M.Sc. Dissertation in Engineering

Sensing as a Service Revisited

A Property Rights Enforcement and Pricing Model for IIoT Data Marketplaces

센싱으로서의 서비스

지적재산권 집행시스템 및 산업용 사물인터넷 데이터 가격산정 모델

August 2019

Graduate School of Seoul National University

College of Engineering

Technology Management, Economics and Policy Major

Jan-Terje Sørlie

Abstract

Sensing as a Service Revisited

Jan-Terje Sørlie

College of Engineering

Technology Management, Economics and Policy Program

Graduate School of Seoul National University

The Industrial Internet of Things (IIoT) has become a valuable data source for products and services based on advanced data analytics. However, evidence suggests that industries are suffering a significant loss of value creation from insufficient IIoT data sharing. We argue that the limited utilization of the Sensing as a Service business model is caused by the economic and technological characteristics of sensor data, and the corresponding absence of applicable digital rights management models. Therefore, we propose a combined property rights enforcement and pricing model to solve the IIoT data sharing incentive problem.

Keywords: Industry 4.0, Industrial Internet of Things, Sensing as a Service, IIoT Data Marketplace, Digital Rights Management, Digital Watermarking

Student number: 2017-21456

Contents

1 Introduction 1

1.1 Background 1

1.2 Problem Description 6

1.3 Research Objective and Question 8

1.4 Methodology 8

1.5 Contributions 9

1.6 Structure 10

2 Literature Review 11

2.1 Sensing as a Service 11

2.2 Economic Characteristics of IIoT Data 14

2.2.1 Property Rights of Data 18

2.2.2 Licensing of IIoT Data 23

2.3 IIoT Data Marketplaces 25

2.3.1 Use-cases and Value Propositions 30

2.3.2 Market Structures and Pricing Models 34

2.4 Digital Rights Management for IIoT 36

3 Model 44

3.1 Assumptions 45

3.2 Watermarking Technique 47

3.2.1 Function 48

3.2.2 Example 50

3.2.3 Robustness 51

3.3 Economic Reasoning 54

3.3.1 The Quality Gap 55

3.3.2 Cost of Watermarking (CoW) 57

3.3.3 Cost of Attacking (CoA) 58

4 Analytical Analysis 60

4.1 Equilibrium Between CoW and CoA 60

4.2 Determining the Optimal Quality Gap 62

4.3 Applicability of the Quality Gap Function.....64

5 Conclusion.....66

5.1 Summary.....66

5.2 Discussion.....66

6 Limitations and Future Research.....68

References.....70

Abstract (Korean).....79

List of Tables

Table 1 – Current (I)IoT data marketplaces.....29
Table 2 – Main watermarking techniques for IIoT data 43
Table 3 – Example of watermarking of an IIoT data stream 51

List of Figures

Figure 1 – The (I)IoT Model and Sensing as a Service 13
Figure 2 – Potential interactions in IIoT data marketplaces.....26
Figure 3 – Generic example of the watermarking process.....50
Figure 4 – The quality gap in product versioning 56
Figure 5 – Equilibrium between CoA and CoW 61
Figure 6 – The quality gap curve 63
Figure 7 – Practical use of the quality gap curve.....64

1 Introduction

1.1 Background

In 2011, the German government launched its *Industrie 4.0* initiative to *"drive digital manufacturing forward by increasing digitization and the interconnection of products, value chains and business models"* (EU, 2017). The fourth industrial revolution has now become widely accepted as the era in which technologies like cyber-physical systems and cognitive computing will enable a significant increase in operational efficiency and productivity (Lu, 2017). Similar to the way in which automation with programmable logic controllers and robotics represented a third industrial revolution in the late 1970s, the fourth industrial revolution is expected to bring smarter manufacturing, products and services (ibid.).

Key characteristics of Industry 4.0 are often found to be digitization, connectivity and interoperability (ibid.). This can be interpreted as if the fourth industrial revolution essentially marks the beginning of an interconnection of the physical and digital worlds (Guth et al., 2016). Digitization of our environment has arguably been an ongoing process for at least the better part of the past century. Much of these efforts can be attributed to our desire to describe the world in a language that computers can understand in order to utilize their superior capabilities.

