Leveraging Edge Computing and Differential Privacy to Securely Enable Industrial Cloud Collaboration Along the Value Chain

Alexander Giehl[1], Michael P. Heinl[1], Maximilian Busch[2]

***Abstract*— Big data continues to grow in the manufacturing domain due to increasing interconnectivity on the shop floor in the course of the fourth industrial revolution. The optimization of machines based on either real-time or historical machine data provides benefits to both machine producers and operators. In order to be able to make use of these opportunities, it is necessary to access the machine data, which can include sensitive information such as intellectual property.**

Employing the use case of machine tools, this paper presents a solution enabling industrial data sharing and cloud collaboration while protecting sensitive information. It employs the edge computing paradigm to apply differential privacy to machine data in order to protect sensitive information and simultaneously allow machine producers to perform the necessary calculations and analyses using this data.

***Index Terms*— Anonymization, chatter analysis, edge computing**

I. INTRODUCTION

Due to the increasing digitization on the shop floor, the amount of data produced in manufacturing constantly increases. In order to unfold the full potential of this trend, data needs to be shared between multiple stakeholders of the value chain. Modern machine tools for example produce a variety of data, not only valuable for machine operators, but also for machine manufacturers. Using this information, the manufacturers can analyze the machines' behavior in various setups and over long periods of time. This allows identifying weaknesses and using the insights to optimize future series of machines.

Although machine tool operators benefit from these optimizations, they hesitate to share their data with third parties due to concerns regarding the loss of intellectual property [1, 2]. This dilemma is underpinned by the increasing emergence of third-party risks, such as attacks on the supply and value chain respectively [3, 4]. While an established trust relationship between the machine operator and the machine manufacturer on the one hand is a necessity, this relationship on the other hand also has been evolved to become a serious attack vector leveraged by malicious actors. In order to mitigate this dilemma, the usage of edge computing in combination with differential privacy (DP) is proposed. The solution consists of different edge and cloud components which have been practically tested in a lab environment with a conventional machine tool.

Originally conceived to mitigate the tension between the gain of knowledge and the protection of personally identifiable information (PII), DP can also be applied to use cases other than those involving PII [5]. DP allows privacy-preserving data analytics by adding noise to data sets to provide a specific level of privacy while maintaining a certain level of meaningfulness to be able to detect patterns and derive statistical conclusions [6]. The proposed solution leverages these characteristics to conceal information potentially able to leak intellectual property while providing enough significance to enable machine producers to identify their machines' weaknesses and potential for optimization.

Edge computing provides the advantage that data obtained from machines can be preprocessed within the company boundaries. This is important, since raw machine data has to be protected against access of third parties as it can contain intellectual property. In addition, edge computing provides the power to apply computationally intensive cryptographic algorithms like DP to the gathered data. In this way, the data can be secured for both the transmission over an insecure channel, such as the Internet, as well as the utilization by third parties, such as the machine tool producer.

In order to enable machine tool users and producers to make use of the data, a privacy-preserving edge computing framework is proposed as this work's major contribution. The framework enables vendor-agnostic collaboration along the industrial value chain [7]. It uses a novel measure for assessing the criticality of industrial data. Furthermore, it enables machine tool users to retain full control over the data after the transfer from the edge to the cloud. Novel privacy models and algorithms were used and a fully implemented software stack was implemented. The approach was validated considering the use case of a cloud-based stability analysis, which enables machine tool operators to continually assess the process quality and enables the machine manufacturers to assess possible deficits of their machine design.

The proposed solution was specifically developed for machine tools but can generally be applied to a wide variety of cyber-physical systems (CPS). For the sake of tangibility, this paper remains with the practically implemented and tested example of machine tools.

