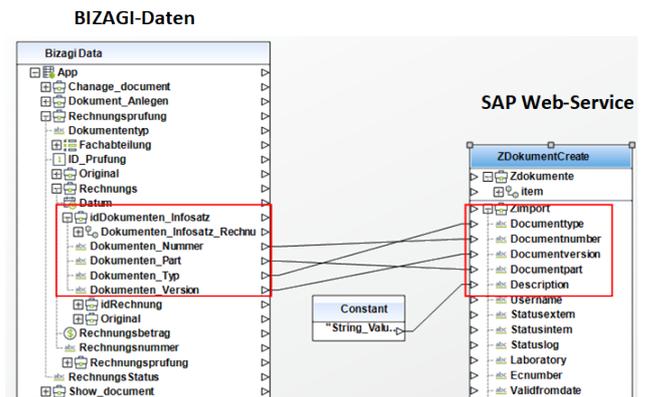


Abbildung 22: Abbildung der Bizagi-Daten auf Web-Service-Schnittstelle

Zur Einbindung des REST-Service kann auf den zuvor beschriebenen Service Z.SHOW_DOC zugegriffen werden. Der HTTP-Link zum Original-File soll anschließend durch eine geeignete Komponente den Nutzinhalt des Dokuments am Applikations-Front-End darstellen.

Hinsichtlich der Anbindung des zugrundeliegenden DMS des ERP-Systems gestaltet sich die Einbindung von RESTful-Services in Bizagi schwieriger, da es initial keinen Austausch von Schnittstelleninformation zwischen den beiden Systemen gibt, wie er beim SOAP-Service mit Hilfe der WSDL-Datei geschieht. Der syntaktisch richtige Aufruf der Services muss in diesem Fall durch den Entwickler selbst zugesichert werden. Dieser muss definieren, in was für einem Format die Antwort des REST-Services erfolgt, damit diese durch Bizagi verarbeitet werden kann. Neben der Content-Formulierung muss der Entwickler die Service-URL so gestalten, dass ein Aufruf mit dynamischen Daten von Bizagi aus möglich ist. Dies wird über die Definition von Parametern erreicht (Abbildung 23 - oberer Teil). Die Bizagi-Daten werden daraufhin auf die Parameter abgebildet, aus denen die Service-URI zur Laufzeit generiert wird (Abbildung 23 - unterer Teil).

Wenn der Service konfiguriert wird, können nach dem Programmstart die Daten eines Dokumenteninfosatzes an den REST-Service geleitet werden. Nachdem die Antwort mit dem HTTP-Link auf das Dokument in Bizagi ankommt, wird diese auf der Ausgangsform durch eine PDF-Komponente angezeigt. Damit sind beide Schnittstellenkonzepte zum Backend validiert.

Zu beachten ist, dass es innerhalb der Studie nicht gelungen ist, aus Bizagi heraus eine Originaldatei in das Backend zu übertragen - es konnten lediglich die Dokumenteninfosätze verarbeitet werden sowie das Original des Backends in der Low-Code Plattform angezeigt werden. Ein entsprechender Geschäftsprozess würde dann erfordern, das Original direkt im Backend einzuchecken. Eine Verarbeitung der Originaldateien (also eine Senden von der Low-Code Plattform ins Backend) erfordert bei den verwendeten Backend-Services

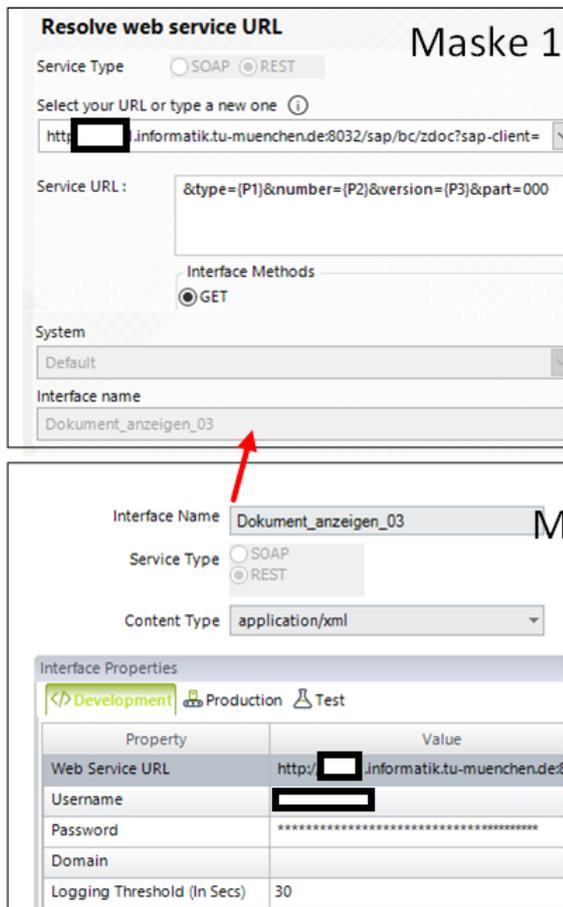


Abbildung 23: Konfiguration des REST-Services

eine Umsetzung der Originaldatei (also z.B. eines PDF) in eine XML-Datei (SOAP) bzw. einen URI (REST), was eine Einbindung eines entsprechenden Codings zur Konvertierung in Base64 erfordert. Für diese Aufgabe existiert zwar in Bizagi eine Funktion zur Enkodierung nach Base64, die als Expression in eine Business-Rule gehängt werden kann ([15]), dann trat aber in dem Zusammenhang ein technisches Problem mit dem HTTP-Request auf. Zur Anzeige des Originals konnte dagegen eine einfache existierende Routine in Bizagi genutzt werden.

Implementierung in Mendix

Wie weiter oben dargestellt, basiert die Entwicklung in Mendix auf User-Stories, die Anwendungsfälle definieren. Es werden daher für die Anwendungsfälle die einzelne Funktionalitäten innerhalb von einzelnen Applikationen abgebildet; die Applikationen können sowohl im Webbrowser aber auch direkt als App (apk) in Android-basierten Geräten ausführbar sein. Der Fokus der Untersuchung liegt in der Anbindung der Services des Backends und deren Einbettung in Mendix' Microflows; auch wenn das Vorgehensmodell eigentlich zuerst eine Gestaltung des Frontends vorsieht.

Aufgrund des objektorientierten Paradigmas in Mendix müssen für die zu verwaltenden Dokumente zuerst über die verwendeten Microflows korrespondierende Objekte erzeugt werden. Konkret kann dies wie folgt gelöst werden:

1. Definition eines Microflows, der ein Input-Dokument erzeugt,
2. Definition eines Microflows, der ein Zusatzdatenobjekt erstellt und eine Referenz dieses Objekts an das erzeugte Input-Dokument bindet,
3. Definition eines Microflows, der die erzeugten Objekte an einen Data-View bindet - das Data-View stellt die Schnittstelle zwischen den Datenobjekten und dem Front-End dar.

Abbildung 24 skizziert das Erstellen eines Zusatzdatenobjekts und die Erzeugung einer Referenz auf ein Input-Dokument, das über eine Parameter-Schnittstelle dem Microflow durch den Data-View übergeben wird. Es werden alle beteiligten Microflows gezeigt mit deren Bindung an das Frontend gezeigt.

Da die Dokumentendaten an das Backend übertragen werden sollen, kann ein Microflow gebaut werden, der beispielsweise an einen Front-End-Button gebunden wird und zur Laufzeit durch eine User-Aktion („Betätigen des Buttons“) ausgeführt wird. Die Konfiguration des REST-Services ähnelt hierbei der von Bizagi und wird deshalb nicht weiter dargestellt.

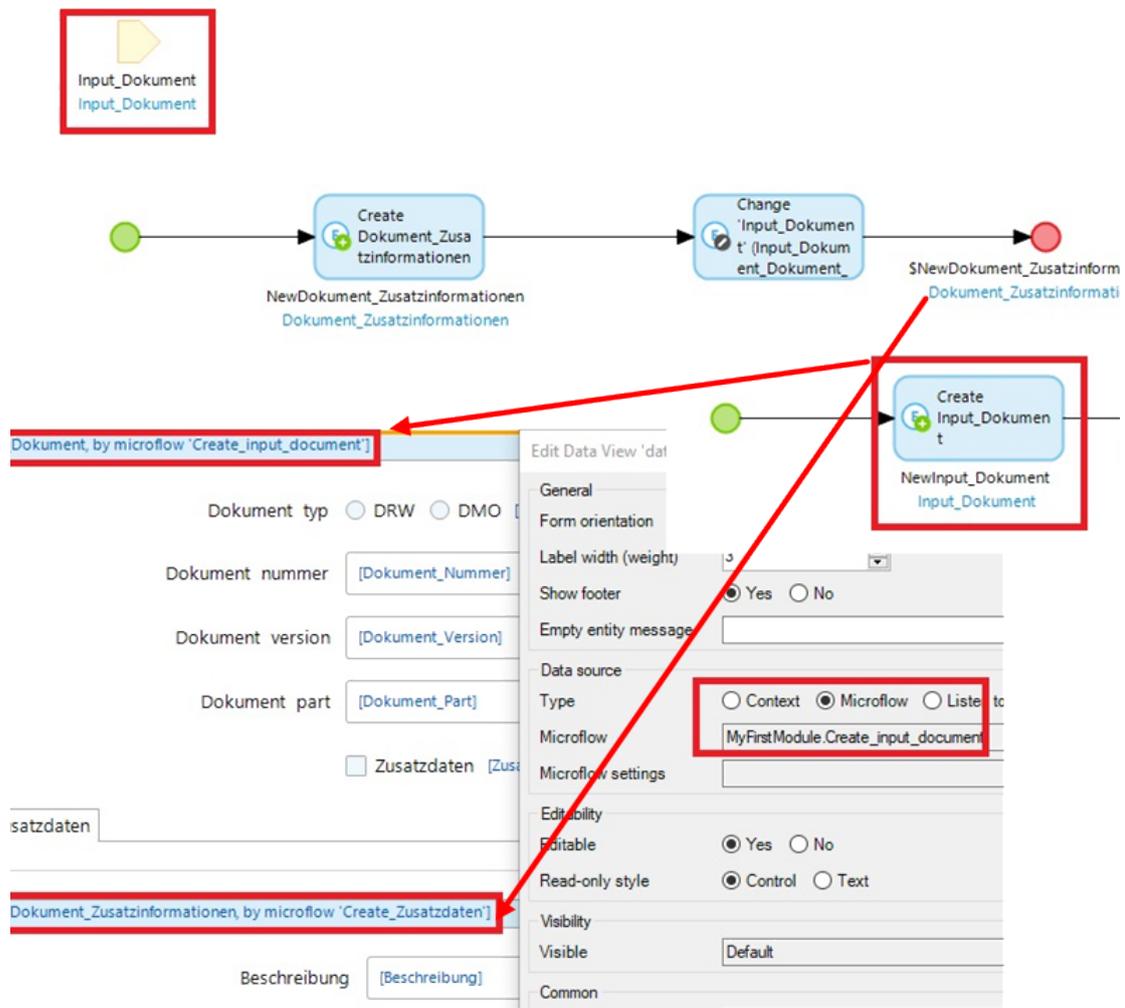


Abbildung 24: Microflows zur Erzeugung der Objekte und Bindung an Frontend

Die wesentliche in der Untersuchung erkannte Komplexität bei Verwendung eines WEB-Services entsteht dann, wenn in Mendix der Web-Service-Aufruf mit der entsprechenden Abbildung der Felder direkt auf Basis der in der NetWeaver-Plattform erzeugten WSDL-Datei erzeugt werden soll. Als Beispiel diene der folgende Ausschnitt der WSDL-Datei die in der NetWeaver-Plattform erzeugt wird:

```
<xsd:import namespace="urn:sap-com:
document:sap:rfc:functions"/>
<xsd:complexType name="DocDraw2">
<xsd:sequence>
<xsd:element name="Documenttype" type="n0:char3"/>
<xsd:element name="Documentnumber" type="n0:char25"/>
<xsd:element name="Documentversion" type="n0:char2"/>
<xsd:element name="Documentpart" type="n0:char3"/>
<xsd:element name="Description" type="n0:char40"/>
</xsd:sequence>
</xsd:complexType>
```

Mendix interpretiert die Aussage der WSDL-Datei und gewährleistet eine Datenvalidität gegenüber der Schnittstelle. Es hat sich jedoch ergeben, dass es NetWeaver-seitig nicht notwendig ist, **alle Attribute** der Entität zu übertragen, um die hinter dem Parameter stehende Tabelle zu ändern bzw. Einträge in dieser zu erzeugen; die Parameter werden auch im Default der WSDL nicht durch NetWeaver erzeugt. Das Problem, das sich nun ergibt, ist, dass Mendix einem Service-Aufruf nur dann zustimmt, wenn die WSDL-Kriterien erfüllt werden. Das NetWeaver eine „unvollständige“ Nachricht im Sinne der WSDL verarbeiten kann, spielt hierbei keine Rolle. Im Fall der Untersuchung war es deshalb notwendig, die WSDL nach der Erzeugung manuell abzuändern und um ein weiteres Attribut aus dem Standard zu ergänzen. Wird einem Element der Zusatz nillable = „true“ hinzugefügt, ist es möglich, ein leeres Typeelement beim Aufruf des Services zu übertragen. Die WSDL wird also manuell um den nillable-Parameter ergänzt; in Bizagi sind die Prüfungen offensichtlich nicht so streng, wie in Mendix und die WSDL kann direkt verarbeitet werden. Mit dem WSDL-Zusatz entsteht die Möglichkeit, einzelne Elemente der SOAP-Nachricht leer zu lassen.

```
<xsd:import namespace="urn:sap-com:
document:sap:rfc:functions"/>
<xsd:complexType name="DocDraw2">
<xsd:sequence>
<xsd:element name="Documenttype" nillable="true"
type="n0:char3"/>
<xsd:element name="Documentnumber" nillable="true"
type="n0:char25"/>
<xsd:element name="Documentversion" nillable="true"
type="n0:char2"/>
<xsd:element name="Documentpart" nillable="true"
type="n0:char3"/>
<xsd:element name="Description" nillable="true"
type="n0:char40"/>
</xsd:sequence>
</xsd:complexType>
```

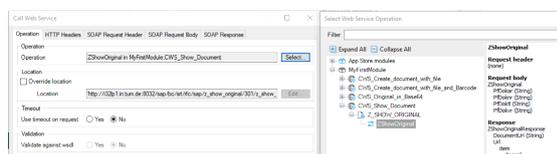


Abbildung 25: SOAP-Service-Aufruf in Mendix

Die nun manipulierte WSDL-Datei kann als Basis für das Datenmodell in Mendix genutzt werden und ermöglicht dem anschließend zu erstellenden Export-Mapping einen flexiblen Aufruf der Web-Service-Schnittstelle. Diese übersendet nach

der Abhandlung der Dienstleistung das Ergebnis des Programmablaufs in Form einer SOAP-Response; Abbildung 25 zeigt den Aufruf. Es wurde prototypisch eine Applikation erstellt, mit der Dokumente über Mendix und die entsprechenden Services angelegt werden können (Beispiel der Android-App in Abbildung 26). Es ist zu beachten, dass in Mendix direkt Android-Applikationen erzeugt werden können - hier kann beispielsweise ein Widget aus dem Mendix-Store verwendet werden, mit dem ein Barcode gescannt werden kann; etwa um den Dokumenteninfosatz direkt mit einem Kundenstamm verknüpfen zu können.

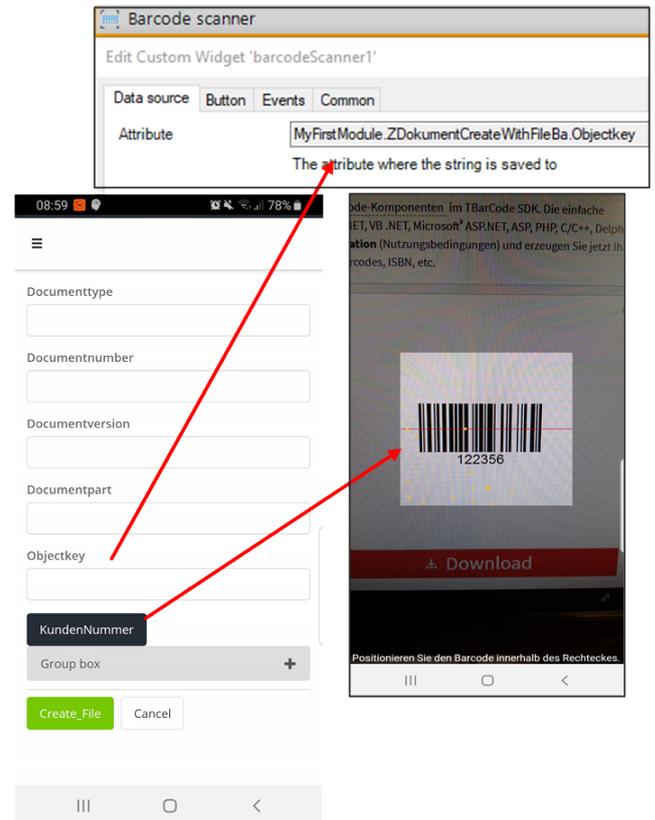


Abbildung 26: Beispiel der Android-Applikation mit Barcode-Scanner

Die Übertragung der Originaldatei von dem Frontend zum Backend in Mendix erfordert eine Kodierung der Datei in die Nachricht. Es kann eine Base64-Enkodierung verwendet werden, wofür Mendix (wie Bizagi) eine Standardfunktion besitzt. Im SAP-Backend existieren ebenfalls Standard-Routinen für die Dekodierung (Funktion „SSFC_BASE64_DECODE“).

RESÜMEE UND AUSBLICK

Mendix erweist sich als Plattform, die über eine Vielzahl an Werkzeugen verfügt, deren Funktionalitätsumfang es ermöglicht, individuellste Probleme zu lösen. Dies zeigt sich insbesondere an der Möglichkeit, trotz der eigentlich offiziell nicht zugreifbaren Front-End-Services von NetWeaver, Originale zwischen Mendix und der Schnittstelle beliebig austauschen zu können. Zu beachten ist jedoch, dass der Mendix-Entwicklungszyklus mit der Definition von User-Stories beginnt, die anschließend in ein Front-End-Modell übersetzt

werden. Erst im Anschluss folgt die Definition des Datenmodells und der Geschäftslogik. Es besteht also die Notwendigkeit, dass bereits im UI-Modell die Prozesslogik in Form von Masken zu berücksichtigen ist. Im Gegensatz zu Bizagi, welches direkt den Geschäftsprozess implementiert, führt eine Verwendung von Mendix somit zu einer anderen Logik der Einbindung der Backend-Services: der Service wird nicht direkt in den Geschäftsprozess eingebunden, sondern durch eine eigene Applikation implementiert, die auf einer User-Story basiert und wiederum in einem Geschäftsprozess verwendet werden kann, welcher aber nicht als solcher abgebildet wird. Somit ist in Mendix im Wesentlichen ein neues Frontend der Backend-Services implementierbar. Das Frontend kann dann von den Nutzern direkt in Prozessen aufgerufen werden. In Bizagi wird dagegen der Service aus dem Geschäftsprozess direkt aufgerufen, da die Applikation direkt auf dem Geschäftsprozess basiert. Dies bedeutet jedoch auch, dass für verschiedene Prozessvarianten potentiell verschiedene Applikationen in Bizagi entwickelt werden müssen, die dann den selben Service aufrufen. Die Behandlung von SOAP-Services hat sich in Mendix als aufwändiger erwiesen, REST-Services waren in der Komplexität vergleichbar. Die Einbindbarkeit komplexer individueller Routinen schien den Autoren im Untersuchungszeitraum aufgrund einer Java-Umgebung in Mendix etwas einfacher zu bewerkstelligen, als in Bizagi - dies hatte eine unmittelbare Auswirkung auf den Transfer der Originaldateien von den Low-Code Plattformen zum Backend.

Weitere Untersuchungsgegenstände könnten nun ODATA-Services, in der speziellen hier betrachteten Domäne ebenso die Verwendung des CMIS-Standards sowie weitere Low-Code Plattformen sein. Die Möglichkeiten, komplexeren Programmcode - etwa Aufrufe umfangreicher Libraries einzubinden oder auch komplexe Algorithmen - sollte speziell für Bizagi ebenfalls weiter validiert werden, da das Programmiermodell den Geschäftsprozess in den Vordergrund stellt.

LITERATUR

- [1] Klaus Goetzer. *Dokumenten-Management*. Dpunkt.verlag, 2008.
- [2] Rinaldo Heck. *Geschäftsprozessorientiertes Dokumentenmanagement*. GalileoPress, 2009.

- [3] SAP. *Definition von Dokumenten*. https://help.sap.com/doc/saphelp_erp60_sp/6.0/de-DE/79/22bf53d25ab64ce10000000a174cb4/content.htm?no_cache=true, 2019. Aufgerufen im Internet am 20.03.2019.
- [4] Pragma Pande. *Architekturkomponenten SAP DMS*. <https://archive.sap.com/documents/docs/DOC-54583>, 2019. Aufgerufen im Internet am 20.03.2019.
- [5] Johann Wagner and Kurt Schwarzenbacher. *Föderative Unternehmensprozesse: Technologien, Standards und Perspektiven für vernetzte Systeme*. John Wiley & Sons, 2007.
- [6] Ingo Melzer. *Service-orientierte Architekturen mit Web Services*, 2. Spektrum, 2007.
- [7] Jim Webber, Savas Parastatidis, and Ian Robinson. *REST in practice: Hypermedia and systems architecture*. O'Reilly Media, Inc., 2010.
- [8] Martin Fowler. *Richardson-Maturity-Model*. <https://martinfowler.com/articles/richardson-MaturityModel.html>, 2019. Aufgerufen im Internet am 20.03.2019.
- [9] Horst Keller. *ABAP Objects*. GalileoPress, 2015.
- [10] Gerhard Rempp, Mark Akermann, Martin Löffler, and Jens Lehmann. *Model Driven SOA : Anwendungsorientierte Methodik und Vorgehen in der Praxis*. Xpert.press, Berlin, Heidelberg, 2011.
- [11] BIZAGI. *BIZAGI Architektur*. <http://help.bizagi.com/bpm-suite/en/index.html?architecture.htm>, 2019. Aufgerufen im Internet am 20.03.2019.
- [12] Mendix. *Mendix Architektur*. <https://www.mendix.com/evaluation-guide/enterprise-capabilities/architecture-intro>, 2019. Aufgerufen im Internet am 20.03.2019.
- [13] Mendix. *Mendix BPMN*. <https://docs.mendix.com/refguide/microflows>, 2019. Aufgerufen im Internet am 20.03.2019.
- [14] Mendix. *Mendix Refguide*. <https://docs.mendix.com/refguide/nanoflows>, 2019. Aufgerufen im Internet am 20.03.2019.
- [15] Bizagi. Convert Files to base64. <http://help.bizagi.com/bpm-suite/en/-index.html?tobase64.htm>. Zuletzt eingesehen am 12.10.2018.

Simulating Dynamic Vehicle Routing Problems with Athos

Benjamin Hoffmann, Michael Guckert,
KITE - Kompetenzzentrum für
Informationstechnologie
Technische Hochschule Mittelhessen, Germany
{benjamin.hoffmann,michael.guckert}
@mnd.thm.de

Kevin Chalmers,
Neil Urquhart
School of Computing
Edinburgh Napier University, Scotland
{k.chalmers,n.urquhart}
@napier.ac.uk

KEYWORDS

Domain-specific language, Agent-based modelling, Evolutionary Algorithms, Vehicle Routing With Time Windows.

ABSTRACT

Complex routing problems, such as vehicle routing problems with additional constraints, are both hard to solve and hard to express in a form that is accessible to the human expert and at the same time processible by a computer system that is supposed to produce a solution of sufficient quality. The formulation must be formal enough to avoid ambiguities and also comprehensible enough to be created, discussed and shared by domain experts. In this paper, we present the domain specific language Athos in which complex routing problems can be expressed in a computationally independent, human-readable form. Athos is then transformed into code that can be run in an adequate target platform. Suitable methods for solving problems are available and applied to the given problem. We present a case study in which we use a genetic algorithm to solve instances of a vehicle routing problem with time windows and demonstrate the end to end process to produce a solution in the Athos environment. Moreover, we show how the Athos system goes beyond optimisation of static routes and can be used as a tool to simulate the impact of traffic and congestion on the tours. We call this extended problem a dynamic vehicle routing problem with time windows.

INTRODUCTION

Even after many years of research, routing problems are still a focus of current research efforts. Generally belonging to the class of NP-hard problems, the search for efficient heuristics remains a challenge. With logistics networks growing in size and complexity, efficiency and sustainability become central issues for the transportation industry.

Rapidly changing requirements need software solutions that can be easily adapted and extended. If using general purpose languages (GPLs), changing software requires systems engineers and domain experts to interact and communicate, which often is a source of misunderstanding leading to weak models and error prone systems (Dalal and Chhillar 2012). A low level of abstraction in GPLs prevents reuse of larger building blocks and therefore implementation has to start from scratch

each time a new problem has to be solved. Using a proprietary platform creates dependencies to standards imposed by a commercial software vendor that may not always be in line with the needs of the current project.

We suggest using a model driven approach to overcome this dilemma. We discuss how domain experts can specify traffic related optimisation problems declaratively with our Domain-Specific Language (DSL) Athos. Athos models are accessible to humans and can be transformed into executable programs to be run in appropriate target platforms (e.g. NetLogo, Repast Symphony, or potentially any other environment).

In this paper, we demonstrate Athos and its features by showing how it can be used to model both a static and a dynamised version of the Vehicle Routing Problem with Time Windows (VRPTW). In the dynamic version of the VRPTW mutual influence of traffic participants is considered. The problem runs in a network of roads defined in the model and uses a genetic algorithm to heuristically compute a solution. The models are as computationally independent as possible and therefore the algorithm and its implementation are not part of the model but of the infrastructure of the target platform.

ATHOS

Athos is a DSL designed for the domain of dynamic transportation problems. The language is implemented in Xtext (<https://www.eclipse.org/Xtext/>). The Athos architecture contains a generator that transforms Platform-Independent Models (PIMs) into Platform-Specific Models (PSMs); i.e. executable code to be run in a target platform. We currently support NetLogo and use it as our main development environment. Prototype support for Repast Symphony has also been implemented and tested. Code of an algorithmic nature (e.g. routing algorithms and heuristics for optimisation) is implemented in Java and made available to the target platform via libraries. The NetLogo implementation, for instance, uses extensions that can be accessed from the generated NetLogo models via the NetLogo-Extension-API. Athos deliberately does not support imperative programming as we aim to provide a purely declarative modelling language. Currently available meta-heuristics libraries in Athos are an implementation of an ACS (Ant Colony System) for solving TSP-like problems (Hoffmann, Guckert, et al. 2018) and an Evolutionary Algorithm for solving more complex vehicle routing problems like the problem presented in this paper.

The VRPTW describes the task of visiting a set of customers for delivery or pick-up of products, depending on context. Each visit must conform to time windows specified for each customer. If the vehicle does not arrive within the limits of the time window, it either has to wait until the window opens (early arrival) or the schedule is not feasible (late arrival). Visits at a customer may also consume a given amount of service time. Vehicles start and end their journeys in one of possibly many depots. If more than one depot is used, the problem is referred to as a multi-depot problem. The problem may be formulated for a fixed number of vehicles or the number of vehicles may be part of the objective function of the optimisation problem. VRPTWs can be used to optimise a single objective (e.g. overall distance travelled) or multiple objectives (e.g. number of vehicles and distance travelled) (Dabia, Demir, and Woensel, van 2014). The VRPTW is highly relevant to real-world problems both in an operational (N. B. Urquhart, Hart, and Judson 2015) and a planing context (N. Urquhart and Fonzone 2017). The user of the VRPTW-solving software will be a domain expert, e.g. a logistics analyst or transport planner. Many industrial contexts will include additional specific constraints defined by the business context, i.e. working conditions of staff, types of vehicle in use, environmental or financial considerations. The implementation of such additional, possibly very versatile, constraints leads to an increased workload for professional programmers using a conventional GPL. A DSL as Athos will potentially reduce development time and give domain experts a tool to easily modify and extend a model without having to access software developers.

Obviously, any problem instance of a VRPTW requires travel time data between customers and depots. Depending on the optimisation criterion used, it may also be necessary to retain distance or emissions data as well. Such data can be held in an origin-destination matrix (see e.g. Dantzig, Ramser, and Hubert 1959), or simply be calculated using the Euclidean distance between customers. However, real-world examples (N. Urquhart and Fonzone 2017; N. B. Urquhart, Hart, and Judson 2015) usually use an underlying street graph and apply path-finding algorithms to find routes between customers.

Summarising the discussion, we see that a DSL supporting the domain of vehicle routing must be capable of expressing the many different formulations of a VRP including differing vehicle types, time windows, service times, capacity constraints and driving time constraints. At the same time, it must also support the modelling of the underlying street graph. The following step by step example illustrates the modelling features of the language and how a VRPTW can be described. The product to be delivered is soap each item having a weight of 1 weight unit. The model contains a single agent type named *delivery* with two states *awt* and *die*. The agents of this type wait for an optimal tour to be computed and then start delivering goods according to the tour received thereby satisfying demands of the customers on that tour. Travel time will be computed using the default duration function that uses length and speed of the agents and counts in the defined congestion factors so that a high amount of traffic in the network has an impact on the delivery.

```

1 model VRPTW_Example
2 world xmin 0 xmax 75 ymin 0 ymax 75
3 products product soap weight 1.0
4 agentTypes
5 agentType staticDelivery congestionFactor 60.0 maxWeight 200.0
6   behaviour awt awaitTour when finished do die;
7   behaviour die vanish;
8 functions
9 durationFunction normal length default

```

Behavioural patterns of agents are defined by means of *behaviour blocks*. For each agent type a single behaviour block with an arbitrary number of *behaviour states* can be defined. Behaviour states correspond to the states of an implicit finite automaton. A state consists of a perceivable action and a specification of transitions. Events can be defined as stimuli that trigger a transition and entail a change of state.

The Athos meta-model reflects all of the elements of the language. Figure 1 shows how agent states are represented there. *AgentTypes* are linked to exactly one *AgentBehaviourBlock*, which contains one or more *AgentBehaviourStates*. The state of an agent corresponds to exactly one observable behaviour that the agent exhibits when being in the respective state. This observable behaviour is an instance of *AgentBehaviourDescription*. Athos has a set of built-in *AgentBehaviourDescriptions* which will continuously be extended in future versions.

Additionally, an *AgentBehaviourState* contains an arbitrary number of *AgentBehaviourTransitions*, which trigger a change to a target state depending on a condition. Possible target states are other named states in the *AgentBehaviourBlock* (in Figure 1 or anonymous states defined in the *AgentBehaviourTransition*).

```

1 agentTypes
2 agentType staticDelivery congestionFactor 60.0 maxWeight 200.0
3   behaviour awt awaitTour when finished do die;
4   behaviour die vanish;

```

The network is a complete graph with edge lengths equal to the Euclidean distance between the respective end nodes. Nodes are defined with their coordinates. As the graph is defined to be complete, the list of edges is empty. The length of the edges is used as travel time. Note that this is only an academic example – any other duration function could be defined and used here.

```

1 complete network
2 nodes
3 node n1 (35.0, 35.0)
4 ...
5 node n51 (47.00, 47.00)
6 edges

```

Sources and demands are defined using nodes of the network. The keyword *ea* indicates that an evolutionary algorithm is to be used to compute the tours for the vehicles. Additional parameters for the algorithm are provided. For each demand quantities, time windows, and service time are defined.

```

1 sources
2 n1 isDepot soap sprouts (staticDelivery) agentsStart route
3   ea (n2, n3, n4, ..., n48, n49, n50, n51) popSize 30
4 demands
5 n1 hasDemand soap absQuantity 0.00 earliestTime 0
6   latestTime 230 serviceTime 0
7 ...
8 n51 hasDemand soap absQuantity 13.00 earliestTime 124
9   latestTime 134 serviceTime 10

```

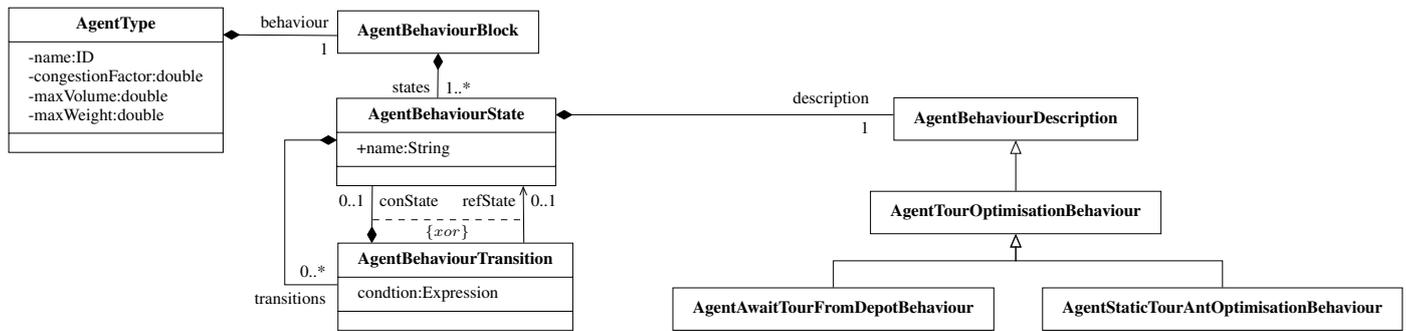


Fig. 1. Athos meta-model for modelling agent behaviour.

Agents of a given type can be monitored by using metrics in which indicators can be defined that can either accumulate values or set values. These indicators are collected for each agent and can be viewed for each single agent or condensed into overall statistics. As any other feature of Athos, metrics have a representation in the meta model analogous to that of the behaviours. However, for the sake of brevity this is not discussed here.

```

1 defineMetrics
2   metrics for staticDelivery (
3     class metric distanceCovered
4       when (isAtCustomer?)
5         add distanceTo last customer
6     individual metric ticksEarly
7       when (isAtCustomer? && earliestTime > currentTime)
8         add earliestTime - currentTime
9   )

```

Beyond the optimisation of tours with the built in evolutionary algorithm, Athos can run simulations of dynamic delivery scenarios with these optimised schedules in which the effect of traffic and congestion in the network can be measured up with the individually defined metrics. While a static VRPTW optimises tours once according to a defined objective function, a dynamic VRPTW goes beyond that and is sensitive to dynamic aspects of traffic in the underlying network. We discuss and compare a static and a dynamic VRPTW in the case study presented in the next section.

CASE STUDY

In this section, we will present a case study that compares the static and dynamic variants of the VRPTW problem. We will analyse how a changed traffic situation influences the success of the planned tours. First we will use Athos to define a static VRPTW. In addition, we will define some metrics to see how the vehicles perform when the traffic situation stays exactly the same throughout the entire simulation. In a second step, we will transform the VRPTW into a dynamic VRPTW by adding additional noise-agents that will induce congestion effects inside the network. We will use the defined metrics to see how congestion influenced the outcome of the calculated VRPTW solution. Note that the complete Athos program, the generated NetLogo programs (together with the required extension) as well as some videos showing the simulation can be obtained from <https://athos.mnd.thm.de/public/ecmscasestudy.html>.

Figure 2 illustrates the graph used in this case study. The graph is an artificially simplified version of an urban area with the following characteristics:

- At its core, the area features a beltway. The roads here are highly dependent on each other so that congestion on one road directly expands to other roads of the beltway.
- The depot is located at the very centre of this beltway. Access roads to the depot also belong to the beltway and thus are also affected by any congestion on the beltway.
- The centre of this area also features a network of highways with high capacity. These highways are independent so that congestion on one highway does not have ripple effects on adjacent highways.
- Suburban areas are accessible through roads of less capacity. These roads are more susceptible to congestion when used by a queue of cars or cars with high congestion factors like heavy-goods vehicles.

```

1 model UrbanArea
2 world xmin 0 xmax 40 ymin 0 ymax 22
3 <<definition of agent types>>
4 functions
5 durationFunction highway
6   length + 1.5 * accCongestionFactor default
7 durationFunction road length * 3 + 4 * accCongestionFactor
8 network
9 <<definition of nodes>>
10 edge undirected e01 from n0 to n1
11   length 0.0 cfactor 1.0 path "cityRing" function highway
12 edge undirected e12 from n1 to n2
13   length 0.0 cfactor 1.0 path "cityRing" function highway
14 edge undirected e16 from n10 to n11
15   length 0.0 cfactor 1.0 function highway
16 edge undirected e17 from n10 to n18
17   length 0.0 cfactor 1.0 function road
18 << definition of other edges >>

```

The first two lines of the listing above give a name to the model and define its global boundaries. The definition of the agent types used in the case study is omitted in the listing but will be presented shortly. Lines 4 to 7 contain the functions that are associated to the two different road types used in the case study. The function associated to edges which represent highways was aptly named “highway”. The meaning of this function is that the length of a highway defines the minimum amount of time a vehicle needs to cross it. For example, a highway of length 200 would require 200 ticks to be crossed by a vehicle, given that the vehicle has a congestion factor of zero

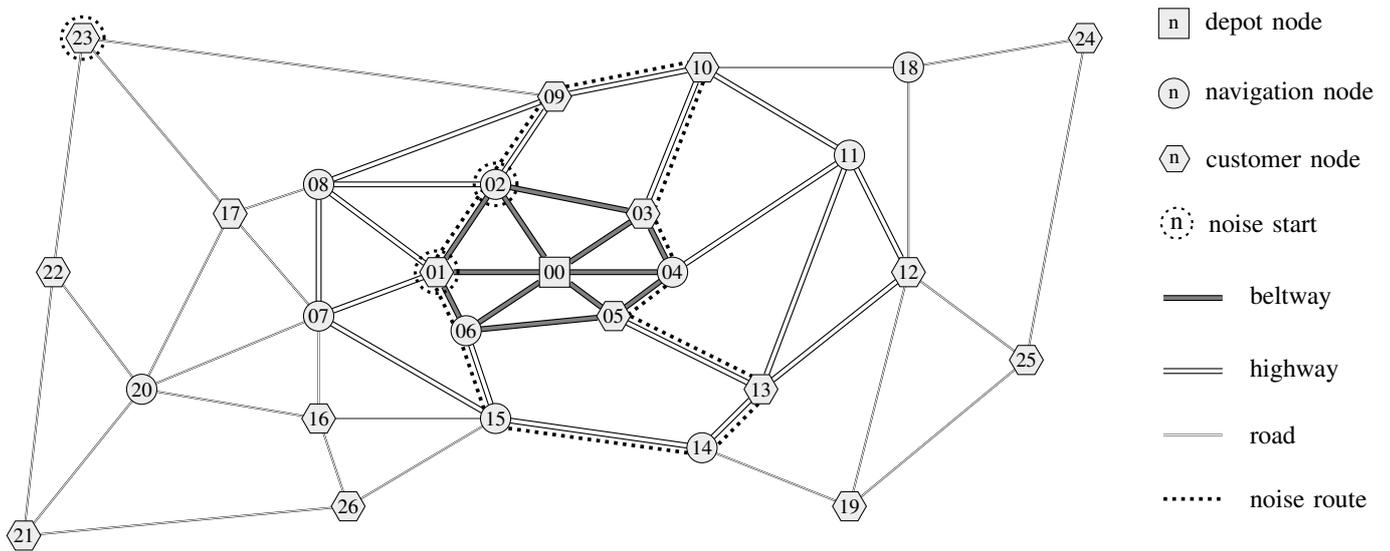


Fig. 2. Artificial graph used in the case study.

and the summed up congestion factor of all other vehicles on the road is also zero. This is due to the fact that the congestion factor of each vehicle on a road additionally increases the time required to cross the respective road.

Roads that belong to the same path share their accumulated congestion factor. In the listing, edge e_{01} and e_{12} belong to the same path. To account for the fact that normal roads take longer to travel and are more susceptible to congestion than highways, the `road` function multiplies the length of the road by three and the accumulated congestion factors by four.

```

1 << environment definition >>
2 agentTypes
3 agentType delivery congestionFactor 0.0 maxWeight 180.0
4   behaviour awt awaitTour when finished do wait;
5   behaviour wait idle for 1000 when finished do die;
6   behaviour die vanish;
7 <<network definition>>
8 sources
9 n0 isDepot soap sprouts (delivery)
10   agentsStart route ea (n3, n5, n9, n10, n12, n13, n16, n17,
11   n19, n21, n22, n23, n24, n25, n26) popSize 30 at 0
12 demands
13 n0 hasDemand soap absQuantity 0.00
14   earliestTime 0 latestTime 500 serviceTime 0
15 n1 hasDemand soap absQuantity 15.0
16   earliestTime 15 latestTime 120 serviceTime 5
17 << more demand specifications >>

```

The above listing shows the specification of the homogeneous fleet of vehicles located at the depot. Lines 3 to 6 specify that these vehicles have a congestion factor of zero and thus do not congest the roads. Each vehicle can provide a customer with 180 units of a given product. Line 4 specifies that this type of vehicle waits at a depot until it is assigned a tour. As the metrics that we will define in the next step do only apply for existing agents, lines 5 and 6 tell a vehicle to idle at the depot for 1,000 ticks before leaving the simulation. In line 9, node n_0 is defined to be a depot. Customers are defined inside the brackets. Note that the code is intention-

ally no longer computationally independent, because the `ea` keyword explicitly specifies the application of an evolutionary algorithm. This allows to specify the parameters used in the respective algorithm – in this case, the population is set to a size of 30. The depot searches for a set of tours for its vehicles at the very beginning of the simulation. This is specified in line 11 with the keyword `at` followed by the value zero.

$$0.001 \cdot \sum_{k \in K} \sum_{i \in N} \sum_{j \in N} t_{ij} x_{ij}^k + 100 \cdot |K| \rightarrow \text{Min!} \quad (1)$$

The depot uses a genetic algorithm (Hoffmann, Chalmers, et al. 2019) that internally builds a complete graph of all customer nodes (Hoffmann, Guckert, et al. 2018) and optimises the function given in (1). Here, K denotes the set of vehicles at the depot, N the set of nodes of the complete graph (including the depot), t_{ij} the travel time from node i to node j and x_{ij}^k a decision variable that is set to 1 if vehicle k travels from node i to node j and 0 otherwise. A parameter value of 0.001 weights the total distance covered by all vehicles. Analogously, 1000 is the weight for the number of vehicles used. The complete set of parameter values is given in Table II.

Finally, lines 12 to 17 contain the demand specifications. Note that the depot appears here in order to define the latest point in time for the return of the vehicles. Table I summarises the constraints related to the customers in this case study.

```

1 << environment and delivery agent type >>
2 agentType noiseAgent congestionFactor 150.0 maxWeight 100.0
3   behaviour roam route exact (n1, n2, n9, n10, n3, n4, n5,
4   n13, n14, n15, n6) repeat 5 times
5   when finished do die;
6   behaviour die vanish;
7 << functions and network definition >>
8 sources
9 << n0 still depot >>
10   n1 sprouts (noiseAgent) at 2
11   n2 sprouts (noiseAgent) at 4
12   n23 sprouts (noiseAgent) at 3

```

The Athos code presented so far allows us to run the simulation in a “noise-free” mode: The depot calculates a solution for the VRPTW and sends out the vehicles to serve the customers. In order to transform the static VRPTW into a dynamic version, we will run a second batch of simulations in which additional noise-agents will increase travel times on the city ring and some of the roads/highways in the network. To this end, the above listing introduces another agent type with a congestion factor of 150. Due to its high congestion factor, this type of agent considerably slows down traffic on its current road (or path of roads). The agent follows a predefined route (specified in lines 3 and 4) 5 times and then disappears. Lines 9 to 11 define the nodes and the exact point in time where an instance of this type of agent appears. Note that even though node `n23` is not part of the specified route of nodes for that type of agent, it can still appear at this node. The agent then uses the fastest way to the first node of the tour specified for this agent type (`n1`).

The final part of the simulation specification for this case study is the definition of a set of metrics. In this case study, we are interested in the cumulative distance travelled by the delivery vehicles. Moreover, the exact time vehicles had to wait due to an early arrival at a customer might give some insight on the efficiency of the calculated tour. Also, it might be important to now the accumulated time by which vehicles arrived to late at a customer and the total number of time windows violated and time windows met. These metrics are specified in the following listing.

```

1 << as before >>
2 defineMetrics updateRate 10
3   metrics for delivery (
4     class metric distanceCovered
5       when ( isAtCustomer? )
6         add distanceTo last customer
7     class metric ticksEarly
8       when (isAtCustomer? && earliestTime > currentTime)
9         add earliestTime - currentTime
10    class metric ticksLate
11      when (isAtCustomer? && latestTime < currentTime)
12        add currentTime - latestTime
13    class metric windowsViolated
14      when (isAtCustomer? && latestTime < currentTime)
15        add 1
16    class metric windowsMet
17      when (isAtCustomer? && currentTime <= latestTime)
18        add 1
19  )

```

Table II and Table III summarise the results of ten simulation runs each for the problem without noise agents and the problem with noise agents that introduce dynamism through reduced travel speeds. As was to be expected, in the simulations without noise-agents, no time windows were violated. In addition to the 16 customers defined for the problem, the metric also counts the timely return of a vehicle to the depot as a met time window. Since the evolutionary algorithm used to solve this problem is not deterministic, some runs feature solutions with three and some with four vehicles resulting in 19 or 20 met time windows.

The introduction of noise-agents changes the situation dramatically. The noise agents effectuate the movement speed on their respective roads in a way that the delivery vehicles do no longer meet all time windows. In fact, nearly half of the defined

TABLE I. CONSTRAINTS OF THE VRPTW.

Cstm	Location	Demand	Earliest	Latest	Service
01	(18.0, 9.0)	15	15	120	5
03	(21.0, 11.0)	20	10	120	7
05	(20.0, 7.5)	50	20	90	10
09	(18.0, 15.0)	45	80	220	15
10	(23.0, 16.0)	25	90	250	10
12	(30.0, 9.0)	30	35	260	5
13	(25.0, 5.0)	60	40	140	7
16	(10.0, 4.0)	35	45	160	8
17	(7.0, 11.0)	40	0	140	50
19	(28.0, 1.0)	15	10	130	9
21	(0.0, 0.0)	50	5	120	5
22	(1.0, 9.0)	30	10	90	15
23	(2.0, 17.0)	40	20	60	5
24	(36.0, 17.0)	35	25	50	8
25	(34.0, 6.0)	20	20	60	9
26	(11.0, 1.0)	15	30	45	10

time windows are violated. In each of the ten simulations with noise-agents the accumulated ticks by which time windows were missed is around 760.