These superior capabilities become prevalent with technologies like big data analysis, which can increase functionality and quality in products and services by revealing patterns and correlations that were previously invisible (Golchha, 2015). Big data is also a key source for machine learning which is the heart of artificial intelligence (Liang et al., 2018). Indeed, advanced data analytics has recently been found to be the most pursued approach to technology innovation, which is yet another indication of data itself becoming an indispensable asset (Ringel et al., 2018).

Although vast amounts of digital information are already being collected and processed, fundamental Industry 4.0 prerequisites such as connectivity and interoperability have been lagging behind due to cybersecurity concerns and technological incompatibility in industrial environments (Kim & Chang, 2014).

The EU Commission supports the theory that interoperability is one of the major barriers to a successful digital economy (Kerber & Schweitzer, 2017). Therefore, projects like BIG IoT have received funding from the EU Horizon 2020 Research and Innovation Program to develop common interfaces to overcome this technological barrier (Bröring et al., 2017). Moreover, much like digital convergence in the telecommunication and media industry was enabled by technological change and convergence on common standards (Mueller 1999), a data-driven convergence of

information technology (IT) and operational technology (OT) is now emerging (Murray, Johnstone, & Valli, 2017). The IT/OT convergence may also contribute to close the remaining technology gap in the Industry 4.0 vision.

However, the levels of information transparency as envisioned for Industry 4.0 cannot be achieved through physical connections and technological compatibility alone. Multiple studies have assessed the economic grounds and associated incentives for sharing digital information in the form of sensor data. For instance, the Sensing as a Service (S^2aaS) business model was anticipated by De Cristofaro, Ding and Tsudik (2009) as an internet-connected sensor network offering commercial data access services. This model has later been covered in great detail by Zaslavsky, Perera and Georgakopoulos (2013), who link the concept to the Internet of Things paradigm. Central to the S^2aaS business model is the concept of IoT data marketplaces (Mišura & Žagar, 2016). The key motivation behind such marketplaces is to create platforms on which raw data streams from different connected devices, which otherwise may remain unexploited or stored in silos, can be traded for increased value creation (Perera, 2017a).

As consensus has yet to be reached on the exact definition of IoT devices, we will consider IoT as physical or virtual sensors capable of exchanging data over the internet – wired or wirelessly. This definition

is inspired by Ashton (2009), who was the first to coin the IoT term in 1999. Some scholars also include actuators under the IoT umbrella (Perera, 2017b), but we will limit our scope to only cover sensors as a part of the Sensing as a Service business model.

The Internet of Things is one of the key frameworks supporting Industry 4.0 (Khan et al., 2017). However, various definitions of IoT allow for many interpretations including personal devices and applications focusing on more user-centric convenience than significant gains in operational efficiency and productivity. Therefore, we will use the term Industrial Internet of Things (IIoT), which can be considered as a subcategory of IoT. Due to the lack of a widely accepted definition of IIoT, we will assign this term to what the EU Commission defines as machine-generated, non-personal raw data. The EU Commission describes such data as being *"created without the direct intervention of a human by computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real"* (Zech, 2017). Thus, we use the term *"industrial"* to emphasize the value propositions for various industries, although the actual data may be collected in non-industrial environments like in private, connected cars (Kerber & Frank, 2017).

Thus, IIoT data may initially involve personal information originating from users and operators through applications like supply chain

optimization, asset tracking and monitoring, predictive maintenance and operational control systems (Lu, 2017). The EU General Data Protection Regulation (GDPR) defines personal data as *"any information relating to an identified or identifiable natural person"*. A natural person is further defined as *"one who can be identified, directly or indirectly, in particular by reference to an identifier"*. Therefore, with IIoT data, we have to assume that any initial personal data has been rendered non-personal through sufficient anonymization. This means that we will not consider privacy concerns in this study.

Prior research highlights many use-cases for data-driven applications across a wide range of industries and underlines the benefits in increased sharing of data as a non-rivalrous good (Lu, 2017; Marjani et al., 2017; Rafaeli & Raban, 2005). However, a recent communication by the EU Commission addressing ownership and access rights of non-personal data, expresses concerns that current limitations in data sharing means that we are not taking full advantage of the emerging data-driven economy (Kerber, 2017).

The EU communication has sparked a debate on the property rights of IIoT data as well as why privately held sensor data is not sufficiently shared (Richter & Slowinski, 2019). This study adds to that debate.