[1]Alexander Giehl and Michael P. Heinl are with the Technical University of Munich, Department of Informatics, and with the Fraunhofer Research Institute for Applied and Integrated Security AISEC, Department of Product Protection and Industrial Security, both Germany. {alexander.giehl,michael.heinl}@aisec.fraunhofer.de

[2]Maximilian Busch is with the Institute for Machine Tools and Industrial Management (*iwb*), Technical University of Munich, Germany. maximilian.busch@iwb.tum.de

The presented work has been funded by the German Federation of Industrial Research Associations (*Arbeitsgemeinschaft industrieller Forschungsvereinigungen* [AiF]) under grant number 20449 N.

 DOI: 10.1109/CASE49439.2021.9551656

II. RELATED WORK

In [1], an edge computing based scheme for industrial ecosystems was presented with an included privacy model. An evaluation was performed on the given use case of the chatter stability analysis as well, whereas the effects of the privacy model on the machine tool data are not discussed. A secure connection is available, however, the machine operators lack control of the released data after its transfer from the edge over the company boundary into the cloud. Privacy-preserving measures are generally not considered in industrial collaboration schemes [8]. However, applications of DP [5] and privacy in general [9] to industrial data sets are discussed in literature without edge computing schemes.

Furthermore, no criticality measure has been established for industrial data and information streams yet. Typical criticality assessments for general information sources which are applied resemble confidentiality classes such as *Secret*, *Confidential*, *Restricted*, or *Unclassified*. Usually, the classification is based on the impact of unauthorized disclosure of the data or information. More advanced privacy classification schemes for personal information classes are proposed by [10] and [11].

III. METHODOLOGY

In this section, the approach aiming at enabling industrial cloud collaboration of machine data is described in detail. The approach provides useful data analytics tools for both user-side monitoring (e.g. for improving of the machining process) and manufacturer-side monitoring (e.g. for improving the machine performance over its lifetime). Nonetheless, the needed information for such analyses can include sensible or critical data for the machine operator.

The approach aims to close this gap by

1) anonymizing critical data streams and
2) ensuring the data sovereignty of the operator's data in the cloud.

For this purpose, the collected machine data needs to be categorized according to the sensitivity and criticality of the implied information first. Afterwards, the anonymization process and the method to enable data sovereignty for the user data in the cloud can be applied.

A. Criticality Analysis

Currently, there are no criticality measures available for the implications in machine and process data, such as sensitive information regarding the machining process. Thus, a new criticality measure was developed which is based on a criticality measure of algorithmic systems published by the *Data Ethics Commission of the Federal Government* [12]. This criticality measure links the criticality of a data set to the potential damage, if this data set would be published without the consent of the data owner (cf. Figure 1). Such a criticality assessment ranges from the least critical Level 1 (*"no or low potential damage"*) up to the most critical Level 5 (*"unacceptable potential damage"*). The criticality assessment can be further utilized as a default level of anonymization as described in the following section.

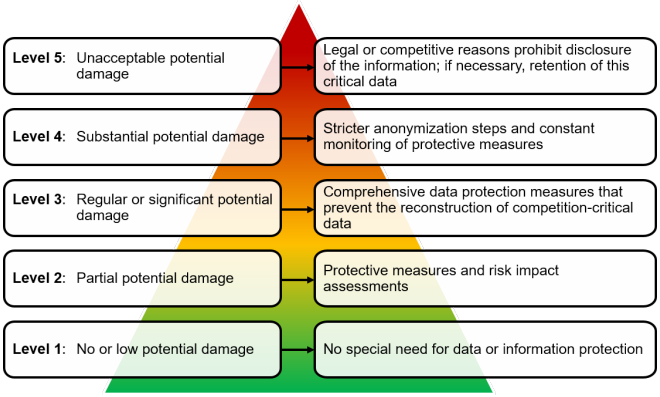

Fig. 1. Criticality measure, depending on the potential damage in case of uncontrolled publication of the data (based on [12]).

