Information Quality Challenges in Industry 4.0

NILS GRUSCHKA, Kiel University of Applied Science Kiel, Germany
JENS LÜSSEM, Kiel University of Applied Science Kiel, Germany

Industrial production is currently undergoing significant changes and this will continue in the next years. The new manufacturing concepts and technologies are often subsumed under the term Industry 4.0. With smart production and cheap customization great benefits for manufacturers and customers are expected. However, like with most new opportunities also new risks will arise.

In this paper, we show how information quality can be affected in a typical Industry 4.0 scenario. Subtle data modification can have significant effects on the production and can be used as a mean to attack modern manufacturing systems. Some of the threats which have been identified can be fended with contemporary technologies while others are hard to control and require new measures.

• Security and privacy → Network security

Additional Key Words and Phrases: data quality, industrial internet.

1. INTRODUCTION

Industry 4.0 is a catch phrase first used in Germany (originally *Industrie 4.0* (Ferber 2012)), which is nowadays used all over for the current (fourth) revolution in industrial manufacturing. It includes aspects like digitalization of production equipment, interconnection of industrial networks with the Internet, dynamic and intelligent manufacturing processes, custom-made products for the same price like mass products (“mass customization”) and close coupling of internal processes with customers and suppliers. The topic is closely connected to other current technologies like Internet-of-things (IoT), cyber physical systems (CPS), machine-to-machine (M2M) communication, cloud computing and big data.

It is a received opinion that security and safety are crucial aspects for Industry 4.0. Cyber-attacks on industrial control systems have been growing over the last years and will continue in the near future (Ashford 2016). Such attacks allow disabling or even destroying manufacturing equipment (BBC News 2014) leading to financial losses. Even human life can be endangered if for example manufacturing robots are remotely controlled. Further, new possibilities for industrial espionage are opened through direct or network access to cyber physical systems.

Until now, little research was done on the effect of Industry 4.0 on data integrity and information quality (e.g. in (Sha and Zeadally 2015) data quality for CPS is discussed, but not for Industry 4.0 overall). In this paper, we take a hypothetical but typical industrial scenario and analyze the role information quality has. More specifically, we show that an attacker has multiple new possibilities to influence information quality and harm industrial production or human’s well-being.

2. EXAMPLE SCENARIO

Figure 1 shows a typical Industry 4.0 scenario taken as an example throughout the whole paper. In this factory, autonomous agents like intelligent robots or intelligent assembly lines are manipulating intermediate products (in our case smart objects) in order to produce end products according to the customer’s needs. The central controlling unit mainly coordinates the autonomous agents.

The flow of goods from the suppliers to the customers similar to pre-Industry 4.0 scenarios. Here however, the customer’s order is influencing directly the production process to create customized products.