1.2 Problem Description

Improved data availability for service providers and developers can represent promising opportunities for machine learning in artificial intelligence-powered applications (Golchha, 2015) or to extend capabilities of existing products and services (Perera, 2017b). In addition, it is expected that academics and research institutions can benefit from broadened access to real-time data (Milham et al., 2017). In more detail, information harvested by IIoT devices can enable a wide range of smart applications within domains like utility metering, logistics, supply chains, agriculture, power grids, traffic and building controls (Marjani et al., 2017) as well as an overall increase in manufacturing efficiency (Lu, 2017).

Despite these business opportunities, the limited presence of data sharing through (I)IoT data marketplaces is evident (Kerber, 2017). While economic articles such as the one by Kerber (2017) suggest that the main barrier to increased data sharing is the lack of knowledge on how data can be exploited, we believe the rapid increase in production and use of (I)IoT data suggests otherwise. In addition, the current presence of online (I)IoT data marketplaces featuring common interfaces indicates that technological incompatibilities have already been overcome (Bröring et al., 2017).

Prior studies on S^2aaS and (I)IoT data marketplaces pay little attention to fundamental prerequisites for any economic transaction, namely pricing, well-defined property rights and associated mechanisms to enforce them. On the other hand, the more economically oriented debate following the EU communication on data ownership does not seem to fully acknowledge the technological aspects of this issue. We believe the uncertainty and disagreements characterizing prior publications on this topic are caused by the lack of an interdisciplinary assessment in this domain.

Although Kerber (2017) assumes that (I)IoT data streams can be classified as an excludable good, we argue that sustainable data sharing is actually inhibited by the current *de facto* non-excludable characteristics of IIoT data streams. This is not due to a lack of technological protection mechanisms, but because such techniques have yet to be incorporated in commercially viable digital rights management models for IIoT data. Thus, in the middle of the heated debate on who should be granted property rights to data, the critical function of enforcing these property rights seems to have been forgotten. No prior research has proposed a model for digital rights management that considers both the property rights enforcement and pricing of IIoT data.

1.3 Research Objective and Question

Based on these shortcomings, our research objective is to combine the related and sometimes even conflicting elements of pricing strategies and property rights enforcement in one digital rights management model for IIoT data, and we ask how this combination can contribute to economically viable IIoT data trading.

1.4 Methodology

This study encompasses the key economic and associated technological mechanisms involved with trading IIoT data. The assessment of prior literature considers supply and demand, property rights allocation and enforcement, and economic characteristics of IIoT data streams as a tradable good.

Based on the findings of our literature review, we complete the following steps to answer our research question:

1. Requirements and relevant scenarios for digital rights management of IIoT data streams are identified:
    - The need for watermarking of IIoT data streams as a measure of property rights enforcement is supported.
    - The need for versioning of IIoT data streams as a measure of profit maximization is supported.

2. A watermarking technique is developed and combined with product versioning in a digital rights management model and used as basis for the economic analysis.
3. The relationship between watermarking robustness and product versions is analytically analyzed by considering:
    i. Revenue maximization from versioning
    ii. Cost minimization of watermark recovery and authentication
    iii. Cost maximization of watermark attacks due to quality reduction
4. Discussion of the results concludes the work.

1.5 Contributions

This study provides a novel approach to remedy the IIoT data sharing incentive problem by combining property rights enforcement and pricing strategy.

The proposed model includes a simplistic watermarking mechanism for sensor data that features a strong relationship with IIoT data versioning. We also show how the so-called optimal quality gap can be quantified for property right holders to achieve profit maximization as an incentive to increase data sharing.

This approach extends the scope of prior literature on S^2aaS and IIoT data marketplaces like Zaslavsky, Perera and Georgakopoulos (2013); Sheng et al. (2013); Mišura and Žagar (2016); and Perera (2017a, 2017b). The academic motivation behind this study is to close the gap between the aforementioned technologically oriented works and the more recent economic articles concerning the EU communication on the property rights of IIoT data (Richter & Slowinski, 2019).

1.6 Structure

The remainder of this study is organized as follows: a literature review on the S^2aaS business model, IIoT data marketplaces and associated economic topics is presented in Chapter 2. In Chapter 3, we summarize the key takeaways from our literature review to support the proposed property rights enforcement and pricing model. We then introduce a digital watermarking technique as a core element for the proceeding economic analyses. In Chapter 4, we utilize our model in an analytical analysis to answer our research question. Chapter 5 summarizes our findings before potential shortcomings of the proposed model are covered in Chapter 6.