B. Anomymization

As a privacy model, DP was chosen. The usage of DP is motivated by an initial requirements analysis [1]. For the presented application, DP offers the advantage that it can be used locally by a data producer, allowing full control over the data before leaving the company's premises. DP aimes at satisfying the Fundamental Law of Data Recovery [6]. Roughly, it states that overly accurate answers to a large number of questions makes any privacy mechanism obsolete. Attackers with sufficient background knowledge about a target (located in a given data set) are able to break any privacy mechanism imposed to protect the target. DP allows the target though to deny its own entry in the data set. Thus, DP works as a formal definition of the principle of plausible deniability.

The definition of DP is given by Equation 1 [13]:

$$\left| ln \frac{P[R(T_1 = U)]}{P[R(T_2 = U)]} \right| \leq \epsilon. \tag{1}$$

T_1 and T_2 are data sets differing in exactly one record. R is a random function (depending on the algorithm, see below) with U being a subspace of possible results of R. The parameter ϵ is the adjustable value in the privacy model of DP allowing for configuration. The left side of Equation 1 quantifies the difference between two (almost identical) data sets T_1 and T_2. This difference is expressed by ϵ that bounds the level of noise added to T_1 and T_2 by R. It is worth noting that DP is a perturbative privacy model, meaning that the original data is modified [14].

It is worth pointing out that DP is a definition rather than an algorithm [6]. This means that for every deployment of DP, a suitable algorithm needs to be chosen. For this solution, an algorithm based on the Laplace distribution is chosen. It offers the advantage of only introducing one additional parameter S for the configuration of the privacy model. S donates the sensitivity of the Laplacian DP algorithm and captures the magnitude by which a single data entry in a dataset can be changed by the randomized statistical distribution.

The parameters ϵ and S, thus, describe the level of perturbation from the original data to the anonymized data. For further reference, a parameter configuration $p = (\epsilon, S)$ is defined. A configuration p is applied to a data set and provides the parameters for privacy-preserving computation of machine data.

C. Data Sovereignty

Even if data is protected by the presented anonymization techniques and state-of-the-art security measures, machine tool operators can still have different requirements regarding critical data. Furthermore, they might decide that they do not want to share already uploaded historical data anymore. The concept of data sovereignty [15] ensures that they have full control over their data at all points in time. In the presented implementation, this is realized by the possibility of fine-grained data sharing configurations via a graphical user interface (cf. Section V-A) as well as the option to remove data based on public/private key pairs which are used as pseudonyms for the operators' machine tools (cf. Section V-C).

IV. ENVIRONMENT

The target infrastructure can be roughly divided into three parts, namely the data and information sources at the *machine tool*, the *edge computing infrastructure*, and the *cloud computing infrastructure* (cf. Figure 2). The following sections further describe each of these these domains as well as their corresponding components.

A. Machine Tool

In order to provide reliable stability analyses, multiple data and information sources at the machine are considered. The data to assess the process stability can be categorized into the four groups presented in Table I.

TABLE I
DATA SETS TO ASSESS THE MACHINING PROCESS STABILITY.

Type	**Data**	**Source**
Discrete	Tool ID	internal (OPC UA)
Semantic	NC program name	internal (OPC UA)
Continuous	X / Y / Z position, spindle speed, torque	internal (OPC UA)
Continuous	Acceleration, sound pressure	external (sensors)

B. Edge Computing Infrastructure

Once the data has been gathered from the machine tool, it is processed by an edge computing device in the form of an industrial personal computer (IPC). The proposed solution was tested with the operating systems Debian GNU/Linux and Microsoft Windows. This demonstrates the system's interoperability with two of the most popular operating systems, helping to avoid a vendor lock-in.

The software managing the connections with the machine's OPC UA server as well as the sensors runs on this IPC. The IPC also contains software components for data processing and secure transfer to the cloud infrastructure as well as a graphical user interface to configure the system. Additionally to sending data to the cloud, the user can also choose to keep a copy on the IPC for their own local analyses and to independently trace what has been transmitted to the cloud. Combining all these functionalities, the IPC constitutes the experimental setup's edge infrastructure.