In both cases the distance travelled by the vehicles was nearly which was to be expected. The slower movement on the roads caused the vehicles to arrive at their customers later than originally calculated which is also reflected in the amount of ticks that vehicles arrived too early at their customers which is reduced by around 6.7 ticks.

In our future work, we will use Athos to further research into dynamic VRPTW to provide strategies that provide satisfactory solutions even in case of sudden traffic surges.

RELATED WORK

Steil et al. (Steil et al. 2011) discuss an approach that encompasses all aspects involved in the domain of patrol routing algorithms. It covers all stages in the development of patrol routes from the specification (expression) and simulation-based assessment (execution and evaluation) to the translation of patrol routes to the real-world (engagement). Accordingly, they call their approach the 4Es approach.

The 4Es approach is similar to the approach presented in this paper in that it integrates the expression, simulation and evaluation in an appropriate environment. Moreover, it also uses a DSL for the expression of routing algorithms. Their DSL, called Turn, allows to define the next destination of an agent in a road network by means of set reduce functions that can be chained to successively reduce the set of all nodes until only one node is left which is then selected as an agent’s next destination. This way the agents in the simulation follow a pre-defined routing strategy. The system evaluates routing strategies by application of four distinct metrics. These metrics provide information on the time it took a first-responding agent to get to the node where an event occurred, the percentage of

TABLE II. RESULTS FOR TEN RUNS OF A NOISE-FREE SIMULATION (PARAMETER SETTINGS: POPSIZE = 30; SIMPLEPERMUTATIONPROB = 0.9; MAXDISTANCE = 4; GENERATIONS = 80; WNOOFTOURS = 100; WTOTALDIST = 0.001; TRMTSIZE = 4; TAKEBESTPROB = 0.8; MUTATIONPROB = 0.1.

Metric/Run	1	2	3	4	5	6	7	8	9	10	Avg.
Total distance	101.90	105.10	105.10	101.90	101.90	107.65	102.79	107.65	101.90	107.65	104.35
Ticks early	27	48	48	27	27	27	25	27	27	27	31
Ticks late	0	0	0	0	0	0	0	0	0	0	0
Windows violated	0	0	0	0	0	0	0	0	0	0	0
Windows met	19	20	20	19	19	19	20	20	19	20	19.5

TABLE III. RESULTS FOR TEN RUNS OF A NOISE-FULL SIMULATION (SAME PARAMETER SETTINGS AS STATED IN TABLE II).

Metric/Run	1	2	3	4	5	6	7	8	9	10	Avg.
Total distance	104.86	101.71	101.71	101.71	101.71	101.71	101.71	104.86	107.42	104.86	103.23
Ticks early	39	18	18	18	18	18	18	39	18	39	24.30
Ticks late	766	764	766	762	760	764	761	764	776	763	764.6
Windows violated	9	9	9	9	9	9	9	9	10	9	9.1
Windows met	11	10	10	10	10	10	10	11	10	11	10.30

nodes in the network visited by agents per day, the number of nodes visited that were in a state in which an event of interest is likely to occur (so-called hot nodes), and the amount of time agents spent at such hot nodes.

Despite the mentioned similarities, the work of Steil et al. differs considerably from our work when looked at in more detail. First of all, the two research efforts target two different domains. As is pointed out by the authors, patrol routing problems share some features with vehicle-routing problems but greatly differ in what practitioners ultimately aim to achieve. In VRPs, in the majority of cases, the objective is the minimisation of a given cost function. By contrast, patrol routing dispatchers often do not search for a solution that optimises any specific value. Instead they search for routes that bring about satisfactory values for certain metrics as long as the routes followed by the agents are somewhat unpredictable and are non-deterministic. In contrast to the Turn DSL, Athos allows an explicit definition of a list of nodes to visit in the exact specified order or the definition of a set of nodes which have to be visited in an order that optimises a user-defined function. Most importantly, in the simulation framework of Steil et al., there is no concept of velocity or congestion. Agents move along the underlying graph among neighbouring nodes one node per time step. Due to the fact that there are no congestion effects or any changes in the movement speeds of agents, their approach cannot be used to simulate dynamic vehicle routing problems where travel times are subject to fluctuations depending on the current traffic situation.

Another platform for traffic and transport simulations is MATSim (Horni, Nagel, and Axhausen 2016). MATSim is a multi-agent microsimulation system based on the co-evolutionary principle. This means that the agents in the system are equipped with a set of plans to follow. Throughout multiple iterations the agents try and evaluate the outcome of the plans at their disposal. In each cycle a plan is selected, applied and evaluated. With a given probability, agents modify different dimensions of their plans. For example, agents can vary the

time they leave a given location, choose a different route or switch to a different mode of transport. Each agent seeks to optimise its individual outcome.

Maciejewski and Nagel present a MATSim-based approach to evaluate algorithms for the DVRP (Maciejewski and Nagel 2012). At the same time, their work intends to plan demand-responsive transport services (DRT services) using the MATSim framework. In their approach MATSim is used to calculate time-dependant travel times for a given scenario. The calculated data is then merged with additional demand and supply data which results in a (dynamic) VRPTW. This way, the authors designed four scenarios. The scenarios were designed to closely resemble traffic situations common to urban environments. Two of the scenarios analysed courier services while the other two scenarios investigated taxi services. They calculated solutions in two ways: First, they solved the problem using time-dependant travel times. Second, they produced solutions based on average travel times. The latter results were then applied to the problem with time-dependant travel times. The authors found that for the courier services the routes calculated on the basis of the time dependent travel times considerably outperformed those based on average travel times. Their explanation is that knowledge of time-dependant travel times allows for the calculation of routes that avoid congested roads. Interestingly, for the taxi services travel times for the solutions of both approaches were rather similar which the authors explain with the nature of the demands to taxi-services which does not lend itself to careful planning. For both scenarios, the solutions based on average times violated the defined time windows when applied to the problem that featured time-dependent travel times.

GAMA (Grignard et al. 2013) is a sophisticated simulation-platform. It features the GAML DSL which was designed for modelling agent-based simulations. The DSL's meta-model elements can be divided into elements for the definition of aspects related to entities, space and time. A species element is to GAML what a class is to object-oriented languages like

Java. A distinct feature of the GAML meta-model is that it allows agents to form containment hierarchies that can be used to model different levels of detail in a simulation. The species definition is also used to equip agents with skills like movement and attributes like movement speed. The species of an agent also defines a set of actions and reflexes. Actions represent behaviour that an agent executes when asked to. By contrast, a reflex represents behaviour that the agent executes in every step of the simulation given that all guard conditions hold. Even though GAML could also be used to model transport and routing problems, it requires modelling in a language not specifically tailored towards this domain. Thus, models are on a less abstract level which makes harder to comprehend and communicate by domain experts.

CONCLUSIONS AND FUTURE WORK

We have demonstrated how Athos can be used to model dynamic vehicle routing problems and how solutions can be computed. Besides the evolutionary algorithm presented here Athos provides other heuristics for solving a variety of complex routing problems (see Hoffmann, Guckert, et al. 2018 and Hoffmann, Chalmers, et al. 2019).

At the moment, we extend the Athos environment with a flexible interface to Open Street Map (www.openstreetmap.org) so that the definition of the underlying network can be generated from OSM data rather than be coded manually. Beyond that our concern is to measure general usability aspects of the language by letting domain experts assess the applicability of Athos. While we currently aim at improving the modeling capabilities of Athos our long-term intention is to develop an integrated instrument that allows a domain expert to describe and solve real world traffic related routing problems without any need for algorithmic decisions. The system will choose appropriate methods and heuristics and generate efficient best-practice code for the target platform.

REFERENCES

- Dabia, S., E. Demir, and T. Woensel, van (2014). *An exact approach for the pollution-routing problem*. English. BETA publicatie : working papers. Technische Universiteit Eindhoven.
- Dalal, Sandeep and Rajender Singh Chhillar (2012). "Case Studies of Most Common and Severe Types of Software System Failure". In: *International Journal of Advanced Research in Computer Science and Software Engineering* 2.8, pp. 341–347.
- Dantzig, George, Bernard Ramser, and John Hubert (1959). "The Truck Dispatching Problem". In: *Management Science* 6.1, pp. 80–91.
- Grignard, Arnaud et al. (2013). "GAMA 1.6: Advancing the art of complex agent-based modeling and simulation". In: *International Conference on Principles and Practice of Multi-Agent Systems*, pp. 117–131.
- Hoffmann, Benjamin, Kevin Chalmers, et al. (2019). "Athos - A Model Driven Approach to Describe and Solve Optimisation Problems: An Application to the Vehicle Routing Problem with Time Windows". In: *Proceedings of the 4th ACM International Workshop on Real World Domain Specific Languages*. RWDSL '19. Washington D. C., DC, USA: ACM, 3:1–3:10.

- Hoffmann, Benjamin, Michael Guckert, et al. (2018). "A Domain-Specific Language For Routing Problems". In: *European Conference on Modelling and Simulation, ECMS 2018, Wilhelmshaven, Germany, May 22-25, 2018, Proceedings*, pp. 262–268.
- Horni, Andreas, Kai Nagel, and Kay W. Axhausen (2016). "Introducing MATSim". In: *The Multi-Agent Transport Simulation MATSim*. Ed. by Andreas Horni, Kai Nagel, and Kay W. Axhausen. Ubiquity Press, pp. 3–8.
- Maciejewski, Michał and Kai Nagel (2012). "Towards Multi-Agent Simulation of the Dynamic Vehicle Routing Problem in MATSim". In: *Parallel Processing and Applied Mathematics*. Ed. by Roman Wyrzykowski et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 551–560.
- Steil, Dana A. et al. (2011). "Patrol Routing Expression, Execution, Evaluation, and Engagement". In: *IEEE Transactions on Intelligent Transportation Systems* 12.1, pp. 58–72.
- Urquhart, Neil B., Emma Hart, and Alistair Judson (2015). "Multi-Modal Employee Routing with Time Windows in an Urban Environment". In: *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM, pp. 1503–1504.
- Urquhart, Neil and Achille Fonzone (2017). "Evolving Solution Choice and Decision Support for a Real-world Optimisation Problem". In: *Proceedings of the Genetic and Evolutionary Computation Conference*. GECCO '17. ACM, pp. 1264–1271.

AUTHOR BIOGRAPHIES



BENJAMIN HOFFMANN is a research assistant at Technische Hochschule Mittelhessen in Friedberg from which he also received his master's degree. He is also a PhD student at Edinburgh Napier University. His research activities are in domain-specific languages, model-driven software development and optimisation problems.



MICHAEL GUCKERT is a Professor of Applied Informatics at Technische Hochschule Mittelhessen and head of KITE - AAC (Kompetenzzentrum für Informationstechnologie - Advanced Analytics Cognitive Computing). He received a degree in Mathematics from Justus Liebig University Giessen and a PhD in Computer Science from Philipps University Marburg. His research areas are multi agent systems, model driven software development and applications of artificial intelligence.



KEVIN CHALMERS is an Associate Professor at Edinburgh Napier University where he leads the Computer Science and Software Engineering subject area. He gained his PhD from Edinburgh Napier University in 2009, examining the application of mobile concurrency models to ubiquitous computing. His research is focused primarily on concurrency and parallelism and how different technologies can support this.



NEIL URQUHART is a lecturer in Computing Science at Edinburgh Napier University where he is Programme Leader for the Computing Science. He gained his PhD from Edinburgh Napier University in 2002, writing a thesis examining the use of Software Agents and Evolutionary Algorithms to solve a real-world routing optimisation problem. His research interests include Evolutionary Computation and Agent-based Systems and their application to real-world optimisation problems

Automatische Qualitätskontrolle mittels Maschinellen Lernen und SAP Leonardo

Professor Dr. Frank Herrmann und Simon Büttner MSc.

Innovationszentrum für Produktionslogistik und Fabrikplanung (IPF)

Ostbayerische Technische Hochschule Regensburg

E-Mail: Frank.Herrmann@OTH-Regensburg.de

Schlüsselwörter

Industrie 4.0, SAP Leonardo, SAP S/4HANA, ERP-System, Machine Learning, durchgängig digitalisierter Produktionsprozess bzw. Unternehmensprozess.

Zusammenfassung

In vielen Unternehmen werden von Menschen durch Ansehen von Produkten Produktionsfehler identifiziert. Die übersehenen Produktionsfehler führen zunehmend häufiger zu teuren Kundenreklamationen. Für die Produktion von Fußballen wird in dieser Arbeit diese manuelle Qualitätssicherung analysiert. Als Alternative werden die Fußballen durch einen Roboter fotografiert und durch maschinelles Lernen werden Produktionsfehler auf diesen digitalen Fotos identifiziert. Dieser Prüfprozess wird in SAP Leonardo implementiert und damit wird der bereits in einem SAP System implementierte Produktionsprozess (bzw. Unternehmensprozess) erweitert, wodurch ein durchgängig digitalisierter Gesamtprozess implementiert wird. Dadurch werden signifikant bessere Prüfergebnisse erreicht und die Durchlaufzeit des Prüfprozesses wird substantiell reduziert.

Einleitung

In nahezu allen Unternehmensbereichen schreiten Digitalisierungsprozesse unaufhörlich voran. Hoher Kostendruck, die Erhaltung der globalen Wettbewerbsfähigkeit und die strategische Unternehmensentwicklung zwingen die Unternehmen zur effizienten Generierung, Einsatz und Verteilung von Informationen (Brettel et al., 2014). In diesem Zusammenhang hat auch der Begriff „Machine Learning“ zunehmend an Bedeutung gewonnen. Mit Machine Learning bestrebt man danach, einem Computer deklaratives Wissen, motorische und kognitive Fähigkeiten und neue Fakten durch Beobachtung und Experimentieren beizubringen (Michalski et al., 1983).

Die SAP AG (kurz SAP) konzentriert sich seit Jahren darauf, die digitale Transformation ihrer Kunden zu unterstützen. Um die Hauptaufgabe der Betreuung von Kunden, Lieferanten, Mitarbeitern und Gütern digitalisiert wahrnehmen zu können, setzt SAP auf S/4HANA, der aktuellen Generation ihrer ERP-Anwendung und SAP Leonardo, einem Portfoliebegriff für die Bündelung von Innovationstechnologien als „Digital Innovation System“.

Ebenso arbeitet SAP daran diesen Aspekt der digitalen Transformation in Unternehmensbereichen wie der Produktion umzusetzen, da dieser Unternehmensbereich viel Potenzial birgt. Der Einsatz von Roboterarmen ist dabei

keine neue Errungenschaft mehr. Allerdings bietet er in Kombination mit einer angebrachten Kamera, welche von den Produkten Bilder via Sensoren an die Cloud übermittelt, neues Potenzial für Automatisierung.

Gegenstand dieser Arbeit ist eine auf realen Daten basierende Fallstudie zur Qualitätssicherung bei der Herstellung von Fußballen, deren Verallgemeinerung zunächst vorgestellt wird. Für unseren Anwendungsfall wird der Prozess zur Qualitätssicherung erläutert. Dem schließt sich die Beschreibung und Analyse des Ist-Zustands an. Anschließend wird ein Konzept zur automatische Qualitätskontrolle mit einer Kamera, SAP Leonardo und maschinellem Lernen vorgeschlagen. Seine Umsetzung und die dabei erzielte Verbesserungen zeigen den Erfolg des Vorgehens. Schließlich fassen wir das Vorgehen zusammen und geben einen Ausblick, gerade auch im Hinblick auf andere Unternehmen.

Problemstellung

Viele Schritte laufen in Produktionsstätten noch immer händisch ab, wobei eine hohe Anzahl an Mitarbeitern benötigt wird, um teilweise noch triviale Arbeiten selbst zu erledigen. Dieser Arbeitsablauf hat zunehmend negativen Einfluss auf die Mitarbeiter. Oftmals wird über Demotivation, Monotonie und körperliche Beschwerden geklagt. Diese Arbeitseinstellung hat auch letztendlich Einfluss auf die Produktqualität. Die zu erledigenden Prüfschritte im Qualitätssicherungsprozess sind zwar in den Arbeitsanweisungen fest geregelt, jedoch führt eine subjektive Bewertung durch die unterschiedlichen Prüf-Mitarbeiter zu unterschiedlichen Bewertungsergebnissen.

Aufgrund dieser unterschiedlichen Ergebnisse nimmt die Reklamationsrate ebenfalls von Jahr zu Jahr zu und die Kunden klagen über eine uneinheitliche Qualität. Die dadurch entstehenden Reklamationskosten tragen zu einem signifikant sinkenden Unternehmenserfolg bei.

Um nun diesem negativen Trend ein Ende zu setzen, wird eine Fallstudie aufgesetzt, welche einzelne strategische Ansätze des Qualitätssicherungsprozesses in Betracht zieht und auf Profitabilität prüft. Durch den Einsatz von Machine Learning Algorithmen soll mithilfe von SAP Leonardo auf eine Alternative umgeschwenkt werden, welche verbesserte Ergebnisse verspricht.

Qualitätskontrolle bei der Herstellung von Fußballen

In dieser Fallstudie wird mit dem Verkauf von 12 Mio. Fußballen eines bestimmten Typs gerechnet. Jeder dieser Fußballen wird am Ende des Herstellungsprozesses einer

Qualitätskontrolle unterzogen. Die Kontrolle sieht vor, die Qualität des Balls hinsichtlich Naht, Form, Aufdruck sowie Farbe auf einer Skala von 1 - 10 zu bewerten. Diese Prüfkriterien sind auf S. 6 in Abb. 2 dargestellt.

Fertiggestellte Fußballbälle kommen zusammen mit einem leeren Qualitätsprüfschein (QPS) aus der Produktion auf ein Fließband in die Qualitätssicherungsabteilung. Dort wird nach jedem Kontrollschritt, der von einem der fünf Mitarbeiter (M1 bis M5) ausgeführt wird, der Schein mit Kennzahlen von 1 (ungenügend) bis 10 (ausgezeichnet) für die entsprechende Prüfung beschriftet. Die Prüfungen werden wie folgt durchlaufen:

Falls eine Wertung von unter sechs Punkten festgestellt wurde, wird der fragliche Ball vom Fließband genommen und ihn in eine nahe gelegene Box (Container für Ausschuss) gelegt. Dieser Ball wird später recycelt. Der QPS wird auf dem Band gelassen, damit am Ende des Bandes der Fehler in eine Statistik eintragen werden kann. Nach den vier Prüfungen folgt ein weiterer Arbeitsplatz, der die übrig gebliebenen Prüfscheine auswertet (M5 – Auswertung). Die Wertungen auf den Scheinen entscheiden darüber, ob der Ball erste oder zweite Wahl ist. Für einen Ball mit Qualität „erste Wahl“ darf kein Kriterium mit einer Wertung unter 8 sein. Ein Ball zweiter Wahl darf keine Wertung unter 6 zugeteilt bekommen haben, muss aber mindestens einmal mit über 8 Punkten bewertet worden sein. Bälle der zweiten Wahl werden auf ein an diesem Kontrollpunkt beginnendes zweites Fließband gelegt.

Die Bälle kommen nun in die Verpackung. Der Ausschuss wurde schon zuvor während der Qualitätsprüfungen in die Boxen gelegt.

In der Qualitätssicherung treten zwei Fehlerklassen auf. Zum einen können fehlerfreie Bälle aus Versehen in den Ausschuss gelangen, was in der Statistik als false positive Rate bezeichnet wird. Zum anderen können fehlerhafte Bälle an die Kunden verkauft werden, obwohl deren Qualität nicht einwandfrei ist. Dies kommt der false negative Rate gleich. Die Tabelle 1 veranschaulicht diese Fehlertypen:

	Ist: Fehlerhaft	Ist: Fehlerfrei
Behauptung: Fehlerhaft	True Positive	False Positive
Behauptung: fehlerfrei	False Negative	True Negative

Tab. 1 Konfusionsmatrix Qualitätssicherung

In der Fallstudie liegen die Kosten für einen false positive Treffer bei 50 Euro. Dies spiegelt die Personal- und Materialkosten pro Ball wider. Die Kosten für einen false negative Treffer werden durch die Formel

$$70\text{€} \cdot y + 10^9 \text{€} \cdot \frac{1}{1 + (e^{-30x + 6,5})^y}$$

näherungsweise beschrieben, wobei x die prozentuale Miss-Rate (false negative Rate, z.B. 2%) widerspiegelt und y die Anzahl der false negative Bälle ist. Y kann also aus x gebildet werden, indem $y = x \cdot 12 \text{ Mio. Bälle}$ berechnet

wird. Diese Formel berücksichtigt folgende Besonderheiten: Werden $x \leq 2\%$ aller Bälle aufgrund der Fehler von den Kunden zurückgeschickt, so sind nur geringe Kosten in Höhe von 70 Euro pro Ball zu tragen. Werden aber mehr als 2% der Kunden unzufrieden, führt ein Schneeballeffekt dazu, dass ein Imageschaden langfristige Umsatzeinbußen fordert, welche die Kosten für die zu ersetzenden Bälle deutlich übersteigen. Diese Variable beginnt für $x = 0,02$ zu steigen und führt schnell zu extremen Verlustbeträgen. Die zugehörige Kostenfunktion ist in der folgenden Abbildung 1 aufgezeichnet.

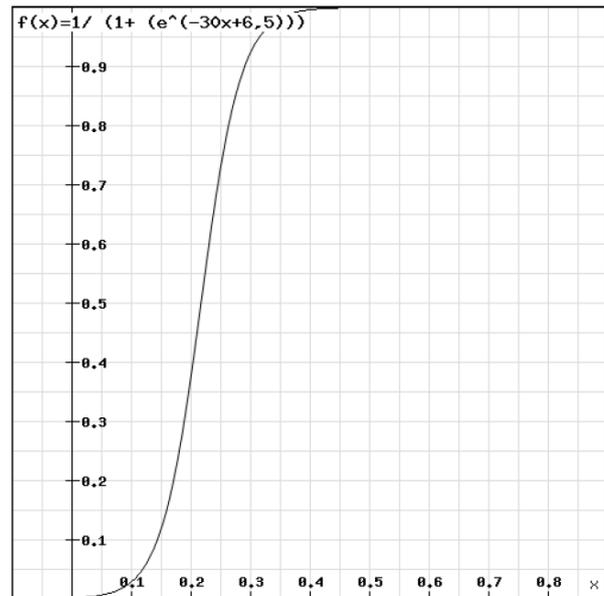


Abb. 1 Graph zur Beschreibung der false negative Kosten

Manuelle Qualitätskontrolle - Ablauf, Dauer und Kosten

Die Bearbeitungszeiten der einzelnen Arbeitsschritte der manuellen Qualitätskontrolle sind folgender Tabelle zu entnehmen:

Prüfung	Dauer	Anmerkung
Nahtprüfung	11 sec	
Formprüfung	10 sec	Mit Hilfsobjekten
Aufdruckprüfung	5 sec	
Farbqualitätsprüfung	5 sec	
Bewertung	5 sec	
Fließbandbewegung	9 sec	
GESAMT	45 sec	

Tab. 2 Dauer manuelle Qualitätssicherung

Es werden 12 Mio. Bälle (24 Mio. Stück $\cdot 0,5$) produziert. Dies bedeutet, dass insgesamt 150.000 Stunden an den Bällen kontrolliert wird ((12 Mio. Stück $\cdot 45\text{sec}/\text{Ball}$) / 3 600). Es kann in einem Jahr an ungefähr 250 Tagen gearbeitet werden, wobei in 3 Schichten gearbeitet wird. Ein Mitarbeiter arbeitet 7 Stunden am Tag in der Qualitätskontrolle. Dies bedeutet, dass ein Mitarbeiter knapp 2 100 Stunden im Jahr arbeiten kann. Für die Qualitätskontrolle am Fließband werden also knapp 72 Mitarbeiter in Vollzeit angestellt (150 000 Std / 2 100 Std).

Da die Bewertung des QPS fünf Sekunden dauert, kann theoretisch ein Ball alle fünf Sekunden aus der Qualitätssicherung kommen. Dies erfordert, dass jeweils zwei Mitarbeiter für die länger andauernden Prüfungen Naht und Form bereitstehen müssten, damit diese Taktung beständig bleibt. Insgesamt werden 12 Mio. Bälle bewertet.

Die Kosten für einen sehr geringen Imageschaden liegen bei kalkulierten 3,2 Mio. Euro. Insgesamt können also Mehrkosten in Höhe von 48,2 Mio. Euro erwartet werden.

Um die Funktion der false negative Formel unter Beweis zu stellen, sei folgender Hinweis angebracht. Wenn jeder zehnte Ball beim Endkunden offensichtliche Mängel

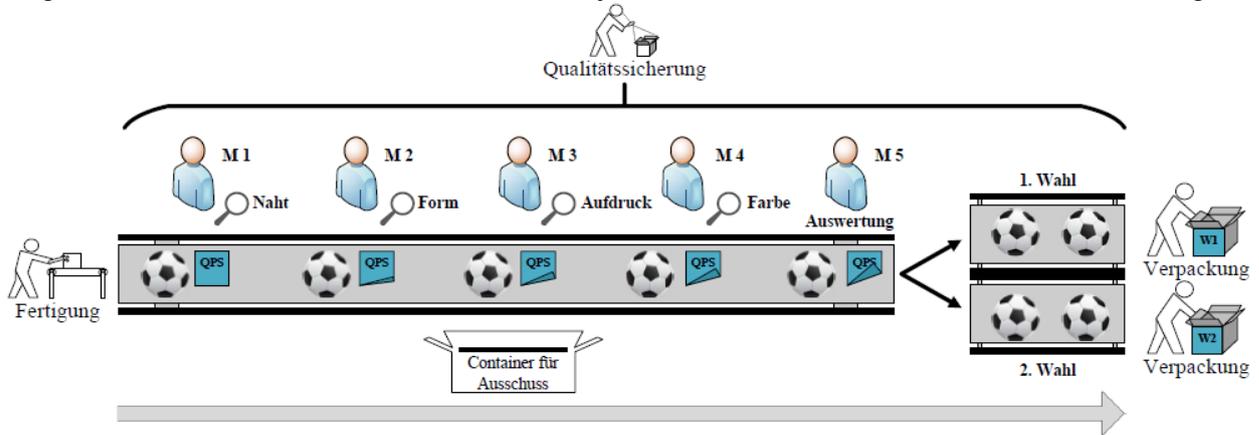


Abb. 2 Manuelle Qualitätssicherung

Somit benötigt die Qualitätssicherung 12 Mio. · 5 Sekunden = 60 Mio. Sekunden Zeit. Dies lässt sich auf 16 667 benötigte Stunden, an denen ein Band laufen muss, umrechnen. Ein Band steht dabei jährlich 5 250 Stunden (21 Std. · 250 Tage) zur Verfügung. Es wird nun allerdings mit einer Maschinenlaufzeit von 5 000 Stunden pro Jahr gerechnet. Dieser Wert ergibt sich aus der branchenüblichen Durchschnittslaufzeit dieses speziellen Bandes und berücksichtigt Ausfälle, Wartung und sonstige Stillstände. Dies wiederum bedeutet, dass mindestens vier parallele Linien in der Qualitätssicherung (5 000 Stunden · 4 > 16 666 Stunden) laufen müssen, um den Aufwand von 12 Mio. zu kontrollierenden Bällen in einem Geschäftsjahr stemmen zu können. Der ungefähre Kostenaufwand pro Stunde für die Instandhaltung der Qualitätskontrolle wird auf 8 Euro geschätzt. Dabei sind beispielsweise die Lohnkosten, die Kosten für Arbeitsbekleidung, die Räumlichkeiten und die Versicherungen enthalten. Daraus entstehen Gesamtkosten in Höhe von 1,2 Mio. Euro (8€/Std · 150 000 Std). Man kann davon ausgehen, dass ein Mitarbeiter in ca. 25 Prüfungen einen ordentlichen Ball versehentlich entfernt. Die vereinfachte false positive Rate liegt also bei 4%. Weiterhin geht ein Ball mit nur einem Qualitätsmangel zu 3% unerkannt in den Versand. Hat dieser Ball jedoch Mängel bezüglich mehrerer Prüfkriterien, sinkt die Wahrscheinlichkeit, dass kein Fehler erkannt wird und der Ball somit in den Handel gelangt. Dies führt zu einer vereinfachten false negative Rate von 2,5%. Diese Raten haben Auswirkungen auf das Kostenmodell. Wenn 4% aller fehlerfreien Bälle aussortiert werden, entstehen Mehrkosten in Höhe von 50 .Euro · 4% · 12 Mio. Bälle = 24 Mio. Euro. Wenn 2,5% der ausgelieferten Bälle zurückgehen, so müssen zusätzliche Kosten von

$$\frac{70\text{€} \cdot y + 10^9 \text{€} \cdot \frac{1}{1 + (e^{-30x + 6,5})}}{y} = 24,2 \text{ Mio €}$$

getragen werden. Hierbei berechnen sich die Rücksendungen und der Ersatz auf insgesamt 21 Mio. Euro.

aufweist und es somit zu einem Schneeballeffekt an Rücksendungen und einem mittleren Imageschaden kommt, so errechnen sich Mehrkosten in Höhe von 29,3 Mio. Euro.

Automatische Qualitätskontrolle mit SAP Leonardo und maschinellem Lernen – Konzept

Prozess

Statt einer manuellen Fehlerprüfung wird nun ein Roboterarm verwendet, an welchem eine Kamera angebracht ist. Diese Vorrichtung schießt pro Ball mehrere Bilder und schickt diese als Binärdaten via der Image Classification API an SAP Leonardo, welches dann die Bilder mittels Machine Learning analysiert. Die Bilder sowie die Machine Learning Analysen werden mittels einer SAP Leonardo Applikation für Endbenutzer sichtbar. Dabei werden alle Kriterien geprüft, indem jedes Bild klassifiziert wird und am Ende eine Entscheidung über die Qualität des Balls getroffen wird. Die Applikation bildet folgende Funktionalitäten im Dashboard ab. Dahinter stehen in Klammern die verwendeten Technologien, die durch den SAP Leonardo Engagement Prozess angeboten wurden:

- Analysedaten aus der Variabilitätsmodellierung zur Klassifizierung der Waben (Analytics, Big Data, ML).
- Zahlen zur Genauigkeit der Klassifikation (Analytics).
- Anzahl der verarbeiteten Bälle in einer Stunde sowie weitere Performance-Daten (Analytics, IoT).
- Benachrichtigungsoptionen bei Ausfall des Roboter-Arms oder anderer abhängiger Komponenten sowie Ausfallquoten (Analytics, IoT).
- Kosteneinsparungen (Analytics).
- Energieverbrauch (IoT).
- Schnittstelle zur Verbreitung des Trainingsmodells auf den Roboterarm (IoT, ML).

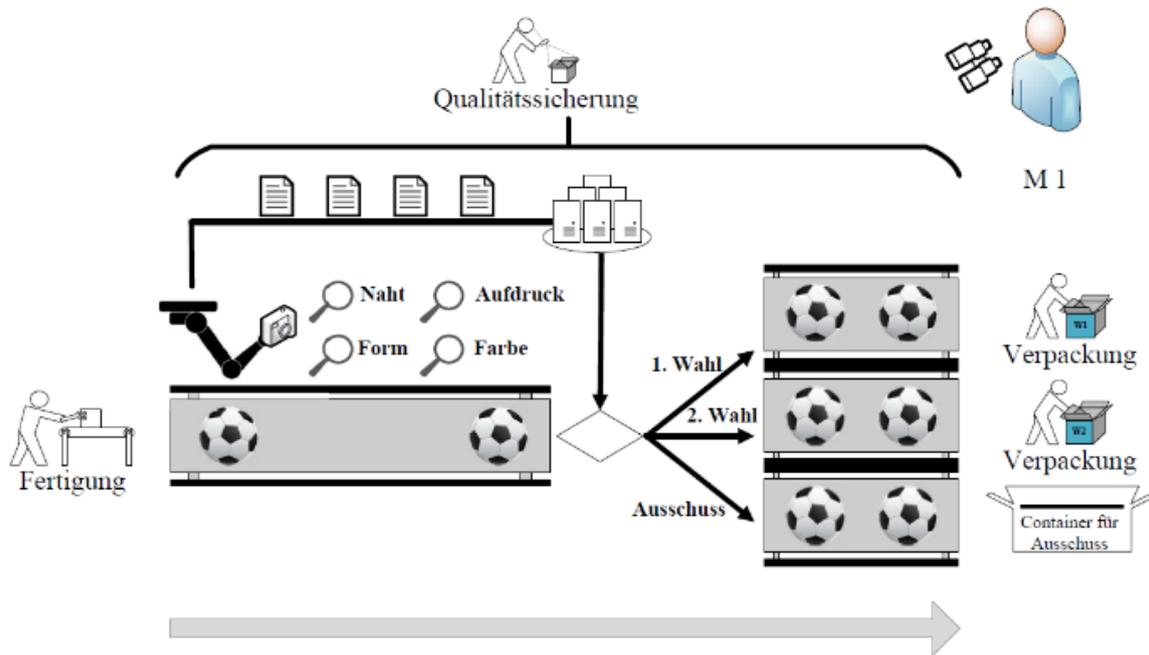


Abb. 3 Qualitätssicherung mit Roboterarm & SAP Leonardo

Der Prozess läuft im Detail – wie in Abb. 3 zu sehen ist – wie folgt ab: Die Produktionsabteilung legt die Bälle auf ein Förderband in unmittelbarer Nähe zur Qualitätssicherung. Das Band stoppt an dem Punkt (linker Ball), an dem der Roboterarm befestigt ist. Der Roboterarm schießt in sehr kurzer Zeit mit einer integrierten Kamera hochauflösende Bilder des Balles. Der Ball wird mit dem Fließband weitergetragen. Dank eines Time-Tracking Systems, das im Hintergrund läuft, kann jede Bilderreihe genau einem Ball zugeordnet werden. Das Fließband muss also nicht stoppen, bis die Analyse ein Ergebnis hervorbringt. Bis zum Ende des Fließbands hat das Softwaresystem Zeit, die Bilder zu analysieren, um eine Entscheidung zu treffen, ob der Ball in Ordnung ist oder in den Ausschuss gelangt. Der Roboterarm muss also die Bilder via WLAN an einen gut performanten Server schicken, auf dem die Bilder via Machine Learning und Bildverarbeitungsalgorithmen analysiert werden können. Das Machine Learning Modell ist in der Lage, die Naht, die Form, den Aufdruck sowie die Farbqualität synchron zu prüfen. Auf die genaue Funktionsweise der Machine Learning Algorithmen wird später im Rahmen der Fallstudie eingegangen. Es werden die gleichen Ratings verwendet, wie vor der Einführung von SAP Leonardo. Wenn die Machine Learning Unit durchlaufen ist, wird dank des Time-Tracking Systems der Ball eindeutig in „in Ordnung“ und „Ausschuss“ klassifiziert. Diese Information kann dann am Ende des Fließbands genutzt werden, um den Ball an den zutreffenden Prozess zu schicken. Zudem hat das IT-Team eine kleinere Idee in die Tat umgesetzt. Zur weiteren Akzeptanzschaffung der Neuerung wurde ein Monitor in der Nähe des Roboterarms aufgestellt, der die Klassifizierung mittels roter und grüner Ampeln symbolisch visualisiert.

Konzept vom maschinellem Lernen

Am Beispiel der Oberflächenprüfung wird nun die Funktionsweise des ML Algorithmus beschrieben. Bei der

Oberflächenprüfung scannt die Kamera jede einzelne der 32 Waben ein und übermittelt das Bild via der Image Classification API an das SAP Leonardo System.

Diese Daten werden in ein Gitternetz aus 418 (22 · 19) Pixel übertragen, wobei jeder Pixel einen Farbwert zwischen 0 (schwarz) und 1 (weiß) hat. Somit lässt sich der Aufdruck jeder Wabe zahlenmäßig genau erfassen. Diese 418 Farbwerte werden nun an die Tabelle 1 der in SAP Leonardo integrierten Datenbank übertragen. Nach der Aufbereitung werden die Daten mit den im Koordinatensystem hinterlegten Daten abgeglichen. Jede dieser 32 Waben soll nun im Optimalfall eine korrekte Zuordnung erhalten. Falls nun aber die Wabe mit Verunreinigungen bzw. mit Oberflächenfehlern behaftet ist, wird es zu keiner Übereinstimmung kommen. Das System wird alarmiert und der Ball wird als qualitativ minderwertig aussortiert.

Mittels dem kNN-Modell und der von (Keyers, 2007) beschriebenen Variabilitätsmodellierung wird nun dieses Bild analysiert und zum Schluss festgestellt, ob sich alles an korrekter Position befindet und keine größeren Druckfehler vorhanden sind.

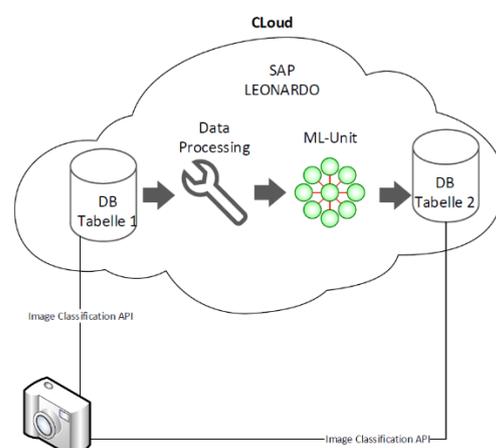


Abb. 4 IT Architektur SAP Leonardo

Die von SAP Leonardo zur Verfügung gestellte IT Architektur setzt sich dabei folgendermaßen zusammen – diese sind in Abbildung 4 visualisiert.

Zu Beginn werden die Rohdaten an Tabelle 1 einer Datenbank (DB) übermittelt. In dieser werden alle von der Kamera übermittelten Daten zusammen mit einem eindeutigen Key gespeichert. Dies wird benötigt, um einen Zugriff auf alle Original-Daten zu ermöglichen.

Von Tabelle 1 aus werden alle Daten automatisch an die Data-Pre-Processing Unit weitergeleitet. Hier werden alle Roh-Datensätze mit weiteren Daten wie z.B. Time stamps versehen und zusätzlich noch für die ML-Unit vorselektiert und normiert. Die Normierung ist wichtig, um mögliche Ausreißer-Werte zu eliminieren, wodurch das Ergebnis nicht zu verfälscht wird. Außerdem werden nicht benötigte Datensätze gelöscht, um die Gesamtperformance des Systems zu steigern.

Die fertig aufbereiteten Daten werden nun an die ML-Unit als Input für das ML Modell übermittelt. Die genaue Funktionsweise dieses Modells wird im darauffolgenden Abschnitt näher spezifiziert.

Das Ergebnis der ML-Unit wird abschließend in einer weiteren Tabelle der Datenbank gespeichert. In Tabelle 2 lassen sich alle Ergebnisse zur Klassifizierung aus dem vorhergehenden Schritt auslesen, evaluieren und visualisieren. Letztendlich wird das Ergebnis der Waben-Klassifizierung in Real-Time zurück an die Terminals des Qualitätsprüfungs-Bandes übermittelt.

SAP Leonardo stellt dabei sicher, dass alle Daten des Kamera-Arms korrekt empfangen werden, zwischen den einzelnen Units gestreamt und letztlich in Real-Time analysiert und an die Qualitätssicherung rückgemeldet werden. SAP stellt somit eine kontinuierliche Weiterverarbeitung bzw. Weiterleitung der Daten zu jedem Zeitpunkt sicher.

In den folgenden Abschnitten wird nun erklärt, wie die Klassifizierung in der ML-Unit abläuft.

Vertiefung: die Nächste-Nachbar-Regel (kNN)

Die nächstgelegene Nachbarn-Regel (kNN) ist eine der ältesten und einfachsten Methoden zur Musterklassifizierung. Dennoch führt sie häufig zu wettbewerbsfähigen Ergebnissen und in bestimmten Bereichen hat sie in Kombination mit Vorwissen den Stand der Technik erheblich verbessert. Die kNN-Regel klassifiziert jedes nicht gekennzeichnete Beispiel anhand der Mehrheitsbewertung unter seinen k-nächsten Nachbarn im Trainingssatz. Seine Leistung hängt daher entscheidend von der Entfernungsmetrik ab, die zur Identifizierung der nächsten Nachbarn verwendet wird. In Ermangelung von Vorwissen verwenden die meisten kNN-Klassifizierer einfache euklidische Distanzen, um die Unähnlichkeiten zwischen den als Vektoreingaben dargestellten Beispielen zu messen (Weinberger & Saul, 2009). Zusammen mit des in Kapitel „Variabilitätsmodellierung“ beschriebenen Verfahren soll die kNN konkret in Bezug auf die Fallstudie beschrieben werden.

Vertiefung: Variabilitätsmodellierung

Es ist eine große Herausforderung, einem Computer das Verstehen von Bildern beizubringen. Durch Einbeziehen typischer Veränderungen in Bildern ist man jedoch imstande, die Erkennung bzw. die Zuordnung von Bildern zu

verbessern. Hierzu wird ein Verfahren beschrieben, welches die Variabilität in der Klassifikation von Bildern untersucht.

Ziel der Klassifikation soll sein, aus vorhandenen Daten eine möglichst optimale Entscheidung abzuleiten. Diese Entscheidungen werden in einer Zuordnung zu einer Klasse $k \in \{1, \dots, K\}$ beschrieben. Optimalität wird erreicht, wenn die Fehlerrate bzw. die falschen Klassenzuordnungen minimiert werden. Zurückgegriffen wird dabei auf Trainings- und unabhängigen Testdaten. Weiterhin wird das Wissen eingesetzt, dass bestimmte Variationen in den Bildern deren Klassenzugehörigkeit nicht verändert. Zur Behandlung der Variabilität wird dabei auf invariante Distanzfunktionen – die Nächste-Nachbar-Regel (kNN) – gesetzt (Keyers, 2007).

In Bezug auf die Fallstudie der Waben kann die Variabilität mit der in Kapitel „Grundlagen“ beschriebenen Qualitätsbewertungen verglichen werden. Bis dato wurden all jene Waben mit Wertung > 8 als erste Wahl eingestuft. Da dadurch jede Wabe einen gewissen Qualitäts-Spielraum hat, innerhalb dessen die Qualität als erste Wahl eingestuft werden kann, muss dies im Rahmen des Machine Learning Algorithmus kNN berücksichtigt werden.

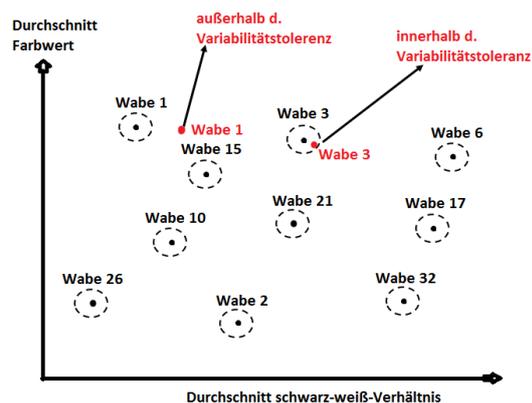


Abb. 5 Visualisierung der Variabilität

Automatische Qualitätskontrolle mit SAP Leonardo und maschinellem Lernen – Umsetzung

Ziel von SAP Leonardo ist es, dass jede von der Kamera aufgezeichnete Wabe vom ML-Algorithmus auf Basis des zuvor erläuterten Verfahrens zu klassifizieren. In diesem Anwendungsfall werden die Entscheidungen in die Klassen $k \in \{1, \dots, 32, \text{Fehler}\}$ vorgenommen. Jede einzelne Wabe fungiert dabei als eigene Klasse. Diese 33 Klassen stehen dabei schon im Vorfeld fest, auf Basis deren das Modell nun mit Trainingsdaten gelernt wird. Hierzu werden die Daten von insgesamt 100 000 Waben mit sehr guter Qualität herangezogen. Diese Durchschnittswerte, basierend auf dem jeweiligen Farbwert und dem schwarz-weiß-Verhältnis, werden nun zur besseren Visualisierung etwas vereinfacht in einem Koordinatensystem dargestellt.

Die grundlegende Voraussetzung hierbei ist, dass jede Wabe unterschiedliche Farb- und Schwarz-Weiß-Werte

hat. Somit kann jeder Wabe im Koordinatensystem ein eigener Bereich zugeordnet werden, ohne sich mit anderen zu überschneiden.

Die schwarzen Punkte repräsentieren hierbei die einzelnen Waben (1,...,32), welche die Trainingsdaten repräsentieren. Der gestrichelte Kreis um die Punkte stellt die Variabilität dar, welche sich aus dem erlaubten maximalen Qualitäts-Spielraum, im Rahmen des kNN ermittelt, ergibt.

Die roten Punkte sind nun neue bzw. unbekannte Testdaten, also eine neu abfotografierte Wabe, welche es jetzt zu klassifizieren gilt. Diese Test-Wabe wird nun entsprechend ihrer Farbwerte und des schwarz-weiß-Verhältnisses an der entsprechenden Stelle im Koordinatensystem platziert. Mittels der euklidischen Distanz wird nun der nächste Nachbar (k=1) ermittelt. Am Beispiel von Wabe 3 (rot) aus Abb. 5 wäre dies ebenfalls Wabe 3 (schwarz). Da der rote Punkt von Wabe 3 innerhalb der gestrichelten Linie liegt und die Distanz zu seinem Nachbarn sehr gering ist, wäre eine entsprechende Klassifizierung erfolgreich. Ein Beispiel eines solchen Falls wären die 32 Waben eines Fußballs, deren Qualität der Waben 24-mal mit 9 und 8-mal mit 10 bewertet wurde. Somit hätte der rote Punkt nur eine sehr kurze Distanz zu seinem schwarzen Trainingspunkt und würde als qualitativ hochwertig eingestuft werden.

Am Beispiel von Wabe 1 (aus Abb. 5), würde man er-

die restlichen 27 mit 9 bewertet worden wären. Dadurch wäre die Qualität einiger Waben ungenügend und der zugehörige Fußball müsste dementsprechend aussortiert werden.

Abb. 6 zeigt etwas vereinfacht den genauen Klassifizierungs-Prozess von Wabe Nr. 24 von der Original-Fotografie bis zu seiner endgültigen digitalen Transformation in einen Feature-Vektor. Weiterhin soll sie dabei den Unterschied zwischen einer Wabe mit guter Qualität (oben) und einer Wabe mit schlechter Qualität (unten) veranschaulichen.