2 Literature Review

2.1 Sensing as a Service

Sensing as a Service emerges from the recent servitization paradigm in information technology (Duan et al., 2015). Centralization and increased availability of computing power in combination with improved infrastructure have fostered a wide range of cloud-oriented ICT services. The most prevalent concepts are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) (Perera et al., 2014). This servitization paradigm is sometimes summarized as Everything as a Service (XaaS) and is largely driven by scalability, accessibility, pay-per-use pricing models and dynamic resource allocation (ibid.). Thus, cloud computing can offer the benefits from low initial investments, economies of scale and high-end services all at once.

More recent contributions to the cloud computing service family include Sensing as a Service (S^2aaS). This is a subcategory of Data as a Service that is shaped by the rapidly expanding ecosystem of (I)IoT devices and platforms. The idea is that billions of (I)IoT devices can provide new insights for value-added services through a multilateral content distribution model (Koutroumpis, Leiponen, & Thomas, 2017; Zaslavsky, Perera, & Georgakopoulos, 2012).

The novelty of Sensing as a Service is not necessarily in the provision of digital information. Such data markets have existed for some time (Muschalle et al., 2012). The distinction between S^2aaS and other means of providing digital commodities as a service is mainly attributed to the raw and real-time properties of IIoT data streams (Mišura & Žagar, 2016). In other words, the data marketplace we assess in this study facilitates exchange of data before value-added services like analytics, aggregation or combination of data have been provided.

In a state-of-the-art literature review on economics and pricing models for (I)IoT, Luong et al. (2016) describe the general (I)IoT model shown in Figure 1. This model is also the basis of other comprehensive reviews in this domain like Yan, Zhang and Vasilakos (2014). The Sensing as a Service interface typically occurs between platform and data processing. In this way, providers of data analyses, applications and other services can access a broad selection of input data.

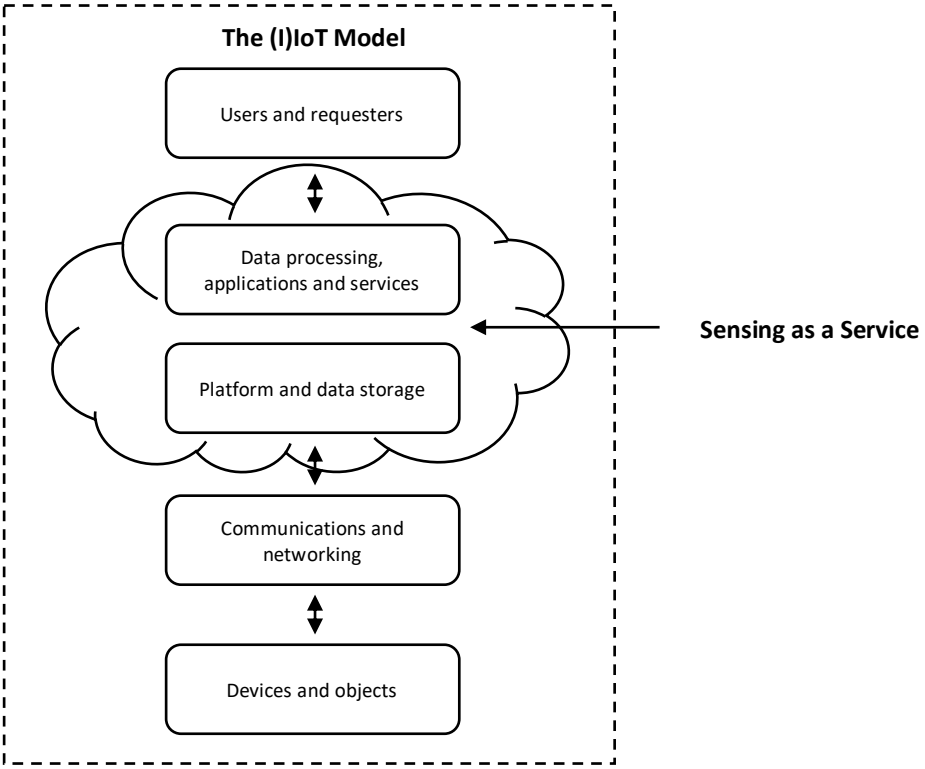

Figure 1 – The (I)IoT Model and Sensing as a Service

Although a single data stream representing an unprocessed and simple physical factor may look worthless at first sight, the ability to sense the environment is allegedly an essential prerequisite for Industry 4.0 technologies (Lu, 2017). When combining multiple IIoT data streams from the environment, a certain combination or collection may construct valuable information. Large collections of IIoT data can also be defined as big data, which is often characterized by its volume, velocity, variety, veracity and value (Jin et al., 2015).