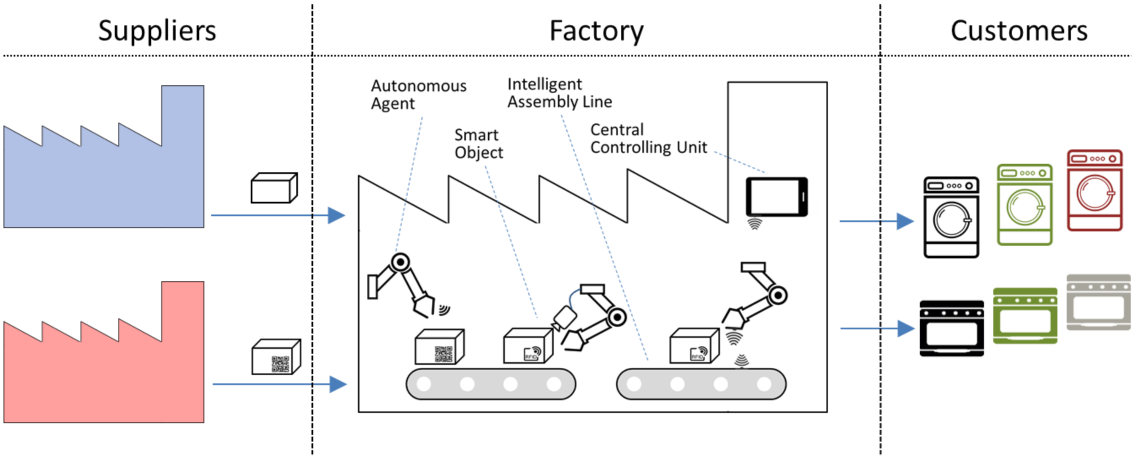

Fig. 1. Industry 4.0 manufacturing scenario

In the following, we would like to focus on the interaction between the autonomous agents and smart objects. In our scenario we like to define these terms as follows:

Definition 1

An *autonomous agent* is an intelligent agent making its own decisions how to act in its system environment.

Definition 2

A *smart object* is an object with enhanced abilities such as storing or processing information or interacting with its environment (here: the autonomous agents).

To be more precise, a smart object consists of a real object and a virtual object, e.g. an RFID tag attached to the real object (see Figure 2).

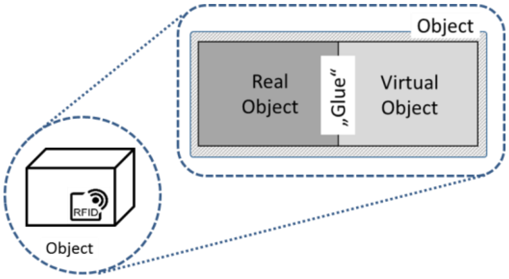

Fig. 2. Composition of smart objects

The virtual object contains information about the corresponding real object (like size of the object, material etc.) and further properties (like producer, envisaged end product etc.).

The autonomous agent can interact with the smart in different ways. Figure 3 shows the 4 most relevant patterns:

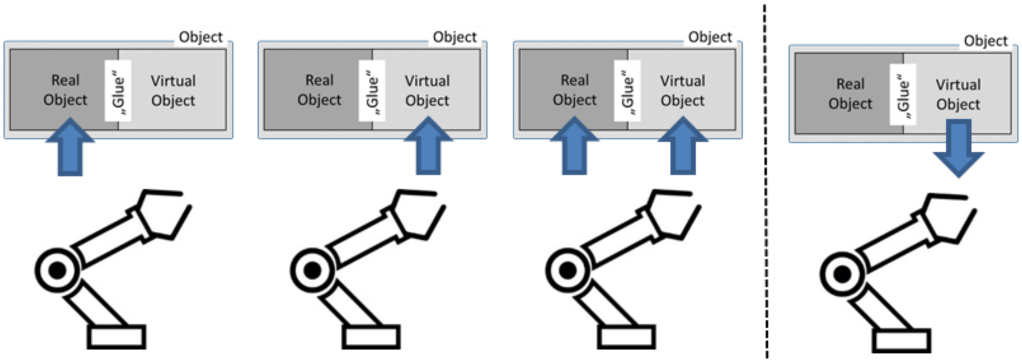

Fig. 3. Interaction between autonomous agents and smart objects

(1) The autonomous agent manipulates the real object (e.g. combining the product with a second one). The result of this manipulation may be stored in the corresponding virtual object or not.
(2) The autonomous agent manipulates the virtual object (e.g. changing some properties).
(3) The autonomous agent manipulates both the real and the virtual object.
(4) The virtual object manipulates the autonomous agent (e.g. sending a construction plan for the envisaged end product)

In the remainder of this paper, the presented model will be assumed. The next section will continue by presenting the data quality dimensions which are relevant for our scenario.

3. DATA QUALITY DIMENSIONS

In literature, data quality is viewed as a multidimensional concept (Wand and Wang 1996; Wang and Strong 1996; De Amicis, Barone, and Batini 2006; Caballero et al. 2007; Burgess, Gray, and Fiddian 2006). Even though a lot of research has been conducted on the dimensions of data quality there is neither an agreement on the data quality dimensions nor on the definition of these dimensions. A quite common choice of data quality dimensions is given in table I.