Ein Feature-Vektor ist dabei ein eindimensionales Array bestehend aus mehreren Nullen und Einsen und stellt eine maschinenlesbare Verschlüsselung der Positions- und Farbdaten jeder einzelnen Wabe dar. Dieser Vektor bildet dabei den Test-Parameter ab. Auf der rechten Seite der Abbildung ist der ebenfalls ein Feature-Vektor zu sehen, welcher die qualitativ sehr guten Trainings-Waben beinhaltet. Mittels des kNN-Algorithmus wird nun der Feature-Vektor der Test-Wabe mit dem Feature-Vektor der Trainings-Waben verglichen. Zum einen wird als Ergebnis ausgegeben, ob die zu testende Wabe

- der Trainings-Wabe am nächsten ist und
- nahe genug an der Trainings-Wabe liegt.

Sind beide Punkte erfüllt – wie bei der oberen Wabe

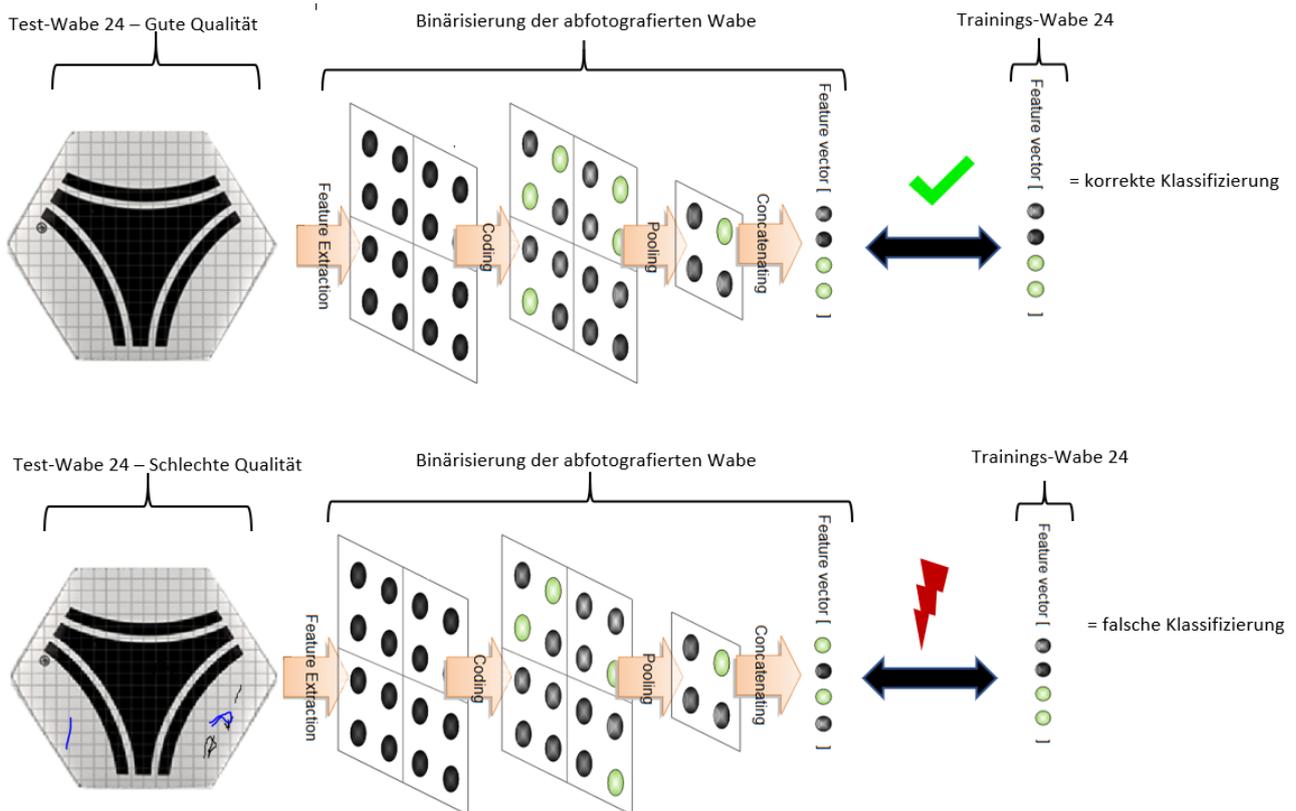


Abb. 6 Binärisierung einer Wabe

kennen, dass auch hier Wabe 1 (schwarz) tendenziell der nächste Nachbar wäre. Da nun allerdings der rote Punkt eine viel zu große Entfernung zu seinem schwarzen Pendant aufweist, wird es zu keiner erfolgreichen Klassifizierung kommen. Dies könnte beispielsweise der Fall sein, wenn 5 der 32 Waben mit einem Qualitäts-Wert < 7 und

aus Abb. 6 – so kommt es zu einer korrekten Klassifizierung und besteht somit den Qualitätstest.

Sind jedoch Beschädigungen auf der Wabe vorhanden, so ist einer der beiden Punkte nicht erfüllt und es kommt

dadurch – wie die untere Wabe in Abb. 6 - zu keiner erfolgreichen Klassifizierung. Die Wabe besteht den Qualitätstest nicht.

Dies weist darauf hin, dass das eben angewandte Verfahren das Erlernen relevanter Feature-Vektoren für die Durchführung einer genaueren Vorhersage, wie bereits von (Essid, Laga & Samir, 2018) beschrieben, evaluiert.

(Keyers, 2007) hat bereits eruiert, dass Klassifizierungen von Bildern imstande sind, eine nur sehr geringe Fehlerrate aufweisen zu können. So konnten von ihm Fehleraten von nur 0,52% statuiert werden.

Die folgenden Zahlen zeigen auf, wieviel Zeit die Qualitätskontrolle eines Balls (32 Waben) nach der Einführung von SAP Leonardo in Anspruch nimmt und welche Schritte der Qualitätsprüfung einen zeitlichen Einfluss darauf haben.

Prüfung	Dauer
Bewegung Roboterarm	9 sec
Klassifizierung d. Bilder	/
Fließbandbewegung	6 sec
GESAMT	15 sec

Tab. 3 Dauer Qualitätssicherung mit SAP Leonardo

Die Klassifizierung der Waben geschieht direkt nach der Aufnahme der Bilder während sich der Roboterarm noch bewegt. Bis der Ball also durch die Fließbandbewegung bei dem Bildschirm, welcher die Klassifizierungsergebnisse anzeigt, angelangt ist, steht schon fest, in welche Kategorie er eingeordnet werden kann. Da die Klassifizierung parallel abläuft, ist die Dauer dieser nicht relevant für die Zeit, die pro Ball benötigt wird. Bezüglich der zusätzlichen Kosten, die durch die Einführung von SAP Leonardo anfallen, wird mit folgenden Kosten kalkuliert:

Wie bereits bei der Qualitätskontrolle vor der Einführung von SAP Leonardo wird auch hier mit einer produzierten Stückzahl von 12 Mio. Bällen gerechnet. Die Qualitätskontrolle dieser dauert nun nur noch (12 Mio. Stück · 15sek/Ball) / 3600 = 50 000 Stunden. Da der Prozess der Prüfung der einzelnen Qualitätsmerkmale nun nicht mehr hintereinander abläuft, sondern alle Kriterien auf einmal geprüft werden, werden mehrere solcher Roboterarme benötigt. Man geht davon aus, dass jede dieser Anlagen ein ganzes Jahr über durchlaufen kann, also 8.760 Std pro Jahr in Betrieb ist. Es werden somit 50 000Std / 8 760 Std = 5,7 - also insgesamt sechs Roboterarme - benötigt. Die Auslastung der Arme beträgt nun weniger als 100. Während dieser Stunden ist jeweils ein Mitarbeiter für zwei Anlagen zur Kontrolle der Prozesse anwesend. Somit sind für diesen Prozess zu jeder Zeit drei Mitarbeiter anwesend. Um faire Arbeits- und Urlaubszeiten zu gewährleisten, werden hierfür 12 Mitarbeiter eingestellt. Für die Instandhaltung aller Anlagen werden zwei Mitarbeiter, einer für die Software, der andere für die Hardware, mit 288 Stunden pro Jahr beauftragt. Vor jeder Saison müssen zwei Mitarbeiter das Datenmodell entwickeln und trainieren. Hierfür werden 160 Stunden angesetzt. Insgesamt werden also circa 16 Mitarbeiter benötigt, welche jedoch alle nicht voll beschäftigt werden. Pro Jahr muss somit für die Qualitätskontrolle dieses Balles für 8.760Std · 3 + 48 Std + 160 Std = 26 488 Stunden aufgekomen werden. Ein Mitarbeiter verdient

nun deutlich mehr, da er entsprechend fortgebildet wurde und höhere Tätigkeiten ausführen muss. Es ist mit einem Kostenaufwand von 15 Euro pro Stunde auf Unternehmenseite zu rechnen. Im ersten Jahr nach der Einführung von SAP Leonardo fielen 200.000 Euro an zusätzlichen Kosten an.

Die jährlichen Kosten ohne Personalaufwand betragen 2.000 Euro + 500 Euro + 3.500 Euro = 6.000 Euro. Somit betragen die Gesamtkosten im ersten Jahr 200.000 Euro + 6.000 Euro + 26.488 Std · 15 Euro/Std = 603.320 Euro. Nachdem die einmaligen Kostenpunkte wegfallen, beträgt der jährliche Aufwand für die Qualitätskontrolle im zweiten Jahr nun 403.320 Euro. Die Personalkosten pro Jahr belaufen sich hierbei auf 397.320 Euro. Um die Kosten besser mit denen des vorherigen Prozesses der Qualitätskontrolle zu vergleichen, ist anzumerken, dass sämtliche zusätzlichen Kosten, wie beispielsweise Betriebskosten, bereits im Stundenlohn integriert sind. Es kann daher davon abgesehen werden, nur die Personalkosten miteinander zu vergleichen. Vielmehr sollten für einen fairen Kostenvergleich die Gesamtkosten des SAP Leonardo Systems, also 403 320 Euro jährlich nach der Einführung, in Betracht gezogen werden.

Für die false positive Rate wird nun die von (Keyers, 2007) ermittelte Fehlerrate von 0,52% herangezogen. In der manuellen Qualitätskontrolle lag die false negative Rate bei 2,5% und die false positive Rate bei 4%. Die false negative Rate liegt unter SAP Leonardo nur noch bei 0,05%, da es unwahrscheinlicher ist, dass ein defekter Ball die Qualitätskontrolle besteht, insbesondere wenn mehr als ein Qualitätskriterium nicht erfüllt ist. Der Algorithmus arbeitet somit wesentlich genauer als das menschliche Auge. Es ist zu berücksichtigen, dass diese Fehlerraten stark davon abhängen, wie gut das Variabilitätsmodell trainiert wird. Es wird großer Wert darauf gelegt, dass dieser Prozess besonders gründlich durchgeführt wird, um die Fehlerraten möglichst gering zu halten.

Die Auswirkungen dieser Raten auf das Kostenmodell lassen sich wie folgt beschreiben: 0,52% aller fehlerfreien Bälle werden trotz ausgezeichneter Qualität aussortiert. Der Schaden, der hierbei entsteht, beläuft sich nur noch auf 50 Euro · 0,52% · 12Mio. Bälle = 3 120 000 Euro. Es werden immer noch 0,05% der ausgelieferten Bälle zurückgesendet. Hieraus entstehen zusätzliche Kosten von

$$70 \cdot 0.0005 + 12 \text{ Mio.} + 10^9 \text{ Euro} \cdot \frac{1}{1+(e^{-30 \cdot 0,0005 + 6,5})}$$

= 1,94 Mio. Euro,

da zusätzlich zu den Kosten eines false positive Fehlers unter anderem Rücksendekosten und Imageschäden getragen werden müssen. Insgesamt ergeben sich daraus Mehrkosten von circa 5,06 Mio. Euro.

Automatische Qualitätskontrolle mit SAP Leonardo und maschinellem Lernen – Erzielte Verbesserungen

Zum Vergleich der manuellen mit der automatischen Qualitätskontrolle werden die Dauer der Qualitätskontrolle eines Balles, Anzahl der benötigten Mitarbeiter, Dauer der Mitarbeiterarbeit, jährliche Prozesskosten, Kosten aus false positive Fehlern und Kosten aus false negative Fehlern gegenübergestellt.

Alle Kennzahlen werden durch die automatische Qualitätskontrolle verbessert; die einzelnen Werte befinden sich in Tabelle 4. So konnte die Dauer der Qualitätskontrolle eines Fußballs durch SAP Leonardo um circa 62.5% gesenkt werden. Die Anzahl der benötigten Mitarbeiter für den Prozess der Qualitätskontrolle konnte ebenfalls deutlich gesenkt werden, konkret ausgehend von der bereits eingeführten Methode um circa 78%. Daraus resultierend sank auch die Dauer der benötigten Mitarbeiterarbeit um 83%. Die jährlichen Prozesskosten konnten um 66% verringert werden. Bei den Kosten für alle false positive Fehler liegt eine beträchtliche Einsparung von 87% vor, welche auf die Erfahrungen der Entwickler des Variabilitäts-Modells zurückzuführen ist. Schließlich liegt bei den Kosten für alle false negative Fehler eine Verringerung um 92% vor. Das größte Verbesserungspotenzial liegt bei den Fehlerkosten – in beiden Fällen über 85%. Diesen Einsparungen stehen erhöhte Einführungskosten von 603320 Euro im ersten Jahr gegenüber. Deswegen lohnt sich eine Einführung von SAP Leonardo. Ein weiterer Vorteil besteht darin, dass die Architektur rund um SAP Leonardo und des Roboterarms für andere bzw. ähnliche Anwendungsfälle verwendet werden kann.

	Manuell	Mit SAP Leonardo
Dauer Qualitätssicherung pro Ball in Sek	45	15
Anzahl Mitarbeiter	72	16
Dauer in Stunden	150 000	26 488
Jährliche Prozesskosten	1.2 Mio	403 320
Kosten false positive	24 Mio	3,12 Mio
Kosten false negative	24.2 Mio	1.94 Mio

Tab. 4 Gegenüberstellung Kennzahlen

Zusammenfassung

Diese auf realen Daten basierende Fallstudie zur Qualitätssicherung demonstriert eine vollautomatische Qualitätskontrolle mit SAP Leonard. Aufgrund den in der Literatur nachgewiesenen Ergebnissen durch maschinelles Lernen dürften sich substantiell geringere Fehlerraten realisieren lassen. Die dann erreichbaren Prozessbeschleunigungen um $\frac{200}{3}$ % und Kostenreduktionen auf einen kleinen Bruchteil sind so gewaltig, dass in naher Zukunft mit tatsächlichen Projekten gerechnet werden kann. Da Bilderkennung beim Kommissionieren von Montagebauteilen sehr wirkungsvoll sein dürfte, mögen viele Anwendungen in Industrie 4.0 sich durch diese Fallstudie inspirieren lassen.

Literaturverzeichnis

- Buxmann, Peter; Schmidt, Holger (Hg.) (2019): Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <https://doi.org/10.1007/978-3-662-57568-0>.
- Essid, O.; Laga, H.; Samir, C. (2018): Automatic detection and classification of manufacturing defects in metal boxes using deep neural networks. In: PloS one. DOI: 10.1371/journal.pone.0203192
- Gurney, Kevin N. (1997): An introduction to neural networks. London: UCL Press. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=182103>.
- Keuper, Frank (Hg.) (2013): Digitalisierung und Innovation. Planung, Entstehung, Entwicklungsperspektiven. Wiesbaden, s.l.: Bearingpoint. Online verfügbar unter <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=580534>.
- Keyers, Daniel (2007): Variabilitätsmodellierung für die Bilderkennung. In: Dorothea Wagner (Hg.): Ausgewählte Informatikdissertationen 2006. Bonn: Ges. für Informatik (Lecture notes in informatics, D-7).
- Malte Brettel; Niklas Friederichsen; Michael Keller; Marius Rosenberg (2014): How Virtualization, Decentralization And Network Building Change The Manufacturing Landscape. An Industry 4.0 Perspective.
- Michalski, Ryszard S.; Carbonell, Jaime G.; Mitchell, Tom M. (Hg.) (1983): Machine Learning. An Artificial Intelligence Approach. Berlin, Heidelberg: Springer (Symbolic Computation). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-662-12405-5>.
- Traeger, M.; Eberhart, A.; Geldner, G.; Morin, A. M.; Putzke, C.; Wulf, H.; Eberhart, L. H. J. (2003): Künstliche neuronale Netze. In: Der Anaesthetist 52 (11), S. 1055–1061. DOI: 10.1007/s00101-003-0576-x.
- Weinberger, Kilian; Saul, Lawrence K. (2009). Distance Metric Learning for Large Margin Nearest Neighbor Classification. *J. Mach. Learn. Res.* 10 (June 2009), 207-244.

Losbildung bei einem Engpass an Lagerfläche im Presswerk eines Automobilzulieferers

Professor Dr.-Ing. Frank Herrmann
Christian Reiter MA

Ostbayerische Technische Hochschule Regensburg
Innovationszentrum für Produktionslogistik und Fabrikplanung
Postfach 120327, 93025 Regensburg, E-Mail: Frank.Herrmann@OTH-Regensburg.de

Schlüsselwörter

Engpass an Lagerfläche, signifikant beschränkter Lagerplatz, ERP-System, SAP System, Losgrößenverfahren, Silver Meal-Verfahren, Feste Losgröße, Wagner-Whitin-Verfahren, Optimierungsmodell mit beschränktem Lagerplatz.

Abstract

Die im Presswerk eines Automobilzulieferers benötigte Lagerfläche stieg in den letzten Jahren so gewaltig an, dass die bisher verwendeten Produktionslose manuell reduziert werden mussten, um überhaupt einlagern zu können. Die dadurch sich ergebenden höheren Kosten werden quantifiziert. Zur Verbesserung werden die strukturellen Unterschiede zwischen den neuen Losen und den durch ein Standardverfahren in ERP-Systemen gelieferten Losen analysiert. Diese Unterschiede zeigen, dass die 2. Stufe in diesem 2-stufigen Vorgehen – 1. Los berechnen und dann 2. dieses geeignet reduzieren – die Entscheidungen der ersten Stufe in der Regel vollständig revidiert. Ein integrativer Ansatz ist erforderlich. Dies erfolgt durch ein lineares Optimierungsmodell mit beschränkter Lagerfläche. Das ILOG-Tool, ein Standardwerkzeug zur Lösung linearer Optimierungsmodelle, löst dieses Modell innerhalb von Minuten. Dass mit diesem optimalen Verfahren die Kosten substantiell verringert werden, wird simulativ nachgewiesen. Seit einigen Monaten wird dieses Verfahren im Rahmen einer rollenden Planung im Presswerk eingesetzt. Eine rollende Planung bewirkt in der Regel einen Verlust in der Lösungsgüte. In diesem Fall wurden nahezu optimale Lösungen erreicht, mit denen die Kosten substantiell verringert wurden. Die Werksleitung und die Mitarbeiter sind mit diesem neuen Verfahren sehr zufrieden.

1. Einleitung mit Problemverallgemeinerung

Seit geraumer Zeit befindet sich die Automobilbranche im Wandel. Die Hersteller reagieren auf neue Markttrends und wollen möglichst alle Kundenwünsche abdecken. Dies führte zu einer enormen Steigerung der Variantenzahl der PKW-Modellreihen, die am Markt erhältlich sind. Wie das Center Automotive Research der Universität Duisburg-Essen bereits 2012 untersuchte, s.

[Statistika19], stiegen die verfügbaren Modellreihen in Deutschland von 227 Modellen im Jahr 1995 auf 415 im Jahr 2015, s. Abbildung 1.

Diese Entwicklung setzte sich in den folgenden Jahren fort, da die Hersteller immer neue Fahrzeuge entwickeln, um auf dem stark umkämpften Markt möglichst jeden Anspruch der Kunden erfüllen zu können und die Lücken im Produktportfolio schließen zu können. Sehr deutlich zeigt sich die am Beispiel der Audi AG. Ende der 1980er Jahre produzierte Audi nur 3 Baureihen mit 10 Modellen. 2015 wurden bereits 12 verschiedene Baureihen mit 59 Modellen angeboten.

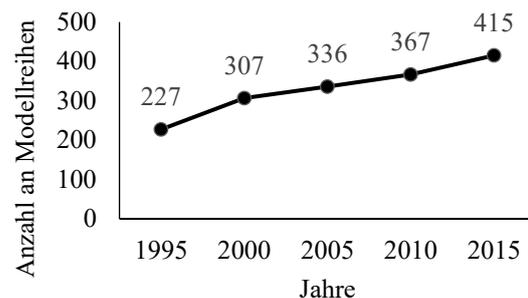


Abbildung 1: Anzahl an PKW-Modellreihen in Deutschland

Seitens der Zulieferer verursacht diese Entwicklung sogar eine noch viel stärker ansteigende Anzahl an Produkten. Dem wurde in den vergangenen Jahren durch ein geschicktes Komplexitätsmanagement, s. [Hirs15], begegnet. Dazu wurde vermehrt auf modulare Baugruppen gesetzt, so dass Teile in möglichst vielen Fahrzeugen verbaut werden können. Obwohl dadurch die Produktanzahl signifikant reduziert werden konnte, s. [Hirs15], gelang nur eine teilweise Kompensation. Deswegen ist trotzdem ein sehr signifikanter Anstieg des Bedarfs an Lagerplatz zu beobachten. Bei dem betrachteten Zulieferer – und dies ist unserer Erfahrung nach repräsentativ – wird dies durch eine Reduktion von Losen gelöst, die durch eine Heuristik bestimmt werden, die in ERP-Systemen typischerweise implementiert ist. Die Reduktion selbst basiert beim betrachteten Zulieferer – und auch dies ist unserer Erfahrung nach ebenfalls repräsentativ – auf Erfahrungswerten beziehungsweise Intuition. Durch die Lösung eines linearen Optimierungsmodells innerhalb von Minuten konnte eine substantielle Kostenreduktion erzielt werden.

Der Artikel ist wie folgt strukturiert: Zuerst wird dieses Problem präzisiert. Danach wird eine Simulation der bisherigen Lösung und der Anwendung von Standardverfahren in ERP-Systemen vorgestellt. Dazu werden die Verfahren beschrieben, die verwendeten Testdaten aus dem realen Unternehmensproblem werden angegeben und die Simulationsergebnisse werden analysiert. Ein lineares Optimierungsmodell mit beschränktem Lagerplatz wird vorgeschlagen. Die Ergebnisse seiner Anwendung auf das Unternehmensproblem – über die gerade genannten Testdaten – werden angegeben und analysiert. Die erzielten Resultate werden zusammengefasst und ihre Anwendbarkeit in anderen Unternehmen wird begründet.

2. Problemstellung beim Beispielunternehmen

Bei dem betrachteten Zulieferer handelt es sich um einen führenden Hersteller im Karosseriebau. Die Untersuchung beschränkt sich auf sein Presswerk. In den letzten fünf Jahren stieg die Zunahme der Anzahl an LKWs pro Jahr zunehmend stärker an, und zwar von ca. 11% auf ca. 26%. Die dadurch verursachte Zunahme an Lagerplatzbedarf wurde noch dadurch verschärft, da, neben Struktur und Verstärkungsteilen, auch viele großvolumige Teile, wie Außenhautkomponenten, die, um Beschädigungen zu vermeiden, in Spezialladungsträgern gelagert (und versendet) werden müssen, produziert werden. Dadurch mussten bereits vor vier Jahren zusätzliche Lagerflächen angemietet werden, wodurch Kosten entstanden, die in den ursprünglichen Preiskalkulationen nicht vorgesehen waren. Durch die bisherige Reduktion der Losgrößen konnte der ansteigende Lagerplatzbedarf lediglich vermindert, aber nicht vermieden werden.

Das Ziel ist ein Verfahren zur Losbildung, welches mit dem existierenden Lagerplatzbedarf auskommt und zugleich möglichst geringe Kosten (i.e. Lager- und Rüstkosten) bewirkt.

3. Aktuelles Verfahren und Standardverfahren

Neben dem bisher in dem realen Anwendungsfall verwendeten Verfahren werden auch Losgrößenverfahren verwendet, die in der Regel in kommerziell verfügbaren ERP-Systemen implementiert sind.

(a) *Verwendetes Verfahren*

Die bisherigen Losgrößen basieren vorwiegend auf dem Verfahren der festen Losgröße von Harris und Andler, die anschließend im laufenden Betrieb durch Erfahrungswerte an die gewachsenen Strukturen und Gegebenheiten angepasst wurden. Dabei wurde bei Überschreitungen der Lagerbeschränkungen reagiert und die Lose entsprechend reduziert. Dadurch erfolgt eine Reduktion ausschließlich aufgrund einer vom Werker noch maximal zu vertretenden Losgröße. Für solche Produkte wurde diese Losgröße immer verwendet – i.e. selbst dann, wenn keine Reduktion von dem berechneten Los nötig ist.

(b) *Ausgewählte Standardverfahren in ERP-Systemen*

Die in kommerziell verfügbaren ERP-Systemen verwendeten Losgrößenverfahren lassen sich in statische und dynamische Verfahren unterscheiden. Statische Verfahren, die in der Regel auf dem von Harris und Andler vorgeschlagenen Vorgehen basieren, berücksichtigen nicht, dass Kundenbedarfe für Perioden wie Tage oder Wochen zu erfüllen sind. Daher werden dynamische Verfahren analysiert. Hiervon sind in kommerziell verfügbaren ERP-Systemen mehrere Verfahren implementiert. Seitens der Forschung werden von diesen die Silver-Meal-Heuristik und die Groff-Heuristik am besten bewertet. Um in dieser Arbeit eine kostenseitige Verbesserung zu betrachten, wird das Wagner-Whitin-Verfahren analysiert, welches die Gesamtkosten über einen Planungszeitraum minimiert. Da die Silver-Meal-Heuristik ebenfalls die Gesamtkosten verwendet (i.e. die Gesamtkosten pro Perioden sind das Entscheidungskriterium) wird sich auf dieses Verfahren beschränkt; nach den Aussagen in der Literatur dürften unsere Ergebnisse sinngemäß auch für die Groff-Heuristik gelten.

Silver-Meal-Heuristik

Nach Wemmerlöv in [Wemm81] und [Wemm82] liefert die 1973 veröffentlichte Silver-Meal-Heuristik, vor allem bei festen Planungshorizonten, sehr gute Ergebnisse. Für eine detaillierte formale Darstellung und Erläuterung anhand eines Beispiels sei auf [Herr11] verwiesen.

Die Silver-Meal-Heuristik, wie auch die meisten anderen dynamischen Verfahren, vergrößern ein bisher gebildetes Los um den Bedarf der nächsten Periode so lange, bis eine Abbruchbedingung erfüllt ist. Nach [Temp08], [Knol87] und [Herr11] bewirkt diese Abbruchbedingung zu kleine Lose, wenn stark volatile Nachfragemengen vorliegen, da dann die durchschnittlichen Gesamtkosten pro Periode oftmals mehrere lokale Minima besitzen. Ebenfalls nach [Temp08], [Knol87] und [Herr11] berechnet die Silver-Meal-Heuristik bei Bedarfen mit einem fallender Trend ungünstige Lose im Vergleich zum Groff-Verfahren. Es sei betont, dass charakteristische Periodenbedarfe existieren, bei denen das Groff-Verfahren schlechtere Lose als die Silver-Meal-Heuristik berechnet. Nach [Temp08], [Knol87] und [Herr11] sind beide Verfahren im Mittel nahezu gleich gut.

Wagner-Whitin (SLULSP)

Von Wagner und Whitin wurde bereits 1958 in [WaWi58] ein lineares Optimierungsproblem formuliert, welches in der Literatur als Single-Level Uncapacitated Lot Sizing Problem (SLULSP) bezeichnet wird, und zur Lösung ein effizientes Verfahren angegeben, welches im Folgenden als Wagner-Whitin-Verfahren bezeichnet wird. Für eine Herleitung und detaillierte Beschreibung sei wiederum auf [Herr09] verwiesen.

Für Bedarfsfolgen mit einem festen Planungshorizont (T) wurde in [Wemm81] und [Wemm82] gezeigt, dass die Gesamtkosten beim Anwenden des Silver-Meal- und des Groff-Verfahrens im Mittel nur um etwa 1% über den durch eine optimale Lösung verursachten Gesamtkosten

liegen, und dass die in der Praxis favorisierten Verfahren der gleitenden wirtschaftlichen Losgröße und das Stückperiodenausgleichsverfahren erheblich schlechtere Lösungen ergeben; diese Zusammenfassung ist auch in [Temp08] angegeben und in [Kno85] wurden ähnliche Ergebnisse publiziert.

Die mittlere Rechenzeit von dem Wagner-Whitin-Verfahren ist für eine Periodenanzahl von T proportional zu T^2 und für Silver-Meal-Heuristik – wie auch für jede andere in der Literatur diskutierte Heuristik, s. [Herr11], ist die Rechenzeit proportional zu $T \cdot \log_2 T$, wodurch aus Sicht des Rechenzeitbedarfs ein optimales Verfahren in der industriellen Praxis einsetzbar sein dürfte.

In der industriellen Praxis ändern sich die Bedarfe im Zeitablauf, wodurch eine erneute Planung in der Regel zu einer geänderten Folge von Losen führt. Um eine Stabilität in der Einlastung von Aufträgen in der Produktion zu erreichen, wird nur eine Bedarfsänderung in (zeitlich) späteren Perioden zugelassen, die durch eine rollende Planung berücksichtigt wird. Durch die schrittweise Fixierung der Lose bei den in der Literatur diskutierten Losgrößenheuristiken, ergibt sich eine Anzahl an Periodenbedarfen, die bekannt sein müssen, so dass Bedarfsänderungen in späteren Perioden zu keiner Änderung von dem ersten Los führen, sofern die jeweilige Losgrößenheuristiken erneut angewendet wird. Ein solcher Entscheidungs-Horizont existiert auch bei einer optimalen Lösung, allerdings ist dieser deutlich länger. Stadler beschreibt in [Stadt00] Modifikationen, mit denen dieser Entscheidungs-Horizont soweit reduziert werden kann, dass er mit dem von Losgrößenheuristiken vergleichbar ist. Damit dürfte ein optimales Verfahren in der industriellen Praxis in der Regel einsetzbar sein.

4. Testdaten

Aus Datenschutzgründen wurde von den realen Unternehmensdaten abstrahiert. Die Ergebnisse entsprechen jedoch denen bei den tatsächlich vorliegenden Daten.

Statt alle Produkte im Unternehmen zu berücksichtigen, wird sich auf eine repräsentative Teilmenge beschränkt. Schließlich werden der Planungshorizont sowie die Rüst- und Lagerkosten festgelegt. Durch die geeignete Wahl der Bedarfe werden die im zeitlichen Verlauf im Unternehmen sich ändernden Auftragsituationen, wie hohe und niedrige Auftragslast, abgebildet.

(a) Repräsentative Teilmenge

Es werden Artikel, die in Eigenfertigung hergestellt werden, berücksichtigt. Diese werden aufgrund der Lagerplatzbeanspruchung und dem Verbrauch klassifiziert. Der Betrachtungszeitraum ging von Mai 2018 bis Mai 2019.

Im einzelnen werden die Artikel im Betrachtungszeitraum nach hohen, mittleren oder niedrigem Lagerflächenbedarf je Mengeneinheit (ME) unterteilt. Dabei schwankt der Lagerflächenbedarf zwischen 0.0001 Flächeneinheiten (FE) je ME bis 1.36 FE je ME. Um Klassen mit ähnlich vielen Elementen zu

erhalten, enthält die Klasse niedrig alle Artikel mit einem Lagerflächenbedarf von 0.0001 FE je ME bis 0.0013 FE je ME, die Klasse mittel enthält alle Artikel mit einem Lagerflächenbedarf von 0.00133 FE je ME bis 0.012 FE je ME und schließlich haben die Artikel in der Klasse hoch einen Lagerflächenbedarf, der zwischen 0.0122 FE je ME und 1.36 FE je ME liegt.

Die Verbräuche werden ebenfalls in hoch, mittel und niedrig unterteilt. Dazu wird für jeden auftretenden Verbrauch der prozentuale Anteil errechnet. Entsprechend der klassischen ABC-Analyse, s. z.B. [Herr11], befinden sich in der Klasse hoch alle Artikel, bei denen der prozentuale Anteil vom Verbrauch zwischen 91% und 100% liegt. Die Klasse mittel enthält alle Artikel, bei denen der prozentuale Anteil vom Verbrauch zwischen 71% und 90% liegt. Und schließlich enthält die Klasse niedrig alle Artikel, bei denen der prozentuale Anteil vom Verbrauch zwischen 0% und 70% liegt. Bei den Artikeln in der Klasse hoch liegen die Bedarfsmengen zwischen 5200 ME bis 72380 ME und in der Klasse mittel bewegen sich die Bedarfsmengen der Artikel zwischen 447 ME und 5142 ME.

Anschließend wurden aus den so klassifizierten Artikeln durch ein Leitungsgremium repräsentative Artikel ausgewählt. Dadurch ergeben sich die in Tabelle 1 eingeteilten Produkte mit ihrer entsprechenden Klasseneinteilung.

		Verbrauch		
		niedrig	mittel	hoch
Lagerplatzbedarf je Artikel	niedrig	Verstärkung links	Stütze Heck	Scharnier rechts A
		Verstärkung rechts		Scharnier links A
				Stütze quer
				Schließblech Mitte
				Schließblech rechts
				Schließblech links
				Scharnier rechts B
				Scharnier links B

Tabelle 1: Klassifizierung relevanter Artikel (wird fortgesetzt)

mittel	Tür vorne rechts B	Verkleidung Heck	Frontklappe innen
	Innenteil Heckklappe	Tür hinten links B	Tür innen vorne links
		Tür vorne links B	Tür innen vorne rechts
		Tür hinten rechts B	Verstärkung Mitte
		Dach A	Heckklappe innen
		Dach B	Frontklappe innen vorne
		Verstärkung Tür	Außenhaut Heck unten
hoch	Tür vorne links A	Tür vorne rechts A	Außenhaut Heck oben A
		Tür hinten rechts A	Außenhaut Heck oben B
		Tür hinten links A	Außenhaut Frontklappe

Tabelle 1: Klassifizierung relevanter Artikel

In der Gruppe der Produkte mit einer hohen Lagerflächenbeanspruchung sind vor allem Außenhautkomponenten zu finden, in der Gruppe mit einer mittlerer Lagerflächenbeanspruchung befinden sich primär Strukturteile und in der mit niedrigem Lagerplatzbedarf sind hauptsächlich Verstärkungsteile.

Bei den Produkten „Tür vorne links A“ und den „Tür vorne rechts A“ handelt es sich um Kuppelprodukte, weswegen ein gleich hoher Verbrauch zu erwarten ist. Verantwortlich für die in Tabelle 1 angegebenen Verbrauchsunterschiede ist beispielsweise eine Verwendung als Ersatzteil. Da im Unternehmen eine mengenmäßig unterschiedliche Produktion nicht möglich ist, ist der Bedarf von solchen Kuppelprodukten das Maximum aus den Einzelbedarfen seiner Endprodukte. Diese zusätzlichen Artikel werden eingelagert und im späteren Verlauf nach Prüfung eventuell entsorgt.

(b) Planungshorizont

Lose werden für die nächsten 50 Wochen berechnet. Allerdings wird bei Verwendung der festen Losgrößen bei Artikeln mit hohen Periodenbedarfen eine mehrfache Produktion pro Woche nötig. Diese Lose decken stets mindestens einen Tagesbedarf ab. Daher besteht für diese Produkte der Planungshorizont aus 350 Tagen.

(c) Rüstkosten

Anhand des Kuppelprodukts „Tür innen vorne links“ und „Tür innen vorne rechts“ werden die zu berücksichtigen Kostenarten zunächst erläutert und

anschließend für alle Unternehmensprodukte verallgemeinert.

Das Kuppelprodukt „Tür innen vorne links“ und „Tür innen vorne rechts“ wird mit 7 Werker und einer Rüstzeit von 0.4 Zeiteinheiten (ZE) auf der Presse 6 gefertigt. Zusätzliches Vorrüsten der Werkzeuge erledigt ein Werker mit einem Hallenkran innerhalb von 1.5 ZE. Ein Werker bereitet mit einem Gabelstapler Ladungsträger an der Anlage innerhalb von 0.25 ZE vor. Diese Aktivitäten verursachen, aufgrund der jeweiligen Stundensätze, Einzelkosten von

- 113.08 Geldeinheiten (GE) für die Presse 6 und 36.24 GE für die dabei eingesetzten Werker,
- 46.46 GE für Vorrüstarbeiten (mit dem Hallenkran),
- 3.67 GE für die Vorbereitung der Ladungsträger, wodurch insgesamt 199.45 GE an Rüstkosten anfallen.

Bei den Artikeln „Verstärkung links“, „Verstärkung rechts“ treten Kosten auf, weil Werkzeuge wegen dem geringen Platz im Presswerk nicht ebendort direkt gelagert werden können. Zum Antransport werden Kranzeiten von 0.5 ZE und Transportzeiten (mit einem Unimog) von 0.25 ZE je Werkzeugsatz benötigt. Wegen der hohen Rüstfrequenz ist eine weitere Schicht für diese Transporte notwendig. Dadurch ergeben sich Kosten von 19.16 GE. Zusätzlich fallen 113.08 GE für eine Presse und 36.24 GE für die dabei eingesetzten Werker an. Damit betragen die Gesamtrüstkosten 218.61 GE.

Im Allgemeinen können folgende Einflussfaktoren auftreten:

- Rüstzeit an einer Presse mit empirischer Berücksichtigung der Störzeiten: zwischen 0.4 ZE und 1.7 ZE.
- Anzahl an benötigten Werkern für einen Rüstvorgang: je nach Größe der verwendeten Presse und dem jeweiligen Auftrag zwischen 2 und 7 Workern.
- Vorrüstzeit der Werkzeuge je nach Anzahl der benötigten Werkzeugsätze. Bei zwei Pressen muss ein Vorrüsten während einem Anlagenstillstand stattfinden, wodurch eine um bis zu 1.5 ZE höhere Rüstzeit an der Anlage verursacht wird.
- Ladungsträgervorbereitung für zum Produktionsstart benötigte Ladungsträger von 0.25 ZE.
- Werkzeugtransport bei Ersatzteilaufträgen über Kranzeiten von 0.5 ZE und Transportzeiten (mit einem Unimog) von 0.25 ZE: also 0.75 ZE.
- Anzahl der benötigten Werkzeugsätze: zwischen 1 und 5.

Für alle Produkte wurden diese Berechnungen durchgeführt; Details sind in [Reit19] angegeben. Ihre Ergebnisse befinden sich in den Tabellen 2 und 3.

(d) Lagerkosten

Wie bei den Rüstkosten wird das Vorgehen anhand dem Kuppelprodukt „Außenhaut Frontklappe“, „Frontklappe innen“ und „Frontklappe innen vorne“ zunächst erläutert und anschließend für alle Unternehmensprodukte verallgemeinert.

Zur Handhabung befinden sich die einzelnen Teile in Ladungsträgern. In einem Ladungsträger für Innenteile

befinden sich 250 ME und in einem Ladungsträger für Außenhäute 40 ME. Aufgrund der Abmessungen der Innenteile und der Berücksichtigung von nicht nutzbarer Fläche, die nahezu zwangsläufig zwischen zwei direkt nebeneinanderliegenden Produkten auftritt – denken Sie beispielsweise an zwei Kugeln – beträgt der Flächenbedarf eines Ladungsträgers für Innenteile 3.12 Flächeneinheiten (FE). Im Fall von Außenhäuten benötigt ein Ladungsträger 5.2 FE an Fläche. Ladungsträger für Innenteile sind dreifach stapelbar und solche für Außenhäute sind zweifach stapelbar. Verantwortlich für solche unterschiedliche Anzahlen an Stapelungen sind die Belastbarkeit der Ladungsträger und deren Beschaffenheit sowie die baulichen Gegebenheiten in den Lagerhallen; letztere bewirkt im Unternehmen eine Reduzierung der umsetzbaren Stapelhöhe um 20 Prozent. Dadurch reduziert sich ihr Flächenbedarf bei Innenteilen auf $\frac{1}{3}$ und bei Außenhäuten auf $\frac{1}{2}$; also auf 1.04 FE für einen Ladungsträger für Innenteile und auf 2.6 FE für einen Ladungsträger für Außenhäute. Bei einer monatlichen Flächenmiete von 0.22 GE je FE und Woche betragen die Lagerkosten 0.23 GE je Woche für einen Ladungsträger für Innenteile sowie 0.57 GE je Woche für einen Ladungsträger für Außenhäute. Benötigt werden Lagerkosten je ME – also GE je ME und Woche – weswegen diese Kosten durch die Füllmenge des jeweiligen Ladungsträgers zu dividieren ist.

Zusätzlich wird die Kapitalbindung der Artikel dadurch berücksichtigt, in dem für ein Innenteil 0.003 GE je Woche und für eine Außenhaut 0.022 GE je Woche an Kosten anfallen.

Schließlich werden noch das Handling und die Administration des Lagers berücksichtigt, in dem für jeden Ladungsträger Kosten von 1.10 GE berechnet werden. Wegen der durchschnittlichen Lagerdauer von 21 Tagen, also 3 Wochen (Perioden), und der Anzahl an Produkten in einem Ladungsträger, werden dadurch Kosten von 0.001467 GE je Woche für ein Innenteil und 0.009167 GE je Woche für ein Außenteil verursacht.

Dadurch betragen die gesamten Lagerkosten 0.0053867 GE je Woche für eine „Frontklappe innen“ und für eine „Frontklappe innen vorne“ sowie 0.0454167 GE je Woche für eine „Außenhaut Frontklappe“. Ihre Summe ergibt die gesamten Lagerkosten für das Kuppelprodukt und betragen 0.0562 GE je Woche und ME. Da dieses Kuppelprodukt hohe Periodenbedarfe besitzt, werden Tage als Periodengröße behandelt und die gesamten Lagerkosten betragen letztlich 0.008 GE je Tag und ME.

Folgende Einflussfaktoren wirken sich dabei im Allgemeinen auf die unterschiedliche Ermittlung der Lagerkosten aus:

- Grundflächen der Ladungsträger zwischen 0.96 FE und 4.00 FE: sie ergeben sich aus den jeweiligen Behältermaßen und der nicht nutzbaren Fläche.
- Stapelhöhen der Ladungsträger zwischen 2 und 6: sie sind durch Belastbarkeit und Beschaffenheit der

Ladungsträger sowie durch die baulichen Gegebenheiten in den Lagerhallen bestimmt.

- Kapitalbindungskosten der Artikel zwischen 0.045 GE und 0.5 GE: sie ergeben sich durch die Komplexität und der Größe der Produkte.
- Füllmenge der Ladungsträger zwischen 14 und 2200 Produkte: sind durch die Art der Ladungsträger und Größe der Produkte bestimmt.

Für alle Produkte wurden diese Berechnungen durchgeführt; Details sind in [Reit19] angegeben. Ihre Ergebnisse sind in den beiden Tabellen 2, im Fall von Tagen als Periodengrößen, und 3, bei Wochen als Periodengröße, angegeben.

Artikel	Rüstkosten in GE	Lagerkosten pro Woche in GE
Innenteil Heckklappe	712.54	0.0097
Tür vorne/hinten links/rechts B 4 Kuppelprodukte	295.23	0.0444
Verstärkung links/rechts 2 Kuppelprodukte	168.04	0.0055
Außenhaut Heck oben B	356.42	0.0407
Tür A vorne/hinten links/rechts 4 Kuppelprodukte	295.23	0.0664
Verkleidung/ Stütze Heck 2 Kuppelprodukte	604.43	0.0097
Tür innen links/rechts 2 Kuppelprodukte	199.45	0.0191
Verstärkung Tür	712.54	0.0161
Scharnier links/rechts A 2 Kuppelprodukte	616.76	0.0090
Dach B	527.81	0.0156
Dach A	527.81	0.0152
Stütze quer	98.26	0.0020

Tabelle 2: Rüst- und Lagerkostenübersicht Artikel mit einem Betrachtungshorizont von 50 Perioden.

	Rüstkosten in GE	Lagerkosten pro Tag in GE
Frontklappe innen/vorne/außen 3 Kuppelprodukte	199.45	0.008
Außenhaut Heck oben/unten A 2 Kuppelprodukte	271.77	0.0603
Heckklappe innen	616.76	0.0011
Scharnier links/rechts B 2 Kuppelprodukte	98.26	0.0015
Verstärkung Mitte	271.77	0.0018
Schließblech links/rechts 2 Kuppelprodukte	98.26	0.0018
Schließblech Mitte	98.26	0.0002

Tabelle 3: Rüst- und Lagerkostenübersicht Artikel mit einem Betrachtungshorizont von 350 Perioden.

(e) Lagerplatzbeschränkungen

Die hergestellten Artikel müssen zu einem Teil in durch den Hersteller bereitgestellten Sonderladungsträgern gelagert und transportiert werden. Diese speziell für diese Produkte entwickelten Behälter stehen nur in einer begrenzter Stückzahl zur Verfügung. Die restlichen Produkte werden auf der verbleibenden Fläche gelagert. Damit kein Produkt unverhältnismässig viel Lagerfläche belegt, wurde für jedes Produkt eine maximal verwendbare Fläche festgelegt.

Die folgende Tabelle 4 enthält alle Produkte, die in Sonderladungsträger zu lagern sind mit ihrer Anzahl und ihrer erlaubten Füllmenge. Das Produkt aus Anzahl und Füllmenge ist folglich die maximale Losgröße.

Bei Kuppelprodukten ist die maximale Losgröße das Minimum der maximalen Losgröße seiner Bestandteile. So ist beispielsweise die maximale Losgröße von dem Kuppelprodukt „Tür vorne links B“, „Tür vorne rechts B“, „Tür hinten links B“ und „Tür hinten rechts B“ das Minimum aus 2660 Mengeneinheiten (ME), 2520 ME, 1960 ME und 840 ME, also 840 ME. Eine Bereinigung einer solchen Situation durch Beschaffung von zusätzlichen Ladungsträgern, aber auch durch Abbau von überschüssigen Ladungsträgern, die in dieser Fallstudie auftreten, ist beabsichtigt; es ändert nichts am grundsätzlichen Problem.

Für die restlichen Produkte wurde die produktspezifische verfügbare Fläche, auf welcher ein Artikel gelagert werden darf, durch den Flächenbedarf dieses Artikel dividiert. Der resultierenden maximale Lagerbestand befindet sich in der folgenden Tabelle 5.