2.2 Economic Characteristics of IIoT Data

There are arguably two main approaches to value creation with IIoT data (Liang et al., 2018): as an intermediate good for the creation of new or enhanced products and services, or as a final information good to support decision-making and optimize existing processes. In this section we will focus on the latter case and the unique properties of digital information goods.

In information economics, information itself obtains its value from the increase in expected payoff or utility from making informed decisions. This can be shown as (Lawrence, 2012):

$$v(x, y) = \pi(x, a_y) - \pi(x, a_0)$$

where:

$\pi(x,a)$ is the profit function at state x with the decision a.

a_y is the decision made with information.

a_0 is the decision made without information.

While some may be surprised to learn the quite recent appreciation of information in the form of raw sensor data as a valuable good, information economics and the power of knowledge in trading has long been appreciated by academics like Stigler (1961). More recently, we have seen the acknowledgement of timely information reaching new

levels. One example of competitive advantage obtained from milliseconds is underlined by Lewis (2014), who explains how high-frequency traders have made billion-dollar investments in infrastructure to reduce latency in communication with stock exchanges. This example highlights an important distinction that needs to be made between static and streaming data as a tradeable good.

A common characteristic of information in general is that determining its price is quite different from pricing physical goods due to the negligible marginal cost of (re)production and low transactional costs. As Shapiro and Varian (1998) proclaimed, *"information is costly to produce but cheap to reproduce"*. In many cases it can also be argued that information is easy to create but hard to trust and easy to spread but hard to control, but disruptive technologies like blockchain and smart contracts for implementing provenance and excludability may challenge the two latter conceptions (Missier et al., 2017; Ølnes, Ubacht, & Janssen, 2017; Özyilmaz, Doğan, & Yurdakul, 2018).

Open data markets for static data like the former Microsoft Azure DataMarket have been around for years. Associated market structures and especially pricing models have therefore been covered in great detail in studies such as Muschalle et al. (2012) and Tang et al. (2013). However, the economic differences between static and streamed data introduce new challenges for data markets. Mišura, Krešimir and Žaga

(2016) emphasize the importance of freshness as well as the fact that a data stream is normally sold before it has been collected. This means that the actual contents and thus the potential value of the data stream may for both parties be highly uncertain at the time it is traded. Moreover, the generic nature of sensor data and its many use-cases can make it challenging to reveal the highly diverse willingness to pay amongst various consumers.

In general, information can be considered a hybrid of public and private goods (Rafaeli & Raban, 2005). It can be relatively non-excludable and non-rivalrous unless technological protection measures prove otherwise. Like a trade secret, a raw data stream could entail sufficient incentives for non-disclosure, but the idea of the open marketplace is to monetize data that may be of value to others and thus exploit gains from trade in the market model. From the perspective of overall welfare maximization, a non-rivalrous good with close to zero marginal cost in production should be shared with any potential stakeholder (Kerber, 2017). However, this approach is not likely to be in the interest of a private owner nor considered a fair distribution for the ones investing in data acquisition. It is also important to keep in mind that although data and information itself is non-rivalrous, the consequences of sharing it may cause rivalrous actions if the information can be used to win a zero-sum game.

That said, a major challenge in marketplaces for goods with close to zero marginal cost of reproduction, negligible transaction costs and no means of absolute protection is the threat of arbitrage (Pantelis & Aija, 2013). A contract preventing a buyer from reproducing and distributing the same data stream is critical to maintain the data producer's economic incentive to collect and share data. Otherwise, a malicious buyer could exploit the arbitrage opportunity and resell the data to multiple buyers at a lower price than the original good. We will look more closely into this aspect in our review of property rights of data.