Table I. Dimensions of Data Quality

Category	**Dimension**
Intrinsic	Believability
	Accuracy
	Objectivity
	Reputation
Contextual	Value-added
	Relevancy
	Timeliness
	Completeness
Representational	Interpretability
	Ease of understanding
	Representational consistency
	Concise representation
Accessibility	Accessibility
	Access security

In this paper, we use the following dimensions and definitions (Wand and Wang 1996; Prat and Madnick 2008):

- *Accessibility*
  Data is accessible if data consumers have access to the data according to their needs.
- *Accuracy*
  Having accurate data implies that an information system represents a real object in the same manner to the one that should have been represented
- *Believability*
  The extent to which data values originate from trustworthy sources
- *Completeness*
  A data set is complete if all necessary values are included.
- *Timeliness*
  Timeliness is defined in terms whether data is up to date or not.

It is obvious, that these dimensions are all relevant to the scenario presented before. As an example, accuracy is always violated, when modifications to the real and the virtual object are not synchronized (see Figure 2). This might happen when an erroneous autonomous agent accesses only the real object, only the virtual object or the real and the virtual object in different ways. Besides these accidental breaches of data quality, such modifications can also be triggered systematically by attackers. This is discussed in the following section.

4. THREATS FOR INFORMATION QUALITY

The scenario presented in section 2 offers multiple possibilities for an attacker to influence the information quality involved in the production process – possibilities which did not exist in pre-Industry-4.0 scenarios. More precisely, the concept of smart objects enables threat on new types of data (i.e. **what** can be attacked). Further, the changes in internetworking and manufacturing process eases the influence of parties involved in the overall system (i.e. **who** can perform attacks).

4.1 Sources of attacks

In the following we will describe first the sources of threats.

Direct attacks

Two “classical” forms of attacks are: first, a (so called) hacker circumventing the perimeter security of the company and gaining access to some systems of the manufacturing company and, second, an employer of the company, who already has some privileges gaining access to systems he should not have. These threat have been present since the 1980s. However, in the “new world” of Industry 4.0 with coupling of office and industrial networks, direct access to manufacturing equipment might be possible. This can lead to clearly observable changes like shutdown of that machine, but also to covert modification of data quality. This will be discusses in detail below.

Attacks from customers

One important goal of Industry 4.0 is massive customization possibilities for customers. Instead of choosing between a small number of fixed variations, customers can freely sketch parts of the product which is produced shortly after submitting the order (to give the customer similar delivery delays like from the current "order of the shelf"). This obviously requires a close coupling of ordering and manufacturing systems without major manual effort or inspection from a manufacturing employee. This interface may offer customers access to production databases or even manufacturing equipment.

One possibility might be a command injection attack. Take the example of a color which can be freely chosen be the customer and which is transferred to the production machine. The customer may enter into the color field of the ordering portal a value like "<command>shutdown()</command>" – instead of a value like "maroon". If this is not adequately filtered and transferred to a programmable production equipment, this can lead (like before) to breach of availability of the machine or safety for workers. Further, also in this case, hidden attacks on data quality are possible (details see below).

Attacks from supply channel

Every manufacturer is highly dependent on his suppliers providing him with raw material and semi-manufactured products. And the manufacturer must trust that the delivered product has actually the characteristics which were ordered respectively which were listed in the accompanying documents. Forging of material or documents (which means low data quality of product properties) is not a new issue. However, with the concept of smart objects and the integration of object properties and manufacturing processes it has new possibilities and more severe impact (Details will be explained below.). Further, it is also possible that modification of products or product information might also happen during transport from the supplier to the manufacturer. Very often shipment is done over thousands of kilometers and takes several week, which gives the shipment company and even complete outsiders a lot of chances to influence the delivery.

4.2 Target of attacks

After describing the "who can attack the data quality", we will discuss what kind of modifications can be done on the data and which data quality dimensions can be influenced.