C. Cloud Computing Infrastructure

The cloud computing infrastructure consists of a Debian GNU/Linux virtual server. It gathers the encrypted data sent by the IPC by providing a representational state transfer (REST) interface. Once received, this interface forwards the data to an Elasticsearch instance indexing the data and providing it to a Kibana instance, which is in turn responsible for the visualization and analysis on the user side. Both Elasticsearch and Kibana run on the same server and are virtualized employing corresponding Docker containers using volumes for data persistence.

V. IMPLEMENTATION

As visualized in Figure 2, the environment presented above is used to run the applications discussed in the following. The communication between the applications is realized by employing the JavaScript Object Notation (JSON) data format if not specified otherwise.

A. Anonymization Manager

In order to enable the machine tool operator to maintain the data sovereignty, the *AnonymizationManager* (AM) had been developed. The AM is written in C++, runs on the IPC and allows controlling the data flow towards the cloud platform. Employing the AM, the data selection and parameterization of the anonymization of the machine tool data can be performed. The GUI of the AM is drag-and-drop enabled to allow for an efficient usage on the shop floor by machine operators.

B. Chatter Analysis Module

In order to provide a reliable stability analysis, an existing chatter detection method is used [16]. By applying this method, the autocorrelation signal of either an acceleration sensor or a microphone is used to detect a process instability. In case of a stable machining process, the autocorrelation of the sensor signal is close to 1, whereas the autocorrelation coefficient drops below a user defined threshold in case the machining process becomes unstable. Figure 3 illustrates the autocorrelation coefficient for such a machining process considering both signals (the microphone's audio signal and the accelerometer's signal). It is worth mentioning, that the autocorrelation coefficient can only quantify the process stability during the tool engagement. In Figure 3, this is the time between $6\,s \leq t \leq 20.0\,s$. As shown in Figure 3, the microphone's audio signal contains more noise than the accelerometer's signal.

Additionally, the autocorrelation coefficient value is continuously monitored based on the *Infinite Horizon Gaussian Process* (IHGP) algorithm [17]. As the IHGP algorithm

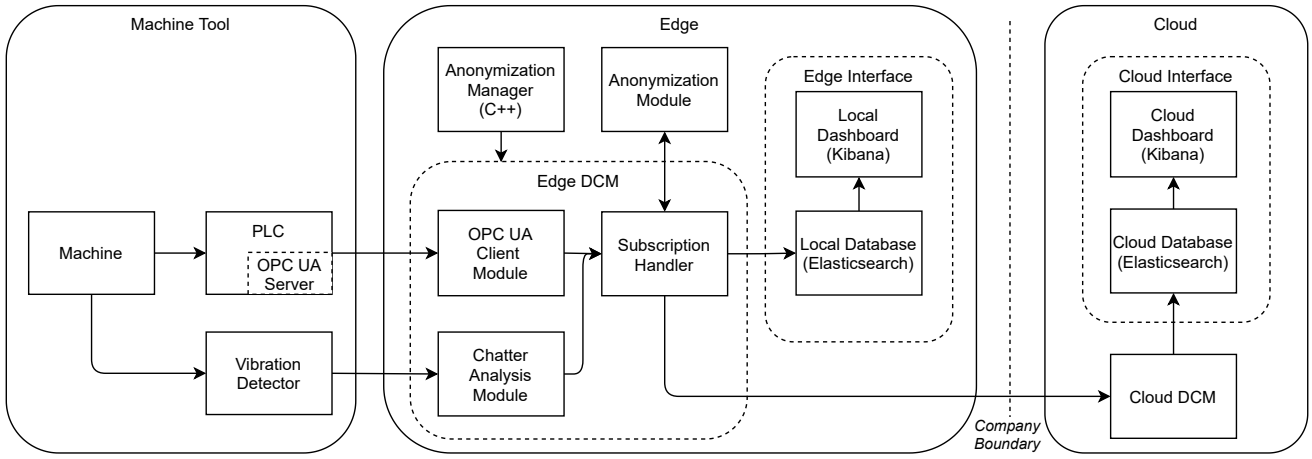

Fig. 2. Architecture comprising the machine tool (left), the edge computing framework (middle) and the cloud infrastructure (right).