Another characteristic of markets for information is that they are known to not exhibit high degrees of transparency, which means that the good needs to be partially revealed before it can be fully evaluated by the consumer. This is obviously not an ideal situation for a close to non-excludable good. However, the importance of freshness in IIoT data streams may allow samples to be a fairly efficient way of signaling quality according to the concept described by Spence (1978).

Apart from sampling, Mišura, Krešimir and Žaga (2016) propose two-way credibility ratings as found in other platforms like eBay and Uber for signaling the quality of a data stream, its provider or consumer. That said, Richter and Slowinski (2019) underline the need for more elaborate trust mechanisms in scenarios where firms consider sharing

critical data than, for example, private consumers evaluating their driver in ride-hailing services like Uber.

2.2.1 Property Rights of Data

According to the Coase Theorem, given well-defined property rights and sufficiently low transaction and bargaining costs, an open market will lead to the most efficient and mutually beneficial outcome (Coase, 1960). Thus, the main goal of the legal system is to assign these property rights and enforce contract law. This is however not always straight-forward for digital goods.

There are two general positions that can be taken when discussing data ownership. One side is driven by increased openness through wide contributions of data assigned to the public domain similar to the open-source ideology. This can for instance be achieved through mandatory transparency from a regulatory point of view. The other side is seeking exclusive property rights and thus private incentives for data creation from profit maximization through the market solution. Due to our focus on the sustainability of the S^2aaS business model, this study will focus on the latter position.

The European Union has gained attention for driving discussions on data governance (Zech, 2017). In 1996, the Database Directive established copyright protection for databases (EU, 1996). A

communication released in 2017 addresses ownership and access rights of non-personal data in relation to the European data economy (Zech, 2017), and the GDPR gave EU residents the right of personal data portability in 2018 (De Hert et al., 2018). Therefore, we will approach the S^2aaS business model from the perspective of the European Union, which seems to be taking the lead on data regulations. Although we will not pursue novel judicial interpretations of data ownership legislations in this chapter, we will highlight certain aspects of EU law and other scholars that are relevant to our research objective.

The EU Copyright Directive is often cited in discussions on the property rights of intangible or digital goods. Therefore, we will first assess the applicability of copyright law and other means of absolute protection of non-personal raw data streams. The digital age has certainly challenged the enforcement of copyright law by enabling highly efficient reproduction, dissemination and storage of digital information (Peters, 2006). As the main goal of copyright law is to stimulate creative activity, the challenge is to balance owners' ability to trade and distribute the work digitally while also preventing infringers from doing the same (ibid.).

At the European level, non-personal data may in general qualify for absolute protection under the law of copyright (EU, 2006), *sui generis* database right (EU, 1996) or the trade secret directive (EU, 2016). Data

expressing literary or artistic works may obtain copyright protection (EU, 2006). We consider such creative elements to be unlikely to occur in IIoT data as covered by this study. Moreover, the database right, which is an acquis of the copyright law, applies to databases that *"by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation"* (EU, 1996). However, copyright protection does not extend to the actual contents of such databases (EU, 1996). Hence, data streams from IIoT devices cannot be individually protected under copyright nor database right law. Lastly, trade secrets must have commercial value and be subject to measures aimed at keeping them secret in order to be protected under EU law (EU, 2016). As the S^2aaS business model implies making IIoT data streams available on a marketplace, we consider trade secret protection to also be irrelevant for this purpose.

Therefore, if legally binding property rights are to be obtained for non-personal IIoT data streams, protection must in most cases be sought through general civil law concepts like contract law. This conclusion is in line with prior assessments of data provenance by Koutroumpis, Leiponen and Thomas (2017) as well as a legal analysis on exclusive rights of data by Wiebe (2016).

Legally binding contracts, which in the EU are based on the Principles of European Contract Law (PECL), can for example be in the form of an

agreement reserving certain rights of the information being exchanged. A contract may represent an adequate solution in certain use cases but poses some obvious shortcomings when compared to automatically granted protection like copyright. For instance, a contract must first fulfill the basic requirements of legal enforceability which involves obtaining acceptance by all parties. This could in some cases require the traded information to be revealed before the contract is legally binding due to low level of transparency of digital assets. Moreover, the doctrine of privity of contract is another common principle differentiating contract law from copyright law in an unfavorable way for data owners. Privity means that the rights or obligations of a contract are only binding to the parties signing the contract. This implies that contract law would normally not prevent a third party from redistributing a data stream against the will of the original creator.