Artikel	Anzahl Ladung- sträger	Füllmenge Ladungsträger in ME	maximale Losgröße in ME
Außenhaut Frontklappe	112	40	4480
Tür vorne links B	38	70	2660
Tür vorne rechts B	36	70	2520
Tür hinten links B	28	70	1960
Tür hinten rechts B	12	70	840
Dach A	6	140	840
Tür vorne links A	17	50	850
Tür vorne rechts A	17	50	850
Tür hinten links A	17	50	850
Tür hinten rechts A	17	50	850
Tür innen vorne links	23	200	4600
Tür innen vorne rechts	23	200	4600

Tabelle 4: Maximale Losgrößen in Mengeneinheiten (ME) für Artikel mit Spezialladungsträgern.

Artikel	Maximaler Lagerbe- stand in ME	Artikel	Maximaler Lagerbe- stand in ME
Dach B	1540	Außenhaut Heck oben B	3240
Innenteil Heckklappe	1400	Scharnier links B	30000
Verkleidung Heck	1500	Scharnier rechts B	30000
Stütze Heck	1500	Verstärkung Mitte	3500
Außenhaut Heck oben A	3780	Schließblec h links	110000
Heckklappe außen unten	3780	Schließblec h rechts	110000
Heckklappe innen	3300	Schließblec h Mitte	60000
Verstärkung Tür	600	Stütze quer	22500
Scharnier links A	13000	Verstärkung rechts	40000
Scharnier rechts A	13000	Verstärkung links	40000

Tabelle 5: Maximale Lagerbestände in Mengeneinheiten (ME) für Artikel mit Lagerplatzbeschränkungen.

(f) Bedarfe

Da der betrachtete Zulieferer sowohl einzelne Pressteile wie auch zusammengesetzte Baugruppen liefert, besteht das Produktportfolio des Presswerks aus Artikeln, die direkt zum Kunden geliefert werden und aus Bauteilen, die zuerst im Rohbau zusammengefügt werden müssen, wobei eine einstufige Produktion vorliegt.

Für Produkte, die direkt für den Versand zum Kunden bestimmt sind, treffen die Bedarfe via „Electronic Data Interchange“ (EDI) nach [KeLS10] und liegen für diese Untersuchung vor.

Bei allen anderen Produkten, die als Komponenten direkt in ein Endprodukt eingehen, entstehen aufgrund der Materialbedarfsplanung (MRP) wie folgt Bedarfe. Es liegen keine Lagerzugänge vor. Alle betrachteten Artikel haben einen Direktbedarfskoeffizienten von 1. Deshalb sind die Endprodukt-Bedarfe zugleich die Nettobedarfe für die einstufigen Losgrößenprobleme. Der Bedarfszeitpunkt ist der Produktionsstart des Endproduktes minus einer Periode für die Vorlaufzeit. Aus den bekannten Bedarfen für die Endprodukte, s.o., wird durch ein am Innovationszentrum für Produktionslogistik und Fabrikplanung vorhandenes Tool zur Materialbedarfsplanung die terminierten Bedarfe für die Komponenten auf die gerade beschriebene Art und Weise ermittelt.

Um allgemeingültige Kennzahlen zu erreichen, wird diese Bestimmung von Periodenbedarfen für die einzelnen Produkte 15-mal wiederholt – zur Anzahl an Wiederholungen s. den nächsten Abschnitt.

In den Tabellen zu den Bedarfen für jedes Produkt im Anhang sind, wegen den 15 Wiederholungen, die mittleren minimalen und maximalen Bedarfe in den Betrachtungszeiträumen sowie deren Mittelwerte und Standardabweichungen angegeben. Daraus lassen sich die Wertebereiche für die zur Berechnung der Losgrößen verwendeten Bedarfe der einzelnen Artikel und Baugruppen sowie deren Schwankungen gut erkennen. Dabei treten zum Teil sehr niedrige Bedarfe auf, bei denen die Mittelwerte im zweistelligen Bereich liegen und auch die Streuungen gering ausfallen, bis zu relativ hohen Bedarfen mit mittleren Bedarfen von über 3000 ME und Streuungen, die mehr als das doppelte des Mittelwertes betragen.

5. Analyse des aktuellen Verfahrens und der Standardverfahren in ERP-Systemen

Gegenstand dieses Abschnitts ist die Analyse der Wirkung der Losgrößenreduktion im betrachteten Unternehmen zur Einhaltung des beschränkten Lagerplatzes. Daher werden, wie im Abschnitt 3 begründet, die drei Verfahren Silver-Meal, Wagner-Whitin sowie das aktuell im Unternehmen verwendete Verfahren gegenübergestellt. Eine Voruntersuchung ergab, dass eine 15-malige Durchführung der Losgrößenberechnung zu statistisch signifikanten Ergebnissen führt.

Gleiche Reichweiten der durch die drei Verfahren bestimmten Losbildungen (i.e. Folge von Losen über den (gleichen) Planungszeitraum) bedeuten identische Lose (es ist eine notwendige und hinreichende Bedingung für identische Lose). Daher werden die auftretenden Reichweiten zunächst untersucht. Solche Unterschiede bewirken oftmals unterschiedliche Kosten sowie verursachte Lagerbestände und vor allem Verletzungen des beschränkten Lagerplatzes. Diese werden daher anschließend untersucht.

Wegen der hohen Anzahl an Messergebnissen, sind ihre einzelnen Kennzahlen zu den Losen in Tabellen je Artikel im Anhang angegeben.

(a) Reichweite der Lose

Da die Reichweite für diese Untersuchung hinreichend genau mit der Anzahl an Losen korreliert und es sich bei der Anzahl an Losen nur um eine Kennzahl handelt, wird nur diese untersucht.

Generell ist die Anzahl der Lose des aktuell verwendeten Verfahrens bei fast jedem betrachteten Artikel deutlich höher als bei den beiden alternativen Standardverfahren. Zwischen dem Silver-Meal-Verfahren und aktuell verwendeten Verfahren ist die prozentuale Erhöhung der Anzahl an Losen aufgeführt, wobei hier und im Weiteren die prozentualen Abweichungen immer bezogen auf den höheren Wert berechnet werden. Diese prozentuale Erhöhung schwankt zwischen 20% und 72%, liegt im Mittel bei 48.7% und hat eine Streuung von 16.3. Die Kennzahlen zeigen ferner bereits eine substantielle Erhöhung (der Anzahl an Losen) vom Wagner-Whitin- zum Silver-Meal-Verfahren. Diese schwankt zwischen 0% und 64.7%, liegt im Mittel bei 35% und hat eine Streuung von 20.1. Daher werden im Unternehmen gegenüber der kostenminimalen Lösung die Anzahl der Lose noch sehr viel höher erhöht (gegenüber dem obigen Vergleich zum Silver-Meal-Verfahren). Sie schanken zwischen 42.9% und 86.4%, haben einen Mittelwert von 67.2% und eine Streuung von 12.4.

(b) Kosten

Da die Anzahl an Losen mit der Anzahl an Rüstvorgängen korreliert, ist nach dem aufgezeigten Anstieg an verwendeten Losen zu erwarten, dass die Rüstkosten entsprechend ansteigen. Im Folgenden wird die prozentuale Erhöhung der Kosten gegenüber dem Verfahren angegeben, das als zweites genannt wird. Die Erhöhung von dem Wagner-Whitin- zum Silver-Meal-Verfahren schwankt zwischen 0% und 64.7%, liegt im Mittel bei 34.1% und hat eine Streuung von 18.9. Von dem Silver-Meal-Verfahren zum aktuellen Verfahren im Unternehmen dominieren die Erhöhungen, es treten auch substantielle Verringerungen auf. Im Einzelnen schwanken diese (auch negativen) Erhöhungen zwischen -42.3% und 72%, liegen im Mittel bei 40.35% und haben eine Streuung von 31.8. Wie aus diesen Werten bereits zu vermuten ist, werden gegenüber der kostenminimalen Lösung die Rüstkosten im Unternehmen noch deutlicher erhöht. Konkret schanken diese zwischen 0% und 86.35%, haben einen Mittelwert von 61.42% und eine Streuung von 21.49. Hohe Rüstkosten implizieren in der Regel geringere Lagerkosten. So sind diese Lagerkosten gegenüber dem optimalen Verfahren im Mittel sowohl beim Silver-Meal-Verfahren als auch bei dem aktuellen Verfahren im Unternehmen deutlich geringer. Gegenüber dem optimalen Verfahren schwanken diese (auch negativen) Erhöhungen durch das Silver-Meal-Verfahren zwischen -64% und 5.9%, liegen im Mittel bei -34% und haben eine Streuung von 21.7. Von dem Silver-Meal-Verfahren zum aktuellen Verfahren im Unternehmen schwanken diese (negativen) Erhöhungen

zwischen -65.3% und 218.17%, liegen im Mittel bei -6.17% und haben eine Streuung von 70.2. Und schließlich schwanken diese (negativen) Erhöhungen vom optimalen Verfahren zum Verfahren im Unternehmen zwischen -82.9% und 38.56%, liegen im Mittel bei -43.64% und haben eine Streuung von 32.86. In Summe schanken dadurch die Erhöhungen der Gesamtkosten vom Wagner-Whitin-Verfahren zum Silver-Meal-Verfahren zwischen 0.4% und 37.55%, liegen im Mittel bei 14.98% und haben eine Streuung von 12.82. Vom Silver-Meal-Verfahren zum Verfahren im Unternehmen dominieren Erhöhung, es treten auch Verringerungen auf. Sie schwanken zwischen -3.3% und 59.62%, liegen im Mittel bei 36.7% und haben eine Streuung von 18.26. Bei allen Produkten liegt ein Gesamtverlust an Kosten vor und die Gesamtkostenerhöhungen schwanken zwischen 10.72% und 72.81%, liegen im Mittel bei 46.09% und haben eine Streuung von 17.3.

(c) *Lagerbestände und Lagerbestandsüberschreitungen*
 Nach der obigen Analyse der Lagerkosten ist klar, dass bereits das Silver-Meal-Verfahren geringere Bestände als das Wagner-Whitin-Verfahren aufweist und für das Verfahren im Unternehmen gilt dies umso mehr. Da die Lagerkostensätze produktspezifisch sind, liegt keine Proportionalität zwischen diesen Beständen und den Lagerkosten vor. Daher sind die prozentualen Verringerungen des Gesamtlagerbestand im folgenden genannt; die prozentualen Abweichungen werden, auch im Weiteren, mit Bezug zu dem Verfahren angegeben, das als zweites genannt ist. Es sei angemerkt, dass die Tabellen im Anhang zusätzlich für die Lagerbestände von jedem Produkt sein Minimum, sein Maximum, seinen Mittelwert und seine Streuung enthalten. Vom Wagner-Whitin-Verfahren zum Silver-Meal-Verfahren schwanken die Verringerungen des Gesamtlagerbestand zwischen 174.92% und -77.11%, liegen im Mittel bei 50.56% und haben eine Streuung von 72.41. Vom Silver-Meal-Verfahren zum Verfahren im Unternehmen schwanken die Verringerungen zwischen 65.31% und -217.99%, liegen im Mittel bei 4.7978% und haben eine Streuung von 71.74. Die deutlichen Verbesserungen der Gesamtkosten (beim Wagner-Whitin-Verfahren) werden somit durch sehr viel höhere Lagerbestände erreicht. Konkret schwankt die substantielle Reduktion der Lagerbestände durch das Verfahren im Unternehmen gegenüber dem Wagner-Whitin-Verfahren zwischen 82.91% und -38.54 %, liegen im Mittel bei 44.43% und haben eine Streuung von 33.69.

Diese deutlichen Verringerungen der Lagerbestände durch das Verfahren im Unternehmen gegenüber dem Silver-Meal- und dem Wagner-Whitin-Verfahren bedeuten nahezu zwangsläufig, dass die vorliegenden Kostensätze zu große Lose präferieren. Die folgende Tabelle 6 enthält für jedes Losgrößenverfahren die über alle Produkte kumulierte Anzahl an Perioden mit einer Überschreitung der Lagerbestandsbeschränkung. Die durch diese Verfahren nicht lagerbaren Mengeneinheiten sind, kumuliert über alle Produkte und Perioden, ebenfalls in dieser Tabelle 6 angegeben und zwar

aufgeschlüsselt nach Minimum, Maximum, Mittelwert und Streuung.

	Anzahl an Perioden mit Überschreitung	min	max	μ	σ
WW	675	0	38646	14403	8375
SM	313	0	24574	1937	3660
Fest	0	0	0	0	0

Tabelle 6: Überschreitungen der Lagerbestandsbeschränkungen, kumuliert nach Produkten und (teilweise) nach Perioden; 3050 ist die Gesamtanzahl an Perioden.

Tabelle 6 zeigt, dass sowohl durch das Wagner-Whitin- als auch durch das Silver-Meal-Verfahren die mögliche Lagermenge substantiell überschritten wird. So tritt beim Wagner-Whitin-Verfahren in 22.13% der Perioden eine Überschreitung der Lagerbestandsbeschränkung auf. Deutlich geringer sind die Überschreitungen durch das Silver-Meal-Verfahren, nämlich in nur 10.26% der Perioden und die mittlerer Anzahl an zu viel zu lagernden Produkte ist nur etwa 13% von der aufgrund des Wagner-Whitin-Verfahrens. Andererseits verursacht das Verfahren im Unternehmen signifikant höhere Gesamtkosten als das Wagner-Whitin-Verfahren und auch als das Silver-Meal-Verfahren. Diese mögliche Kosteneinsparung erfordert eine Erweiterung der Lagerfläche, die nicht wirtschaftlich ist, auch deswegen, weil angenommen wurde, dass durch eine bessere Planung der beschränkten Lagerfläche deutlich geringere Gesamtkosten als durch das aktuell im Unternehmen verwendeten Verfahren realisiert werden können.

Eine weitere Analyse des aktuellen Verfahrens zeigt, dass die „Reduktion aufgrund einer vom Werker noch maximal zu vertretenden Losgröße“, s. Abschnitt 3, das Auftreten von Lagerbestandsüberschreitungen nicht immer verhindert. Zur Vermeidung wird ein Tool (in Excel) eingesetzt, das die zu planenden Perioden kennzeichnet, in denen eine Überschreitung der Lagerkapazität auftritt, und zu verschiebende oder reduzierende Lose anhand der geringsten Rüstkosten vorschlägt. Dieser Fall tritt bei 2 von 39 untersuchten Artikeln auf. So beträgt beispielsweise bei dem Produkt „Heckklappe innen“ die feste Losgröße nach dem Verfahren im Unternehmen 3200 Mengeneinheiten (ME) und der maximal mögliche Lagerendbestand beträgt 3300 ME. In einer Periode liegt jedoch ein Bedarf über 3640 ME vor. Da der Lagerendbestand in der Vorperiode nur 275 ME beträgt, kann der Bedarf nicht durch ein einziges Los gedeckt werden, weswegen zusätzlich in der Vorperiode produziert werden muss. Da eine Produktion der festen Losgröße zu einer Überschreitung des maximalen Lagerendbestands führen würde, schlägt das (genannte) Tool vor, das Los soweit zu reduzieren, so dass der maximale Lagerendbestand komplett genutzt wird. Die grundsätzliche Bedienung von solchen hohen Abrufen ist dabei aber nicht problematisch, da die Artikel im Versandprozess bereits einer Fläche zur

Vorkommissionierung im Versandbereich zugeführt werden und dadurch eine solche Lagerplatzüberschreitung behoben wird. Sowohl das aktuell im Unternehmen verwendete Verfahren als auch das Silver-Meal-Verfahren berechnen so große Lose, so dass keine Fehlmengen auftreten. Fehlmengen sind bei dem Optimierungsmodell ohne beschränkter Kapazität (i.e. dem Wagner-Whitin-Verfahren) und mit beschränkter Kapazität, s. den nächsten Abschnitt, generell ausgeschlossen.

Die analysierten Ähnlichkeiten und Unterschiede zwischen den Losen der drei Verfahren belegen, dass die beschränkte Lagerkapazität und die durch Kostensatz präferierten Losgrößen simultan berücksichtigt werden müssen. Folglich dürfte sich das Vorgehen im Unternehmen nur mit hohem Aufwand, in Form von Rechenzeit und Zeit für die Entwicklung einer (besseren) Losgrößenreduktion, entscheidend verbessern lassen. Diesen Aufwand in ein simulatives Verfahren zu investieren, dürfte zu deutlich geringeren (Gesamt-) Kosten führen.

6. Verbesserung durch eine optimale Lösung

Das Wagner-Whitin-Verfahren löst das lineare Optimierungsmodell SLULSP effizient. In ein SLULSP lassen sich beschränkte Kapazitäten, wie einem maximalen Lagerbestand oder einer maximalen Losgröße, durch Restriktionen integrieren. Leider ist das resultierende Problem NP-vollständig. Dies bedeutet, dass es kein Verfahren mit einer polynomial beschränkten Laufzeit gibt. Jede Ausprägung von diesem linearen Optimierungsmodell mit beschränkter Lagerfläche wurde in IBM ILOG CPLEX Optimization Studio in der Version 12.9.0.0 implementiert. Sowohl für diese hohe Anzahl an Perioden als auch für alle Produkte im Presswerk – und nicht nur für die Testdaten – berechnet ILOG innerhalb von Minuten eine optimale Lösung. Damit ist diese Alternative im Presswerk einsetzbar und deswegen wird sie nun genauer analysiert.

Zunächst wird die Erweiterung des SLULSP um eine Beschränkung der Lagerkapazität angegeben. In [Herr09] ist dies unter der Bezeichnung Inventory Capacitated Single-Level Lot Sizing Problem (ICSLSP) erläutert und analysiert. Ebendort, i.e. in [Herr09], finden sich weitere Erweiterungen um verschiedene Ausprägungen von beschränkter Kapazität.

(a) Optimierungsmodell

Das Optimierungsproblem (OPT-I) wird nun vorgestellt. Um das Modell einfach zu halten, ist lediglich eine Version für ein Produkt k angegeben:

Parameter:

T : Länge des Planungszeitraums.

$d_{k,t}$: Nettobedarfsmenge in Periode $t \forall 1 \leq t \leq T$.

h_k : Lagerkostensatz.

M : große Zahl.

s_k : Rüstkostensatz.

y_k^{max} : maximaler Lagerbestand.

Variablen:

$q_{k,t}$: Losgröße in Periode $t \forall 1 \leq t \leq T$.

$y_{k,t}$: Lagerbestand am Ende der Periode $t \forall 0 \leq t \leq T$.

$\gamma_{k,t}$: binäre Rüstkostenvariable mit

$$\gamma_{k,t} = \begin{cases} 1, & \text{falls } q_t > 0 \\ 0, & \text{falls } q_t = 0 \end{cases} \quad \forall 1 \leq t \leq T.$$

Zielfunktion:

$$Z = \sum_{t=1}^T (s_k \cdot \gamma_{k,t} + h_k \cdot y_{k,t}).$$

Restriktionen:

$$y_{k,t-1} + q_{k,t} - d_{k,t} = y_{k,t} \quad \forall 1 \leq t \leq T \quad \text{Lagerbilanzgleichung.}$$

$$q_{k,t} - M \cdot \gamma_{k,t} \leq 0 \quad \forall 1 \leq t \leq T \quad \text{Rüstbedingung.}$$

$$y_{k,0} = 0 \text{ und } y_{k,T} = 0 \quad \text{Lageranfangs- und Endbestand.}$$

$$q_{k,t} \geq 0 \text{ und } y_{k,t} \geq 0 \quad \forall 1 \leq t \leq T \quad \text{Nichtnegativität.}$$

$$\gamma_{k,t} \in \{0,1\} \quad \forall 1 \leq t \leq T \quad \text{binäre Rüstvariable.}$$

$$y_{k,t} \leq y_k^{max} \quad \forall 1 \leq t \leq T \quad \text{maximaler Lagerbestand.}$$

Minimiere Z .

Für alle Produkte kann ein einheitlicher Wert für M gewählt werden, nämlich 100000 und T ist 50 oder 350.

Zur Anwendung von OPT-I auf ein Produkt, sind die Parameter mit konkreten Werten zu belegen. Z.B. lauten die konkreten Belegungen für das Produkt $k =$ „Tür innen vorne links und rechts“: $h_k = 0.0191$, $s_k = 199.45$, $y_k^{max} = 4600$ und die in der folgenden Tabelle 8 angegebenen Bedarfe ($d_{k,t}$).

Periode	1	2	3	4	5
Bedarf	2600	4000	3000	2800	3200
Periode	6	7	8	9	10
Bedarf	4200	3000	3000	3200	2600
Periode	11	12	13	14	15
Bedarf	4400	0	3600	4600	3600
Periode	16	17	18	19	20
Bedarf	4400	2600	3600	3800	4400
Periode	21	22	23	24	25
Bedarf	3800	3400	2600	2800	4200
Periode	26	27	28	29	30
Bedarf	3760	3640	3760	3440	3680
Periode	31	32	33	34	35
Bedarf	4440	2520	2160	0	4000

Tabelle 7: Darstellung der Periodenbedarfe ($d_{k,t}$) zum Beispiel: Tür innen vorne links und rechts. (wird fortgesetzt)

Periode	36	37	38	39	40
Bedarf	2000	4000	5000	4320	2160
Periode	41	42	43	44	45
Bedarf	2160	2160	1550	1549	3098
Periode	46	47	48	49	50
Bedarf	1549	3400	3400	3800	2800

Tabelle 8: Darstellung der Periodenbedarfe ($d_{k,t}$) zum Beispiel: Tür innen vorne links und rechts.

Im tatsächlich verwendeten Modell (OPT-I) werden alle Produkte verwendet; es enthält für jedes Produkt k_i obige Restriktionen (durch Ersetzen von k durch k_i) und die Summe aller einzelnen Zielfunktionen ist die neue Zielfunktion.

(b) Analyse der optimalen Lösung

Grund für diese Untersuchung war die Vermutung, dass das aktuelle Verfahren die beschränkten Lagerbestände einhält, aber zu lasten zu hoher Gesamtkosten geht. Dies wird durch die einzelnen Gesamtkosten zu den Losen in den Tabellen je Artikel im Anhang belegt. Bei 7 von 18 (i.e. in 38.89%) Produkten bewirkt OPT-I die gleichen Gesamtkosten wie SLULSP; auch die Rüst- und Lagerkosten sind gleich. Hierfür verantwortlich ist, dass bei diesen Produkten der Platzbedarf gegenüber der verfügbaren, produktspezifischen Fläche klein ist. Bei allen Produkten sind die Gesamtkosten geringer als durch das aktuelle Verfahren im Unternehmen. Konkret schwanken die substantiellen Reduktionen der Gesamtkosten durch OPT-I gegenüber dem aktuellen Verfahren zwischen 2.65% und 72.81%, liegen im Mittel bei 34.57% und haben eine Streuung von 18.59 – hier und im Folgenden werden die prozentualen Werte mit Bezug zu dem Verfahren angegeben, das als zweites genannt ist. Die kumulierten Kosten sind bei dem aktuellen Verfahren um 49895 Geldeinheiten (GE) höher als bei OPT-I; konkret steigen sie von 96924 GE auf 146819 GE – SLULSP hat gegenüber OPT-I um 26,96 % geringere kumulierte Kosten.

Das aktuelle Verfahren reduziert in 8 von 18 Fällen den Lagerbestand stärker als das optimale Verfahren OPT-I. Über alle 18 Fälle betrachtet, schwanken die Reduktionen des Lagerbestands durch das aktuelle Verfahren gegenüber OPT-I zwischen zwischen 5751.5 und einer Erhöhung von 44.26, liegen im Mittel bei 346.31% und haben eine Streuung von 1350.8.

OPT-I verbraucht den Lagerplatz komplett und das aktuelle Verfahren verwendet ca. 75.48% des Lagerplatzes. Durch OPT-I werden 37164266 Produkte gelagert, während durch das aktuelle Verfahren lediglich 4617853 Produkte gelagert werden; also 2546413 Produkte weniger. Diese Werte korrelieren mit den Anzahlen an Losen bei diesen beiden Verfahren.

Folgende Zahlen demonstrieren die sehr starke Einschränkung der begrenzten Lagerfläche für eine kostenminimale Losbildung. Durch SLULSP müssten 11437024 Produkte eingelagert werden, wodurch die vorhandene Fläche um 73% erhöht werden müsste.

Gegenüber OPT-I würde SLULSP 4272758 Produkte mehr lagern - also vermindert OPT-I die Anzahl an eingelagerten Produkte um 37.36 %.

(c) Rollende Planung

Wie im Abschnitt 3 bereits erläutert wurde, führt im Allgemeinen die Verwendung von SLULSP in einer rollenden Planung zu nicht optimalen Ergebnissen. In [Stadt00] gab Stadtler eine Korrektur an. Daher wurde untersucht, ob OPT-I in einer rollenden Planung nicht kostenminimale Lose liefert und wie hoch etwaige Kostenerhöhungen sind.

Wie im Abschnitt 3 erläutert wurde, ist der sogenannte Entscheidungshorizont relevant. Nach der Literatur, s. [Herr09] und [Stadt00], ist seine Länge entscheidend und diese ist sensitiv gegenüber Parameteränderungen. Daher wurde der bisherige Planungszeitraum T – für alle Produkte – verdreifacht und die Bedarfe der zusätzlichen Perioden haben den gleichen Mittelwert und die gleiche Streuung wie die Bedarfe in den bisherigen Perioden; es sei daran erinnert, dass T entweder 50 oder 350 ist. Jedes solche Planungsproblem wird sowohl komplett geplant wie auch rollend geplant, und zwar mit drei direkt aufeinanderfolgenden Planungszeiträumen über jeweils T Perioden. Identische Ergebnisse ergaben sich für alle Produkte.

50 bzw. 350 Perioden ohne Bedarfsänderungen im Zeitablauf sind für die meisten Produkte zu viel. Untersucht wurde eine rollende Planung über 25 Perioden statt 50 Perioden und der Möglichkeit vor der zweiten Planung Bedarfe ab der 26. Periode zu ändern. Dies führte zu einer Verschlechterung der Gesamtkosten um 0.94% – es sei angemerkt, dass die Verschlechterung der Gesamtkosten beim SLULSP lediglich 0.66 % beträgt.

Im Unternehmen besteht der Planungshorizont einer solchen rollenden Planung stets aus wenigstens 5 Perioden. Die Idee von dem Verfahren von Stadtler, s. [Stadt00], wurde auf OPT-I übertragen. Dadurch konnte erreicht werden, dass für jedes Produkt die Verschlechterung der Gesamtkosten gegenüber der Betrachtung von 50 (bzw. 350) Perioden stets höchstens 0.55% beträgt.

Analog zum aktuellen Vorgehen im Unternehmen steht für jedes Produkt eine feste Lagerfläche (eben die sogenannten Kanäle) zur Verfügung. Wird auf diese Zuordnung verzichtet, so darf die Summe der benötigten Lagerflächen lediglich die Gesamtfläche nicht überschreiten. Diese beiden Änderungen führen zum Optimierungsmodell OPT-II. Seine optimale Lösung hat folgende strukturelle Unterschiede gegenüber der optimalen Lösung durch OPT-I:

- Viele Produkte schöpfen die bisher zugeteilte feste Lagerfläche zugunsten von anderen Produkten nicht aus.
- Deutlich weniger Lose haben die gleichen Produktionszeitpunkte – m.a.W. die Produktionszeitpunkte sind zeitlich versetzt.

Diese flexiblere Nutzung der Lagerfläche (durch OPT-II) bewirkt eine Reduktion der Gesamtkosten um weitere 26% (gegenüber OPT-I). Diese Kostenreduktion wird

durch andere Lose erreicht. Der komplette Lagerplatz wird weiterhin genutzt. Dürfen in OPT-II die Gesamtkosten von OPT-I maximal verwendet werden, so wird nur 73% der Gesamtfläche verbraucht. Es sei betont, dass für diese Lösung wie auch für die anderen keine Fehlmengen auftreten.

7. Zusammenfassung

Durch die substantiell zunehmende Variantenvielfalt bei den PKW-Herstellern und die dadurch noch viel stärker ansteigende Anzahl an zu lagernden Produkten bei den Zulieferern der Automobilindustrie reichen bei vielen Unternehmen die vorhandene Lagerkapazität nicht aus, um kostengünstige Lose zu produzieren. Oftmals, wie bei dem hier betrachteten Presswerk eines Automobilzulieferers, erfolgt die Losreduktion ohne Berücksichtigung der ökonomischen Implikationen. Konkret liegt eine mittlere Erhöhung der Gesamtkosten um 46.09% vor. Eine optimale Lösung ist bei dem hier betrachteten Presswerk innerhalb von Minuten durch ein kommerziell verfügbares Tool, wie ILOG, lösbar. Bei anderen Unternehmen dürften ähnliche Rechenzeiten ausreichen. Beim Presswerk können für die Testdaten dadurch die Gesamtkosten im Mittel um 34.57% reduziert werden. Ist die Lagerfläche beliebig nutzbar, so beträgt die mittlere Reduktion der Gesamtkosten sogar 43.31%. Der Einsatz von diesem Verfahren im realen Betrieb seit einigen Monaten bestätigt eine Kosteneinsparung in dieser Größenordnung; wie analysiert sind die Lose aufgrund einer notwendigen rollenden Planung nur nahezu kostenminimal. Damit belegt diese Arbeit, dass durch die Nutzung von Verfahren des Operations Research substantiell bessere Ergebnisse erzielt werden und sie häufig einen so geringeren Rechenzeitbedarf haben, dass sie auch unter industriellen Randbedingungen anwendbar sind.

Literatur

[Herr09] Herrmann, Frank: Logik der Produktionslogistik Oldenbourg Verlag, Regensburg, 2009.

[Herr11] Herrmann, Frank: Operative Planung in IT-Systemen für die Produktionsplanung und -steuerung, Vieweg+Teubner Verlag, Wiesbaden, 2011.

[Herr18] Herrmann, Frank, Übungsbuch Losbildung und Fertigungssteuerung Aufgaben zur operativen Produktionsplanung und -steuerung, Springer Fachmedien, Wiesbaden, 2018.

[Hirs15] Hirschberg, Carsten; Variantenvielfalt; IPL Magazin; Ausgabe 30, Januar 2015; <https://ipl-mag.de/ipl-magazin-rubriken/scm-daten/390-variantenvielfalt>; Zugriff: 17.06.2019.

[KeLS10] Kenning, Peter; Lackes, Richard; Siepermann Markus: Electronic Data Interchange, <https://wirtschaftslexikon.gabler.de/definition/electronic-data-interchange-edi-33656>, Zugriff: 29.05.19.

[Knol87] Knolmayer, Gerhard: The performance of lot sizing heuristics in the case of sparse demand patterns,

In: A. Kusiak (Hrsg.), Modern Production Management Systems (1987), S. 265-279. New York: North-Holland.

[LaSi18] Lackes, Richard; Siepermann, Markus; Definition MRP; <https://wirtschaftslexikon.gabler.de/definition/mrp-39273/version-262686>; Zugriff: 11.07.19

[Reit19] Reiter, Christian: Interne Berechnungen, 2019.

[Stadt00] Stadler, Hartmut: Improved rolling schedules for the dynamic single-level-lot-sizing problem. In: Management Science 46 (2000), S.318-326.

[Statistika19] Anzahl der Modellreihen im deutschen Pkw-Markt in den Jahren 1995 bis 2015 aus <https://de.statista.com/statistik/daten/studie/224036/umfrage/pkw-modellreihen-in-deutschland/>, Zugriff: 23.04.2019.

[Temp08] Tempelmeier, Horst: Materiallogistik, Springer Verlag, 7. Auflage, Köln 2008.

[WaWi58] Wagner, H. M.; Whitin, T. M.: Dynamic version of the economic lot size model. In: Management Science 5, 1958, S. 89-96.

[Wemm81] Wemmerlöf, U.: The ubiquitous EOQ – its relation to discrete lot sizing heuristics. In: Journal of Operations & Production Management 1 (1981), S.161-179.

[Wemm82] Wemmerlöf, U.: A comparison of discrete single stage lot-sizing heuristics with special emphasis on rules based in the marginal cost principle. In: Engineering Costs and Production Economics 7 (1982), S.45-53.

Anhang

Die folgenden Tabellen enthalten Kennzahlen im Mittel, wegen den 15 Wiederholungen, zu den verwendeten Bedarfen und den (möglichen) Unterschieden in den Losen.

Bedarfe:

Legende	
min:	Minimaler Periodenbedarf im Mittel.
max:	Maximaler Periodenbedarf im Mittel.
μ :	Mittelwert der Periodenbedarfe im Mittel.
σ :	Standardabweichung der Periodenbedarfe im Mittel.

Komponente Innenteil Heckklappe			
Periodenbedarfe			
Min	Max	μ	Σ
0	563	67	76

Komponente Verstärkung links/rechts Kuppelprodukte			
Periodenbedarfe			
Min	max	μ	σ
0	272	84	68

Komponente Außenhaut Heck oben B			
Periodenbedarfe			
Min	max	μ	σ
0	822	515	207

Komponente Frontklappe innen/vorne/außen 3 Kuppelprodukte			
Periodenbedarfe			
Min	Max	μ	σ
0	3121	468	541

Komponente Außenhaut Heck oben/unten 2 Kuppelprodukte			
Periodenbedarfe			
min	Max	μ	σ
0	3656	366	767

Komponente Heckklappe innen			
Periodenbedarfe			
min	max	μ	σ
0	3656	366	767

Komponente Scharnier links/rechts 2 Kuppelprodukte			
Periodenbedarfe			
min	max	μ	Σ
0	3656	366	767

Komponente Verstärkung Mitte			
Periodenbedarfe			
min	max	μ	σ
0	3656	366	767

Komponente Schließblech links/rechts 2 Kuppelprodukte			
Periodenbedarfe			
min	max	μ	σ
0	3656	366	767

Endprodukt Tür A vorne/hinten links/rechts 4 Kuppelprodukte			
Periodenbedarfe			
min	max	μ	σ
0	48	34	14

Endprodukt Verkleidung/ Stütze Heck 2 Kuppelprodukte			
Periodenbedarfe			
min	max	μ	σ
0	286	132	30

Endprodukt Tür innen links/rechts 2 Kuppelprodukte			
Periodenbedarfe			
min	max	μ	σ
0	5000	3155	1057

Endprodukt Verstärkung Tür			
Periodenbedarfe			
min	max	μ	σ
0	71	20	18

Komponente Scharnier links/rechts 2 Kuppelprodukte			
Periodenbedarfe			
min	max	μ	σ
0	4080	1763	861

Endprodukt Dach B			
Periodenbedarfe			
min	max	μ	σ
0	432	192	123

Endprodukt Dach A			
Periodenbedarfe			
min	max	μ	σ
0	691	175	128

Endprodukt Stütze quer			
Periodenbedarfe			
min	max	μ	σ
540	11610	2679	2339

Komponente Schließblech Mitte			
Periodenbedarfe			
min	max	μ	σ
0	3656	366	767

Reichweite der Lose:

Legende	
WW:	Wagner-Whitin-Verfahren.
SM:	Silver-Meal-Verfahren im Mittel.
Fest:	Bisherige feste Losgrößen im Mittel.
OM:	Optimierungsmodell im Mittel.
min:	Minimale Reichweite im Mittel.
max:	Maximale Reichweite im Mittel.
μ :	Mittelwert der Reichweiten im Mittel.
σ :	Standardabweichung der Reichweiten im Mittel.

Komponente Innenteil Heckklappe					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	50	50	50	0	1
SM	18	32	25	7	2
Fest	2	17	13	6	4
OM	50	50	50	0	1

Komponente Verstärkung links/rechts 2 Kuppelprodukte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	24	26	25	1	2
SM	10	20	17	5	3
Fest	3	11	8	3	6
OM	24	26	25	1	2

Komponente Außenhaut Heck oben B					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	5	11	7	2	7
SM	3	11	6	2	8
Fest	2	6	3	1	19
OM	4	8	6	1	9

Komponente Frontklappe innen/vorne/außen 3 Kuppelprodukte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	7	29	13	6	26
SM	4	18	9	3	38
Fest	2	26	7	5	51
OM	4	28	9	6	39

Komponente Außenhaut Heck oben/unten 2 Kuppelprodukte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	7	35	15	5	23
SM	4	23	8	4	42
Fest	1	28	6	4	56
OM	7	28	12	4	30

Komponente Heckklappe innen					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	51	68	58	6	6
SM	13	42	21	8	17
Fest	1	28	9	4	41
OM	7	33	11	5	31

Komponente Scharnier links/rechts 2 Kuppelprodukte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	10	35	19	6	18
SM	5	28	10	5	36
Fest	1	28	8	4	45
OM	10	35	19	6	18

Komponente Verstärkung Mitte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	21	42	29	6	12
SM	7	28	13	6	26
Fest	1	24	7	4	49
OM	7	33	12	5	30

Komponente Schließblech links/rechts 2 Kuppelprodukte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	10	35	18	6	20
SM	5	23	9	4	37
Fest	4	33	13	6	26
OM	10	35	18	6	20

Endprodukt Tür A vorne/hinten links/rechts 4 Kuppelprodukte					
	Reichweite				
Verfahren	min	max	μ	σ	Anzahl Lose
WW	14	19	17	2	3
SM	3	18	13	6	4
Fest	15	19	17	2	3
OM	14	19	17	2	3

Endprodukt Verkleidung/ Stütze Heck 2 Kuppelprodukte					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	25	25	25	0	2
SM	19	31	25	6	2
Fest	7	11	10	2	5
OM	10	10	10	0	5

Endprodukt Tür innen links/rechts 2 Kuppelprodukte					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	2	4	3	1	18
SM	2	5	3	1	20
Fest	1	2	1	0	42
OM	1	3	1	1	36

Endprodukt Verstärkung Tür					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	50	50	50	0	1
SM	17	33	25	8	2
Fest	11	24	17	5	3
OM	25	25	25	0	2

Komponente Scharnier links/rechts 2 Kuppelprodukte					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	6	13	8	2	6
SM	6	8	7	1	7
Fest	1	5	2	1	25
OM	6	8	7	1	7

Endprodukt Dach B					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	13	22	17	4	3
SM	9	14	13	2	4
Fest	2	8	4	2	12
OM	6	12	8	2	6

Endprodukt Dach A					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	25	25	25	0	2
SM	1	18	13	7	4
Fest	1	8	4	2	12
OM	1	8	5	2	11

Endprodukt Stütze quer					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	4	9	6	2	8
SM	2	7	5	1	10
Fest	1	4	2	1	21
OM	4	9	6	2	8

Komponente Schließblech Mitte					
Verfahren	Reichweite				Anzahl Lose
	min	max	μ	σ	
WW	51	68	58	6	6
SM	14	42	21	8	17
Fest	1	32	8	5	44
OM	51	68	58	6	6

Kosten der einzelnen Losbildungen:

Legende	
WW:	Wagner-Whitin-Verfahren.
SM:	Silver-Meal-Verfahren.
Fest:	Bisherige feste Losgrößen.
OM:	Optimierungsmodell.
Rüstkosten:	Gesamte Rüstkosten.
Lagerkosten:	Gesamte Lagerkosten.
Gesamtkosten:	Summe aus Rüst- und Lagerkosten.

Komponente Innenteil Heckklappe			
Verfahren	Kosten		
	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	296	283	579
SM	433	184	617
Fest	1234	100	1334
OM	296	283	579

Komponente Verstärkung links/rechts 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	336	250	586
SM	504	173	677
Fest	1008	107	1116
OM	336	250	586

Komponente Außenhaut Heck oben B			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	2495	2697	5192
SM	2851	2364	5215
Fest	6772	1307	8079
OM	3208	2044	5252

Komponente Frontklappe innen/vorne/außen 3 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	5186	4139	9325
SM	7579	2754	10333
Fest	10172	3230	13402
OM	7779	2848	10627

Komponente Außenhaut Heck oben/unten 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	6251	4837	11087
SM	11414	1850	13264
Fest	15219	3452	18672
OM	8153	3320	11473

Komponente Heckklappe innen			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	3701	3685	7385
SM	10485	1340	11825
Fest	25287	641	25928
OM	19120	488	19608

Komponente Scharnier links/rechts 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	1769	1281	3049
SM	3537	486	4023
Fest	4422	890	5312
OM	1769	1281	3049

Komponente Verstärkung Mitte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	3261	2243	5505
SM	7066	1044	8110
Fest	13317	889	14206
OM	8153	743	8896

Komponente Schließblech links/rechts 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	1965	1302	3267
SM	3636	567	4203
Fest	2555	1804	4359
OM	1965	1302	3267

Endprodukt Tür A vorne/hinten links/rechts 4 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	886	839	1724
SM	1181	814	1995
Fest	886	1045	1931
OM	886	839	1724

Endprodukt Verkleidung/ Stütze Heck 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	1209	755	1964
SM	1209	800	2009
Fest	3022	372	3394
OM	3022	282	3304

Endprodukt Tür innen links/rechts 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	3590	2444	6034
SM	3989	2151	6140
Fest	8377	1962	10339
OM	7180	1094	8274

Endprodukt Verstärkung Tür			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	713	417	1129
SM	1425	247	1672
Fest	2138	188	2326
OM	1425	176	1601

Komponente Scharnier links/rechts 2 Kuppelprodukte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	3701	2605	6306
SM	4317	2243	6560
Fest	15419	825	16244
OM	4317	2090	6407

Endprodukt Dach B			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	1583	1119	2702
SM	2111	736	2847
Fest	6334	361	6695
OM	3167	483	3650

Endprodukt Dach A			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	1056	1303	2359
SM	2111	920	3031
Fest	6334	319	6653
OM	5806	218	6024

Endprodukt Stütze quer			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	786	613	1399
SM	983	441	1423
Fest	2063	338	2401
OM	786	613	1399

Komponente Schließblech Mitte			
	Kosten		
Verfahren	Rüst-kosten	Lager-kosten	Gesamt-kosten
WW	590	614	1204
SM	1670	221	1891
Fest	4323	105	4428
OM	590	614	1204

Lagerbestände durch die einzelnen Losbildungen:

Legende	
WW:	Wagner-Whitin-Verfahren.
SM:	Silver-Meal-Verfahren.
Fest:	Bisherige feste Losgrößen.
OM:	Optimierungsmodell.
min:	Minimaler Lagerbestand pro Periode im Mittel.
max:	Maximaler Lagerbestand pro Periode im Mittel.
μ :	Mittelwert der Lagerbestände pro Periode im Mittel.
σ :	Standardabweichung der Lagerbestände pro Periode im Mittel.
Σ :	Summe aller Lagerbestände pro Periode im Mittel.

Komponente Innenteil Heckklappe					
	Lagerbestand pro Periode				
Verfahren	min	max	μ	σ	Σ
WW	0	3248	1596	875	79776
SM	0	1958	770	468	38496
Fest	8	1130	550	349	27481
OM	0	3248	1596	875	79776

Komponente Verstärkung links/rechts 2 Kuppelprodukte					
	Lagerbestand pro Periode				
Verfahren	min	max	μ	σ	Σ
WW	0	2136	911	596	45528
SM	0	1664	630	480	31512
Fest	16	760	390	226	19512
OM	0	2136	911	596	45528

Komponente Außenhaut Heck oben B					
	Lagerbestand pro Periode				
Verfahren	min	max	μ	σ	Σ
WW	0	3520	1325	1076	66268
SM	0	3520	1162	1028	58084
Fest	1	1353	642	419	32119
OM	0	2952	1004	854	50224

Komponente Frontklappe innen/vorne/außen 3 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	7921	2124	1930	739180
SM	0	5208	1413	1344	491832
Fest	25	3240	1647	904	576715
OM	0	4328	1462	1156	508649

Komponente Schließblech links/rechts 2 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	7752	1958	1918	685343
SM	0	5080	853	1267	298580
Fest	26	5027	2713	1415	949444
OM	0	7752	1958	1918	685343

Komponente Außenhaut Heck oben/unten 2 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	5866	1607	1676	562386
SM	0	4836	615	1054	215096
Fest	1	3260	1147	698	401452
OM	0	3760	1103	1208	386026

Endprodukt Tür A vorne/hinten links/rechts 4 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	632	253	171	12632
SM	0	616	245	172	12264
Fest	5	605	315	167	15745
OM	0	632	253	171	12632

Komponente Heckklappe innen					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	21078	8773	6415	3070476
SM	0	13618	3191	3346	1116872
Fest	4	3300	1526	942	534060
OM	0	3165	1163	1149	407004

Endprodukt Verkleidung/ Stütze Heck 2 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	3120	1556	932	77805
SM	0	3835	1650	1076	82485
Fest	59	1439	766	401	38305
OM	0	1170	581	369	29055

Komponente Scharnier links/rechts 2 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	8320	2287	2226	800395
SM	0	5624	868	1375	303692
Fest	8	2897	1589	808	556135
OM	0	8320	2287	2226	800395

Endprodukt Tür innen links/rechts 2 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	7000	2559	2388	127936
SM	0	9120	2253	2500	112637
Fest	83	3600	2055	1018	102744
OM	0	4000	1145	1229	57268

Komponente Verstärkung Mitte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	11848	3770	3272	1319630
SM	0	9424	1754	2270	614041
Fest	22	3500	1495	811	523130
OM	0	3448	1249	1164	437100

Endprodukt Verstärkung Tür					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	994	518	342	25881
SM	0	675	307	223	15354
Fest	12	387	234	97	11708
OM	0	538	218	144	10906

Komponente Scharnier links/rechts 2 Kuppelprodukte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	19200	5789	4946	289440
SM	0	13840	4985	4014	249240
Fest	60	3580	1832	1034	91620
OM	0	11880	4644	3733	232200

Endprodukt Dach B					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	3564	1435	1010	71730
SM	0	2664	943	753	47160
Fest	43	869	463	245	23159
OM	0	1530	619	450	30942

Endprodukt Dach A					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	4188	1715	1223	85752
SM	0	3060	1210	791	60516
Fest	49	785	420	219	20994
OM	0	720	287	211	14352

Endprodukt Stütze quer					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	17100	6128	4894	306390
SM	0	14220	4408	4034	220420
Fest	22	8894	3376	2030	168798
OM	0	17100	6128	4894	306390

Komponente Schließblech Mitte					
Lagerbestand pro Periode					
Verfahren	min	max	μ	σ	Σ
WW	0	21078	8773	6415	3070476
SM	0	13618	3151	3298	1102754
Fest	0	3523	1499	841	524732
OM	0	21078	8773	6415	3070476

FAILURE-HANDLING STRATEGIES FOR MOBILE ROBOTS IN AUTOMATED WAREHOUSES

Thomas Lienert

Ludwig Stigler

Johannes Fottner

Chair of Materials Handling, Material Flow, Logistics

Department of Mechanical Engineering

Technical University of Munich

Boltzmannstrasse 15, 85748 Garching, Germany

Email: thomas.lienert@tum.de, ludwig.stigler@tum.de, j.fottner@tum.de

KEYWORDS

Automated Warehouses, Mobile Robots, Failure Handling, Time Window Routing Method, Discrete Event Simulation

ABSTRACT

Automated warehouses operated by a fleet of robots not only offer great flexibility, as fleet size can be adjusted easily to throughput requirements, they also provide higher redundancy compared to common solutions for automated storage and retrieval systems. In case a single robot fails, the remaining fleet of robots is able to continue working within the system, so that throughput is only slightly affected.