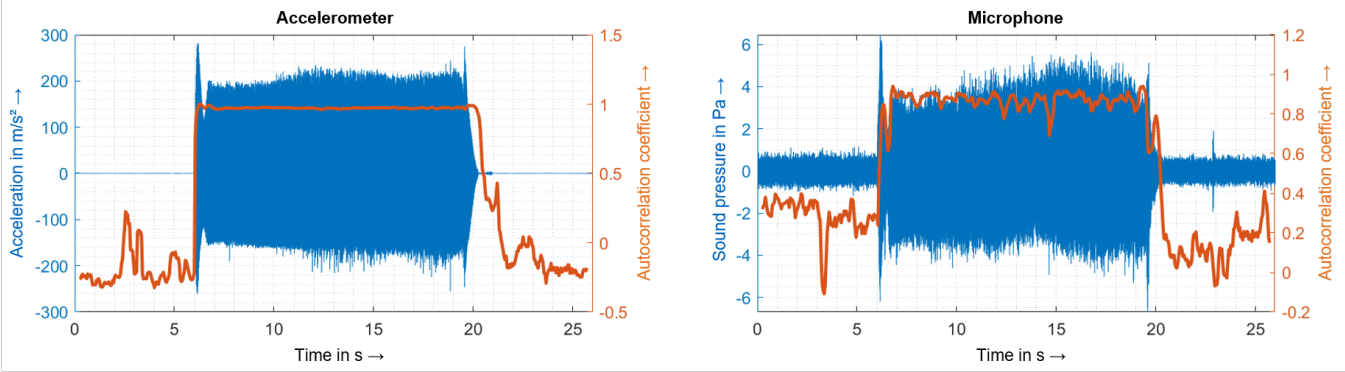

Fig. 3. Comparison between the two signals and their corresponding autocorrelation coefficient as a measure of the process stability.

is based on real-time capable Gaussian process regression, the algorithm provides an uncertainty estimation of the autocorrelation coefficient as well. For efficiency reasons, the algorithm is implemented in C++, based on [17].

It is worth highlighting that the autocorrelation signal itself is not anonymized, as this signal does not contain any sensitive information on the machining process.

C. Data Control Module

The Data Control Module (DCM) is the system's focal point coordinating all the other components. It comprises an edge and a cloud component, both implemented in Python 3.

The Edge DCM runs on the IPC and is responsible for the coordination of communication between the machine tool, the cloud, as well as other components running on the IPC.

As an input, it takes a JSON based configuration file generated by the AM. Among other, this configuration file contains information such as IP addresses, ports, and credentials necessary to subscribe to the UPC UA server and to connect to the local Elasticsearch instance as well as to the Cloud DCM. Furthermore, it contains information about which OPC UA variables to subscribe for and whether these variables have to be anonymized before sending them to the cloud. This is the result of the mapping done by the user employing the AM.

Each machine tool possesses its own unique public/private key pair to sign data before the transmission. The combination of a public key and a signature sent with each datagram serves multiple purposes. First, it ensures that data has not been tampered with during the transmission. Second, it provides a pseudonym for the machine tool so that cohesive data can be identified by the cloud. By utilizing a newly generated public/private key pair, a new pseudonym theoretically unlinkable to the former one can be generated by the machine tool operators. Third, issuing a respective deletion datagram signed by the corresponding private key, the machine tool operator is able to delete all data related to this pseudonym.

The Cloud DCM's main task is to verify incoming requests sent by the Edge DCM and forward them to the Elasticsearch and Kibana stack running in the cloud. In case of an ordinary datagram, its signature is checked using the corresponding public key. If the verification is successful, the data is written into the database. In case of a deletion datagram, an additional timestamp verification is applied in order to prevent replay attacks. Once the signature as well as the timestamp are verified, the entire data record related to this very public key is removed from the database.