Direct attacks

For the direct attacks (either internally or externally) we can assume that the objective is either espionage (which has little connection to data quality) or some kind of sabotage. Beside the direct acts of sabotage which were discussed before (like disabling equipment or hurting workers) changes of data quality allow subtle attack. The attacker can influence the producing equipment either to write incorrect information in the virtual object (violation of *accuracy*) or to abandon from updating the information while the product is changed during the manufacturing process (violation of *timeliness*) or to incorrectly combine the individual component information into a composite product (violation of *completeness*). In all these cases, either the real object is modified (compared to the expected outcome) and the virtual object shows the expected values or vice versa or even both values are changed.

Depending on the concrete scenario, this can have different effects. First, these deviations might be detected by quality checks before the final product leaves the manufacturing plant. Then the hunt for the source of the inconsistency will start, possibly stopping the complete production process. The second possibility is, the changes are not detected and the product is delivered to end consumers. Here the changed (possibly low quality) product can work unreliable leading to outage (complains, lawsuits) or even to hurting people.

Attacks from customers

For a customer a reasonable motivation for an attack is probably gaining a better or more valuable product than he has paid for. Thus, he might program the manufacturing agent to produce a product from material A (high value) but writing material B (what he has paid for) into the virtual object. When other agents read the virtual object they verify that the information matches the order information and the modification of the real object remains undetected. From the viewpoint of information quality, the *accuracy* is violated.

Also *accessibility* might be relevant in this attack scenario. When a product is created from different components the customer might program an agent to select other (again more valuable) components for his final product than he has paid for. Here, the information regarding the deviation from the order exists inside the product. However, it might not be available, as the components tags are hidden inside the final product or even shielded from radio waves.

Attacks during shipment

Modifications of products during shipment from the supplier to the manufacturer might also aim for sabotage. In contrast to the "direct attacks" we assume no attacks on the producing equipment are possible, but by degrading the *accuracy* of the product information (by either modifying the real product or the virtual product) similar effects on the production or the end consumer can be caused. Additionally, the *believability* of the supplier's data quality can be diminished, if the manufacturer detects the modifications and blames the supplier.

Attacks from supplier

The most likely motive for attacks from the supplier is shipping low-grade products to the manufacturer and trying to disguise this. So the supplier might send products made from material A and is storing material B (which the manufacturer is expecting) in the data tag (violation of *accuracy*). If the product send by the supplier is a compound product the supplier might also misuse the *accessibility* problem discussed above: he is composing the product from other component as the data tag on the compound product is claiming. As the data tags of the components are not accessible to the manufacturer this attack is hard to detect. Finally, the supplier must take into account, that any attacks he is performing might influence the *believability* of all data he is sending to the manufacturer in the future.

Table II. Affected data quality dimensions for different attacks

	Direct Attack	**Customer**	**Supplier**	**Shipment**
Accessibility		X	X	
Accuracy	X	X	X	X
Believability			X	X
Completeness	X			
Timelines	X			

Table II gives an overview of the data quality dimensions involved in different attacks. One can see, for example, that accuracy is included with all attacks. However, a lack of accuracy might also result from faulty manufacturing equipment. This makes these intentional attacks hard to distinguish from other errors.

4.3 Countermeassures for Attacks

The question, how to fend the attacks presented above, is partly still open. Some threats can be mitigated using existing technologies.

Direct attacks

To avoid direct attacks on manufacturing equipment, the production network should be treated like "normal" enterprise networks. This includes: perimeter protection (firewall, IDS), security management and incident procedures, security audits and penetration testing. Special attention should be paid to the security of remote management interfaces, e.g. enforcing strong authentication.

Attacks from customers

The Web application used by the customers must be tested for typical Web application flaws like XSS, command and SQL injection. Here, special care must be taken to the fact, that the Web application is not only communicating with a data base in the backend must possibly also directly with manufacturing machines.