However, adequate strategies are required for this scenario. In this contribution, we present four different approaches to cope with robot downtimes, which are based on the routing of the robots. These strategies are compared by performing a simulation study in which a robotic mobile fulfilment system is considered.

INTRODUCTION

In addition to common stacker-crane-based automated storage and retrieval systems, a new type of automated warehouse has been deployed within the past few years that is used for part-to-picker order picking. This type basically consists of a rack system containing storage items and a fleet of vehicles moving within the storage area, fulfilling storage and retrieval requests.

There are two main categories that can be distinguished. In the first category, vehicles move on a single storage tier. The products are stored on racks arranged in storage aisles on the ground (figure 1). Whenever a certain storage item is needed, a vehicle travels towards the rack containing the item, lifts the rack and brings it to the picking zone, which is located somewhere near the storage area. The item is picked by an operator and the rack is brought back to an empty storage location. Vehicles travel along the storage aisles as well as underneath the racks, insofar as they are not carrying one of these.

In literature, these systems are denominated as robotic mobile fulfilment systems (RMFS) and the vehicles as robots (Azadeh et al. 2018).

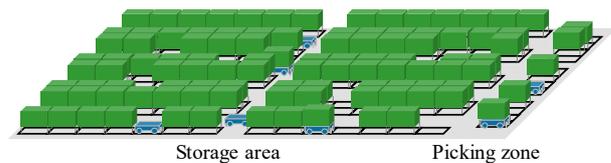


Figure 1: Robotic mobile fulfilment system

The second type of system consists of several storage tiers, connected by lifts (figure 2). In some configurations, these lifts are used only for the vertical transport of the storage items, whilst the vehicles always operate on the same storage tier. In other configurations that we consider, lifts are used to transfer vehicles between the storage tiers and the input/output (I/O) locations. On every storage tier, vehicles use a grid of storage and cross-aisles to reach the storage locations. In literature, these systems are called shuttle systems and the vehicles are denominated as shuttles (Tappia et al. 2018).

We refer to both of these types – RMFS and shuttle systems – as mobile-robot-based warehouses. One of the main characteristics is that every robot can reach every storage location. The system can therefore theoretically be operated with a single robot. Depending on the required throughput, the fleet size (i.e. the number of robots working in the system) can be adjusted. Furthermore, in the event that a single robot fails, the remaining robots are able to continue responding to storage and retrieval requests and the system's throughput is only slightly affected.

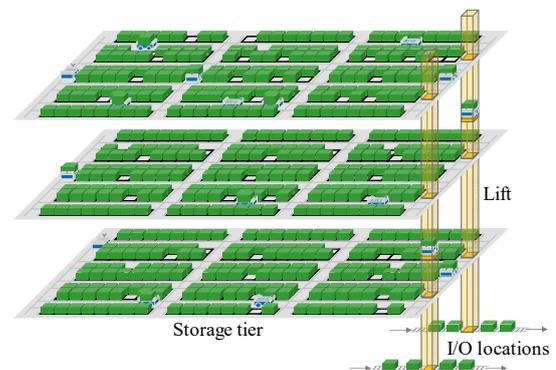


Figure 2: Shuttle system

In literature, RMFS as well as shuttle systems are widely discussed. There are several decision problems that need to be addressed during the planning phase as well as during the operation, such as layout design (Lienert et al. 2018), storage assignment (Boysen et al. 2018), order batching (Boysen et al. 2017), dispatching (Yuan and Gong 2017), traffic management (Merschformann et al. 2017), battery charging and swapping (Zou et al. 2017) as well as dwelling strategies for idle robots (Roy et al. 2016).

In general, different strategies are developed and compared to each other by using analytical models, often based on queuing networks, or by conducting simulation studies. However, robot downtimes are not considered, although these have an impact on the attainable throughput. Reasons for downtimes are manifold – for instance the interruption of the power supply, contaminations, insufficient maintenance and unstable loads.

In this work, we present different strategies for handling downtimes and compare these by performing a simulation study.

The reminder of this paper is organized as follows. Next, we briefly introduce the time window routing method, as our strategies are based on this conflict-free routing approach. Subsequently, we describe four different failure-handling strategies. We apply these to an RMFS and compare the results before we conclude our work.

TIME WINDOW ROUTING METHOD

To run mobile-robot-based warehouses robustly and efficiently, complex control strategies are needed. Amongst others, traffic needs to be managed to avoid congestion, blocking and collisions among robots. One option for coping with these challenges is routing based on time windows: Before a robot starts moving, the whole path is reserved – from its current position towards the destination. To apply this method, the layout (of each storage tier) is modelled as a graph. There is a timeline with reserved and free time windows for each node representing a layout section (figure 3).

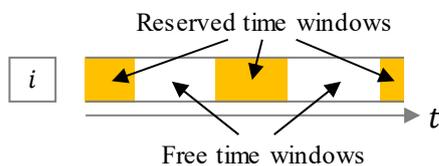


Figure 3: Reserved and free time windows on a node i

If a route has to be calculated, the procedure searches for a conflict-free route through the free time windows using an A*-algorithm. The required time windows are reserved and the robot can start executing the computed route.

The idea of this method was introduced first by Kim and Tanchocco (Kim and Tanchocco 1991) and has been applied in different contexts – for instance routing

automated guided vehicles in container terminals (Stenzel 2008), organizing aircraft taxi traffic at airports (Bussacker 2005) or in general moving agents over an infrastructure (ter Mors 2010). Furthermore, the concept has been used for organizing a fleet of robots both in RMFS (Hvězda et al. 2018) and shuttle-systems (Lienert and Fottner 2017).

The absence of deadlocks can even be guaranteed, in the case where some robots are late and do not meet their reserved time intervals. The crucial point is that the node's crossing order of the robots, based on the conflict-free schedule, is maintained (Maza and Castagna 2005).

In a previous work, we modified the method to incorporate acceleration and deceleration processes, which are usually neglected. During the planning, so-called “segments” are created, which describe a movement of a robot over several nodes in a straight line. The computed route is executed segment by segment, respecting the node's crossing order (Lienert and Fottner 2018).

Figure 4 shows a fragment of a layout graph (nodes i, \dots, m) and the corresponding timelines. For the robot r_1 there is a segment planned that comprises a movement over five nodes. Furthermore, there are some more reservations that belong to the routes of other robots (r_2, \dots, r_4). If robot r_2 is delayed, then robot r_1 is forced to decelerate and shorten the segment by introducing an intermediate stop on node l in order to meet the sequence of reservations on node m (robot r_2 before robot r_1). Note, that reservations not only comprise occupation time themselves (marked by the trajectories), but also additional buffer times in advance of the reservations, so as to prevent collisions should delays occur.

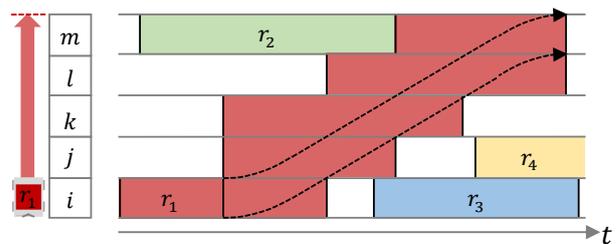


Figure 4: Segment and reservations of other robots

There are two reasons why the routing procedure might not lead to a successful result, and thus no route towards the destination being available at the given time. The first reason is an endless reservation of an idle robot placed on the destination node or on a node that belongs to the only available path towards the destination node.

The second reason is that the time window from which the routing starts is bounded by another reservation on the same node. In figure 5, routing towards the destination node k , starting at timestamp t_{start} , is unsuccessful because the current free time window on start node i cannot be left before the reservation of robot r_3 starts.

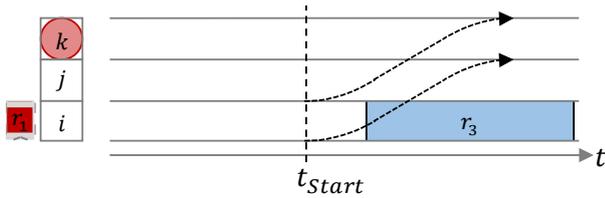


Figure 5: Routing is unsuccessful due to another reservation on the start node

The failure-handling strategies presented in the following section are based on the time window routing method.

FAILURE-HANDLING STRATEGIES

In this section, we describe four different strategies to cope with downtimes. We assume that a downtime only occurs before a robot starts with the execution of the next segment of its route. This limits the state space of the system and the strategies can be simplified – for instance, a breakdown can never occur during a loading or picking process. Furthermore, we assume that a robot resumes working after a certain time span, the mean time to repair (MTTR) has elapsed.

All presented strategies are generic in the sense they are not designed to fit a certain layout. The strategies only manipulate the routes. Orders – retrieval or storage tasks – that are already assigned to robots are not modified.

Strategy 1: Ignore

The first strategy is simple and straightforward: robot failures are strictly ignored by the control. As soon as a failure occurs, the corresponding robot stops and remains stationary on the node where it is currently located. As a consequence, this node is blocked and other robots are prevented from passing that node. These robots are forced to interrupt the execution of their routes. Furthermore, all robots that reserved a time window on one of the nodes which are part of the remaining route of the broken robot have to stop, to the extent that they are supposed to pass the node after the broken robot.

With reference to figure 4, if robot r_1 breaks down and thereby blocks its current node i , it is not only robot r_3 which is affected, but also robot r_4 . Note, that these robots might block yet others.

As soon as the MTTR has elapsed, the broken robot resumes the execution of its originally computed route, and blocked robots are triggered to continue their routes as well.

Strategy 2: Pause

The second approach is as simple as the first one, but more restrictive. As soon as a failure occurs, all robots operating within the system are forced to interrupt the execution of their routes. More specifically, if there is a breakdown present in the system, a robot will not start with the execution of the next segment of its route. As

soon as the MTTR elapsed, all robots resume with the execution of their originally computed routes. Note, that all robots will be significantly delayed.

This strategy does not take advantage of the higher redundancy mentioned above. However, it serves as a lower bound for the comparison of the throughput.

Strategy 3: Restart

The idea behind the third strategy is to stop all robots once again, but then to perform a restart where all routes are recalculated avoiding the node blocked by the broken robot.

As soon as all robots have come to a standstill, their remaining routes – and more specifically their reserved time windows – are deleted. Next, each robot reserves an endless time window on its current node. That prevents any other robot from routing over that node. All robots (besides the broken one) are added to a list of robots that need to be routed.

Next, the restart procedure as described by the flowchart in figure 6 commences.

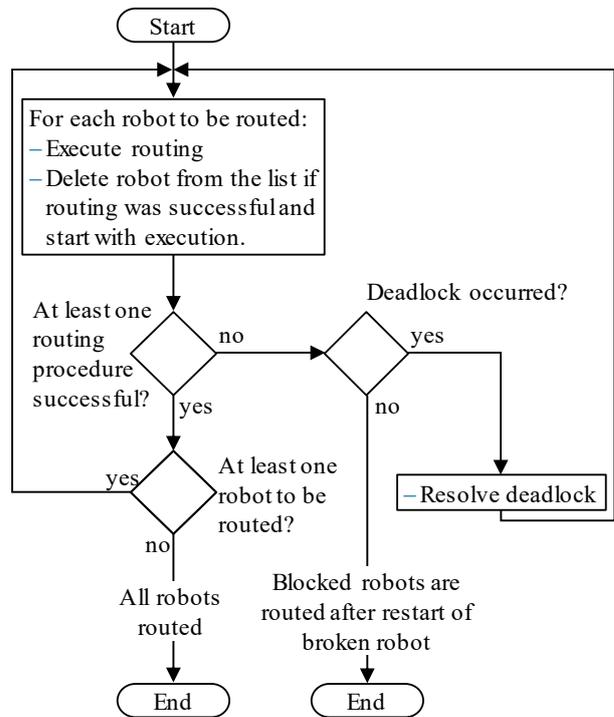


Figure 6: Procedure: restart of all robots to be routed

Routing is executed for each robot on the list. If routing is successful, the robot is removed from the list, the endless time window is deleted so the node can be accessed by other robots, and the execution of the newly constructed route starts. Otherwise, the robot will remain on the list and will be routed once again in the subsequent iteration.

If, in an iteration, at least one robot could be routed successfully, the routing will restart with the first robot remaining in the list again. The iterations continue until all robots have been routed successfully.

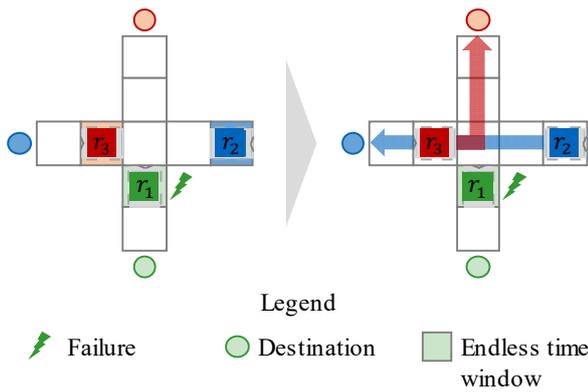


Figure 7: Robot r_2 and r_3 are restarted

In the example in figure 7, the routing of robot r_2 will fail during the first iteration and is only possible after robot r_3 has been routed.

There are two reasons why it is not possible to route a single robot during an iteration. First, a robot is prevented from reaching its destination by the endless reservation of the broken robot. In that case, the robot remains blocked until the broken robot resumes working.

As soon as the MTTR has elapsed, the flowchart in figure 8 is executed.

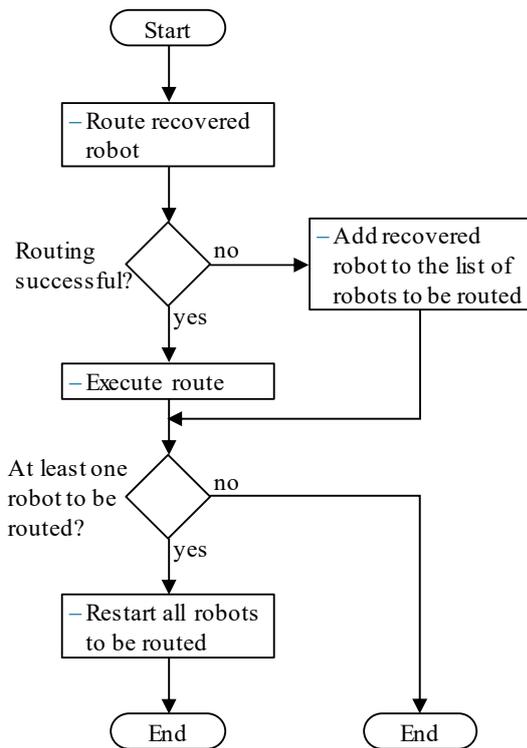


Figure 8: Procedure: restart of a single robot

After the recovered robot has been routed, blocked robots are dealt with by performing the restart procedure (flowchart in figure 6) once again.

Note that the recovered robot is added to the list of robots to be routed if routing fails.

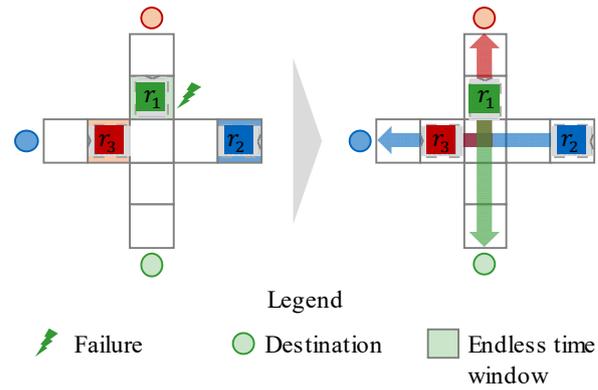


Figure 9: Robot r_2 and r_3 can only be routed after broken robot r_1 resumes working

In the example in figure 9, robot r_3 cannot be routed successfully, as the broken robot r_1 blocks the only available path. Robots r_3 as well as r_2 will be restarted after the MTTR has elapsed and a route for robot r_1 has been recalculated, such that the endless time window has been deleted. The second reason is that two or more robots block each other for endless time (Figure 10, a). In that case, an intermediate destination will be assigned to one of the robots involved in that deadlock (Figure 10, b). As soon as the robot reaches that intermediate destination (Figure 10, c) the routing towards the original destination takes place (Figure 10, d).

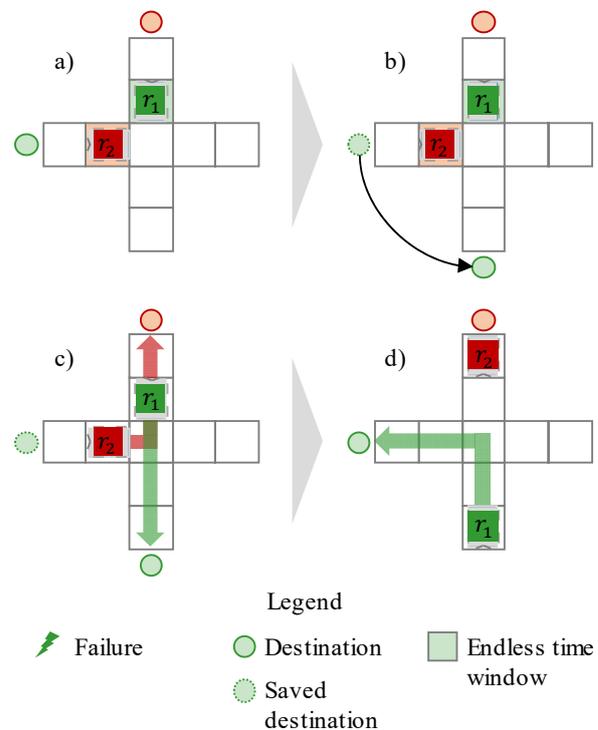


Figure 10: Deadlock resolution

The strategy *Restart* can also be used to initialize a system without failures, adding all robots located somewhere in the system to the list of robots to be routed.

Strategy 4: Reroute

The fourth strategy consists in rerouting only robots that are directly affected by the broken robot.

If a failure occurs, the execution of the computed route of the robot is stopped. As the flowchart in figure 11 describes, all robots that reserved a time window on that node are identified and added to the list of robots to be rerouted.

Referring to figure 4, if robot r_2 breaks down on node m , robot r_1 has to be rerouted. Next, the route of the broken robot is deleted as well as all reserved time windows on its current node. An endless time window on the node is reserved.

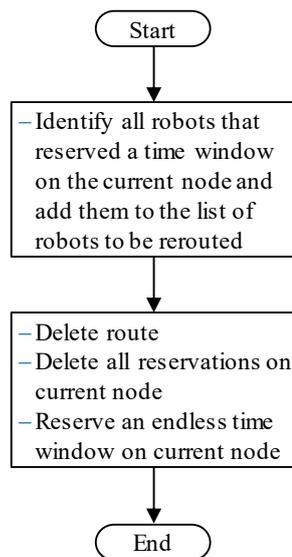


Figure 11: Procedure: end execution

Whenever a robot is ready to execute the next segment of its route, it will be checked whether that robot must be rerouted. In that case, the remaining route of the robot is deleted and the routing is again executed. If routing is successful, an alternative route without the blocked node is found and the execution of the newly calculated route starts.

If there is no route available due to an endless reservation, the execution is ended as described previously. The robot is treated as if it were broken, and other robots which routed using the robot's current node are affected, and need to be rerouted. As soon as the MTTR has elapsed, the recovered robot is restarted as in the previously described strategy (flowchart in figure 8), which leads to a restart of all blocked robots to be routed (flowchart in figure 6).

If another time window on the same node impedes a successful routing as shown in figure 5, the execution is once again ended, though this is immediately followed by a restart (flowchart in figure 8). Note that ending execution deletes the limiting time window.

The flowchart in figure 12 summarizes the described procedure.

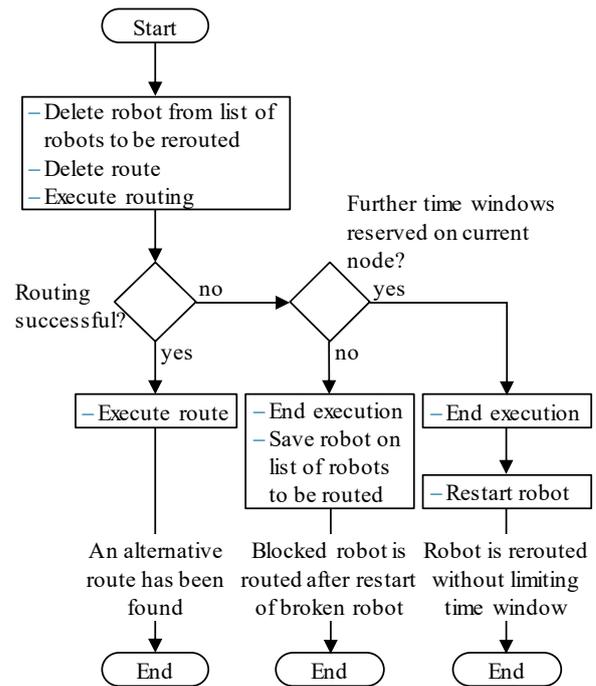


Figure 12: Procedure: reroute

SIMULATION STUDY

In this section, we compare the previously described strategies by performing a simulation study, considering an RMFS.

Considered System

We apply the strategies to a fleet of robots moving within an RMFS with 336 storage locations, which are arranged in seven double rows divided by storage aisles. There are two cross-aisles located at one third and at two thirds of the aisle length. All aisles can be used for bi-directional traffic.

There are four picking areas with five picking places each arranged before the storage system. In front of these places, there are two uni-directional cross-aisles (see figure 13). A replenishment area, where empty racks are refilled, is located on the opposite side of the storage area.

The robots are dedicated to a picking zone and perform three different cycles to maintain the material flow between storage locations, picking area and replenishment area. For a more detailed description of the system, we refer to (Lienert et al. 2018).

We implemented the RMFS using the discrete event simulation environment *Tecnomatix Plant Simulation*. Figure 13 shows a screenshot of the simulation model, comprising the picking area and part of the storage aisles. Note that loaded robots must use aisles and cross-aisles whereas unloaded robots are also free to use storage locations for navigation towards their destination. Each rectangle represents a node in the layout graph that is used for the time window routing.

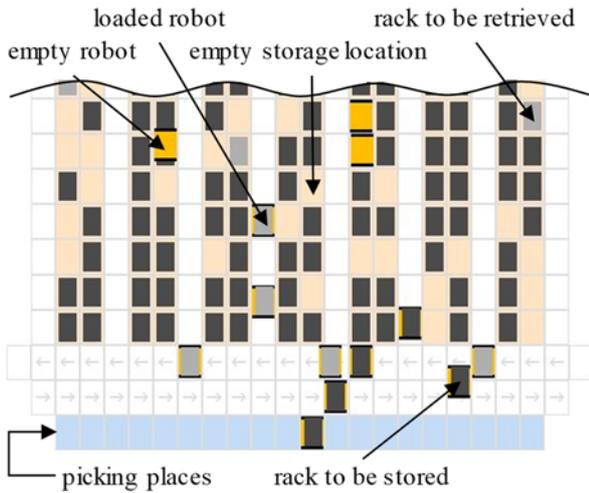


Figure 13: Screenshot of the simulation model

Parameter Settings

In a first experiment, availability is set to 100 percent – so no downtimes occur – so as to determine maximum system throughput. Subsequently, availability is set to 99 percent and MTTR to three minutes.

Failures are generated based on these parameters by the software using an *Erlang* distribution for the duration of a failure and an exponential distribution for the time between them. As failures are generated for each robot independently, several robots can be affected by a breakdown at the same time.

We vary the number of robots, starting with four robots (one for each picking zone), and going up to 60 robots working in the system in steps of four, and repeat the experiments for each strategy. All the remaining parameters, such as robot’s acceleration and maximum speed, remain the same. Simulation time is set to 72 hours. No warm-up time is taken into account. We conduct five replications for each parameter setting.

Results

The chart in figure 14 shows the throughput measured in cycles per hour. Regarding the idealized system, performance scales (in a quasi-linear manner) with the number of robots until saturation is reached. The curves for the idealized systems *Reroute* and *Restart* show a small “knee” between 16 and 20 robots. As each robot is assigned to a certain picking zone, with 20 robots, the number of robots per picking zone equals the number of picking places, and a different strategy for the supply of the picking zone is used (see Lienert et al. 2018).

The more robots are operating in the system, the clearer the differences between the strategies. As expected, *Pause* shows the worst performance. Throughput peaks at just 20 robots, and then starts decreasing. With an increasing number of robots, the probability of a breakdown grows and the whole fleet is prevented from working. *Ignore* shows a similar behaviour, although throughput is slightly higher. The more robots operating in the system, the more robots affected by a breakdown.

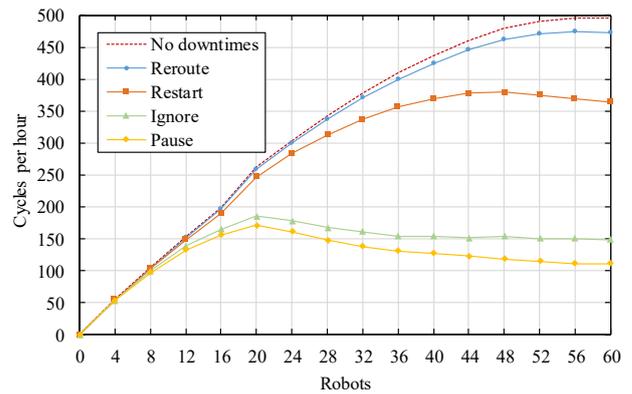


Figure 14: Throughput reached by the strategies

In the worst-case scenario, all robots have to interrupt the execution of their routes in order to maintain the node’s crossing order.

Restart reaches a significantly higher throughput than *Ignore*, which peaks with 48 robots before throughput decreases slightly. However, the highest throughput is reached by *Reroute*. The curve is similar to the one for the idealized system, but saturation is reached at a lower number of cycles.

In a second experiment, we vary the availability of the robots from 97.5 percent in steps of 0.5 percent, and repeat the experiments applying only *Reroute*. The chart in figure 15 shows the corresponding throughput.

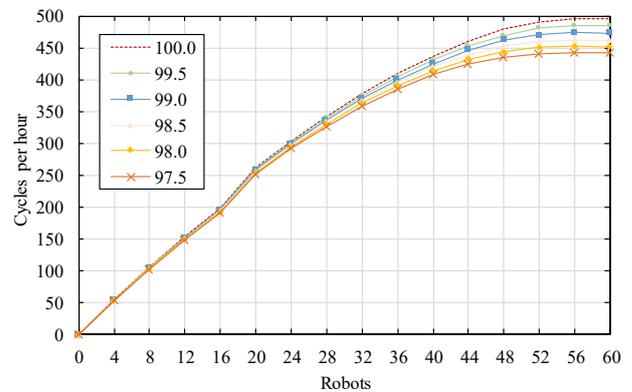


Figure 15: Throughput for different availabilities

As can be observed, all curves are similar and differ mainly in the maximum throughput that can be reached. The chart in figure 16 provides a closer view of the loss of throughput compared to the idealized system without downtimes. In the case of high availability (99.5 percent), the loss of throughput remains relatively small and closed to the lower bound of 0.5 percent. However, with an increasing number of robots, the loss of throughput generated by the failures also increases. With 60 robots, the loss of throughput is above 2 percent.

The more evident this behaviour is, the lower the availability. With an availability of 97.5 percent, loss of throughput reaches as high as 11 percent.

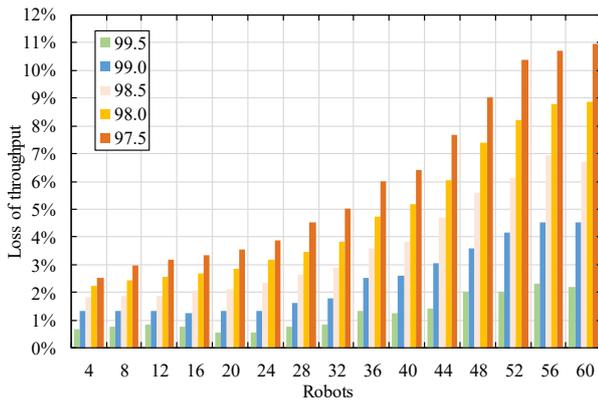


Figure 16: Loss of throughput for different availabilities

CONCLUSION

In this contribution, we considered mobile-robot-based warehouses. We presented four different strategies for dealing with failures that are based on the time window routing method. We applied these strategies to an RMFS and conducted a simulation study to compare the performance. Rerouting robots showed the best results and should be used for further investigations. Furthermore, we showed how throughput is affected for different levels of availability.

By way of next research, we suggest introducing safety corridors for accessing broken robots. In this respect, two questions have to be answered. First: Which path should be taken by an operator to access the broken robot? And second: How can the corridor be established in order to guarantee that no robot will enter that corridor while an operator is accessing the broken robot?

One option is routing an operator through the system applying the time window routing as well. However, all reservations must be endless to ensure no robot enters the corridor as long as the operator moves within the system.

REFERENCES

Azadeh, K.; de Koster, R. and Roy, D., 2018, "Robotized and Automated Warehouse Systems: Review and Recent Developments." Available at SSRN.

Boysen, N.; Briskorn, D. and Emde, S., 2017, "Parts-to-picker based order processing in a rack-moving mobile robots environment." *European Journal of Operational Research* 262, No.2, 550-562.

Boysen, N.; de Koster, R. and Weidinger, F., 2018, "Path planning for robotic mobile fulfillment systems.", "Warehousing in the e-commerce era: A survey." *European Journal of Operational Research*, in press.

Busacker, T. 2005. *Steigerung der Flughafen-Kapazität durch Modellierung und Optimierung von Flughafen-Boden-Rollverkehr – Ein Beitrag zu einem künftigen Rollführungssystem*. Dissertation. Technische Universität Berlin.

Havězda, J.; Rybecký, T.; Kulich, M. and Přeučil, L., 2018, "Context-Aware Route Planning for Automated Warehouses." In *Proceedings of the 21st International Conference on Intelligent Transportation Systems*, 2955-2960.

Kim C. W. and Tanchoco J. M. A., 1991, "Conflict-free shortest-time bi-directional AGV routing." *International Journal of Production Research* 29, No.12, 2377-2391.

Lienert, T. and Fottner, J., 2017, "No more deadlocks – applying the time window routing method to shuttle systems." In *Proceedings of the 31st European Conference on Modelling and Simulation*, 169-175.

Lienert, T.; Wenzler, F. and Fottner, J., 2018, "Robust integration of acceleration and deceleration processes into the time window routing method." In *Proceedings of the 9th International Scientific Symposium on Logistics*, 66-86.

Lienert, T.; Staab, T.; Ludwig, C. and Fottner, J., 2018, "Simulation-based Performance Analysis in Robotic Mobile Fulfillment Systems." In *Proceedings of the 8th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, 383-390.

Maza, S. and Castagna, P., 2005, "A performance-based structural policy for conflict-free routing of bi-directional automated guided vehicles." *Computers in Industry* 56, No.7, 719-733.

Mershformann, M.; Xie, L.; Erdmann, D., 2018, "Path planning for robotic mobile fulfillment systems", arXiv preprint arXiv:1706.09347

Roy, D.; Krishnamurthy, A.; Heragu, S. and Malmborg, C., 2015, "Queuing models to analyze dwell-point and cross-aisle location in autonomous vehicle-based warehouse systems", *European Journal of Operational Research* 242, No. 1, 72-87.

Stenzel, B. 2008. *Online Disjoint Vehicle Routing with Application to AGV Routing*. Dissertation. Technische Universität Berlin.

Tappia, E.; Roy, D. de Koster, R. and Melacini, M., 2018, "Modeling, Analysis, and Design Insights for Shuttle-Based Compact Storage Systems", *Transportation Science* 51, No.1

Ter Mors, A. W. 2010. *The world according to MARP*. Dissertation. Delft University of Technology

Yuan, Z.; Gong, Y.Y., 2017, "Bot-In-Time Delivery for Robotic Mobile Fulfillment Systems." *IEEE Transactions on Engineering and Management* 64, No.1, 83-93

Zou, B.; Xu, X.; Gong, Y.Y., de Koster, R., 2017, "Evaluating battery charging and swapping strategies in a robotic mobile fulfillment system." *European Journal of Operational Research* 267, No. 2, 733-753.

THOMAS LIENERT has been working as a research assistant at the Chair of Materials Handling, Material Flow and Logistics, Technical University of Munich, since 2014. His research deals with the simulation of mobile-robot-based warehouses. His email address is: thomas.lienert@tum.de.

LUDWIG STIGLER graduated from the Technical University of Munich. His master's thesis deals with failure-handling strategies for mobile robots. His email address is: ludwig.stigler@tum.de.

JOHANNES FOTTNER is professor and head of the Chair of Materials Handling, Material flow, Logistics at the Technical University of Munich. His email address is: j.fottner@tum.de.

VOLLAUTOMATION EINES LABORS UNTER ANWENDUNG VON INDUSTRIE 4.0

Daniel Marczoch
CSL Plasma GmbH
Business Technology
35041 Marburg
daniel.marczoch@cslbehring.com

Prof. Dr. Harald Ritz
Technische Hochschule Mittelhessen
Fachbereich MNI
Wiesenstr. 14, 35390 Gießen
harald.ritz@mni.thm.de

ABSTRACT

Der vorliegende Artikel wurde in Zusammenarbeit mit der globalen Business-Technology-Organisation und dem Labor in Göttingen der CSL Plasma GmbH in Marburg erarbeitet und behandelt die Konzeption eines vollautomatisierten serologischen Labors. Das Unternehmen führt derzeit einen Umbau des Labors und eine damit verbundene Kapazitätserweiterung durch. Im Zuge dessen werden sämtliche Prozesse untersucht und eine mögliche Automation bisher manuell durchgeführter Tätigkeiten überprüft.

In einem pharmazeutischen Betrieb, der sich auf Medikamente bestehend aus menschlichem Blutplasma spezialisiert hat, existieren besondere Anforderungen im Bereich der Qualitätskontrolle. Der Rohstoff Blutplasma wird von freiwilligen Spendern gewonnen und muss vor der Verarbeitung in der Produktion auf Viren und Krankheiten getestet werden.

Eine Automation kann diesbezüglich viele Vorteile bieten, wie z. B. verbesserte Wettbewerbsfähigkeit durch Kosten- und Zeiteinsparungen, weniger direkten Kontakt zu den Proben und dadurch niedrigere Ansteckungsgefahr für das Laborpersonal und eine zuverlässigere Qualität der Messungen.

SCHLÜSSELWÖRTER

Cyber-physische Systeme, Digitalisierung, digitale Transformation, Industrie 4.0, Internet of Things, LIMS, Schnittstellen, Software.

CSL PLASMA

Die Behringwerke wurden im Jahr 1904 durch Emil von Behring (erster Nobelpreisträger für Physiologie und Medizin) in Marburg gegründet, um Impfstoffe und Seren zur Bekämpfung von Infektionskrankheiten herzustellen. CSL Plasma agiert als Schwestergesellschaft und gleichzeitig als Zulieferer der aus den Behringwerken hervorgegangenen CSL Behring. Die Muttergesellschaft der CSL Plasma, CSL Limited, wurde 1916 für die Versorgung mit biologischen Arzneimitteln in Australien gegründet ([7] CSL Behring 2013).

Der Hauptsitz von CSL Plasma befindet sich in Boca Raton/Fl in den USA. Die Aufgabe CSL Plasmas ist die Beschaffung, Testung und Lagerung von menschlichem Blutplasma für CSL Behring zur Herstellung von Biotherapeutika. Die Beschaffung erfolgt durch freiwillige Spender, die in eigenen Plasmacentern in den USA, Deutschland und Ungarn ihr Blutplasma spenden können, oder durch den Kauf von Blutplasma von Drittlieferanten. In Europa betreibt CSL Plasma elf Plasmacenter, ein Labor in Göttingen und ein Plasmalogistikzentrum in Schwalmstadt. Der Sitz der Verwaltung von CSL Plasma Europa befindet sich in Cölbe. Die gesamte CSL-Organisation beschäftigt weltweit über 22.000 Mitarbeiter und hat im Geschäftsjahr 2017/18 einen Umsatz von

ca. 7,9 Mrd. US \$ erwirtschaftet ([6] CSL Behring 2018; [8] CSL Limited 2018).

Heute ist CSL ein weltweit führendes Unternehmen für die Herstellung von Gerinnungsfaktoren, Produkten für die Intensivmedizin und Immunglobulinen für seltene und schwere Krankheiten. Mit CSL Plasma betreibt CSL eines der größten Netzwerke von Plasmaspendezentren weltweit ([9] CSL Plasma 2019).

Der IT-Bereich des Unternehmens, in welchem dieser Artikel erarbeitet wird, wird Business Technology (BT) genannt und ist eine Teileinheit der Global Business Technology der CSL Limited. Es handelt sich hierbei um eine globale Organisation, die von vernetzten Teams aus dem gesamten Konzern mit dem Ziel betrieben wird, IT-Lösungen und -Services für die einzelnen Geschäftsbereiche bereitzustellen, um dem Unternehmen einen effizienten Geschäftsbetrieb zu ermöglichen.

AUSGANGSSITUATION

In einem pharmazeutischen Betrieb, der sich auf Medikamente bestehend aus menschlichem Blutplasma spezialisiert hat, existieren besondere Anforderungen im Bereich der Qualitätskontrolle. Der Rohstoff Blutplasma wird von freiwilligen Spendern gewonnen und muss vor der Verarbeitung in der Produktion auf Viren und Krankheiten in einem serologischen Labor getestet werden.

Eine Testung ist essentiell und kann nicht ausgelassen werden. Allerdings ist diese zeitaufwendig und birgt im

Falle einer infizierten Spende das Risiko einer Ansteckung der Arbeitskräfte im Labor. Für diesen Fall ist eine vollautomatisierte Testung menschlichen Blutplasmas in der Qualitätskontrolle eines pharmazeutischen Betriebes sinnvoll, da hierdurch Ansteckungsgefahr und zeitlicher Aufwand minimiert werden können.

Ziel dieses Artikels ist es, ein Konzept für ein vollautomatisiertes Labor unter Einbezug von Industrie 4.0 aufzuzeigen, das den Anforderungen an verschiedenste Regularien gerecht wird. In diesem Zusammenhang stellen sich folgende Fragen:

- Welche Technologien werden für die Umsetzung eines solchen Konzeptes benötigt?
- Wie wird die Automation softwareseitig unterstützt?
- Wie sind die Verantwortlichkeiten in einem vollautomatisierten Labor verteilt?

Daraus leitet sich als zentrale Fragestellung ab, wie ein vollautomatisiertes Labor unter Verwendung von Industrie 4.0 konzipiert und realisiert werden kann.

Hierzu wird zunächst die Art des Labors definiert, für das ein Prozess nach Industrie 4.0 konzipiert werden soll. Hierfür wird ein bisher regulärer Laborprozess analysiert und eine Systemarchitektur in einem Labor nach aktuellem Standard vorgestellt.

Anschließend wird ein vollautomatisierter Laborprozess unter Einbezug von Industrie 4.0 mit BPMN 2.0 modelliert. Der Prozess beginnt mit der Anlieferung von Proben des menschlichen Blutplasmas und endet mit der Entsorgung dieser Proben.

Weiterhin werden die Verantwortlichkeiten der verschiedenen Beteiligten und Technologien während des Prozesses erläutert. Anknüpfend hieran wird der modellierte Prozess mit dem bisher etablierten Labor verglichen und die Vor- und Nachteile erörtert. Abschließend folgt in diesem Artikel ein Beispiel aus der Praxis, welches sich dem vorgestellten Konzept annähert.

HINTERGRUNDINFORMATIONEN

Laborautomation

Im folgenden Abschnitt soll der Begriff „Laborautomatisierung“ näher erläutert werden. Um ihn zu beschreiben, muss zuvor definiert werden, was Automatisierung bedeutet. In der DIN IEC 60050-351 wird der Begriff „automatisch“ aus der technischen Sicht als „ein[...] Prozess oder eine Einrichtung bezeichne[t], der oder die unter festgelegten Bedingungen ohne menschliches Eingreifen abläuft oder arbeitet“ ([10] DIN 2014). Aus betriebswirtschaftlicher Sicht ist Automatisierung die „[...] Substitution des Faktors Arbeit durch den Faktor Betriebsmittel“ ([15] Gabler 2015).

Automatisierung entsteht durch den Transfer von Aufgaben bzw. Tätigkeiten eines Menschen auf Maschinen (künstliche Systeme). Es wird nicht nur die Ausführung, sondern auch die Steuerung und Koordination eines Prozesses auf eine oder mehrere Maschinen übertragen. Das Verhältnis von automatisierten Prozessen zum Gesamtprozess wird durch den Automatisierungsgrad beschrieben. Je nach Automatisierungsgrad wird zwischen Teil- und Vollautomatisierung unterschieden. Das ursprüngliche Ziel von Automatisierung ist die Steigerung der Produktivität ([15] Gabler 2015).

Laborautomatisierung beschreibt die Automatisierung von Prozessen und Aufgaben in einer Laborumgebung. Die meisten Arbeitsschritte wie Messungen, Ansteuerungen von Aktoren und Datenverarbeitung können durch computergestützte Systeme automatisiert werden ([1] Analytik News 2019).

Durch steigende Komplexität der Prozesse in einem Labor und durch die wachsenden Anforderungen durch Behörden ist das Automatisieren von Laborgeräten zwingend notwendig. Ohne eine Automatisierung, oder zumindest eine Teilautomatisierung, können die genannten Herausforderungen nicht effizient bewältigt werden. Die Laborautomatisierung ermöglicht eine wirtschaftlich effiziente Kontrolle dieser Probleme ([13] Fraunhofer IPA 2012).

Laborinformationssysteme

Ein LIS (Laborinformationssystem), oder auch LIMS (Laborinformations- und Management-System), ist eine IT-Lösung, die das Probenmanagement und Verarbeiten der Analysedaten in Laboren unterstützt. Da Labore in der Regel sehr individuell aufgebaut sind, ist es essenziell für ein LIMS hochgradig flexibel und somit für jedes Labor geeignet zu sein. Typische Charakteristika eines LIMS sind Standardfunktionalitäten wie das Anlegen und Registrieren von Proben, das Zuweisen der durchzuführenden Analysen und die Erfassung und Anzeige der Testdaten in Form eines Reports. Darüber hinaus bieten die Hersteller eines LIMS typischerweise die Möglichkeit der zusätzlichen Konfiguration. Diese ist notwendig, um die unterschiedlichen Anforderungen der verschiedenen Kunden abzudecken. Aus diesem Grund wird ein LIMS im Laufe seines Lebenszyklus im Labor häufig von einer standardisierten Lösung zu einer hoch spezialisierten Software, die nur individuell für ein Labor anwendbar ist ([18] LIMS 2018).

Strukturell ist ein LIMS in ein Serversystem und einen oder mehrere Arbeitsplatzrechner unterteilt, die über die Netzwerkinfrastruktur miteinander verbunden sind. Auf dem Serversystem befinden sich die LIMS-Datenbank und auch die Applikation selbst. Oftmals werden Datenbank und Applikation in der Praxis auf mehrere Server verteilt. Die Arbeitsplatzrechner beherbergen den Client, worüber die Anwender auf die Applikation auf dem Server zugreifen können ([18] LIMS 2018).

Ein LIMS bietet auch die Möglichkeit an andere Komponenten und Systeme angebunden zu werden. Beispielsweise können im Prozess übergeordnete Systeme mit dem LIMS vernetzt oder auch Analysegeräte zur Übertragung von Testergebnissen an ein LIMS angebunden werden ([18] LIMS 2018).

Industrie 4.0

Der Begriff „Industrie 4.0“ wurde durch den von der Bundesregierung beauftragten Arbeitskreis Industrie 4.0 definiert. Das Ergebnis dieses Arbeitskreises ist die Beschreibung der Auswirkungen des „Internet der Dinge“ (dt. für „Internet of Things (IoT)“) auf die Produktion in deutschen Unternehmen ([20] Reinheimer 2017). Die Namensgebung Industrie 4.0 bezieht sich auf die vierte industrielle Revolution ([3] Andelfinger u.a. 2017).

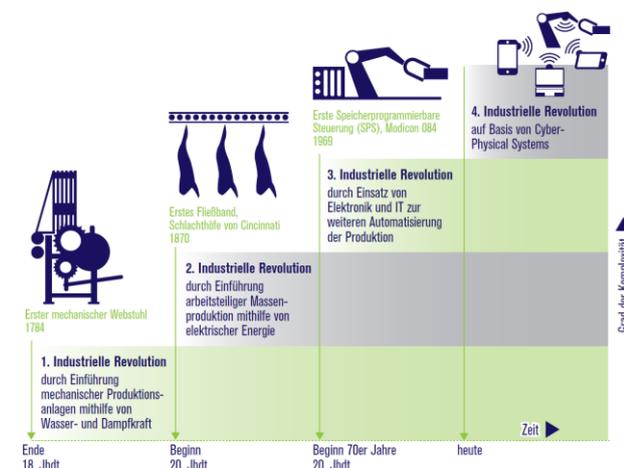


Abbildung 1: Die vier Stufen der industriellen Revolution ([19] Promotorengruppe 2013)

Um zunächst Industrie 4.0 in die industrielle Entwicklung einzuordnen, werden die vorhergehenden industriellen Revolutionen kurz erläutert. In Abbildung 1 wird die industrielle Entwicklung veranschaulicht. Die erste industrielle Revolution wurde mit der Erfindung der Dampfmaschine und dem maschinell betriebenen Webstuhl Ende des 18. Jahrhunderts eingeleitet. Die Wertschöpfung wurde hierdurch gesteigert, allerdings verschlechterten sich die Lebens- und Arbeitsbedingungen der Menschen durch die neue Art von Arbeit beträchtlich. Die zweite industrielle Revolution folgte knapp 100 Jahre später mit der Einführung der Fließbandarbeit nach Henry Ford. Der Mensch übernahm nicht mehr die gesamte Fertigung eines Produktes, sondern war nun für einzelne Teilfertigungsschritte verantwortlich. Die einzelnen Aufgaben waren somit weniger komplex, neue Arbeiter hatten kürzere Anlernzeiten und benötigten keine spezielle Berufsausbildung. Auch die zweite industrielle Revolution führte zu einer Wertschöpfungssteigerung. Die dritte industrielle Revolution begann in den siebziger Jahren mit der Verbreitung des Computers. Der Fokus stand nun bei Automatisierung und Optimierung von Prozessen ([3] Andelfinger u.a. 2017).