D. Anonymization Module

The Anonymization Module takes raw data provided by the Edge DCM as input. If the data has been configured to be anonymized, the Anonymization Module applies the DP algorithm to this data. Subsequently, it passes the data back to the Edge DCM's subscription handler which in turn transmits it to the Cloud DCM using a connection secured by Transport Layer Security (TLS).

E. Edge & Cloud User Interfaces

Elasticsearch and Kibana run on both the edge as well as the cloud device (cf. Figure 2). While Elasticsearch is used for data storage and indexing, Kibana provides the actual user interface enabling data visualization and analyses. Besides employing the most recent Docker images at the time of the implementation, additional security measures provided by Elasticsearch's X-Pack are employed in order to protect both data in transit as well as data in rest. These measures include identity management in the form of role-based access control (RBAC), TLS for the connection between Elasticsearch and Kibana instances, as well as front and back end transmission via encrypted and server-side authenticated Hypertext Transfer Protocol Secure (HTTPS). As an additional line of defense, devices used to access these interfaces should only have trustworthy digital certificates installed in their browsers for the purpose of verification [18, 19].

VI. EVALUATION

For most applications, the anonymization is a two-step process [20]. The first step is the de-identification, i.e. the removal of any directly identifying data fields. This means that discrete and semantic data, such as the variables presented in Table I, are replaced. The continuous data fields are then used in the next step. This second step is the actual anonymization. For this, the presented privacy model DP with an implementation of the Laplace distribution based on [21] is applied to the machine data.

In order to find a suitable parameter configuration $p = (\epsilon, S)$, multiple experiments were conducted using the data of a productive machine tool (GROB G350). The variance, which is induced by the randomization of DP, needs to be handled for usable data analytics whilst still preserving privacy.

Figure 4 shows the maximum deviation from 100 anonymization operations on the original machine data value for different sets of p. The deviation is measured absolutely and depicted as a heatmap with darker fields experiencing less deviation than brighter fields. There is a low deviation in the quadrant located at the lower right corner (large values for ϵ, small values for S) and a high deviation in the top left corner (small ϵ, large S). Thus, a balanced set with $p_b = (0.5; 0.5)$ is chosen for the following analysis.

The practical anonymization capabilities are evaluated for spatial data points of the toolpath. Figure 5 shows the original data set and the corresponding toolpath (spelling out the letters *VDW*) in black on the left. The anonymized data

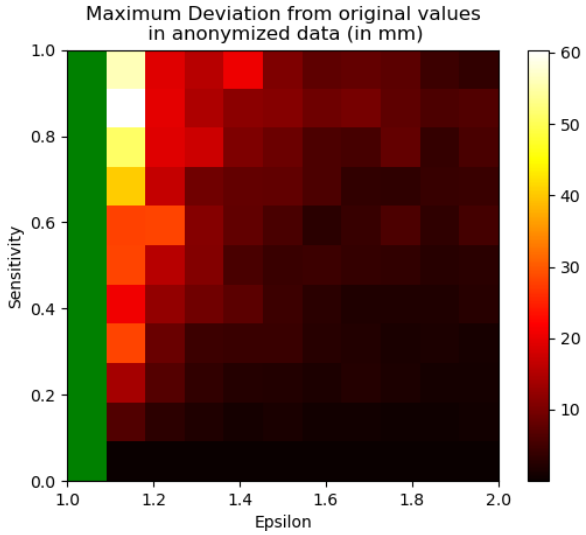

Fig. 4. Maximum deviation of the anonymization from the real data values (depicted over ranges $\epsilon = [0; 1]$ and $S = [0; 1]$). Green fields show numerically unstable sets of p.

points accessible by the cloud are shown in red. A detailed view on the left shows the anonmyization in effect. An attacker cannot distinguish the correct polynomial order of the anonymized data points without meeting strong assumptions. Thus, it can be concluded, that a position-dependent chatter analysis without reconstruction of the exact toolpath is possible.