Attacks from supplier or during shipment

Attacks on the virtual objects are new and, thus, little research was done on this topic, yet. To avoid modification of data stored in a virtual object, all data can be digitally signed. This requires a public key infrastructure for the involved public keys and increases the effort for verifying data, but it ensures integrity of information stored in an information tag. However, a large number of attacks are still possible. First, the supplier still can forge material and/or data stored on the product. As he is in possession of a legitimate signing key, he can create valid signatures for incorrect information. But also for an outsider there are still possibilities. He can remove an information tag (containing data with a valid signature) from one product and apply it to a different product. Thus, a strong binding between a product and its information tag must be ensured. For some products a physical binding can be gained, i.e. the information tag is embedded into the product in a way that prevents any later removal. For other products, the data on the original tag might contain (signed by the original producer) a fingerprint of the product which can be easily calculated by the manufacturer receiving this product. This topic will be covered in future research.

5. CONCLUSION

Modern industrial processes open new business possibilities but also new attack possibilities. Besides direct attacks with immediate and apparent effect, also subtle influences on the data quality can be used, which are much harder to detect. The effects of the latter ones range from slightly inferior end products and fraud performed by customers or suppliers to production shutdown and personal injury.

Fending such threats will be a huge challenge for Industry 4.0. First, awareness for these problems is required. This will enable system designs which take these threat into account. Then, some attacks can be fended using approved information technology countermeasures. Other problems (like binding of virtual objects to real objects) are still open and need further research.

Further, organizational countermeasures must be developed. Enterprises need new policies and employee trainings in order to cope with the new situation. This might also require special IoT/I4.0 security standards.

Finally, the next big challenge will come with the next generation of production agents: in the future, virtual objects will not only contain an identifier and some properties, but complete construction plans and even software for the production equipment. This information is uploaded on-the-fly from the product and executed on the agent. How can be ensured that this program does no harm to human or material?

REFERENCES

Ashford, Warwick. 2016. “Industrial Control Systems a Growing Target for Cyber Attack.” *ComputerWeekly*. Accessed April 2. http://www.computerweekly.com/news/4500272123/Industrial-control-systems-a-growing-target-for-cyber-attack.

BBC News. 2014. “Hack Attack Causes ‘Massive Damage’ at Steel Works.” December 22. http://www.bbc.com/news/technology-30575104.

Burgess, Mikhaila SE, W. Alex Gray, and Nick J. Fiddian. 2006. “Quality Measures and the Information Consumer.” *Challenges of Managing Information Quality in Service Organizations, Idea Press Group, Hershey, USA*, 213–242.

Caballero, Ismael, Eugenio Verbo, Coral Calero, and Mario Piattini. 2007. “A Data Quality Measurement Information Model Based On ISO/IEC 15939.” In *ICIQ*, 393–408.

De Amicis, Fabrizio, Daniele Barone, and Carlo Batini. 2006. “An Analytical Framework to Analyze Dependencies Among Data Quality Dimensions.” In *ICIQ*, 369–383.

Ferber, Stefan. 2012. “Industry 4.0 – Germany Takes First Steps toward the next Industrial Revolution.” *Bosch ConnectedWorld Blog*. October 16. http://blog.bosch-si.com/categories/manufacturing/2012/10/industry-4-0-germany-takes-first-steps-toward-the-next-industrial-revolution/.

Prat, Nicolas, and Stuart Madnick. 2008. “Measuring Data Believability: A Provenance Approach.” In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, 393–393. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4439098.

Sha, Kewei, and Sherali Zeadally. 2015. “Data Quality Challenges in Cyber-Physical Systems.” *J. Data and Information Quality* 6 (2–3): 8:1–8:4. doi:10.1145/2740965.

Wand, Yair, and Richard Y. Wang. 1996. “Anchoring Data Quality Dimensions in Ontological Foundations.” *Commun. ACM* 39 (11): 86–95. doi:10.1145/240455.240479.

Wang, Richard Y., and Diane M. Strong. 1996. “Beyond Accuracy: What Data Quality Means to Data Consumers.” *J. Manage. Inf. Syst.* 12 (4): 5–33. doi:10.1080/07421222.1996.11518099.