Während die dritte industrielle Revolution noch nicht vollständig abgeschlossen ist, bahnt sich mit der Digitalisierung als Grundlage bereits die vierte industrielle Revolution an. Diese ist geprägt durch cyber-physische Systeme, IoT und Kommunikation von Maschinen untereinander ([3] Andelfinger u.a. 2017).

Industrie 4.0 sieht die Vernetzung von räumlich unabhängigen Produktionsressourcen vor. Diese sind beispielsweise Maschinen, Roboter, Betriebsmittel, Fördersystem und Lagersysteme. Die Produktionsressourcen steuern und konfigurieren sich situationsabhängig selbst. Gestützt wird die autonome Organisation durch Sensoren und übergreifende Wissensbereitstellung ([14] Fraunhofer IESE 2019).

Um die Idee von Industrie 4.0 zu verstehen und ein Konzept für dessen Umsetzung zur Automatisierung eines Labors zu entwickeln, sind Einblicke in verschiedene Themengebiete notwendig. In den folgenden Abschnitten werden diese Themen näher erläutert.

Grundlagen zur Umsetzung von Industrie 4.0

Die Grundlagen für eine Umsetzung von Industrie 4.0 bilden die Konzepte des IoT und der cyber-physischen Systeme ([12] FM 2014).

Das Konzept von IoT sieht eine Vernetzung jeglicher Gegenstände miteinander vor. Maschinen, Fahrzeuge, Haarselekttronik, PCs, jegliche Sensoren, Videokameras oder sogar Verpackungen, sind vernetzt und kommunizieren miteinander. Die Vernetzung wird durch die permanent voranschreitende Miniaturisierung von Computerbauteilen ermöglicht, wie es z. B. bei Smartphones zu beobachten ist. Das Ausstatten nahezu aller Gegenstände mit WLAN-Empfängern oder Radio Frequency Identification (RFID)-Chips ermöglicht es, dass diese Gegenstände kabellos und somit ungebunden miteinander kommunizieren ([2] Andelfinger u.a. 2015). Ein RFID-Chip kann Informationen speichern und in Form von Radiowellen übertragen, welche von einem RFID-Scanner ausgelesen werden ([24] Welt 2012). RFID-Chips sind bereits in viele Karten, Ausweise und Etiketten integriert und ermöglichen eine kontaktlose Übermittlung von Daten. Der Chip kommt ohne Batterien oder Akkus aus. Die benötigte Energie kann aus den erzeugten Radiowellen des RFID-Scanners erzeugt und in einem Kondensator abgespeichert werden ([25] Wirtschaftslexikon24.com 2018).

RFID-Chips, Sensoren und Aktoren spielen eine entscheidende Rolle im IoT. Die Daten (verwendet wird auch der Begriff „Big Data“, der später erläutert wird), welche die jeweiligen Komponenten erheben, werden allerdings nicht nur lokal für den einzelnen Gegenstand genutzt, sondern global zur Verfügung gestellt. Das kann beispielsweise über eine Cloud geschehen. Die gespeicherten Daten können dadurch standortunabhängig und von jeder Person verarbeitet werden ([3] Andelfinger u.a. 2017). Der Nutzen dieser Vernetzung wird von Carolyn

Gramling in einem Artikel über frühzeitige Erdbebenwarnungen mit Hilfe von Smartphones beschrieben. Hierbei erklärt sie, wie die Bewegungssensoren von Smartphones Auskunft darüber geben, ob sich ein Erdbeben in einem Gebiet anbahnt. Ändern sich beispielsweise Bewegungsdaten mehrerer Smartphones in einem bestimmten Bereich gleichzeitig, können durch die öffentliche Bereitstellung dieser Daten Frühwarnsysteme versorgt werden. Mögliche Gefahren werden dadurch eventuell früher erkannt, als nur durch einen feststehenden Sensor eines Instituts ([16] Gramling 2016).

Cyber-physische Systeme (CPS) sind, ebenso wie das IoT, essentiell für eine Umsetzung von Industrie 4.0. CPS bauen auf dem Konzept des IoT auf und werden durch die Verbindung verschiedener Komponenten (z. B. Maschinen, Werkzeuge, etc.) charakterisiert. Sowohl physische, als auch biologische oder bautechnische Komponenten sind informationstechnologisch miteinander verbunden. Die Komponenten werden über eine Recheneinheit eingebunden, überwacht und gesteuert. CPS werden insbesondere dadurch ausgezeichnet, dass verschiedene IT-Komponenten mit mechanischen oder auch elektronischen Elementen untereinander kommunizieren. Die Kommunikation bzw. der Datenaustausch findet hierbei in Echtzeit statt, beispielsweise über das Internet. Sogenannte eingebettete Systeme analysieren über Sensoren ihre Umwelt und können diese über Aktoren beeinflussen. Rechen- und Speicherkapazitäten oder sogar Funktionalitäten werden den CPS in einer Cloud über Internet/Intranet zur Verfügung gestellt. Die Schnittstelle zum Menschen befindet sich an den Komponenten selbst oder wird über eine GUI realisiert ([17] Janiesch 2019).

Eine Cloud wird für sogenanntes Cloud-Computing verwendet. Cloud-Computing beschreibt das zur Verfügung stellen von elektronischen Diensten im Web, wodurch diese Dienste standortunabhängig erreichbar sind. Dienste können das Bereitstellen von IT-Infrastruktur, Plattformen oder Anwendungen sein. Charakterisierend für eine Cloud ist, dass Ressourcen virtualisiert und die Dienste dynamisch anpassbar sind. Das hat zur Folge, dass Ressourcen optimal genutzt werden, da ohne weitere Aufwendungen auf nicht genutzte Kapazitäten zurückgegriffen werden kann. Eine Cloud kann eigenständig betrieben oder von einem Cloud-Anbieter (z. B. Google oder Amazon) zur Verfügung gestellt werden ([4] Baun u.a. 2011).

Aus dem IoT, CPS und Cloud-Computing ergibt sich ein weiterer wichtiger Begriff, der bei Industrie 4.0 eine zentrale Rolle spielt. Es handelt sich um den Begriff „Big Data“. Unter Big Data werden Daten bezeichnet, die aufgrund von Umfang und Diversität nicht mit konventionellen Methoden verarbeitet und analysiert werden können ([11] Fasel u.a. 2016). Big Data kann umfassende Informationen geben und ist bei Industrie 4.0 im Bereich Wartung und Instandhaltung essentiell. In diesem Zusammenhang ist das Themengebiet „Predictive Maintenance“ ebenfalls zu erwähnen.

Hierunter wird das Voraussagen von Instandhaltungsaktionen aufgrund von erhobenen Big Data und vorliegenden Berechnungen verstanden. Das bedeutet, dass anhand von umfangreich zur Verfügung stehenden Daten aus der Vergangenheit und Gegenwart erwartete Werte in der Zukunft präzise berechnet werden können. Somit werden beispielsweise Verschleißteile von Maschinen gewechselt, bevor es zu einem Stillstand dieser Maschine kommt ([11] Fasel u.a. 2016).

Anforderungen an die Systemarchitektur

Um eine Automatisierung nach Industrie 4.0 umsetzen zu können, müssen Unternehmen von Grund auf die Architektur ihrer Systeme überdenken. Systemarchitekturen orientieren sich in der heutigen Zeit größtenteils an der Automatisierungspyramide (siehe Abbildung 2).

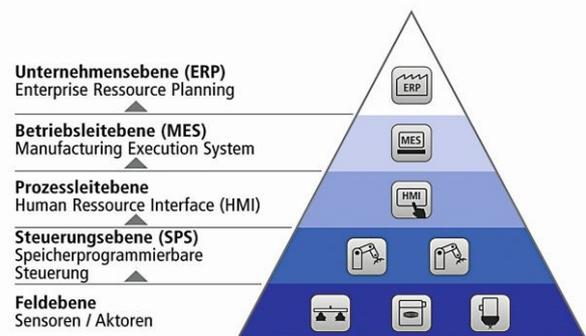


Abbildung 2: Automatisierungspyramide ([5] Blum 2015)

Die Automatisierungspyramide umfasst fünf Ebenen der Automatisierung. In der obersten Ebene findet die Grobplanung einer Produktion mit Hilfe von ERP-Systemen statt. Die feinere Planung wird in der zweiten Ebene, der Betriebsleitebene, durchgeführt, mit Hilfe eines MES. Auf der Prozessleitebene finden die Überwachung und Kontrolle der Prozesse durch sogenannte Supervisory Control and Data Acquisition (SCADA)-Systeme statt. Auf der Steuerungsebene werden die Maschinen und Geräte mit Hilfe von SPS-Systemen gesteuert. Die unterste Ebene ist die Feldebene, welche mit Aktoren und Sensoren die Schnittstelle zum Produktionsprozess ist. Darin werden Daten erfasst und die letztendliche Prozesssteuerung findet auf dieser Ebene statt.

Der Datenfluss verläuft in diesem Modell ausgehend von der untersten bis zur obersten Ebene. Das hat zur Folge, dass beispielsweise ein Prozessleitsystem erst zu einem späteren Zeitpunkt von einer aufgetretenen Abweichung im Prozess erfährt. Auf Betriebsleitebene würden diese Informationen noch später auftreten. Eine verzögerte Aktualität von Daten und Informationen ist in diesem bisher angewendeten Modell zu beachten ([3] Andelfinger u.a. 2017).

Für eine Umsetzung von Industrie 4.0 müssen die soeben beschriebenen Hierarchieebenen der Automatisierungspyramide überwunden werden. Durch die digitale Transformation wird dies ermöglicht. Die senkrechte Vernetzung wird durch eine vertikale Vernetzung von Maschinen und Systemen abgelöst. Außerdem wird die horizontale Vernetzung erweitert. Das bedeutet, dass die Hierarchie der Systeme flacher wird und die Daten weniger Systeme nacheinander durchlaufen. Vielmehr werden die Daten zeitgleich mehreren Systemen direkt bereitgestellt ([3] Andelfinger u.a. 2017).

Zur Veranschaulichung existieren diverse Modelle, die eine überarbeitete Vernetzung beschreiben. Das Modell in Abbildung 3 spiegelt eine Möglichkeit eine Systemarchitektur für Industrie 4.0 zu gestalten wider.

Das Modell ist im Gegensatz zur Automatisierungspyramide (siehe Abbildung 2) nur in drei Ebenen unterteilt. Die erste Ebene wird auch Produktionsebene genannt und ist die Grundlage für einen automatisierten Ablauf einer Produktion. In dieser Ebene ist die Vernetzung mehrerer CPS innerhalb eines Produktionsstandortes beschrieben. Diese Vernetzung wird auch als cyber-physisches Produktionssystem (CPPS) bezeichnet. Die verschiedenen CPS kommunizieren innerhalb des CPPS miteinander.

Die Vernetzung der CPS erfolgt über LAN oder WLAN. Bei den Informationen, die von den CPS untereinander ausgetauscht werden, handelt es sich um Daten zur Verbzw. Bearbeitung der Werkstücke. An den Werkstücken selbst befinden sich RFID-Chips, welche den CPS Infor-

auf, wie z. B. Temperatur, Luftfeuchtigkeit und Erschütterungen. Diese werden dann zusammen mit den Informationen aus den RFID-Chips über eine standardisierte Schnittstelle vom lokalen Netzwerk innerhalb des CPPS in eine Cloud übertragen. Durch das Ablegen der Informationen in der Cloud können Nutzer diese Daten standortunabhängig über ein mobiles Endgerät aufrufen ([3] Andelfinger u.a. 2017).

Die zweite Ebene kann auch als Cloud- oder Serviceebene bezeichnet werden. Hier befinden sich eine oder auch mehrere Clouds, in denen sich die Systeme aus Abbildung 2 befinden und die die Daten aus den CPPS erhalten. Allerdings werden diese als Services implementiert. Ein Service kann beispielsweise in Form einer mobilen Applikation für ein Smartphone oder Tablet etabliert werden. Innerhalb der Cloud bzw. zwischen mehreren Clouds sind die verschiedenen Services miteinander vernetzt. Dadurch können die Services über die Hierarchie der Automatisierungspyramide hinweg Daten austauschen. Die Hierarchie reduziert sich somit in diesem Modell ([3] Andelfinger u.a. 2017).

Auf der dritten Ebene befindet sich die Unternehmensebene. Die Ebene wird durch die verschiedenen Bereiche eines Unternehmens dargestellt, wie z. B. Vertrieb, Einkauf, Marketing oder Forschung und Entwicklung. Über Endgeräte können Mitarbeiter dieser Bereiche die Services in der Cloud verwenden. Die Endgeräte selbst sind über standardisierte Schnittstellen mit der untersten und der mittleren Ebene verbunden. Die Nutzer sehen über dieses Verfahren Daten in Echtzeit ein und passen Prozesse ohne Zeitverzögerung an, ohne dass der Prozess gestoppt werden muss. ([3] Andelfinger u.a. 2017).

Das Modell beschreibt ein agiles System, welches modular und dezentral aufgebaut ist. Es ist somit zeitsparend und kostengünstig bei der Einbindung neuer Komponenten oder der Veränderung bestehender Elemente. Außerdem ermöglicht die Cloud-Technologie eine unkomplizierte Einbindung von Partnerunternehmen oder Lieferanten. Auch der Einsatz unterschiedlicher Hardware- und Softwarekomponenten wird dadurch simplifiziert, was zusätzlich durch die standardisierten Schnittstellen unterstützt wird.

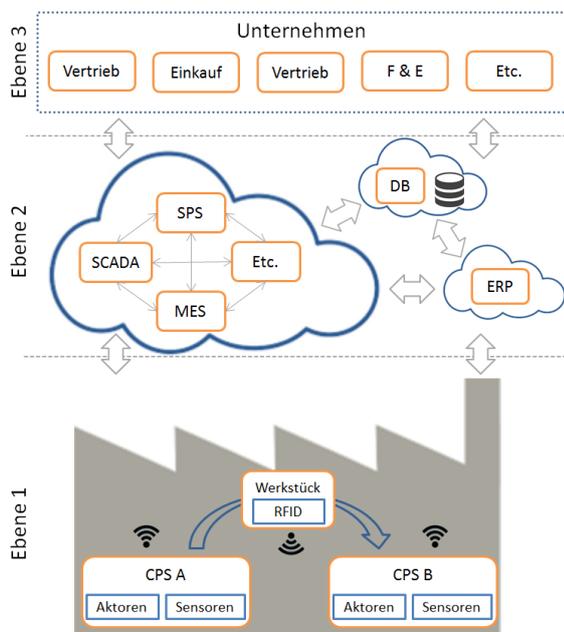


Abbildung 3: Systemarchitektur für Industrie 4.0, in Anlehnung ([3] Andelfinger u.a. 2017)

mationen für die Fertigung und Montage übermitteln. Sensoren nehmen Informationen aus der lokalen Umwelt

KONZEPTION ZUR ANWENDUNG VON INDUSTRIE 4.0 IN EINER LABORUMGEBUNG

Beispielhafter Aufbau eines Labors

Es existieren viele verschiedene Arten von Laboren, die alle einen unterschiedlichen Zweck erfüllen. Im folgenden Abschnitt wird zunächst definiert, welche Art von Labor in diesem Artikel zur Konzepterstellung verwendet wird.

Im Folgenden wird auf ein serologisches Labor bezogen, wie es bei der CSL Plasma GmbH vorzufinden ist. Das Labor hat den Zweck Proben von Plasmaspenden zu testen, die zur Verarbeitung zu Medikamenten verwendet werden. Das Serum wird in Probenröhrchen angeliefert und in zwei verschiedenen Analysegeräten auf unterschiedliche Viren bzw. deren Antikörper getestet. Das Labor besteht aus drei Bereichen: Präanalytik, Analytik und Postanalytik. In der Präanalytik werden die Proben vor Analysen vorbereitet, d. h. empfangen, geschwenkt und geöffnet. In der Analytik werden verschiedene Analysen an den Proben durchgeführt. Bei CSL Plasma sind es insgesamt acht Analysen an zwei verschiedenen Geräten. Hierbei ist ein Gerät für die „Standardtestung“ zuständig, während an dem zweiten Gerät nur Zusatztestungen durchgeführt werden. Nach der Analytik folgt die Postanalytik. Die Proben werden hier nachbehandelt. Darunter fällt beispielsweise eine Kühlagerung und die Entsorgung. Unterstützt wird das Probenmanagement im Labor durch das Laborinformations- und Management-System LabWare LIMS, welches sich auf einem internen Server befindet. Über einen Citrix-Server wird die Anwendung den Labormitarbeitern zur Verfügung gestellt. Die Daten, die aus den Proben in der Analytik erhoben werden, müssen anschließend durch einen Freigabeprozess validiert und daraufhin an andere computergestützte Systeme von CSL Plasma exportiert werden.

Im Labor selbst liegt kein physisches Produkt vor, welches produziert wird. Ein Labor liefert in erster Linie Testergebnisse. Der zentrale Gegenstand in einem Labor sind demnach nicht die Erzeugnisse, also die Ergebnisse, sondern die Proben, aus denen die Ergebnisse gewonnen werden.

Systemarchitektur des Labors

Die Systemarchitektur des derzeitigen Labors ist in Abbildung 4 dargestellt. Im Labor befinden sich drei verschiedene Geräte, die direkt oder über eine Middleware an das LIMS angebunden sind. Die zwei Analysegeräte sind über eine ASTM-Schnittstelle (d.h. gemäß Standarddefinition der American Society for Testing and Materials für den Datenaustausch zwischen Labor-IT-System und Analysegeräten) und der Probensortierer ist über eine FTP-Schnittstelle angebunden. Das LIMS befindet sich auf einem Server im Firmennetzwerk. Auf diesem Server befindet sich auch die LIMS-Datenbank. Die Applikation wird über einen Citrix-Server bereitgestellt und

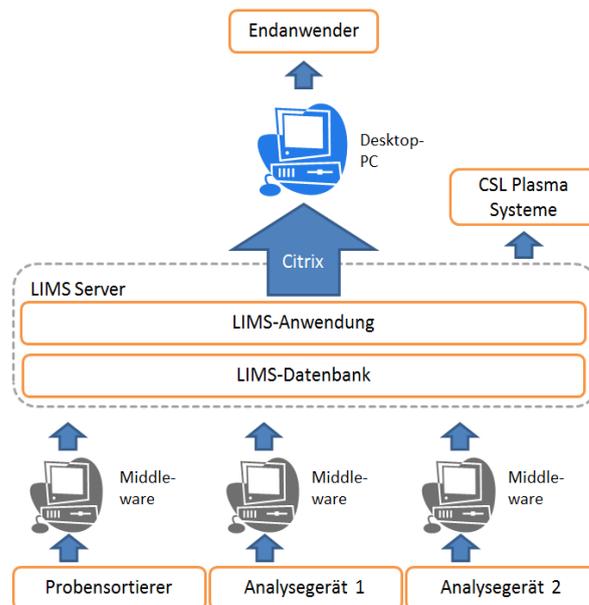


Abbildung 4: Systemarchitektur des derzeitigen Labors

kann somit von den Rechnern im Labor aufgerufen werden, ohne dass eine lokale Installation notwendig ist.

Die Analysegeräte stellen dem LIMS über die Schnittstellen Daten zur Verfügung, z. B. Testergebnisse. Diese werden wiederum vom LIMS über Schnittstellen anderen Systemen bei CSL Plasma bereitgestellt.

Voraussetzungen und Technologien für die Umsetzung von Industrie 4.0

Um ein serologisches Labor nach Industrie 4.0 zu realisieren, müssen gewisse Voraussetzungen erfüllt werden. Deshalb wird folgend festgelegt, welche Anforderungen an ein Labor nach Industrie 4.0 bestehen und welche Technologien angewandt werden. Dies wird bezogen auf das Labor aus dem vorhergehenden Abschnitt und auf Grundlage der Erkenntnisse aus dem Abschnitt zu Industrie 4.0 durchgeführt. Zur Veranschaulichung sind das Labor als CPPS und die neue Systemarchitektur in den Abbildungen 5 und 6 skizziert.

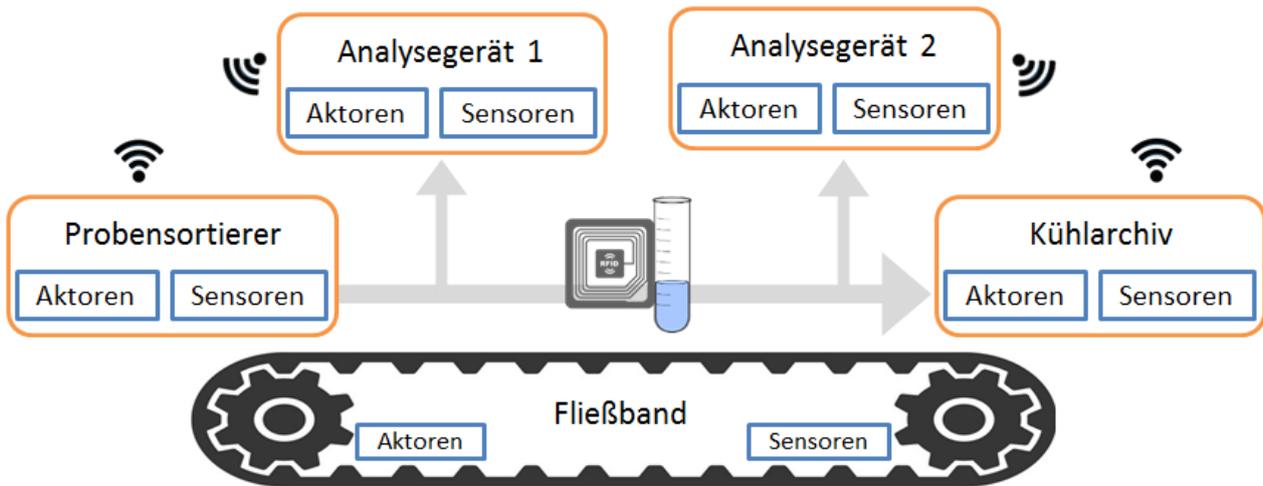


Abbildung 5: Labor als Cyber-physisches Produktionssystem

In dem in Abbildung 3 beschriebenen Modell wird das Werkstück zur Identifizierung mit einem RFID-Chip ausgestattet. Auch die Probe wird mit einem RFID-Chip versehen. Der Chip enthält Informationen über die Probe, die für die Verarbeitung im Labor benötigt werden. Diese sind z. B.

- eine Probenidentifikationsnummer,
- das Datum an dem die Probe erhoben wurde,
- die Analysen, für die Proben durchgeführt werden sollen,
- das Material, welches sich in der Probe befindet und
- das Verfallsdatum der Probe .

Für die Präanalytik benötigt das Labor ein Gerät zum Öffnen und Sortieren der Proben. Die Sortierung muss anhand der Informationen auf dem RFID-Chip durchgeführt werden. Hierzu muss das Gerät mit einem RFID-Scanner ausgestattet sein, um die Daten auslesen zu können.

In der Analytik werden zwei Geräte für die unterschiedlichen Tests benötigt. Die Geräte müssen die erforderlichen Funktionalitäten für die Durchführung der serologischen Testung aufbringen. Weiterhin werden beide Geräte mit einem RFID-Scanner ausgestattet, um die Proben zu identifizieren.

In der Postanalytik wird ein Gerät benötigt, welches die Proben archivieren kann. Das beinhaltet eine sichere und kühle Lagerung. Zudem muss das Gerät in der Lage sein aufgrund empfangener Daten Proben vorschriftsgemäß in gesonderten Behältern zu entsorgen. Auch hier wird zur Identifizierung ein RFID-Scanner benötigt.

Ein zusätzliches Gerät stellt ein Fließband dar, welches sich von der Präanalytik über die Analytik bis zur Postanalytik erstreckt. Das Fließband besitzt die Aufgabe, die Proben über den gesamten Prozess von einem

Gerät zum nächsten zu transportieren, ohne dass ein Mitarbeiter eingreifen muss. An verschiedenen Knotenpunkten müssen RFID-Scanner angebracht werden, um die Proben zu identifizieren und somit die aktuelle Position zu ermitteln und geeignete Transportmaßnahmen zu veranlassen.

Alle Geräte sollen darüber hinaus mit Sensoren und Aktoren ausgerüstet sein, um auf der einen Seite Daten während der einzelnen Vorgänge zu sammeln, und auf der anderen Seite auf Veränderungen reagieren zu können. Zusätzlich benötigt jedes Gerät einen Netzwerkzugang, um alle erhobenen Daten den anderen Geräten und Services via Cloud bereitzustellen. Bei jedem Gerät handelt es sich somit um ein cyber-physisches System. Es liegen dementsprechend fünf CPS vor, die gemeinsam das cyber-physische Produktionssystem darstellen. Durch standardisierte Schnittstellen ist das CPPS zusätzlich modular aufgebaut. Das vereinfacht eine In- oder Außerbetriebnahme neuer CPS oder Services.

Neben dem Labor müssen auch die Systeme gewisse Voraussetzungen erfüllen. Grundsätzlich kann weiterhin ein LIMS oder LIS verwendet werden. Das LIMS wird allerdings als Service in einer Cloud zur Verfügung gestellt. In der Cloud befinden sich ebenfalls die anderen Systeme von CSL Plasma als Services. Auch die LIMS-Datenbank befindet sich in der Cloud. Diese kann entweder als separate Datenbank betrieben oder in eine gemeinsame Datenbank für alle Services integriert werden. Die Cloud muss über eine standardisierte Schnittstelle an das CPPS angebunden werden, um somit die dort erhobenen Daten zu erhalten und Daten den einzelnen CPS zur Verfügung zu stellen. Über eine Applikation greifen die Labormitarbeiter auf die Services von mehreren Endgeräten (Tablets, PCs, Smartphones) zu.

Die anfallenden Big Data können in der Cloud gespeichert und in Services, z. B. LIMS, verwendet werden. Über den Service ist es dem Labormitarbeiter möglich

Ad-hoc-Analysen aufrufen und auch Prozesse direkt anpassen. Außerdem kann Big Data verwendet werden, um „Predictive Maintenance“ anzuwenden. Somit kann der Laborbetrieb aufrecht gehalten werden, da Störungen und Fehlfunktionen präziser vorausgesagt und präventiv behandelt werden.

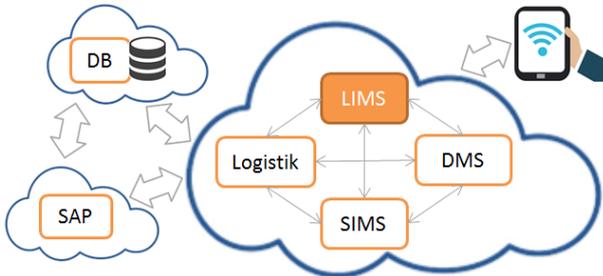


Abbildung 6: Systemarchitektur des Labors nach Industrie 4.0

Automatisierter Laborprozess

Der Prozess beginnt mit der Ankunft der Proben im Labor. Hier werden die Proben manuell auf das Fließband gelegt und direkt über einen Schütteeingang an den Probensortierer übergeben. Dieser meldet der Cloud, dass er Proben erhalten hat. Anschließend werden die RFID-Chips der Proben gescannt und der Cloud wird mitgeteilt, dass die Proben mit der entsprechenden ID im Labor angekommen sind. Jede Probe wird dann nacheinander auf dem Fließband weitergeleitet. Das Fließband gibt seinen aktuellen Status an die Cloud weiter und scannt daraufhin den RFID-Chip der Proben. Die aktuelle Position der Probe wird der Cloud mitgeteilt. Weiterhin wird anhand der Informationen auf dem Chip entschieden, ob die Probe getestet werden muss oder ob es sich um eine nicht zu testende Probe handelt.

Sofern die Probe nicht getestet werden muss, wird sie direkt über das Fließband an das Kühlarchiv weitergeleitet. Muss eine Probe getestet werden, identifiziert ein RFID-Scanner anhand der Daten auf dem RFID-Chip, ob die Probe im Analysegerät 1 getestet werden muss. Wenn keine Anforderung für einen Test im Analysegeräte 1 vorliegt, wird die Probe auf dem Fließband weitergeleitet. Falls eine Anforderung vorliegt, wird über direkte Kommunikation zwischen dem Fließband und dem Analysegerät 1 überprüft, ob die Kapazitäten aktuell ausreichend für eine Analyse sind. Wenn keine Kapazitäten frei sind, wird die Probe ebenfalls auf dem Fließband weitergeleitet. Andernfalls wird die Probe in das Analysegerät 1 transportiert. Die Position der Probe und der Status des Analysegeräts werden der Cloud weitergegeben. Anschließend wird die Analyse durchgeführt. Die erhobenen Daten werden nun in der Cloud gespeichert.

Die Testergebnisse liegen vor und die Probe wird auf das Fließband geleitet. Hier wird der Cloud wieder der Status des Fließbands und die Position der Probe mitgeteilt. Im nächsten Schritt wird wieder der RFID-Chip gescannt. Wenn keine Anforderungen für das Analysegerät 2 vorliegen, bleibt die Probe auf dem Fließband. Muss die

Probe im Analysegerät 2 getestet werden, wird wie zuvor überprüft, ob das Analysegerät freie Kapazitäten hat. Wenn die derzeitige Kapazität nicht ausreicht, verlässt die Probe das Fließband nicht. Sind Kapazitäten verfügbar, wird die Probe an das Analysegerät 2 übergeben. Wieder werden Positionsdaten der Probe und der Status des Analysegeräts an die Cloud weitergegeben. Die Probe wird daraufhin getestet. Die aktuellen Umgebungsdaten und die Daten aus der Testung werden an die Cloud übermittelt.

Für das Analysegerät 2 liegen nun die Testergebnisse vor und die Probe kann wieder an das Fließband übergeben werden. Die Positionsdaten der Probe und der Status des Fließbandes werden der Cloud mitgeteilt. Die Probe hat das Ende des Fließbandes erreicht. An diesem Punkt wird überprüft, ob noch Tests für die Probe vorgesehen sind. Ist das der Fall, wird die Probe wieder auf den Fließbandabschnitt vor dem Analysegerät 1 geleitet. Falls zu diesem Zeitpunkt keinen weiteren Tests ausstehend sind, wird die Probe in das Kühlarchiv geleitet.

Die Probe wird archiviert bzw. eingelagert und daraufhin wird der RFID-Chip gescannt. Der Cloud wird mitgeteilt, dass sich die Probe im Archiv befindet. Außerdem wird überprüft, ob noch Tests ausstehen. Falls noch Analysen für die Probe durchgeführt werden müssen, wird diese wieder auf das Fließband weitergeleitet. Wenn keine Analysen mehr ausstehen wird in der Cloud abgefragt, ob die Probe für die Entsorgung freigegeben ist. Wenn die Probe noch nicht freigegeben ist, wird sie vorerst archiviert. Ist die Probe für die Entsorgung autorisiert, wird sie durch das Kühlarchiv entsorgt und der Cloud wird der Probenstatus mitgeteilt. Anschließend ist der automatisierte Laborprozess beendet.

In Abbildung 7 ist das BPMN-Modell des automatisierten Laborprozess abgebildet. Der Pool stellt das serologische Labor dar. Dieses unterteilt sich in fünf sogenannte Swimlanes: Probensortierer, Fließband, Analysegerät 1, Analysegerät 2 und Kühlarchiv.

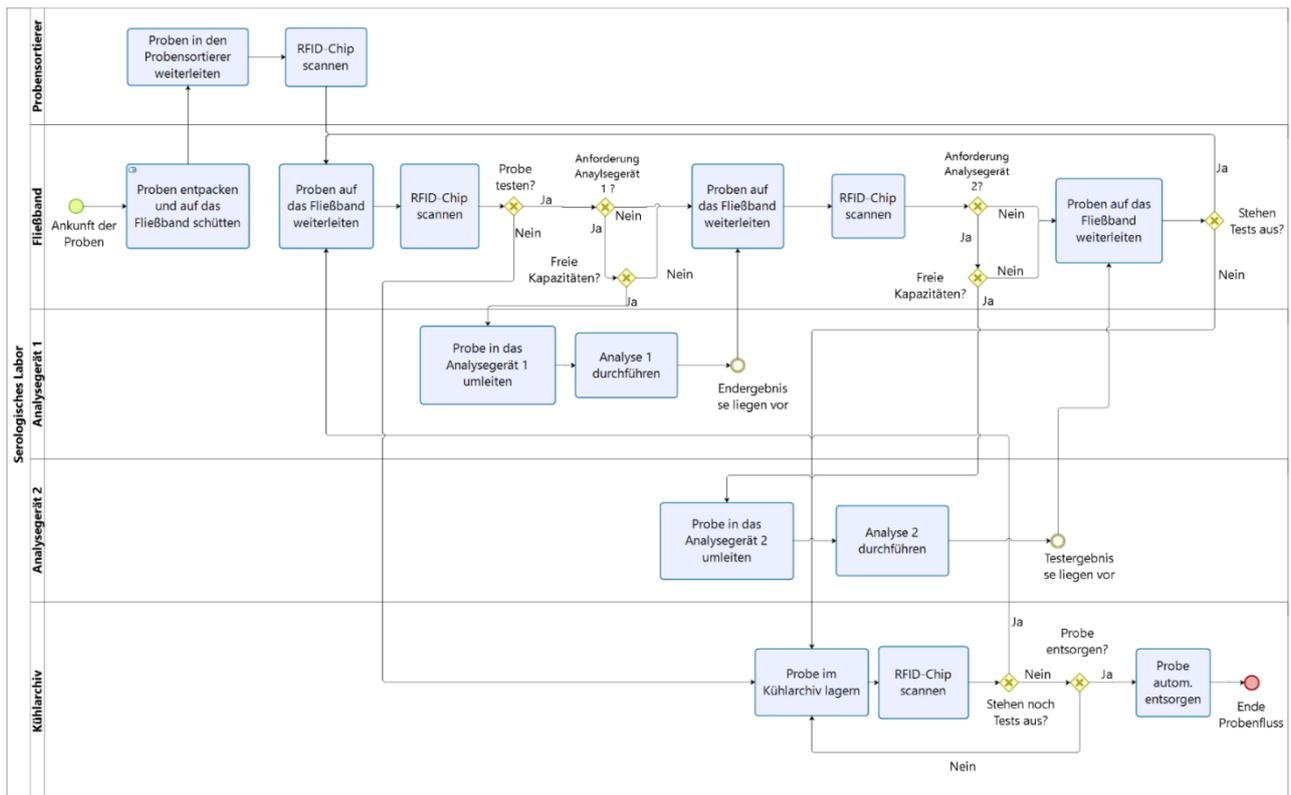


Abbildung 7: Laborprozess nach Industrie 4.0

Gegenüberstellung

Nachdem in den vorhergehenden Abschnitten die zentrale Fragestellung behandelt wurde, werden anschließend der Nutzen und eventuelle Nachteile durch ein Labor nach Industrie 4.0 verdeutlicht. Hierfür werden der Prozess und die Systemarchitektur des serologischen Labors von CSL Plasma mit dem Konzept verglichen.

Zunächst werden die Prozesse untersucht. Im Standardlaborprozess wird deutlich, dass viele Prozessschritte manuell durch den Labormitarbeiter durchgeführt werden. Unter anderem der Probentransport, das Be- und Entladen der Geräte und das Abgleichen einzelner Proben mit Listen aus dem LIMS müssen manuell durchgeführt werden. Diese manuellen Tätigkeiten sind zeitintensiv und insbesondere letztere Aufgabe ist sehr fehleranfällig.

Der vollautomatisierte Laborprozess sieht nur zu Beginn ein manuelles Eingreifen vor. Ansonsten werden alle Aufgaben bezüglich des Transports, der Be- und Entladung und die Entscheidungsfindung selbstständig durch die CPS vorgenommen. Das selbstständige Arbeiten der CPS ist weniger zeitintensiv und minimiert Fehler, da die abzugleichenden Informationen anhand von eindeutigen IDs informationstechnologisch verglichen werden.

Die Systemarchitektur beim Standardlabor ist so aufgebaut, dass das LIMS durchgehend verfügbar ist und ein Datentransfer vom Gerät zum LIMS und vom LIMS zu

den anderen Systemen gewährleistet wird. Die Daten werden allerdings nach einer klaren Hierarchie weitergeleitet. Sie werden zunächst in den Geräten erhoben, dann an die jeweilige Middleware geleitet. Daraufhin werden sie über eine Schnittstelle an das LIMS übertragen. Dieses stellt dann die Daten den anderen Systemen (Logistiksystem, Spendenmanagementsystem) zur Verfügung. Dadurch sind Daten nur zeitverzögert verfügbar. Durch das Bereitstellen über den Citrix-Server ist das LIMS ohne lokale Installation verfügbar. Allerdings kann das LIMS nur über ein Endgerät abgerufen werden mit installiertem Citrix-Receiver. Außerdem ist die Anwendung selbst nicht optimiert für eine mobile Ansicht, was eine problemlose Bedienung nur am PC ermöglicht.

Die Systemarchitektur beim automatisierten Labor ist so aufgebaut, dass die CPS die erhobenen Daten direkt in der Cloud zur Verfügung stellen. Dadurch ist es möglich, dass alle Services direkt auf die Daten zugreifen. Da auch die Services ihre Daten in einer Cloud speichern, können alle Services und CPS in Echtzeit auf alle Daten zugreifen. Das beschleunigt unter anderem Entscheidungsprozesse im Labor. Das Bereitstellen der Services erfolgt über Apps. Realisiert als Web-Applikation sind die Services mit einem Browser über nahezu jedes Endgerät abrufbar. Dadurch sind die Nutzer in der Lage, flexibel Daten abzurufen und sind nicht an feste Geräte für die Nutzung der Services gebunden.

Das Durchführen des Standardprozesses wird zum großen Teil durch den Labormitarbeiter ermöglicht. Dieser führt zwar keine Analysen eigenständig durch, ist aber

durchaus für das Treffen von Entscheidungen verantwortlich. Er nimmt aktiv am Prozess teil und muss in seinen Tätigkeiten auch geschult sein.

Der automatisierte Prozess wird zu kaum einem Zeitpunkt vom Labormitarbeiter beeinflusst. Er startet lediglich den Prozess, indem er die Proben aus den Kartons entpackt. Ansonsten kann der Prozessverantwortliche bei Bedarf einschreiten, was aber prinzipiell nicht notwendig ist.

Es ist zu erkennen, dass sich aus Industrie 4.0 einige Vorteile für ein Labor ableiten lassen. Prozesse sind durch die Automation und die selbstständige Entscheidungsfindung durch die CPS weniger fehleranfällig und zeitsparender. Durch die vertikale und horizontale Vernetzung der Services sind Daten, wie z. B. Testergebnisse, in Echtzeit vorhanden und somit schneller verfügbar. Dadurch können abhängige Prozesse früher starten und präventive bzw. korrigierende Maßnahmen frühzeitig eingeleitet werden. Die Bereitstellung der Services über Web-Applikationen ermöglicht dem Prozessverantwortlichen die Prozesse flexibel von jedem Standort der Welt aus zu steuern. Somit ist dieser nicht an einen festen Arbeitsplatz gebunden und kann seine Arbeit flexibel gestalten. Durch die Automation und die selbstständige Organisation des CPPS werden außerdem weniger Mitarbeiter im Laborbereich benötigt. Das minimiert unter anderem das Unfallrisiko im Labor. Hinzu kommt, dass der Labormitarbeiter selbst keine speziellen Qualifikationen benötigt.

Industrie 4.0 birgt allerdings auch einige Nachteile. Wie aus den Prozessmodellen hervorgeht, ist der Prozess im automatisierten Labor wesentlich komplexer. Zusätzlich wird durch den hohen Automatisierungsgrad und die Verknüpfung aller Komponenten und Services eine höhere Anzahl an technischem Fachpersonal benötigt. Die Cloud enthält als zentraler Datenspeicher ebenfalls Risiken. Die Daten sind in der Theorie global verfügbar. Dadurch steigt die Wahrscheinlichkeit, dass vertrauliche Firmendaten von externen Personen abgegriffen werden können. Die Daten in der Cloud müssen besonders abgesichert sein gegen Zugriffe von außen. IT-Sicherheit spielt dementsprechend eine größere Rolle im IoT, was wiederum zusätzlichen Aufwand verursacht. Ein letzter Aspekt ist die abnehmende Bedeutung des Laborpersonals. Die Labormitarbeiter hatte zuvor Verantwortung und Entscheidungen wurden selbst getroffen. Im automatisierten Prozess hat der Labormitarbeiter diese Verantwortung nicht mehr und er führt nur simple, sich wiederholende Tätigkeiten aus.

PRAXISBEISPIEL

Der bisher beschriebene Prozess ist ein Konzept für die Umsetzung von Industrie 4.0. Derzeit existieren bereits Lösungen von einigen Geräteherstellern ein vollautomatisiertes Labor umzusetzen. Im folgenden Abschnitt wird ein System von der Firma Roche Diagnostics vorgestellt

werden, welches die vollautomatisierte Testung von Proben in einem serologischen Labor ermöglicht.

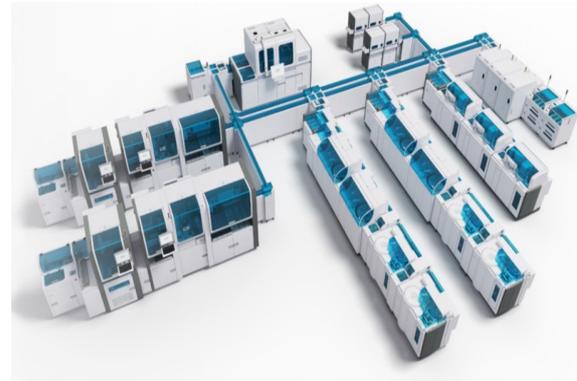


Abbildung 8: cobas® 8100 automated ([22] Roche 2018)

In Abbildung 8 ist das cobas® 8100 System von Roche Diagnostics abgebildet. Es beschreibt eine vollautomatisierte Lösung um Proben mit verschiedenen Analysemethoden zu testen. Das System vereint Präanalytik, Analytik und Postanalytik und ist modular aufgebaut. Für die Präanalytik (linker Teil des Bildes) lassen sich je nach Bedarf die Probenverteiler P512 oder P612 einsetzen. In der Analytik (zentral im Bild) kann ein breites Spektrum an Roche-Analysegeräten verwendet werden. Für die Postanalytik (rechts außen) existieren drei verschiedene Archive, die sich durch ihre Archivierungskapazität und -dauer voneinander unterscheiden. Verbunden sind die Geräte durch ein mehrspuriges Transportband, welches einen bidirektionalen Probentransport zwischen den Geräten ermöglicht. Die Testung erfolgt in jedem Gerät automatisch. Der Mitarbeiter hat trotzdem die Möglichkeit über Zugänge am Gerät manuelle Schritte durchzuführen ([21] Roche 2017).

Die Systemarchitektur des cobas® 8100 Systems baut sich folgendermaßen auf. Jedes modular eingebundene Gerät wird durch seine eigene Middleware gesteuert. Diese ist direkt am Gerät über ein Terminal abrufbar. Hier werden Konfigurationen für die jeweilige Testung durchgeführt. Sind die Geräte bzw. die Terminals in ein Netzwerk eingebunden, können diese auch innerhalb des Netzwerks über eine Remotedesktopverbindung erreicht und konfiguriert werden ([21] Roche 2017).

Das Fließband und der Probenverkehr innerhalb des Systems werden zentral über eine weitere Middleware gesteuert. Roche verwendet hierfür die selbstentwickelte cobas-IT-Middleware. Auch diese ist über ein separates Terminal oder über eine Remotedesktopverbindung innerhalb des Netzwerks verfügbar. Damit die cobas-IT-Middleware einen reibungslosen Probenfluss ermöglichen kann, ist diese mit ASTM- oder HL7-Schnittstellen an die Middleware jedes Geräts angebunden. Zusätzlich kann die cobas-IT-Middleware an ein LIMS angebunden werden, um somit Ergebnisse und Daten zur Testung zu übertragen ([21] Roche 2017).

Roche setzt bereits einzelne Elemente von Industrie 4.0 im cobas® 8100 System um. Das System ist modular aufgebaut, was die Anbindung neuer Komponenten ermöglicht. Prozesse können vollautomatisiert ablaufen, ohne dass Menschen aktiv eingreifen müssen. Das System organisiert sich selbst, indem es Daten auswertet, die innerhalb der Geräte erhoben werden.

Für eine Umsetzung nach Industrie 4.0 müssten am System noch einige Änderungen vorgenommen werden. Zum einen sollten die Geräte direkt miteinander kommunizieren können, ohne dass eine Middleware verwendet werden muss. Die cobas-IT-Middleware sollte nicht als lokale Anwendung verfügbar sein, sondern über eine Cloud-Struktur als Service angeboten werden. Der Service ist wiederum über eine Applikation verfügbar. Dadurch lässt sich der Prozess von mehreren Standorten aus überwachen.

RESULTAT UND ZUSAMMENFASSUNG

Wie eingangs formuliert war Ziel des Artikels, ein Konzept für ein vollautomatisiertes Labor unter Anwendung von Technologien und Prozessen nach Industrie 4.0 zu erstellen. Dies gelang unter anderem indem eine neue Systemarchitektur erarbeitet und auch neue Technologien auf das Labor zugeschnitten wurden. Systeme wie das LIMS werden in dem vorgelegten Konzept in der Cloud als Service angeboten, liefern Echtzeitdaten und sind standortunabhängig erreichbar. Außerdem wurden die CPS im CPPS als Voraussetzung definiert. So kann sich das Labor selbst steuern und die Testung der Proben im Labor vollautomatisch durchgeführt werden.

Weiterhin wurde deutlich, welche Technologien für Industrie 4.0 notwendig sind, wie die Automation durch Software und IT unterstützt wird, und welche Rollen und Aufgaben zu erfüllen sind. Somit liegt ein realisierbares Konzept vor, wie ein Labor nach Industrie 4.0 umgestaltet und optimiert werden kann. Ein weiterer Schritt im Rahmen des digitalen Prozesstransformationsprogramms des Unternehmens ([23] Staufenberg u.a. 2018).