VII. CONCLUSION

In this paper, an approach to mitigate the dilemma of cloud-based data processing of potentially sensitive machine tool data was presented. For this purpose, an edge and cloud based infrastructure has been designed and implemented. On the edge device, real-time machine data is captured, analyzed, and anonymized based on its criticality. In order to support the user in the configuration process, a scale to assess the data criticality was developed. The practicability of this setup was tested using a productive machine tool with a stability analysis as use case.

VIII. FUTURE WORK

The approach to ensure data sovereignty can be enriched by cryptographic guarantees provided by trusted execution environments [22] employed in the cloud. Also, integration into existing manufacturing processes needs to be considered [23].

The process for estimating the anonymization parameters can be further extended. Possible approaches are related to the automated estimation of anonymization parameters [24] or using the inherent noise in a given data set [25].

While the presented approach is located in the context of a machine tool use case, its application to other domains that share similar properties, e.g. real-time requirements or extensive usage of embedded systems, is encouraged. This includes, but is not limited to, smart energy networks, building automation, or medical technologies.

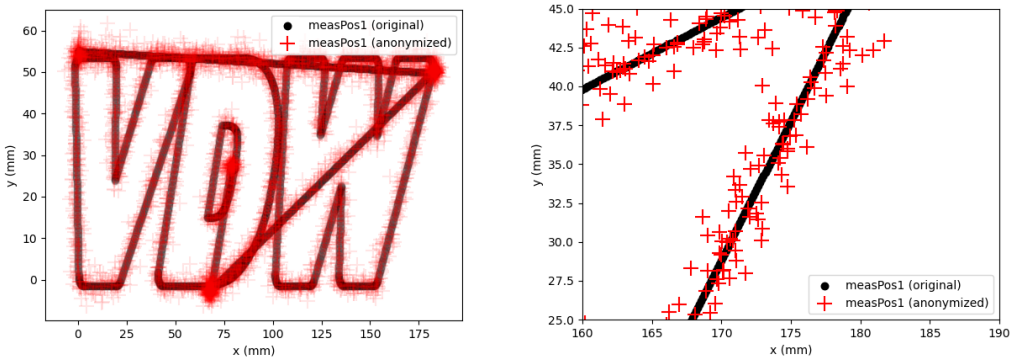

Fig. 5. Comparison between original spatial data and anonymized spatial data (left) and detailed view (right) with parameters $\epsilon = 0.5$ and $S = 0.5$.

REFERENCES

[1] A. Giehl, P. Schneider, M. Busch, F. Schnoes, R. Kleinwort, and M. F. Zaeh, "Edge-computing enhanced privacy protection for industrial ecosystems in the context of SMEs", in *12th CMI Conference 2019*, Copenhagen, Denmark: IEEE, 2019.

[2] J. Mahler, B. Hou, S. Niyaz, F. T. Pokorny, R. Chandra, and K. Goldberg, "Privacy-Preserving Cloud-Based Grasp Planning in the Cloud", in *IEEE 12th International Conference on Automation Science and Engineering Proceedings*, Dallas, TX, USA: IEEE, 2016, pp. 1–8.

[3] European Union Agency for Network and Information Security (ENISA), *ENISA Threat Landscape: Emerging Trends*, 2020.

[4] Ponemon Institute, *Data Risk in the Third-Party Ecosystem: Third Annual Study*, 2018.

[5] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey", *arXiv preprint arXiv:1812.02282*, 2018.

[6] C. Dwork, A. Roth, *et al.*, "The algorithmic foundations of differential privacy", *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[7] A. Giehl and N. Wiedermann, "Security verification of third party design files in manufacturing", in *10th International Conference on Computer and Automation Engineering Proceedings*, Brisbane, Australia: ACM, 2018.

[8] Y. Wang, O. Anokhin, and R. Anderl, "Concept and use case driven approach for mapping it security requirements on system assets and processes in industrie 4.0", *Procedia CIRP*, vol. 63, pp. 207–212, 2017.

[9] A. Fahad, Z. Tari, A. Almalawi, A. Goscinski, I. Khalil, and A. Mahmood, "Ppfscada: Privacy preserving framework for scada data publishing", *Future generation computer systems*, vol. 37, pp. 496–511, 2014.

[10] I. A. Goldberg, "A Pseudonymous Communications Infrastructure for the Internet", Ph.D. Thesis, University of California at Berkeley, 2000.

[11] X. Lu, Z. Qu, Q. Li, and P. Hui, "Privacy information security classification for internet of things based on internet data", *International Journal of Distributed Sensor Networks*, vol. 2015, 2015.

[12] B. Schmucker, M. Busch, R. Kleinwort, and M. Zäh, "Edge and cloud-based process monitoring and control: System identification and control of machine tools under consideration of data protection", *WT Werkstattstechnik*, vol. 110, no. 3, pp. 113–118, 2020.

[13] C. Dwork, "Differential privacy", *Encyclopedia of Cryptography and Security*, pp. 338–340, 2011.

[14] R. Chen, B. C. Fung, N. Mohammed, B. C. Desai, and K. Wang, "Privacy-preserving trajectory data publishing by local suppression", *Information Sciences*, vol. 231, pp. 83–97, 2013.

[15] M. Jarke, "Data sovereignty and the internet of production", in *Advanced Information Systems Engineering*, S. Dustdar, E. Yu, C. Salinesi, D. Rieu, and V. Pant, Eds., Cham: Springer International Publishing, 2020, pp. 549–558.

[16] M. Zaeh, F. Schnoes, B. Obst, and D. Hartmann, "Combined offline simulation and online adaptation approach for the accuracy improvement of milling robots", *CIRP Annals*, vol. 69, no. 1, pp. 337–340, 2020.

[17] A. Solin, J. Hensman, and R. E. Turner, "Infinite-horizon Gaussian processes", *Advances in Neural Information Processing Systems*, vol. December, pp. 3486–3495, 2018.

[18] M. P. Heinl, A. Giehl, N. Wiedermann, S. Plaga, and F. Kargl, "MERCAT: A Metric for the Evaluation and Reconsideration of Certificate Authority Trustworthiness", in *CCSW'19: Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop*, London, United Kingdom: Association for Computing Machinery, 2019, pp. 1–15.

[19] M. P. Heinl, "A Metric to Assess the Trustworthiness of Certificate Authorities", Master's Thesis, University of Ulm, Ulm, Germany, 2019.

[20] B. C. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments", *ACM Computing Surveys (Csur)*, vol. 42, no. 4, pp. 1–53, 2010.

[21] N. Holohan, S. Braghin, P. Mac Aonghusa, and K. Levacher, "Diffprivlib: The IBM differential privacy library", *arXiv preprint arXiv:1907.02444*, 2019.

[22] M. Sabt, M. Achemlal, and A. Bouabdallah, "Trusted execution environment: What it is, and what it is not", in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1, 2015, pp. 57–64.

[23] A. Giehl, N. Wiedermann, M. Tayebi Gholamzadeh, and C. Eckert, "Integrating security evaluations into virtual commissioning", in *IEEE 16th International Conference on Automation Science and Engineering Proceedings*, Hong Kong: IEEE, 2020, pp. 1193–1200.

[24] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud, "Differential privacy in linear distributed control systems: Entropy minimizing mechanisms and performance tradeoffs", *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 118–130, 2017.

[25] J. Giraldo, A. Cardenas, and M. Kantarcioglu, "Security and privacy trade-offs in cps by leveraging inherent differential privacy", in *2017 IEEE Conference on Control Technology and Applications (CCTA)*, IEEE, 2017, pp. 1313–1318.