Ausblickend kann man das erarbeitete Konzept für eine Realisierung in Erwägung ziehen und somit zu einer Optimierung eines Laborprozesses beitragen. Beispielsweise kann CSL Plasma das Konzept als Grundlage für eine vollständige oder teilweise Einführung von Industrie 4.0 in das Laborumfeld verwenden. Es kann zunächst die Systemarchitektur angepasst werden, um Applikationen standortunabhängig verfügbar zu machen.

Weiterhin können die derzeitigen Analysegeräte durch fortschrittlichere Geräte ersetzt werden. Hierbei bietet sich unter anderem die Lösung von Roche aus dem vorhergehenden Praxisbeispiel an. Dementsprechend werden durch das Management die Rollen und Verantwortlichkeiten für das Labor neu definiert.

Die Technologien für die Realisierung existieren bereits, dennoch ist eine solche Umsetzung sehr aufwendig und somit eher in kleineren Abschnitten durchzuführen. Besonders Themen bezüglich IT-Sicherheit, die in diesem Artikel außen vor gelassen wurden, spielen hierbei eine große Rolle. Eine eventuelle Realisierung ist demnach zwar zeitintensiv aber möglich, was bedeutet, dass Industrie 4.0 zukünftig auch in einem Laborumfeld bei CSL Plasma angewendet werden kann und sollte.

LITERATUR

- [1] Analytik News (2019): *Laborautomation – Grundlagen*, Online im Internet: URL: <https://analytik.news/Links/Anwendungsgebiete/Laborautomation.html> [Abruf: 14.04.2019].
- [2] Andelfinger, Volker P. und Till Hänisch (2015): *Internet der Dinge – Technik, Trends und Geschäftsmodelle*, 1. Auflage, Springer Vieweg Verl., Wiesbaden.
- [3] Andelfinger, Volker P. und Till Hänisch (Hrsg.) (2017): *Industrie 4.0 – Wie cyber-physische Systeme die Arbeitswelt verändern*, 1. Auflage, Springer Vieweg Verl., Wiesbaden.
- [4] Baun, Christian, Kunze, Marcel, Nimis, Jens und Stefan Tai (2011): *Cloud Computing – Web-basierte dynamische IT-Services*, 2. Auflage, Springer-Verl., Berlin, Heidelberg.
- [5] Blum, Florian (2015): *Mit intelligenter Automatisierungstechnik zu Industrie 4.0*, online im Internet, URL: <https://www.fluid.de/unbekannte-kategorien/mit-intelligenter-automatisierungstechnik-zu-industrie-4-0-296.html> [Abruf: 14.04.2019].
- [6] CSL Behring (2018): *Fact-Sheet*, online im Internet: URL: <https://www.cslbehring.de/-/media/shared/documents/factsheet85x11ger.pdf> [Abruf: 14.04.2019].
- [7] CSL Behring GmbH, 35041 Marburg (2013): *Unternehmensbroschüre*, online im Internet: URL: http://www.cslbehring.de/docs/369/30/CSL_Unternehmensbroschüre,0.pdf[Abruf: 22.08.2017].
- [8] CSL Limited, Parkville, Australien (2018): *CSL Limited – 2018 Full Year Results*, online im Internet: URL: <https://www.csl.com/-/media/shared/documents/results/2018-fy-analyst.pdf> [Abruf: 14.04.2019].
- [9] CSL Plasma (2019): *About CSL Plasma*, online im Internet: URL: <https://www.cslplasma.com/about-csl-plasma> [Stand: 07.03.2019].
- [10] DIN IEC 60050-351:2014-09 (2014): *Internationales Elektronisches Wörterbuch – Teil 351: Leittechnik*, Deutsche Norm vom September 2014.
- [11] Fasel, Daniel und Andreas Meier (Hrsg.) (2016): *Big Data – Grundlagen, Systeme und Nutzungspotentiale*, 1. Auflage, Springer Vieweg Verl., Wiesbaden.
- [12] FM – das Logistik-Magazin (2014): *Enabler für Industrie 4.0*, online im Internet: URL: https://www.fm-online.de/PDF/2014/FM_9_14/Software_9.pdf [Abruf: 13.02.2018].

- [13] Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA (2012): *Laborautomatisierung und Bioproduktionstechnik*, online im Internet: URL: https://www.ipa.fraunhofer.de/content/dam/ipa/de/documents/Kompetenzen/Laborautomatisierung-und-Bioproduktionstechnik/Abteilungsbrochure_Laborautomatisierung_und_Bioproduktionstechnik.pdf [Abruf: 14.04.2019].
- [14] Fraunhofer-Institut für Experimentelles Software Engineering IESE (2019): *Industrie 4.0*, online im Internet: URL: https://www.iese.fraunhofer.de/de/innovation_trends/industrie4_0.html [Abruf: 14.04.2019].
- [15] Gabler Wirtschaftslexikon (2015): *Automatisierung*, online im Internet: URL: <https://wirtschaftslexikon.gabler.de/definition/automatisierung-27138/version-187010> [Abruf: 14.04.2019].
- [16] Gramling, Carolyn (2016): *Smartphones may soon provide earthquake warnings*, in: Science, online im Internet: URL: <https://www.sciencemag.org/news/2016/02/smartphones-may-soon-provide-earthquake-warnings> [Abruf: 19.04.2019].
- [17] Janiesch, Christian (2019): *Cyber-physische Systeme*, in: Enzyklopädie der Wirtschaftsinformatik, online im Internet: URL: <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/informationssysteme/Sektorspezifische-Anwendungssysteme/cyber-physische-systeme>, GITO Verlag [Abruf: 14.04.2019].
- [18] LIMS (2018): *Grundlagen*, online im Internet: URL: <https://www.lims.de/grundlagen.htm>, IMCOR GmbH, Filderstadt [Abruf: 14.04.2019].
- [19] Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft und aca-tech (Hrsg.) (2013): *Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0*, online im Internet: URL: https://www.bmbf.de/files/Umsetzungsempfehlungen_Industrie4_0.pdf, S.17 [Abruf: 14.04.2019].
- [20] Reinheimer, Stefan (Hrsg.) (2017): *Industrie 4.0 – Herausforderungen, Konzepte und Praxisbeispiele*, 1. Auflage, Springer Vieweg Verl., Wiesbaden.
- [21] Roche Diagnostics (2017): *cobas® 8100 automated workflow series – 3-D intelligence in lab automation*, online im Internet: URL: <https://www.roche.de/res/content/7785/cobas-8100-brochure-oct2017.pdf> [Abruf: 14.04.2019].
- [22] Roche Diagnostics (2018): *Vollautomations-Systeme*, online im Internet: URL: <https://www.roche.de/diagnostics/systeme/praeanalytik-postanalytik/vollautomations-systeme.html#cobas-8100-automated-workflow-series> [Abruf: 14.04.2019].
- [23] Staufenberg, Manuel; Nordwald, Benjamin und Harald Ritz (2018): *Entwicklung einer Migrationsstrategie für SAP-Schnittstellen aus SAP Process Orchestration*, in: Anwendungen und Konzepte der Wirtschaftsinformatik (AKWI), Nr. 8, online im Internet: URL: <http://akwi.hswlu.ch>, S.28-36 [Abruf: 14.04.2019].
- [24] Welt (2012): *So funktioniert ein RFID-Chip*, online im Internet: URL: https://www.welt.de/print/die_welt/wissen/article109536379/So-funktioniert-ein-RFID-Chip.html [Abruf: 14.04.2019].
- [25] Wirtschaftslexikon24.com (2018): *Automatisierung*, online im Internet: URL: <http://www.wirtschaftslexikon24.com/d/automatisierung/automatisierung.htm> [Abruf: 14.04.2019].

HYBRID FLOW SHOP SCHEDULING OF AUTOMOTIVE PARTS

Tuanjai Somboonwiwat
Chatkaew Ratcharak
Department of Production Engineering
King Mongkut's University of
Technology Thonburi (KMUTT)
Thailand
E-mail: tuanjai.som@kmutt.ac.th

Tuangyot Supeekit
Department of Industrial Engineering
Mahidol University
Thailand
Email: tuangyot.sup@mahidol.edu

KEYWORDS

Automotive parts, Hybrid flow shop scheduling,
Optimal production schedule

ABSTRACT

Flow shop scheduling problem is a type of scheduling dealing with sequencing jobs on a set of machines in compliance with predetermined processing orders. Each production stage to be scheduled in typical flow shop scheduling contains only one machine. However, in automotive part industry, many parts are produced in sequential flow shop containing more than one machines in each production stage. This circumstance cannot apply the existing method of flow shop scheduling. The objective of this research is to schedule the production process of automotive parts. The feature production is hybrid flow shop which consists of two-stages. In each stage, there are several manufacturing machines and each machine can produce more than one product. Thus, production scheduling is a complex problem. This paper, therefore, develops mathematical model to solve the hybrid flow shop production scheduling under different constraints of each machine. The setup time and production time of each machine can be different for each part. The solution for the experimental data sets from an automotive part manufacturer reveals that the process time can be reduced by 34.29%.

INTRODUCTION

Automotive industry is a very important sector for the country's economy since it generates trade and financial inflows to the country. Automotive part (auto part) manufacturers play an important role in the industry to supply parts for vehicle manufacturers. The response time of auto part manufacturers, which is the total amount of time the manufacturers takes to respond to the orders of auto parts, greatly affects the vehicle production. Responding to vehicle manufacturer demands, then is the goal of automotive parts manufacturers. They have to plan their productions and schedule the machine operations to ensure the shortest total completion time for all orders. Typically, the production type of the automotive parts manufacturers is flow shop where the processes are in a predetermined processing order; one process must be completed before another. The machine

scheduling in the auto part manufacturer requires flow shop scheduling.

In flow shop scheduling the jobs must be produced through the first, second and the following stages. This scheduling problem is considered easy if there is only one machine for each production stage. The typical objective of flow shop scheduling is to minimize the makespan, i.e. to find the minimum total time needed to finish all of the production orders. Hence, the sequencing is decided for the scheduling problem. Typically, the flow shop scheduling deals with scheduling a number of jobs on different stages which contain only one machine on each stage. If each stage consists of many machines working in parallel, this problem is called hybrid flow shop scheduling (Choi et al. 2009). The scheduling problem becomes complicated which assigning and sequencing are required. The previous studies regarding the hybrid flow shop scheduling employ heuristics approaches to schedule the production. For example, Vignier et al. (1996) applies a branch and bound based algorithm to schedule jobs in multi-stages flow shop to minimize the makespan. Watanakich (2001) studies a two-stage hybrid flow shop scheduling with machine setup time, and solved the problem using a heuristic. He presents a two phase heuristic approach; constructing a schedule and assigning jobs with setup time consideration. This represents a difference between regular and hybrid flow shop scheduling. Wong et al. (2001) propose a genetic algorithm to schedule cutting and sewing operations in a manufacturer. Mallikarjuna et al. (2013) apply tabu search algorithm to complete flow shop scheduling. Puck-In (2014) tries to solve the scheduling problem by applying genetic algorithm and hybrid local search to minimize the makespan. It can be seen that most of previous studies apply heuristic algorithm to solve the hybrid flow shop scheduling in order that the makespan are minimized. However, the heuristic approaches do not typically guarantee the optimal solution for the problem.

This paper intends to present a mathematical formulation to solve a two-stage hybrid flow shop scheduling with the job and time constraints in order to achieve minimum makespan of all customer orders. Then the mathematical formulation is validated by applying the formulation to solve the hybrid scheduling in a case manufacturer.

The organization of this paper is as the following. The next section describes the hybrid flow shop scheduling problem. Then the generic mathematical formulation for hybrid flow shop scheduling is presented as a binary integer programming. After that, a numerical example of a case auto part manufacturer is presented to illustrate an application of the formulation to assign jobs to facilities and sequence the jobs. Finally, the conclusion and future research are presented.

PROBLEM DESCRIPTION

This scheduling problem is a two-stage hybrid flow shop. In this flow processes, the jobs can be different types but they must be produced through the first and second stages. There are several non-identical parallel machines in each stage which some jobs cannot be produced at some machines. Also, the processing time of each job at each stage can be different when it is produced at different machines.

This problem studies the assigning and sequencing of jobs for each stage of two flow processes. The job must be accomplished and produced at a particular machines and specific sequence. The job i must be produced through the first stage using machine j in the sequence l and the second stage using machine k in the sequence l as shown in Figure 1.

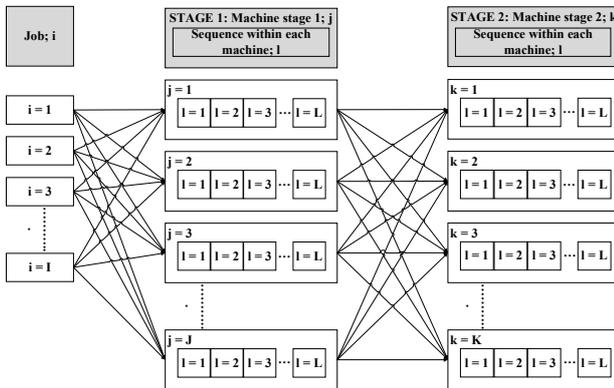


Figure 1: Hybrid Flow Shop Scheduling

MATHEMATICAL FORMULATION

The section describes the mathematical model formulation for this hybrid flow shop with various constraints. The objective is to minimize the makespan in order to find the optimal scheduling consisting of assigning and sequencing.

Indices

- i job : $i = 1, 2, 3, \dots, I$
- j machine at j stage 1: $j = 1, 2, 3, \dots, J$
- k machine k at stage 2: $k = 1, 2, 3, \dots, K$
- l sequence of job: $l = 1, 2, 3, \dots, L$

Parameters

- P_{ij} processing time of job i processed at machine j in stage 1
- P_{ik} processing time of job i processed at machine k in stage 2
- S_{ij} setup time of job i processed at machine j in stage 1
- S_{ik} setup time of job i processed at machine k in stage 2
- ST_{ijl} starting time of job i processed at machine j on sequence l
- ST_{ikl} starting time of job i processed at machine k on sequence l
- ET_{ijl} completion time of job i processed at machine j on sequence l
- ET_{ikl} completion time of job i processed at machine k on sequence l
- C_{ijl} total time of job i processed at machine j on sequence l
- C_{ikl} total time of job i processed at machine k on sequence l

Decision Variables

- $X_{ijl} = 1$ if job i is assigned at machine j on sequence l ; and 0 otherwise
- $Y_{ikl} = 1$ if job i is assigned at machine k on sequence l ; and 0 otherwise

Dependent Variables

- $(C_{max})_{ik}$ total time of job i processed at the last sequence of machine k

Objective Function

The objective function for the scheduling is to minimize the makespan for all jobs.

$$\text{Minimize } Z = C_{max} \quad (1)$$

$$\text{where } C_{max} = \text{Max}\{E_{ikl}\} ; \forall i, \forall k$$

Constraints

1. Constraints regarding the job assigned:

For each job, there should be only one job to be processed at machine j in sequence l .

$$\sum_{i=1}^I X_{ijl} = 1; \forall j, \forall l \quad (2)$$

Similarly, for each job, there should be only one job to be processed at machine k in sequence l .

$$\sum_{i=1}^I Y_{ikl} = 1; \forall k, \forall l \quad (3)$$

All the decision variables are binary.

$$X_{ijl}, Y_{ikl} \in \{0, 1\}; \forall i, \forall j, \forall k, \forall l \quad (4)$$

If there is any job that cannot be processed at a particular machine, that decision variable equals to 0. For example, if the jobs number 1 to 5 cannot be processed at machine 5 or stage 1, the decision variable X_{i5l} equals to 0:

$$X_{ijl} = 0; \forall l, i = 1, 2, 3, 4, 5 \quad (5)$$

2. Constraints related to time

The starting time of job i processed at machine j on the first sequence in stage 1 equals to 0.

$$ST_{ij1} = 0; \forall i, \forall j \quad (6)$$

The starting time of job i processed at machine j in sequence l equals to the completion time of its immediate predecessor job i processed at machine j .

$$ST_{ijl} = ET_{ij(l-1)}; \forall i, \forall j, l = 2, \dots, L \quad (7)$$

The starting time of job i processed at machine k in sequence l must greater than or equal to the completion time of job i processed at machine j in sequence l .

$$Y_{ikl}ST_{ikl} \geq E_{ijl}X_{ijl}; \forall i, \forall k, \forall l \quad (8)$$

The completion time of job i processed at machine j in sequence l equals to the summation of starting time of job i , set up time and processing time at machine j .

$$E_{ijl} = (ST_{ijl}X_{ijl}) + (S_{ij} + P_{ij})X_{ijl}; \forall i, \forall j, \forall l \quad (9)$$

The completion time of job i processed at machine k on sequence l equals to the summation of starting time of job i , set up time and processing time at machine k .

$$E_{ikl} = (ST_{ikl}X_{ikl}) + (S_{ik} + P_{ik})X_{ikl}; \forall i, \forall k, \forall l \quad (10)$$

NUMERICAL EXAMPLE

The aforementioned mathematical formulation can be applied to the case of an automotive part manufacturer to solve the scheduling problems in the factory. The main processes in the case factory are metal cutting processes which machine automotive parts as per customer orders including washer and washer 5th gear thrust. The production of the two parts into the production flow shop is currently scheduled by assigning the job to the idle machines without scheduling plan. This results in long makespan and tardy jobs. The parts of washer and washer 5th gear thrust are often tardy. This needs to be change by planning the scheduling in advance.

The major machining processes for washers and washer 5th gear thrusts to be studied consists of 2 stages; Cutting and Turning processes. Cutting and Turning contain 5 and 3 machines, respectively. The problem can be depicted in Figure 2.

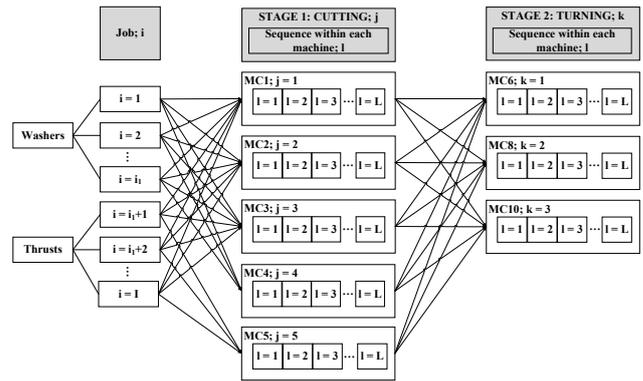


Figure 2: The case scheduling problem

The machines in each stage are interchangeable. There are only little exception regarding the selection of machines as presented in Table 1.

Table 1: Exceptions of machine selection

Stage	Machine no.	Product A Washer	Product B Thrust
Stage 1 Cutting	1	✓	✓
	2	✓	✓
	3	✓	✓
	4	✓	
	5		✓
Stage 2 Turning	6	✓	✓
	8	✓	✓
	10	✓	✓

In order to schedule these jobs, the job orders for the products must be grouped and assigned the job numbers. The example of the case contains 14 jobs; 7 jobs for washers and the rest for washer 5th gear thrust. Jobs of washers are assigned the job numbers 1 to 7, while jobs of thrusts are assigned numbers 8 to 14. The information regarding orders, number of pieces and processing time of each order on each machine are presented in Table 2. The setup time for each machine is 1 hour per a changeover.

Table 2: Jobs and their processing times

Product	Job (i)	Pieces	Processing time (hours)												
			Cutting					Turning							
			1	2	3	4	5	6	8	10					
Washer	1	450	7	3	6	6	0	5	3	1					
	2	552	3	8	7	4	0	6	3	2					
	3	487	7	2	6	3	0	5	3	5					
	4	650	10	10	9	2	0	7	4	3					
	5	500	8	8	7	2	0	3	5	5					
	6	480	7	7	6	2	0	3	5	5					
	7	378	6	6	5	1	0	4	2	4					
Washer 5th gear thrust	8	442	7	5	2	0	6	5	2	4					
	9	398	6	5	5	0	4	5	2	4					
	10	375	3	5	3	0	5	2	4	3					
	11	426	7	4	2	0	6	2	4	3					
	12	500	8	6	3	0	7	2	5	4					
	13	415	6	5	2	0	6	4	2	3					
	14	387	6	4	2	0	5	4	1	3					

Though the machines for each stage are interchangeable, the processing time on different machines are different. For example, Job 1 of 450 washers can be processed on machine no. 1, 2, 3, and 4. It takes 7 hours to complete 450 washers on machine no.1, while it takes only 3, 6, and 6 hours on machine 2, 3, and 4, respectively. Therefore, the selection of machines affects the makespan for all jobs. And it eventually affects the utilization of machines.

The previous scheduling technique used in this case factory yielded 35 hours makespan of scheduling for 14 jobs. The gantt chart to present the makespan of previous scheduling technique can be depicted in Figure 3.

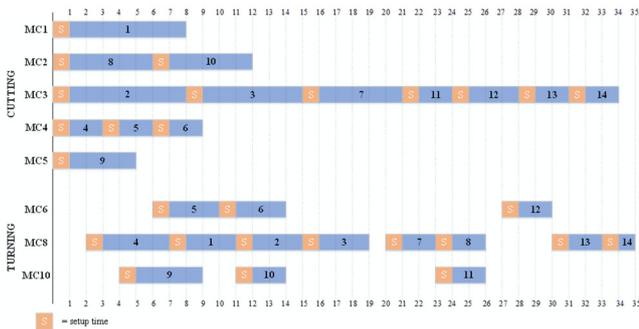


Figure 3: Scheduling applying previous scheduling technique

The aforementioned formulation can be used to shorten the makespan for all the 14 jobs.

Results

The mathematical formulation of the case is applied to the case to make a decision for the scheduling of 7 orders of washers and 7 orders of washer 5th gear thrusts over one-week period (Table 2). The formulation is then solved by the Premium Excel Solver. Using the data of processing time in Table 2, the suitable machine for each job can be selected. The selection of machines yields the total makespan of 23 hours which is the minimum numbers of makespan for the 14 jobs. The result of machine selection can be presented in Table 3.

Table 3: Jobs and selections of machines

Decision variables Job (i)	Processing time (hours)							
	Cutting				Turning			
	1	2	3	4	5	6	8	10
X_{1j}	1	0	I	0	0	0	0	0
	2	I	0	0	0	0	0	0
	3	0	I	0	0	0	0	0
	4	0	0	0	I	0	0	0
	5	0	0	0	I	0	0	0
	6	0	0	0	I	0	0	0
	7	0	0	0	I	0	0	0
	8	0	0	I	0	0	0	0
	9	0	0	0	0	I	0	0
	10	0	0	I	0	0	0	0
	11	0	0	I	0	0	0	0
	12	0	0	I	0	0	0	0
	13	0	0	I	0	0	0	0
	14	0	0	I	0	0	0	0

Table 3: Jobs and selections of machines (cont.)

Decision variables Job (i)	Processing time (hours)							
	Cutting				Turning			
	1	2	3	4	5	6	8	10
Y_{ik}	1	0	0	0	0	0	0	I
	2	0	0	0	0	0	0	I
	3	0	0	0	0	0	I	0
	4	0	0	0	0	0	0	I
	5	0	0	0	0	0	I	0
	6	0	0	0	0	0	I	0
	7	0	0	0	0	0	I	0
	8	0	0	0	0	0	I	0
	9	0	0	0	0	0	I	0
	10	0	0	0	0	0	I	0
	11	0	0	0	0	0	I	0
	12	0	0	0	0	0	I	0
	13	0	0	0	0	0	0	I
	14	0	0	0	0	0	I	0

I = Selected machine for the particular job
0 = Cannot be considered

From Table 3, it can be seen that the mathematical formulation can be used to select proper machines that yield the minimum makespan. Job 1 is to be cut on machine 2 and turned on machine 10; Job 2 is to be cut on machine 1 and turned on machine 10; and so on. Then, the jobs that need to be processed on the same machine need to be sequenced. The sequence of jobs in all cutting and turning machines are presented in Table 4.

Table 4: Sequence of jobs in each machine

Job (i)	MC No.	Sequence						
		1	2	3	4	5	6	7
X_{ij} Cutting	1	2	I	0	0	0	0	0
	2	1	I	0	0	0	0	0
	3	2	0	I	0	0	0	0
	4	4	I	0	0	0	0	0
	5	4	0	I	0	0	0	0
	6	4	0	0	I	0	0	0
	7	4	0	0	0	I	0	0
	8	3	I	0	0	0	0	0
	9	5	I	0	0	0	1	0
	10	3	0	I	0	0	0	0
	11	3	0	0	I	0	0	0
	12	3	0	0	0	I	0	0
	13	3	0	0	0	0	I	0
	14	3	0	0	0	0	0	I
Y_{ik} Turning	1	10	0	I	0	0	0	0
	2	10	0	0	I	0	0	0
	3	8	0	0	0	I	0	0
	4	10	0	0	0	I	0	0
	5	6	0	0	0	0	I	0
	6	6	0	0	0	0	0	I
	7	8	0	0	0	0	I	0
	8	8	0	I	0	0	0	0
	9	8	0	0	I	0	0	0
	10	6	0	0	I	0	0	0
	11	6	0	0	0	I	0	0
	12	6	0	0	0	0	0	I
	13	10	0	0	0	0	0	I
	14	8	0	0	0	0	0	0

Following the sequence of jobs processed at each machine presented in Table 4, the Gantt Chart to present the makespan for all 14 jobs can be depicted in Figure 4.

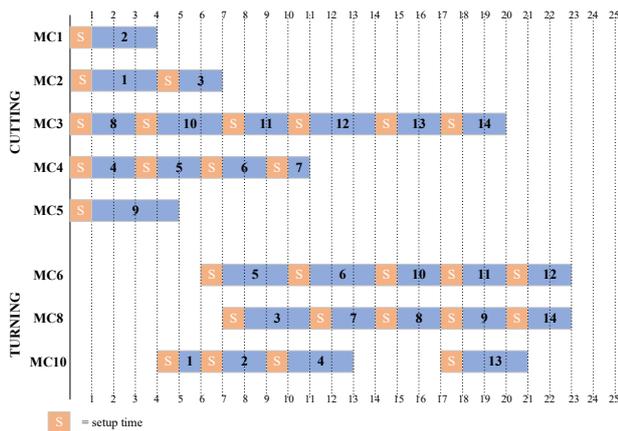


Figure 4: New scheduling applying created mathematical formulation

The result shows the optimal schedule of all the jobs in the two-stage hybrid flow shop by minimizing the makespan. The total makespan of all the 14 jobs is 23 hours. The scheduling in Figure 4 also informs starting, ending, and processing times for each job on each machine. The makespan of scheduling applying the mathematical formulation initiated in this paper is shorter than the 35 hours makespan of previous scheduling method (Figure 3) or 34.29% reduction in the makespan, regardless of longer total time in some jobs.

CONCLUSION

This paper attempts to formulate the scheduling technique for hybrid flow shop production. It can be seen from the case that this mathematical model has a potential to create the optimal schedule for the two-stage flow shop that contains multiple machines in each process. This scheduling methodology concurrently considers the processes of 2 stages to ensure the minimum makespan of all the jobs. It is simply because the final makespan of all the jobs depends on the completion time of the last job

AUTHOR BIOGRAPHIES



TUANJAI SOMBOONWIWAT is an Associate Professor in the Industrial Management section, Department of Production Engineering Faculty of Engineering, King Mongkut's University of Technology Thonburi, Thailand. She received her M. Eng. in Industrial Engineering from Chulalongkorn University, Thailand and Ph.D. in Industrial Engineering from Corvallis, Oregon State University, USA. Her research interests include green supply chain and logistics, business process and applications of operations research. She can be reached at her e-mail address: tuanjai.som@kmutt.ac.th.



TUANGYOT SUPEEKIT is an Assistant Professor in the Department of Industrial Engineering, Faculty of Engineering, Mahidol University in Thailand. He received his Master of Engineering Management from University of Technology, Sydney. His research interests include business process improvement, performance measurement and logistics and supply chain management. He can be reached at his email address: tuangyot.sup@mahidol.edu.

CHATKAEW RATCHARAK was a graduate student at the Department of Production Engineering, King Mongkut's University of Technology Thonburi, Thailand. Her research interests are in supply chain management and applied operations management. Her e-mail address is: ratcharak_chatkaew@hotmail.com.

on stage 2, whereas the starting time of stage 2 relies on the completion time of stage 1. The 2 stages must be considered simultaneously in order to obtain the optimal solution for the scheduling. This mathematical model can be considered useful since it can assign jobs to proper machines and sequence the jobs for each machine to create the optimal production schedule. For future research, this scheduling technique can be expanded to incorporate 3 or more stage production which is typical cases in many manufacturing industry, especially the automotive part industry.

REFERENCES

- Choi, H.S., Kim, H.W., Lee, D.H., Yoon, J., Yun, C.Y. and K.B. Chae. 2008. "Scheduling algorithms for two-stage reentrant hybrid flow shops: minimizing makespan under the maximum allowable due dates," *International Journal Advanced Manufacturing Technology*, Vol.42, pp. 963–973.
- Mallikarjuna, K., Rajashekarpatil, Naweem Kumar, H.M, and Hanumanthappa, S. 2013. "Performance heuristic method over traditional method for solving the job sequencing problem", *International Journal of Engineering Research & Technology*, Vol. 3, pp.3095 – 3099.
- Puck-In, A. 2014. "Solving sequence of job scheduling problem by Genetic Algorithm with local search", *Industrial Technology Lampang University Journal*, Vol. 7, pp. 111 – 126.
- Vignier, A., Dardilhac, D., and Proust, C. 1996. "A branch and bound approach to minimize the total completion time in a k-stage hybrid flowshop", *Proceedings of 1996 IEEE Conference on Emerging Technologies and Factory Automation*, Vol. 1, pp. 215 – 220.
- Watanakich, P., 2001, Scheduling for a Two-Stage Hybrid Flow Shop with Machine Setup Time, *Master of Engineering Thesis, Industrial Engineering*, Faculty of Engineering, Kasetsart University, pp. 1-55.
- Wong, W.K., Chan, C.K. and Ip, W.H. 2001. "A hybrid flowshop scheduling model for apparel manufacturer, *International of Clothing Science and Technology*, Vol. 13, pp. 115 – 131.

CONCEPTS FOR GDPR-COMPLIANT PROCESSING OF PERSONAL DATA ON BLOCKCHAIN: A LITERATURE REVIEW

Florian Zemler

Faculty of Computer Sciences and Mathematics

OTH Regensburg
Prüfeninger Straße 58
93049 Regensburg, Germany

ABSTRACT

Blockchain is an emerging technology that is currently highly discussed in academia and practice. It offers a new approach of sharing data with participants in a network without the need to fully trust them. Among other reasons, this can be achieved because data once stored on a Blockchain is immutable. The recently introduced General Data Protection Regulation (GDPR) standardizes the European privacy regulations and brings major changes regarding how to handle personal data. Literature agrees that processing personal data on Blockchain is not compliant with the requirements of the GDPR. The present paper conducts a structured literature review and identifies four possible concepts for potentially GDPR-compliant data processing using Blockchain: Encryption & Key Destruction, Legal Argumentation, Off-Chain Storage, Redactable Blockchain. Each concept is described and analyzed in view of GDPR's requirements. It becomes clear that the concepts Legal Argumentation and Encryption & Key Destruction might at the moment not be totally compliant with the GDPR.

KEYWORDS

Blockchain, GDPR, Personal Data, Compliance

INTRODUCTION

Blockchain is an emerging technology and many different use-cases have been identified by scholars and practitioners. One of the first applications building upon that technology was Bitcoin, a peer-to-peer cash system (Eberhardt and Tai 2017). Besides the use of Blockchain for cryptographic currencies, the idea of an immutable and distributed ledger is getting more interesting for a large range of different industries and business sectors. Therefore, it is reasonable that applications are and will be designed which process personal data on Blockchain. Especially since late May 2018, when the General Data Protection Regulation (GDPR) came to effect, companies are sensitized how personal data is processed because unlawful handling of it can be heavily prosecuted. Especially when personal data is stored on Blockchain, the literature, e.g., Finck (2018) and Marnau (2017), already identified a conflict with GDPR's requirements. The goal of the paper at hand is to identify concepts and solutions for GDPR-compliant processing of personal data on Blockchain. Thus, it poses the following research question:

RQ: What are possible concepts to enable GDPR-compliant processing of personal data on Blockchain?

To answer this research question, the paper conducts a structured literature review, because it is an appropriate method and creates a "firm foundation for advancing

knowledge" (Webster and Watson 2002). In addition, this review helps later researchers to analyze what is already known in this research area.

The paper first provides an overview of the Blockchain technology and the GDPR. Then, the conflict between both is outlined which occurs when personal data is processed using Blockchain. The next part describes the research methodology to allow replicability of this review and provides a descriptive analysis of the identified literature. The next chapter presents the findings and evaluates the results. It starts with a concept-centric matrix of the literature. Then, all identified concepts for GDPR-compliant processing of personal data on Blockchain are presented. Each concept is described according to the literature and is analyzed in view of the GDPR's requirements. In the last section, the concepts are evaluated, and the advantages and disadvantages are compiled. Finally, the results of the literature review are summarized and a proposal for further research based on the results of the present paper is provided.

The paper at hand contributes to current research by summarizing the existing knowledge and providing a first in-depth analysis of potentially GDPR-conform concepts for processing personal data on Blockchain in light of GDPR's requirements. To the author's best knowledge, this is the first structured literature review addressing this aspect.

BACKGROUND

The following chapter provides a basic overview of the Blockchain technology and the GDPR. The first section

briefly presents the concept of Blockchain. The second part provides an overview of the scope of the GDPR. The last part outlines the conflict between both.

Blockchain

Blockchain was first introduced by Nakamoto in 2008 in form of the peer-to-peer cash system Bitcoin (Marnau 2017; Azaria et al. 2016; Eberhardt and Tai 2017).

With Blockchain, data is distributed among all participants (nodes) in the network (Viriyasitavat and Hoonsopon 2018). Data is stored as a transaction and several transactions are summarized to a single block (Dinh et al. 2018). Every block holds a reference to its precursor. Blocks can only be added to the end of Blockchain which leads to an append-only data structure (Dinh et al. 2018).

The connection between two blocks is established by the cryptographic hash of a block which is stored in its successor (Zheng et al. 2017). The following Figure 1 provides an overview of this procedure.

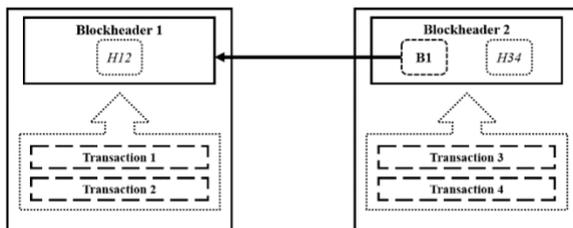


Figure 1: Simplified representation of two blocks in a Blockchain. In dependence on Drescher 2017.

The fact that each block is connected to its precursor by its cryptographic hash leads to the immutability of a Blockchain. Every change of a transaction, which is already stored on-chain, would modify the hash of every block after it and thereby the hash values of the whole Blockchain.

The special feature of Blockchain is that the participants in the network do not have to fully trust each other, because the technology ensures that every participant agrees on a common consensus (Dinh et al. 2018). This is accomplished by so-called consensus algorithms (Dinh et al. 2018).

The Blockchain technology can be divided into three main types:

- **Public Blockchain:** Completely decentralized and open to everyone (Zheng et al. 2017; Viriyasitavat and Hoonsopon 2018)
- **Private Blockchain:** Completely centralized and only selected participants can read and create transactions (Zheng et al. 2017; Vo et al.)
- **Consortium Blockchain:** Partly decentralized and managed by several organizations (Zheng et al. 2017; Ateniense et al. 2017)

GDPR

The GDPR came to effect on May 25th, 2018 and replaced the existing Data Protection Directive from 1995. The GDPR is relevant for every automated processing of personal data in the EU or from subjects located within the EU. In Art. 4 (1) GDPR defines personal as:

any information relating to an identified or identifiable natural person [...]; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name [or] an identification number [...].

The processing of personal data must meet certain requirements. Art. 5 GDPR prescribes that every processing must obey the “principles relating to processing of personal data”. These principles state, among others, that personal data must be correct, up to date, and that incorrect data must be erased immediately. Moreover, data should only be stored as long as it is necessary.

The GDPR grants the data subject, i.e., the person whose personal data is processed, several additional rights regarding the processing of its data. Art. 16 GDPR allows the data subject to have incorrect personal data corrected. The right to be forgotten of Art. 17 GDPR grants the right that under certain conditions the data subject can have its personal data deleted. Additional rights are, e.g., the right to object (Art. 21 GDPR) or the right of access of the data subject by Art. 15 GDPR.

Security and privacy of the processed personal data is also an important part of the GDPR. In Art. 25 GDPR demands appropriate technical and organizational measures to guarantee the principles of processing personal data as well as procedures to only process data of a data subject that is necessary for a certain scenario. Art. 32 GDPR deals with the security of personal data processing. It requires the implementation of technical and organizational measures to secure the processing. These measures are, among others, the use of encryption and pseudonymization as well as methods of ensuring the confidentiality, integrity, and availability of personal data.

Finally, Art. 44 GDPR manages the transfer of personal data to third countries. The transfer of these kinds of data is, in general, only permitted when the receiving country ensures a proper level of protection. If this is not the case, there are some additional conditions which however permit the transfer. According to Art. 49 GDPR this can be for example the explicit consent of the data subject.

Conflict between Blockchain and GDPR

After presenting the basic concept of the Blockchain technology and an overview of the GDPR-requirements of processing personal data, it becomes clear that conflict situations may arise when personal data is processed on Blockchain.

Distributing data among all participants in a network and making changes impossible are basic concepts of the

Blockchain technology. But in view of privacy regulations, this leads to several problems. As mentioned before the processing of personal data must follow certain principles according to Art. 5 GDPR.

Marnau (2017) compared the requirements of Art. 5 with the Blockchain technology and came to the conclusion that especially the principles of Accuracy (Art. 5 (1) lit. d) and Storage limitation (Art. 5 (1) lit. e) cannot be easily fulfilled. Due to its immutability, personal data stored on Blockchain cannot be updated or deleted if necessary. This also contradicts the right to be forgotten and the right of rectification of the data subject.

Therefore, it must be avoided to store personal data on Public Blockchain because, first, it cannot be guaranteed that the data is used in proper ways (Marnau 2017) and, second, it might be possible that the data is transferred to countries which do not have a certain level of protection according to Art. 45 (1) GDPR.

It seems obvious that with the basic concept of Blockchain, it is not possible to fulfill these requirements for lawful processing of personal data of the GDPR.

METHODOLOGY

The following chapter provides an overview of the methodology applied in this paper.

As recommended by Webster and Watson (2002) the review starts with collecting relevant peer-reviewed literature by searching in commonly known databases of high-quality journals and conferences.

The analysis starts with a keyword search consisting of the terms “Blockchain” AND “GDPR” AND “Personal Data” in all fields in the current top 10 Information System and Management journals¹. This approach generated no results. A second research in the commonly known conference proceedings of the AMCIS, ECIS, HICSS, ICIS, and PACIS lead to the same result of no records.

Because of the lack of high-quality literature, the search scope was extended to peer-reviewed articles in further journals and conferences as well as non-peer-reviewed and practitioners’ sources. For this approach, the same keywords were used in commonly known academic databases. The following Table 1 provides an overview of the databases that were consulted and the corresponding number of results.

Table 1: List of databases and number of results for the keyword search

Nr.	Database	Results
1	Google Scholar	867
2	ScienceDirect	32
3	IEEE Xplore	5
4	EBSCOhost	2
5	ProQuest	0
		906

In the first step, the articles were reviewed for relevancy by reading their titles. Then, the abstracts of the remaining 60 articles were analyzed for relevancy. The remaining 40 articles were reviewed for relevancy by skimming and scanning their full texts. In the end, a total number of 16 relevant sources remained. Then, as suggested by Webster and Watson (2002) a forward and backward search was performed to identify additional relevant literature. This procedure lead to three more articles. To identify concepts from practitioner and further sources a Google search with the same keywords was performed and the first 100 results were inspected. This procedure revealed an additional amount of 9 articles and papers for the literature review. In summary, the keyword search returned 28 sources. Figure 2 provides an overview of the literature selection process.

The following Figure 3 shows the distribution of the identified literature over time as well as their appearance

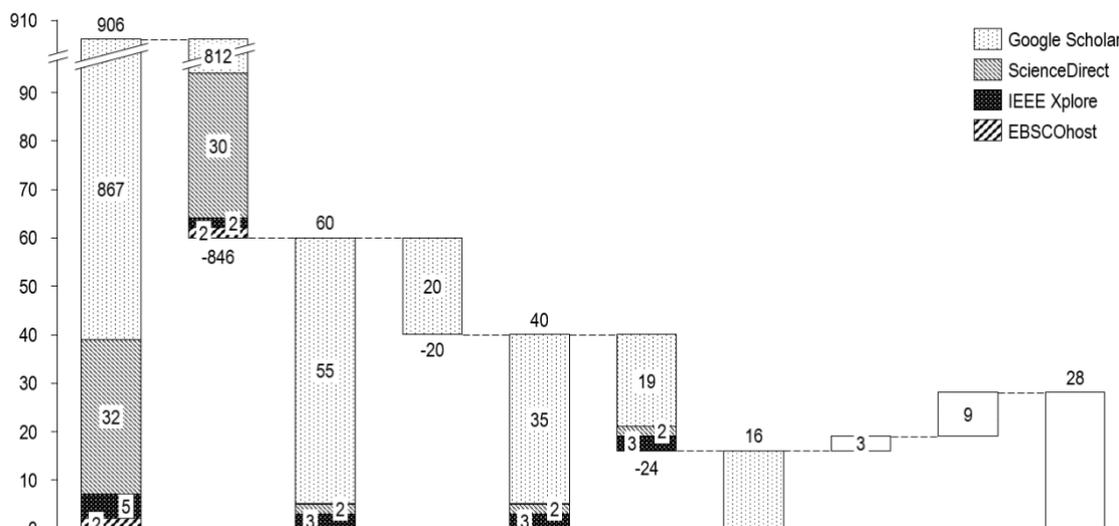


Figure 2: Waterfall diagram representing the literature selection process. Own figure.

¹ <https://www.scimagojr.com> (last access on 28.12.2018). Journals sorted by h-index.

in a journal, conference, or other sources like internet documents.

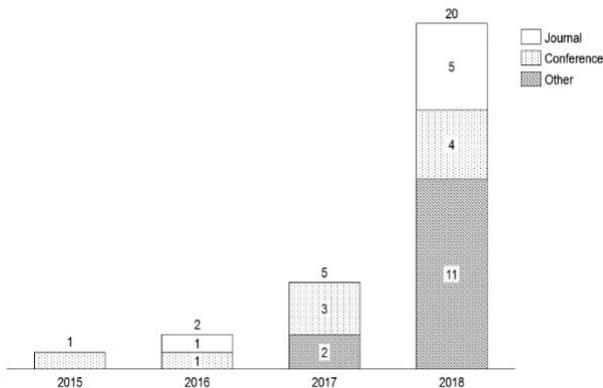


Figure 3: Overview of the distribution of the publications by year and type. Own figure.

It seems to be reasonable that most of the relevant literature (20; 71%) was published in 2018 when the GDPR came into effect. Many of the papers (13; 46%) come from non-peer-reviewed sources like internet documents or government (related) facilities' reports. This shows the need for more scholarly literature in this specific research area.

As recommended by Webster and Watson (2002) the results were structured concept-centric. The review of the literature identified two main concepts:

1. Off-Chain Storage
2. Redactable Blockchain

Beside these two concepts above, two others, the Legal Argumentations, and the Encryption & Key Destruction are mentioned in the literature. Each of these concepts is described according to the reviewed literature and checked against the requirements of the GDPR. In the end, the concepts are evaluated, and their advantages and disadvantages are outlined.

FINDINGS

This chapter presents the findings of the literature review. First, an overview of the analyzed literature and the corresponding concepts is presented. Then, the identified concepts from the literature are introduced and described. Each concept is analyzed in view of the GDPR. In the end, a final evaluation of the concepts is presented, which outlines their individual advantages and disadvantages.

Overview of the analyzed literature

The review of the papers and practitioner reports revealed two main concepts: The Off-Chain Storage and the Redactable Blockchain. In addition, further concepts were identified which are summarized under the item "Other". Each concept is explained in detail below.

The following Table 2 provides an overview of the analyzed articles in alphabetical order and the concepts they are representing.

Table 2: Overview of the analyzed papers and their corresponding concepts

Author(s)	Off-Chain Storage	Redactable Blockchain	Other	Peer-reviewed?
Ateniese et al. (2017)		✓		✓
Azaria et al. (2016)	✓			✓
Berberich and Steiner (2016)	✓		✓	✓
Cichosz et al. (2018)	✓			✓
CNIL (2018)	✓		✓	✗
Eberhardt and Tai (2017)	✓			✓
Eichler et al. (2018)	✓			✗
Esposito et al. (2018)	✓			✓
Fan et al. (2018)	✓			✓
Finck (2018)	✓	✓		✓
Gräther et al. (2018)	✓			✓
Ibáñez et al. (2018)	✓	✓	✓	✗
Jensen (2018)	✓		✓	✗
Jussila (2018)	✓	✓	✓	✗
Katuwal et al. (2018)	✓			✗
Kuner et al. (2018)		✓		✓
Lima (2018)	✓		✓	✗
Michels (2018)	✓		✓	✗
Moser (2017)	✓			✗
Pagallo et al. (2018)	✓	✓	✓	✓
Salmensuu (2018)			✓	✗
Steichen et al. (2018)	✓			✓
The European Union Blockchain Observatory & Forum (2018)	✓		✓	✗
van Geelkerken and Konings (2017)			✓	✓
van Humbeeck (2017)	✓			✗
Wirth and Kolain (2018)		✓		✓
Zhangy et al. (2018)	✓			✗
Zyskind et al. (2015)	✓			✓
	23	7	11	15

Table 2 shows that most of the available literature deals with Off-Chain Storage as a possible solution for the Blockchain and personal data conflict. It is followed by

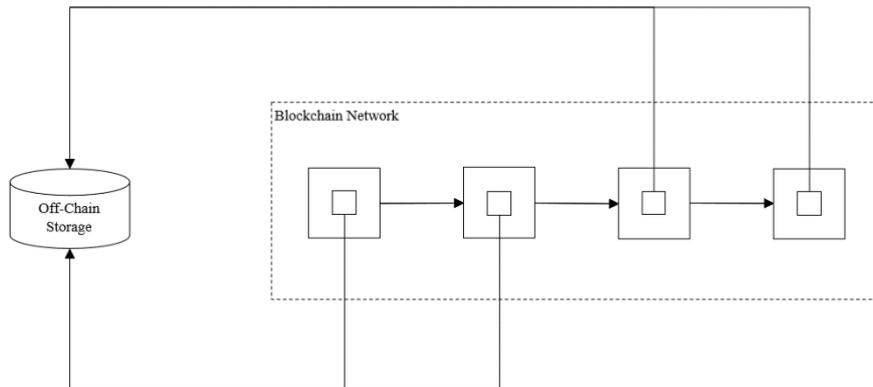


Figure 4: Off-Chain Storage architecture. Own figure.

the Redactable Blockchain as a sole concept and two additional concepts summarized under the term “Other”.

Identified concepts

This section details the identified concepts for storing personal data using Blockchain. It starts with Off-Chain Storage, which is the most discussed concept. In the next step, the Redactable Blockchain concept is introduced. In the last section, two more concepts are presented.

Off-Chain Storage

Off-Chain Storage is the most discussed concept in the reviewed literature for GDPR-compliant processing of personal data on Blockchain. Storing data "off-chain" means, that the data, in this case, the personal data or in general the payload, is not kept inside the Blockchain network, but stored outside, e.g., in a traditional database (Esposito et al. 2018; Ibáñez et al. 2018). Only a reference (for example, a hash value) to the outside storage location where the actual data is stored, is saved on the Blockchain (Katuwal et al. 2018; Zyskind et al. 2015; Steichen et al. 2018; Pagallo et al. 2018; van

Humbecck 2017). Figure 4 provides a simplified overview of an Off-Chain Storage architecture.

In general, storing larger data sets outside of a Blockchain network is highly recommended, because storage capacity on a Blockchain is relatively cost-intensive (Zhangy et al. 2018) and currently not very performant (Jensen 2018). Omaar (2017) identified the cost of storing one Gigabyte of data on the Ethereum Blockchain to be approximately 17.500 Ethereum per Gigabyte which was worth approximately two million euros in March 2019

As Figure 4 illustrates, off-chain data is stored outside of the Blockchain network. That often implies the reintroduction of a trusted third party (TTP) which guarantees the confidentiality and integrity of the data. This provides a certain degree of control to a centralized party that seems to be a violation of the principles of Blockchain (Ibáñez et al. 2018). For this reason, Eberhardt and Tai (2017) recommend the use of a content addressable storage for off-chain location, which stores the files not by their names, but by their hash values. This has the advantage that it is now possible to trustless outsource data because a change of the data would lead

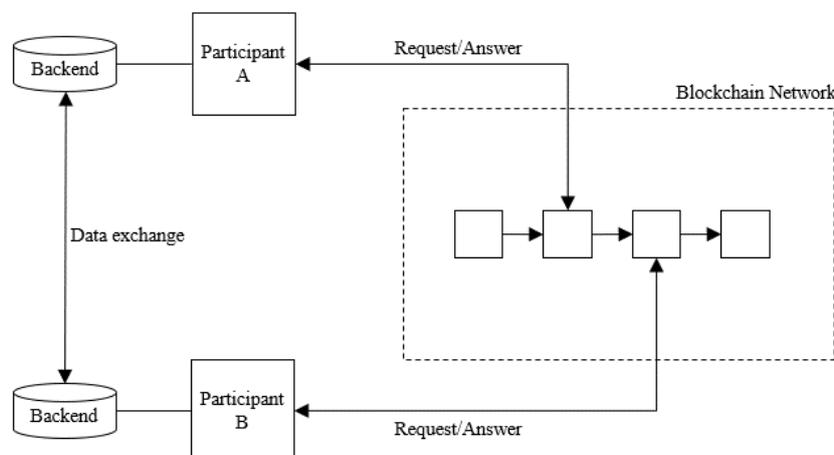


Figure 5: Off-Chain Storage concept by van Humbecck (2017). Own figure.

to an adjustment of its hash value and therefore to its storage location (Eberhardt and Tai 2017).

In the reviewed literature Zyskind et al. (2015) came up first with a solution to use Blockchain in combination with a distributed file system and to only store the reference on-chain to avoid a centralized storage location (Zyskind et al. 2015).

In practice, Steichen et al. (2018) and Gräther et al. (2018) make use of this technique in their applications. They both use the decentralized Interplanetary File System as their content addressable storage system. That means, that no centralized location is needed, and trust is ensured by the way the data is stored on the file system. Van Humbeeck (2017) presents a slightly different concept. In his solution, data is also stored off-chain, but not in a centralized location or in a content addressable storage, instead in the backend system of each participant of the Blockchain network. The Blockchain itself only contains links to the off-chain locations and the hash values of the requested data (van Humbeeck 2017). If a participant has the privilege and the need to access a certain data set, s/he receives the reference (for example, in form of an access token or database string) to where the data is stored and the corresponding hash value. Then, the requestor can directly fetch the data from the backend system where it is stored. Figure 5 provides an overview of the architectural design.

In the context of the GDPR, storing personal data off-chain brings, at first sight, many benefits.

Some argue that with this procedure no personal data stays on the Blockchain and therefore the requirements of the GDPR can be met (Steichen et al. 2018; Katuwal et al. 2018; Ibáñez et al. 2018). However, in most cases,

the hash value of the personal data is still kept on-chain and works as a reference or as proof-of-correctness. At the moment it cannot be said with absolute certainty that hash values of personal data can be regarded as anonymous data or should more likely be treated as pseudonymous data (Eichler et al. 2018). This objection is based on a report by the Article 29 Data Protection Working Party (2014), a former advisory body of the European Commission, which clearly states that hashing techniques must be considered as pseudonymization. The fact that hashed personal data should be considered as pseudonymous data is heavily discussed at the moment (The European Union Blockchain Observatory & Forum 2018) but seems to be widely accepted in the reviewed literature (Finck 2018; Ibáñez et al. 2018; Jensen 2018; Jussila 2018). The proposition, whether the hash value of personal data should be considered as pseudonymized personal data depends on the linkability, i.e. the possibility to connect the hashed data with the original data, between them. Article 29 Data Protection Working Party (2014) argues, that even if a hash cannot be reversed, it can simply be recalculated if the range of input values and the hash function are known. The risk of hash recalculation can be minimized by adding additional information to the dataset, for example, a secret key (Article 29 Data Protection Working Party 2014; The European Union Blockchain Observatory & Forum 2018).

The French National Commission on Informatics and Liberty, CNIL (2018), published a report on the responsible use of Blockchain and personal data. They recommend to only store personal data on a Blockchain as a cryptographic commitment². When that is not

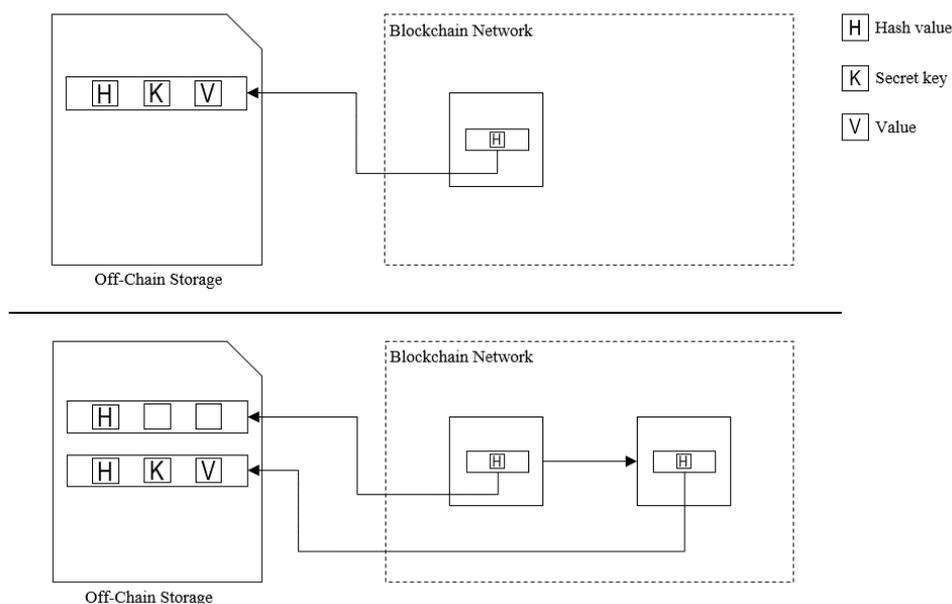


Figure 6: Example of rectification of data with Off-Chain Storage. Own figure

² Describes a method where it is possible to commit to a certain value but keeping this information hidden to others. Later, it is possible to

reveal this information. The procedure seems to be firstly described by Blum1983 in form of the “coin flipping by telephone” problem.

possible the personal data should be stored as a keyed-hash value on Blockchain. And only if that again is not possible, state of the art encryption algorithms should be applied.

Especially when looking at the data subjects' right to be forgotten and the right of rectification, it is technically not possible to fulfill them. The CNIL (2018) however argues that with the use of state-of-the-art keys and algorithms it is possible to move closer to the desired effects. The erasure of personal data can be performed by deleting the data off-chain and the corresponding key, used for generating the hash value, which is stored on-chain. In this case, it is not possible to prove or verify which data has been hashed (CNIL 2018) and the data staying on-chain could be considered as anonymous data (Eichler et al. 2018). In view of the right of rectification, the old dataset can be deleted as described before and a new transaction containing the corrected data can be submitted to the Blockchain. Figure 6 above provides an overview of this technique.

In summary, it can be said, that at the moment there is no legal guarantee that working with cryptographic references based on personal data on-chain is a GDPR-compliant concept for personal data processing using Blockchain. Even referenced data may still be perceived as pseudonymized personal data from a legal perspective and erasure or rectification is technically not possible. Of course, these circumstances can change over time with changing legal perspectives on this procedure.

Redactable Blockchain

Redactable Blockchain was firstly mentioned in the reviewed literature by Ateniese et al. (2017) and is a relatively new concept. By "redactable" the authors mean to rewrite one or more blocks that were already written on Blockchain, to compress any number of already existing blocks to a smaller number and to insert one or more blocks to the existing chain (Ateniese et al. 2017). At first, this seems to contradict the immutability of Blockchain, one of its basic concepts. However, Ateniese et al. (2017) argue that immutability may not be appropriate for all new applications based on Blockchain technology. This can, i.e., be the storage of files or the management of personal health records. This data should be able to be deleted if they contain errors or it is required by law (Ateniese et al. 2017). This law could be the GDPR.

The immutability of a Blockchain comes from the collision resistance of the hash values that connect each block to its precursor. To make Blockchain mutable the concept makes use of a special form of a so-called "chameleon hash function" (Ateniese et al. 2017). A chameleon hash function works like any other hash function with the difference that it has something like a trapdoor which can be used to generate collisions (Ateniese et al. 2017). These collisions can, e.g., be used to alter transactional data without changing the corresponding hash value of the block and by this maintain the connection to its successor. Ateniese et al.

(2017) describe the procedure like adding a lock to the connection between two blocks which can be opened with the right key.

Figure 7 presents the three different phases of a Redactable Blockchain. In (1) the Redactable Blockchain behaves like any other Blockchain and no modification is possible because all locks are secured. At (2) the connection between B2 and B3 is opened with the secret key and modification is possible. In (3) the modifications at B2 are finished and result in form of the block B2'. The connection between B2' and B3 is locked again and no more modification is possible.

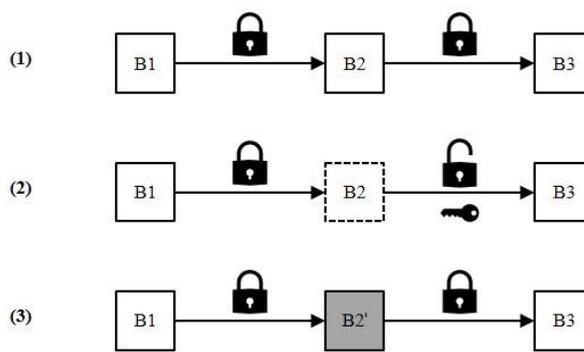


Figure 7: Principles of the Redactable Blockchain. Own figure. Adapted from Ateniese et al. 2017

It is important to notice that when the key to the lock of the hash function is lost or gets destroyed it is not possible to modify the blocks and the Blockchain is immutable again (Ateniese et al. 2017). Therefore, the management of the trapdoor key is an essential part of this concept. Ateniese et al. (2017) describe that, e.g., in a Private Blockchain network the key could be given to the central authority or in a Consortium Blockchain it could be shared among all the participants of the network. In the analyzed literature only a few authors, like Finck (2018), Ibáñez et al. (2018), or Pagallo et al. (2018), identified the potential use of the Redactable Blockchain concept in the context of the Blockchain and GDPR conflict. A real-world application with a Redactable Blockchain could not be identified in the reviewed literature.

The Redactable Blockchain could be an interesting solution for the described conflict (Jussila 2018). The idea of directly removing blocks containing personal data would solve many problems. However, the concept also faces some problems. First, adding redactability to an existing Blockchain is not possible, that means that the decision for this concept must be made before the network is set up (Ibáñez et al. 2018). Secondly, old copies of Blockchain would still contain the redacted data (Ateniese et al. 2017; Finck 2018) but a compliant Blockchain node will accept the redacted data (Ibáñez et al. 2018) and delete the old copies (Ateniese et al. 2017). Finally, there is always the risk that a party redacts the Blockchain to its favor (Ibáñez et al. 2018).

In the context of the GDPR, this concept seems to solve the problems of the conflict between storing personal

data on Blockchain and the data privacy regulation. The ability to delete or alter data after they have been stored on-chain is a huge benefit. Of course, it must be ensured that all the participants in the Blockchain network are operating in compliance with the GDPR that means that redactions to the Blockchain are performed immediately and old copies are deleted trustworthy. This procedure can, for example, be audited by an external party. Critics may argue that giving the possibility to redact data on the Blockchain to a single entity is a violation of the basic principles of the technology (Pagallo et al. 2018). Even the general opportunity of altering data stored on the Blockchain contradicts the basic idea of Blockchain. Ateniese et al. (2017) argue that the immutability of Blockchain should be reconsidered as applications, based on this technology, develop from simple lab experiments to real-world deployments.

Other

The review of the literature also identified two additional proposals for solutions which could not be assigned to one of the two previously described concepts. This section briefly describes these concepts and discusses them in the context of GDPR.

Legal Argumentation

Legal Argumentation deals with the imprecise definition of some parts of this regulation. Due to the immutability of the Blockchain technology, it is a real challenge to delete personal data once stored on Blockchain. From a legal point of view, the literature review identified two different ways of arguing against this proceeding.

Ibáñez et al. (2018) bring forward that the right to be forgotten is not an absolute right and the concept of erasure leaves room for interpretation. In their example, they present that erasure or amendment is performed by adding a new transaction to Blockchain which contains a reference to the obsolete entry and invalidates it semantically (Ibáñez et al. 2018).

Another solution is presented by Berberich and Steiner (2016) who illustrate that with the help of Article 17 (1) lit. b it might be possible to argue that personal data stored on-chain is required for processing because Blockchain needs a persistent chain to function correctly. Therefore, the right to be forgotten of the data subject could not be applied here.

All these legal arguments should currently be treated with prudent care. Because of the lack of judgments on this field, this concept operates in a grey area. It is very likely that these procedures could be considered illegal by a court in the future. Then, it is not possible to remove the personal data from an existing Blockchain, even if required by law.

Encryption & Key Erasure

The last concept identified in the course of the literature review deals with encrypting data on a Blockchain to reach GDPR compliance. Some studies suggest encrypting the data that is stored on a Blockchain and, when it must be deleted, simply destroy the encryption

key (van Geelkerken and Konings 2017; Jussila 2018; Pagallo et al. 2018; Michels 2018; Jensen 2018; Ibáñez et al. 2018). This concept assumes that with the use of state-of-the-art encryption techniques data gets inaccessible when the encryption key is not available anymore. Following their argumentation, this procedure comes close to the erasure of data.

In view of the GDPR, this concept must be seen critically. First, it is possible that today's encryption algorithms are no longer considered secure in the future so that it might be possible to decrypt the data without the knowledge of the original encryption key (Michels 2018; Ibáñez et al. 2018; Eichler et al. 2018). Second, Article 29 Data Protection Working Party (2014) clearly states that encryption must be considered as a form of pseudonymization and it does not automatically anonymize the data. It is important to notice that encryption only guarantees confidentiality over a certain time period but anonymization should last indefinitely (Article 29 Data Protection Working Party 2014).

In general, it can be said that the use of Encryption & Key destruction should not be considered as the main technique for a GDPR-compliant concept.

Evaluation

In the previous part, the different concepts from the literature were described and critically reflected. This section summarizes the results and provides a final evaluation.

The literature review revealed two main concepts of processing personal data on Blockchain in a GDPR-compliant way. Two additional, but not as strongly represented concepts, were also among the findings. Each of these concepts is for itself considered more or less suitable to fulfill the requirements of the GDPR.

The following Table 3 provides an overview of the concepts discussed in the paper at hand and summarizes the advantages and disadvantages of each concept.

It becomes clear that every concept described below has its own advantages and disadvantages. Currently, it seems that the Redactable Blockchain concept can fulfill the requirements of the GDPR because it enables the change of data directly on a Blockchain. On the downside, this contradicts the immutability of Blockchain, and it could be abused by participants in the network.

Off-Chain Storage may require the reintroduction of a TTP, which, in some aspects, contradicts the basic idea of the Blockchain technology. There are ways of avoiding a central storage location and thus a TTP, for example, by distributing the data among the participants in a Blockchain. However, this cannot always be feasible depending on the use-case. In view of privacy regulations, it seems to be possible to process personal data on a Blockchain with the Off-Chain Storage concepts when no personal data is stored on the blocks.

Table 3: Advantages and disadvantages of the identified concepts

Concept	Advantages	Disadvantages
Off-Chain Storage	<ul style="list-style-type: none"> - Personal data is stored off-chain - Only a reference to the off-chain storage location and the hash of the data is kept on-chain - Enables the processing of large data 	<ul style="list-style-type: none"> - May require the reintroduction of a TTP - Depending on the way personal data is stored off-chain, the hash on-chain must be handled as personal data - May require major technical modifications depending where the data is held
Redactable Blockchain	<ul style="list-style-type: none"> - Provides a way of altering and deleting personal data directly on Blockchain - Does not require a change in the way Blockchain is used 	<ul style="list-style-type: none"> - Requires technical adaptations - Redactability cannot be added to an existing Blockchain - If the secret key is lost, the Blockchain remains immutable - Nodes are required to delete old copies of the redacted blocks - Nodes can abuse this feature to their favor - May require off-chain storage for large data
Legal Argumentations	<ul style="list-style-type: none"> - No changes of the Blockchain technology required - Data which is obsolete can simply be invalidated with a new transaction 	<ul style="list-style-type: none"> - Concept must be handled very carefully because it operates in legal grey area - Judgment can declare this concept as not compliant with the GDPR - May require off-chain storage for large data
Encryption & Key Erasure	<ul style="list-style-type: none"> - No changes of the Blockchain technology required - Data on-chain is stored encrypted - When data needs to be removed from Blockchain the encryption key is simply destroyed 	<ul style="list-style-type: none"> - Should not be used as the main concept for GDPR-compliance - Encrypted personal data must be handled as pseudonymized data - It might be able to decrypt data in the future with the use of up-to-date technology - Judgment can declare this concept as not compliant with the GDPR - May require off-chain storage for large data

Besides privacy concerns, keeping data outside of the Blockchain network is often reasonable, otherwise, there might be a loss of performance when the datasets get larger. Therefore, Redactable Blockchain concepts might reach their limits quickly depending on the use case. In such situations, a combination of redactability and off-chain storage could make sense.

The concepts, summarized under the term “other”, are discussed to a lesser extent. The Legal Argumentations and the Encryption & Key Destruction concept both make use of interpretations of the GDPR which may be declared invalid by a court of law. Therefore, it seems reasonable to wait for judgments in this field before investing further research in this specific topic with the risk of a court ruling against these procedures.

CONCLUSION

The paper at hand conducted a structured literature review and identified four potential concepts for GDPR-compliant processing of personal data using Blockchain in academic literature and practitioner sources. A first analysis of the literature revealed a lack of scholarly research papers in this specific research field. Many of the sources (13; 46%) for this review did not undergo a peer-review process. A reason for this might be the rapid development of Blockchain technology which is mainly documented in internet sources as well as the recent legal validity of the GDPR.

The concept that is discussed the most in the reviewed literature is Off-Chain Storage. With this concept, personal data is stored outside of the Blockchain network and only references the off-chain location on Blockchain. The problem is that in most cases a “proof-of-correctness” in form of the hash of the data is kept on-chain and depending on the hashed data this could as well be personal data. This concept is already applied in practice.

The second identified concept, Redactable Blockchain, makes use of a special hashing algorithm, which allows data altering of a block without changing its corresponding hash. This enables altering and deleting data on Blockchain and is in view of the requirements of the GDPR a very interesting solution. On the downside, this concept is often criticized as a violation of the immutability of Blockchain, one of its fundamental characteristics.

The next concept, Legal Argumentations, makes use of interpretations of the GDPR. On the one hand, it can be argued that the right to be forgotten is not an absolute right and data on Blockchain could be altered or deleted by adding a new transaction to the block which declares the old one invalid. On the other hand, it could be possible to claim that consistent data processing is required for Blockchain to function properly. Therefore, no data can be altered or deleted. This argumentation can be based on Art. 17 (1) GDPR.

The last concept, Encryption & Key Erasure, describes a technique in which the personal data is encrypted before storing the data on Blockchain and the key is stored apart. If the data is not required anymore, the encryption key is

simply deleted, and the data is thereby not accessible anymore. The main critic of this method is that in the near future it might be possible to decrypt the data without the need of the decryption key and the data gets accessible again.

In general, it is currently strongly recommended to not store personal data on a Public Blockchain. If personal data was stored on a Public Blockchain, it could not be guaranteed that the data stays within the territorial scope of the GDPR and it might violate the principles of transferring personal data to third countries. The Legal Argumentation concept acts in grey area and so it is always possible that this method is declared illegal. The situation is similar to the concept of Encryption & Key Erasure. In this concept, the personal data stays encrypted on Blockchain, as long as the Blockchain exists. Even if the decryption key is destroyed, the data currently counts as personal data.

It seems that only the Off-chain Storage and the Redactable Blockchain concepts can guarantee compliant processing under data protection regulations. When using the Off-Chain Storage concept it should be paid attention that no recalculating of the hash value, which often stays on-chain as a "proof-of-correctness", is possible. If the Redactable Blockchain concept should be applied, it is important to manage the secret key very securely. If the key is lost, it is not possible to redact data stored on Blockchain.

Referring to the research question, the review revealed four possible concepts for GDPR-compliant processing of personal data in the literature. Based on the analysis in view of the GDPR, only the Off-Chain Storage and Redactable Blockchain concept can fulfill the data privacy requirements at the moment.

Based on the results of the present paper it is now possible for further researchers to develop Blockchain applications which allow processing of personal data in compliance with the GDPR.

Limiting factors of this literature review are that nearly half of the sources are not peer-reviewed and their significance might not meet all academic standards. As mentioned before the Blockchain technology is in rapid development and most of this knowledge is shared on internet platforms. Due to the research methodology, the paper at hand focused on research papers and less on online articles from developers or independent researchers. In addition, the paper did not examine the practical applicability of the identified concepts. Hence, it is possible that a concept cannot be realized with current state of the art technology.

REFERENCES

Article 29 Data Protection Working Party (2014): Opinion 05/2014 on anonymisation techniques. 0829/14/EN WP216. Edited by Article 29 Data Protection Working Party. Available online at <https://ec.europa.eu/justice/article->

29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, checked on 12/30/2018.

- Ateniese, Giuseppe; Magri, Bernardo; Venturi, Daniele; Andrade, Ewerton (2017): Redactable Blockchain – or – Rewriting history in Bitcoin and friends. In : 2017 IEEE European Symposium on Security and Privacy (EuroS&P). 2017 IEEE European Symposium on Security and Privacy (EuroS&P). Paris, France, 26.04.2017 - 28.04.2017: IEEE, pp. 111–126.
- Azaria, Asaph; Ekblaw, Ariel; Vieira, Thiago; Lippman, Andrew (2016): MedRec: Using Blockchain for medical data access and permission management. In : 2016 2nd International Conference on Open and Big Data (OBD). 2016 2nd International Conference on Open and Big Data (OBD). Vienna, Austria, 22.08.2016 - 24.08.2016: IEEE, pp. 25–30.
- Berberich, Matthias; Steiner, Malgorzata (2016): Blockchain technology and the GDPR – How to reconcile privacy and distributed ledgers? In *European Data Protection Law Review* 2 (3), pp. 422–426. DOI: 10.21552/EDPL/2016/3/21.
- Blum, Manuel (1983): Coin flipping by telephone a protocol for solving impossible problems. In *ACM SIGACT News* 15 (1), pp. 23–27.
- Cichosz, Simon Lebech; Stausholm, Mads Nibe; Kronborg, Thomas; Vestergaard, Peter; Hejlesen, Ole (2018): How to use Blockchain for diabetes health care data and access management: An operational concept. In *Journal of diabetes science and technology*. DOI: 10.1177/1932296818790281.
- CNIL (2018): Blockchain. Solutions for a responsible use of the Blockchain in the context of personal data. Available online at <https://www.cnil.fr/sites/default/files/atoms/files/blockchain.pdf>, checked on 12/28/2018.
- Dinh, Tien Tuan Anh; Liu, Rui; Zhang, Meihui; Chen, Gang; Ooi, Beng Chin; Wang, Ji (2018): Untangling Blockchain: A data processing view of blockchain systems. In *IEEE Trans. Knowl. Data Eng.* 30 (7), pp. 1366–1385. DOI: 10.1109/TKDE.2017.2781227.
- Drescher, Daniel (2017): Blockchain basics. A non-technical introduction in 25 steps. Place of publication not identified, New York, NY: Apress; Distributed to the book trade worldwide by Springer Science+Business Media New York. Available online at <http://proquest.tech.safaribooksonline.de/9781484226049>.
- Eberhardt, Jacob; Tai, Stefan (2017): On or off the Blockchain? Insights on off-chaining computation and data. In Flavio de Paoli, Stefan Schulte, Einar Broch Johnsen (Eds.): Service-oriented and cloud computing, vol. 10465. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 3–15.
- Eichler, Natalie.; Jongerius, Silvan; McMullen, Greg; Naegele, Oliver; Steininger, Liz; Wagner, Kai (2018): Blockchain, data protection, and the GDPR.

- Blockchain Bundesverband. Available online at http://www.bundesblock.de/wp-content/uploads/2018/05/GDPR_Position_Paper_v1.0.pdf.
- Esposito, Christian; Santis, Alfredo de; Tortora, Genny; Chang, Henry; Choo, Kim-Kwang Raymond (2018): Blockchain: A Panacea for healthcare cloud-based data security and privacy? In *IEEE Cloud Comput.* 5 (1), pp. 31–37. DOI: 10.1109/MCC.2018.011791712.
- Fan, Kai; Wang, Shangyang; Ren, Yanhui; Li, Hui; Yang, Yintang (2018): MedBlock: Efficient and secure medical data sharing via Blockchain. In *Journal of medical systems* 42 (8), p. 136. DOI: 10.1007/s10916-018-0993-7.
- Finck, Michèle (2018): Blockchains and data protection in the European Union. In *European Data Protection Law Review* 4 (1), pp. 17–35. DOI: 10.21552/edpl/2018/1/6.
- Gräther, Wolfgang; Kolvenbach, Sabine; Ruland, Rudolf; Schütte, Julian; Torres, Christof; Wendland, Florian (2018): Blockchain for education: Lifelong learning passport. In W. Prinz & P. Hoschka (Ed.): Proceedings of the 1st ERCIM Blockchain Workshop 2018, Reports of the European Society for Socially Embedded Technologies.
- Ibáñez, Luis-Daniel; O'Hara, Kieron; Simperl, Elena (2018): On Blockchains and the General Data Protection Regulation. Available online at https://eprints.soton.ac.uk/422879/1/Blockchain_GDPR_4.pdf.
- Jensen, Greg (2018): Reconciling GDPR rights to erasure and rectification of personal data with Blockchain. Available online at <https://blogs.oracle.com/cloudsecurity/reconciling-gdpr-rights-to-erasure-and-rectification-of-personal-data-with-blockchain>, checked on 12/28/2018.
- Jussila, Jani-Pekka (2018): Reconciling the conflict between the 'immutability' of public and permissionless Blockchain technology and the right to erasure under Article 17 of the General Data Protection Regulation.
- Katuwal, Gajendra J.; Pandey, Sandip; Hennessey, Mark; Lamichhane, Bishal (2018): Applications of Blockchain in healthcare: Current landscape & challenges. Available online at <http://arxiv.org/pdf/1812.02776v1>.
- Kuner, Christopher; Cate, Fred; Lynskey, Orla; Millard, Christopher; Ni Loideain, Nora; Svantesson, Dan (2018): Blockchain versus data protection. In *International Data Privacy Law* 8 (2), pp. 103–104. DOI: 10.1093/idpl/ipy009.
- Lima, Claudio (2018): Blockchain-GDPR privacy by design. How decentralized Blockchain Internet will comply with GDPR data privacy. Blockchain Engineering Council. Available online at <https://blockchain.ieee.org/images/files/pdf/blockchain-gdpr-privacy-by-design.pdf>, checked on 12/28/2018.
- Marnau, Ninja (2017): Die Blockchain im Spannungsfeld der Grundsätze der Datenschutzgrundverordnung. In Maximilian Eibl, Martin Gaedke (Eds.): Informatik 2017 - Bände I-III. Tagung vom 25.-29. September 2017 in Chemnitz. Bonn: Gesellschaft für Informatik (GI-Edition Proceedings, 275), pp. 1025–1036.
- Michels, Dave (2018): Can Blockchain operators comply with EU data protection law? Available online at <https://journal.binarydistrict.com/can-blockchain-operators-comply-with-eu-data-protection-law/>, checked on 12/28/2018.
- Moser, Jana (2017): The application and impact of the European General Data Protection Regulation on Blockchains. R3. Available online at https://www.r3.com/wp-content/uploads/2018/04/GDPR_Blockchains_R3.pdf, checked on 12/28/2018.
- Nakamoto, Satoshi (2008): Bitcoin: A peer-to-peer electronic cash system. Available online at <https://bitcoin.org/bitcoin.pdf>, updated on 1/10/2019.
- Omaar, Jamila (2017): Forever isn't free: The cost of storage on a Blockchain database. With assistance of Simon Schwerin, McMullen Greg. Available online at <https://medium.com/ipdb-blog/forever-isnt-free-the-cost-of-storage-on-a-blockchain-database-59003f63e01>, checked on 2/28/2019.
- Pagallo, Ugo; Bassi, Eleonora; Crepaldi, Marco; Durante, Massimo (2018): Chronicle of a clash foretold: Blockchains and the GDPR's right to erasure. In M. Palmirani (Ed.): Legal Knowledge and Information Systems: JURIX 2018: The Thirty-first Annual Conference: IOS Press, pp. 81–90.
- Salmensuu, Cagala (2018): The General Data Protection Regulation and Blockchains. University of Helsinki Law Faculty. Available online at <https://ssrn.com/abstract=3143992>.
- Steichen, Mathis; Fiz, Beltran; Norvill, Robert; Shbair, Wazen; State, Radu (2018): Blockchain-based, decentralized access control for IPFS. In : 2018 IEEE International Conference on Blockchain. Halifax, Canada.
- The European Union Blockchain Observatory & Forum (2018): Blockchain and the GDPR. The European Union Blockchain Observatory & Forum. Available online at https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf, checked on 12/28/2018.
- van Geelkerken, F.W.J.; Konings, K. (2017): Using Blockchain to strengthen the rights granted through the GDPR. In : 7th International youth science forum «Litteris et Artibus». Lviv, Ukraine, pp. 458–461.
- van Humbeeck, Andries (2017): The Blockchain-GDPR paradox. Available online at <https://medium.com/wearetheledger/the-blockchain-gdpr-paradox-fc51e663d047>, checked on 12/28/2018.
- Viriyasitavat, Wattana; Hoonsopon, Danupol (2018): Blockchain characteristics and consensus in modern business processes. In *Journal of Industrial*

- Vo, Hoang Tam; Kundu, Ashish; Mohania, Mukesh: Research Directions in Blockchain Data Management and Analytics.
- Webster, Jane; Watson, Richard T. (2002): Analyzing the past to prepare for the future: Writing a literature review. In *MIS quarterly*, pp. xiii–xxiii.
- Wirth, Christian; Kolain, Michael (2018): Privacy by BlockChain design: A Blockchain-enabled GDPR-compliant approach for handling personal data. In W. Prinz & P. Hoschka (Ed.): Proceedings of the 1st ERCIM Blockchain Workshop 2018, Reports of the European Society for Socially Embedded Technologies.
- Zhangy, Shifa; Kim, Anne; Liu, Dianbo; Nuckchady, Sandeep C.; Huangy, Lauren; Masurkary, Aditya et al. (2018): Genie: A secure, transparent sharing and services platform for genetic and health data. Available online at <http://arxiv.org/pdf/1811.01431v1>.
- Zheng, Zibin; Xie, Shaoan; Dai, Hongning; Chen, Xiangping; Wang, Huaimin (2017): An overview of Blockchain technology: Architecture, consensus, and future trends. In : 2017 IEEE International Congress on Big Data (BigData Congress). 2017 IEEE International Congress on Big Data (BigData Congress). Honolulu, HI, USA, 25.06.2017 - 30.06.2017: IEEE, pp. 557–564.
- Zyskind, Guy; Nathan, Oz; Pentland, Alex 'Sandy' (2015): Decentralizing privacy: Using Blockchain to protect personal data. In : 2015 IEEE Security and Privacy Workshops. 2015 IEEE Security and Privacy Workshops (SPW). San Jose, CA, 21.05.2015 - 22.05.2015: IEEE, pp. 180–184.

SAP CoPilot – Untersuchung des digitalen Assistenten hinsichtlich seiner Funktionsweise und Anwendungsmöglichkeiten

Silas Blieder	Prof. Dr. Harald Ritz	Tim Förster
advades GmbH	Technische Hochschule Mittelhessen	advades GmbH
SAP Development	Fachbereich Mathematik, Naturwissenschaften und Informatik	SAP Development
Kaiserstraße 7 35745 Herborn s.blieder@advades.com	Wiesenstraße 14 35390 Gießen harald.ritz@mni.thm.de	Kaiserstraße 7 35745 Herborn t.foerster@advades.com

Kategorie Masterarbeit

Schlüsselwörter

Digitale Assistenten, Chatbots, Conversational AI, NLP, SAP CoPilot, User Experience, Digitalisierung

Zusammenfassung

Die Masterarbeit wurde in Kooperation mit dem SAP-Partner advades GmbH in Herborn erarbeitet. Das Unternehmen ist spezialisiert auf neue SAP-Technologien und hat bei der Erstellung dieser Arbeit vor allem mit technischem Know-How und der erforderlichen IT-Infrastruktur unterstützt.

„Bots are like new applications, human language is the new UI layer. Digital assistants are meta apps, or like the new browsers. Intelligence is infused into all your interactions. That’s the rich platform that we have.“ Dieses Zitat stammt aus einer Präsentation des Microsoft-CEOs Satya Nadella im Jahr 2016. Frei übersetzt bedeutet es, dass Chatbots und digitale Assistenten, unterstützt durch künstliche Intelligenz, die zukünftige Form der Interaktion zwischen Mensch und Computer sind und bisherige Apps sowie Browser ersetzen werden.

Das Ziel dieser Arbeit war die Untersuchung von digitalen Assistenten. Dazu wurden Erkenntnisse über die Funktionsweise, die technischen Grundlagen und die grundlegenden Konzepte dargelegt. Der Fokus liegt außerdem auf den möglichen Anwendungsgebieten und Einsatzszenarien von digitalen Assistenten in Unternehmen. In diesem Bereich sind Chatbots und digitale Assistenten zwar ebenfalls im Vormarsch, aber noch lange nicht etabliert. Dabei steht im Mittelpunkt der digitale Assistent von SAP: CoPilot. Unternehmen stehen momentan vor den Herausforderungen der Digitalisierung. Digitale Assistenten sollen ihnen dabei helfen, Geschäftsprozesse zu digitalisieren und zu automatisieren. Für Unternehmen ist es vorteilhaft zu wissen, wie digitale Assistenten funktionieren und wie sie an die Unternehmensbedürfnisse angepasst werden können. Daher wurden in dieser Arbeit auch praktische Ansätze für die Erweiterung der Fähigkeiten von

digitalen Assistenten am Beispiel von SAP CoPilot betrachtet.

SAP hat mit CoPilot im Jahr 2016 einen digitalen Assistenten veröffentlicht, der eine neue Art der User Experience für Unternehmensanwendungen bieten soll. Der Anwender kann mit SAP CoPilot über natürliche Sprache interagieren, ähnliche wie mit Amazon Alexa, Apple Siri, Google Assistant und Co. SAP möchte jedoch nicht mit diesen Produkten in Konkurrenz treten, sondern setzt bewusst auf die Anforderungen von Geschäftsanwendungen und die Integration in die entsprechenden Geschäftsprozesse. SAP CoPilot ist ein komplexes Softwaresystem, welches Algorithmen des maschinellen Lernens (Machine Learning, ML) für die Verarbeitung von natürlicher Sprache (Natural Language Processing, NLP) nutzt. Eine Schlüsselfunktion von SAP CoPilot ist die Erweiterbarkeit um zusätzliche Funktionalitäten, Dadurch können Unternehmen ihre Geschäftsprozesse mit dem digitalen Assistenten abbilden.

Im praktischen Teil der Masterarbeit wurden SAP CoPilot mit eigenentwickelten „Skills“ erweitert. Dabei hat sich herausgestellt, dass vor allem eher einfache Verwaltungsprozesse wie z.B. Employee-Self-Services (u.a. Urlaubsanträge stellen) das Potenzial haben, mit digitalen Assistenten implementiert zu werden. Komplexe Prozesse mit vielen Eingabefeldern und mehreren Prozessschritten eignen sich eher weniger für die Umsetzung. Das liegt u.a. daran, dass der Anwender keinen genauen Überblick über den Prozessverlauf und die bereits eingegebenen Daten hat.

Die zentrale Erkenntnis ist, dass letztendlich die Art des Prozesses darüber entscheidet, welche Benutzungsoberfläche sich am besten für dessen Ausführung eignet. Nicht immer ist ein digitaler Assistent die beste Wahl, es kommt auf den konkreten Anwendungsfall an. An den grundlegenden Konzepten von Chatbots und digitalen Assistenten hat sich seit Jahrzehnten nicht viel verändert: Ein digitaler Assistent (bzw. Chatbot) nimmt Befehle entgegen und führt Aktionen aus oder sucht nach Informationen im Internet. Die nächste Entwicklungsstufe ist ein digitaler Assistent, der auch Entscheidungen treffen kann – wenn die entsprechende Rahmenbedingungen für dessen Handeln gesetzt wurden.

Implementierung eines Scoringverfahrens zur Bewertung der gegenwärtigen IT-Sicherheitslage im Unternehmen

Philipp Korzendorfer

Technische Hochschule
Mittelhessen

Fachbereich Mathematik,
Naturwissenschaften und
Datenverarbeitung
Wilhelm-Leuschner-Str. 13
61169 Friedberg (Hessen)
philipp.korzendoerfer@mnd.
thm.de

Prof. Dr. Harald Ritz

Technische Hochschule
Mittelhessen

Fachbereich Mathematik,
Naturwissenschaften und
Informatik
Wiesenstraße 14
35390 Gießen
harald.ritz@mni.thm.de

Prof. Dr. Frank Kammer

Technische Hochschule
Mittelhessen

Fachbereich Mathematik,
Naturwissenschaften und
Informatik
Wiesenstraße 14
35390 Gießen
frank.kammer@mni.thm.de

Kategorie

Bachelorarbeit

Schlüsselwörter

Scoringverfahren, IT-Sicherheit,
Sicherheitsbedrohungen, Schutzmaßnahmen

Als zentrales Ergebnis ergibt sich, dass in einigen Monate eine signifikante Erhöhung des Scorewertes sichtbar ist, dessen Klärung des Ursprungs in der Verantwortung der IT-Sicherheitsspezialisten liegt.

Zusammenfassung

Für Personen, ohne Fachkenntnisse aus dem Bereich der IT, sind operative Kennzahlen von beispielsweise Antiviruslösungen nur schwer zu deuten.

Die vorliegende Bachelorarbeit, die in einem mittelständischen Unternehmen in Hessen durchgeführt wurde, dient der Entwicklung und Implementierung eines Scoringverfahrens zur Bewertung der gegenwärtigen IT-Sicherheitslage in einem Technologieunternehmen.

Der zu entwickelnde Scoringwert fasst dabei alle IT-sicherheitsrelevanten Kennzahlen zu einem Wert zusammen und verschafft somit einen Überblick über die gesamte Bedrohungslage. Im Vordergrund steht dabei die Optimierung der IT-Sicherheitsüberwachung, um u.a. bislang unerkannte Bedrohungen schneller detektieren und bekämpfen zu können. Vergleichbar ist diese Kennzahl mit einem Aktienindex, der hingegen die wirtschaftliche Lage eines Konzerns widerspiegelt.

Zur Berechnung des Scoringwertes werden im Verlaufe dieser Arbeit, Verfahren aus dem Finanzwesen sowie statistische Verfahren herangezogen.

Anhand der daraus entstehenden Grafiken lassen sich Trendverläufe über einen Zeitraum erkennen, die auf ihre entsprechenden Ursachen hin analysiert werden können.

Analyse der Nutzenaspekte einer BI-as-a-Service-Lösung im Vergleich zu individuellen Software-as-a-Service-Komponenten - Untersucht am Beispiel eines Personalmarketing-Dashboards

Isabel Schlund

Technische Hochschule
Mittelhessen

Fachbereich MND
Wilhelm-Leuschner-Straße 13
61169 Friedberg
E-Mail:
isabel.schlund@mnd.thm.de

Prof. Dr. Harald Ritz

Technische Hochschule
Mittelhessen

Fachbereich MNI
Wiesenstraße 14
35390 Gießen
E-Mail:
harald.ritz@mni.thm.de

Arthur Arendt

OPITZ CONSULTING
Deutschland GmbH

Norsk-Data-Straße 3
61352 Bad Homburg v. d. H.
E-Mail:
arthur.arendt@opitz-
consulting.com

Kategorie

Bachelorarbeit

Schlüsselwörter

Business Intelligence, Cloud-BI, Software as a Service, Business Intelligence as a Service, Microsoft Azure, Domo

Zusammenfassung

Klassische Data-Warehouse- und Business-Intelligence-Systeme, basierend auf On-Premises-Architekturen, verlieren mehr und mehr an Bedeutung. Die Hersteller solcher Anwendungen setzen zunehmend auf cloudbasierte Technologien, neue Anbieter drängen auf den Markt, etablierte Hersteller stellen entsprechende Lösungen bereit. Dabei geht die Spannweite von individuell kombinierbaren Software-as-a-Service-Komponenten bis hin zu cloudbasierten BI-Gesamtlösungen.

Es stellen sich verschiedene Fragen: Ist der Betrieb von BI in der Cloud aus technischer und funktionaler Sicht heute bereits vollständig möglich und sinnvoll? Können altbekannten Vorgehensweisen beibehalten werden? Gibt es Alternativen? Möglicherweise sogar Verbesserungen? Darüber hinaus kommt die zentrale Frage auf, ob durch das Beziehen von BI als Cloud-Service, sei es in Form von unabhängigen SaaS-Komponenten oder einer vollständigen BIaaS-Lösung, ein Mehrwert im Vergleich zu bewährten On-Premises-Produkten entsteht.

Als Ziel dieser Arbeit soll eine begründete Aussage getroffen werden, welcher Ansatz für den spezifischen Fall der Entwicklung eines Dashboards für das Personalmarketing zu empfehlen ist, inwiefern bei der Nutzung einer Cloud-BI-Lösung ein Mehrwert entsteht und durch welche Kriterien die untersuchten Anwendungen dabei überzeugen können.

Beide Ansätze werden dazu gegenübergestellt und anhand von ausgewählten Nutzenaspekten im Rahmen

des vorliegenden Anwendungsfalles verglichen. Die betrachteten Kriterien werden in ausgewählten Beispielen anschaulich dargestellt, um einen Überblick über die Möglichkeiten der beiden Lösungen zu verschaffen.

Der Fokus liegt dabei auf der Implementierung des Backends. Dazu gehören u.a. die Datenspeicherung und die für die Durchführung des Extraktions-, Transformations- und Ladeprozesses notwendigen funktionalen Aspekte. Darüber hinaus werden ebenfalls einige Punkte rund um die Visualisierung, Handhabung, Performance sowie das zugrunde liegende Kostenmodell betrachtet, da diese in der Gesamtbetrachtung durchaus auch von Bedeutung sind.

Der Mehrwert einer cloudbasierten BI-Lösung macht sich hauptsächlich in Form von eingesparten Kosten, verkürzten Entwicklungszeiten, erleichterten Arbeitsschritten oder einer benutzerfreundlicheren Handhabung bemerkbar. Um die vielfältigen Aspekte zu bewerten und übersichtlich darzustellen, wird eine Nutzwertanalyse durchgeführt.

Beide Cloud-BI-Ansätze können die Anforderungen des vorliegenden Anwendungsfalles zufriedenstellend erfüllen. Während die SaaS-Lösung ihren Fokus auf einen ausführlichen Funktionsumfang legt, konzentriert sich die BIaaS-Lösung sehr auf benutzerfreundliche Handhabung.

Insbesondere aufgrund des Kostenaspektes sollten hier individuell zusammenstellbare SaaS-Komponenten gewählt werden. Eine vollständig abgerundete BIaaS-Lösung hat aus funktionaler Sicht jedoch großes Potential für Anwendungsfälle mit vergleichbaren Anforderungen. Daher ist an dieser Stelle die Betrachtung weiterer Produkte empfehlenswert.

Sei es eine BIaaS- oder eine individuelle SaaS-Lösung – cloudbasierte BI-Anwendungen stellen durchaus eine Alternative zu herkömmlichen Ansätzen dar.