The lack of absolute protection of IIoT data from non-personal sensors is in line with the general consensus of factual information not being copyrightable. Despite the challenges this may introduce for IIoT data owners, the principle is arguably still reasonable as the converse would naturally pose a major hindrance to most activities in science, media, culture and more. However, in a commission staff working document on the free flow of data and emerging issues of the European data economy, the EU has proposed a data producer's right where the *"owner or long-term user of a device"* or *"persons or entities that*

operate sensor-equipped machines, tools or devices at their own economic risk" would be granted the property right for machine-generated, non-personal raw data (Zech, 2017).

In an economic article discussing the above-mentioned EU communication, Kerber (2017) summarizes the intentions of the EU Commission as facilitating increased data creation and sharing through reduced transaction costs for trading and licensing. The plan is to achieve this by establishing the producer's right as well as what may become a mandatory access right to what we in this study define as IIoT data. However, Kerber (2017) argues that it is not necessary with a producer's right because the situation would be no different from today's solution where a party will have *de facto* rights to data. Property rights can still be licensed out with a contract regardless of any exclusive rights. Kerber (2017) emphasizes that this is only the case if the data is in fact excludable through proper technological measures, which is a disputed matter as argued in this study.

On the other hand, Kerber (2017) is more positive to the proposed mandatory data access right because it may relieve the monopolistic control and unequal bargaining power that exclusive or *de facto* property rights are likely to cause. For instance, non-rivalrous, privately held data could be accessed in the public interest. Also, certain mandatory access rights could prove helpful in complex multi-

stakeholder situations such as the example of connected cars which we will return to later. Nevertheless, regulations such as the producer's right could, if enacted, answer a relevant question of data ownership in industrial scenarios where machine suppliers collect data from their products to provide additional services to its users or in other ways monetize such data.

Although the proposed producer's right is initially exclusive, it is so as codified information – that is, not on the semantic level. The EU Commission is thus still within the principles behind non-copyrightable factual information. However, a recurring question in copyright disputes is the level of commonality between original works and infringing copies. Such uncertainties could therefore also arise with data streams protected under a producer's right, which is likely to introduce many peculiar claims; for instance, how many similarities must an allegedly replicated data stream share with the original stream for it to infringe the producer's right? That said, only a limited number of studies have been conducted in this domain and future research should explore consequences of a data producer's right in more detail (Kerber & Frank, 2017).

2.2.2 Licensing of IIoT Data

As we have seen, distributed digital content can be regulated by law with absolute rights such as copyright, or through a civil contract. Any

granted rights of use are normally described in a product license. Such license terms are often incorporated as a semantic component of a Digital Rights Management system, called Rights Expression Languages (Nadah, de Rosnay, & Bachimont, 2007).

From the perspective of IIoT data and other use-cases where the final product is likely to be composed of data from heterogeneously distributed sources, a challenge arises when composite licenses compliant with all their sub-licenses are to be described. A typical example of this is when a digital product is a result of multiple data streams from different sources. This is known as the license attribution stacking problem. In this regard, Governatori et al. (2013) have proposed an automated framework for composing such licenses based on deontic logic. Standardized licensing terms that are machine-readable also facilitate automated trading in marketplaces for digital goods.

According to an online catalog for data and analysis, www.data.world, these are the common license types for published data sets: public domain, attribution, share-alike, non-commercial, database only, and no derivates. Data is dedicated to the public domain by waving all rights to the extent allowed by law. An attribution license means that one must give credit to the creator and indicate if any changes have been made. ‘Share-alike’ means that the user is obligated to distribute any

transformation or derivate works under the same license as the original work. A non-commercial license requires the user to not exploit the contents for commercial purposes. 'Database only' specifies that the license only applies to the database and not the contents, and 'no derivates' prevents any work being derived from the contents. Non-profit organizations such as Creative Commons have released copyright license models to the public in order to simplify the legal aspects related to reserving and waiving copyrights. The final license is often a combination of the above-mentioned types, such as Attribution-Non-Commercial.

In this study, we will consider the profit maximization approach to data sharing through the IIoT data marketplace, and we will therefore not consider open data nor other works assigned to the public domain.

2.3 IIoT Data Marketplaces

Open marketplace platforms for IIoT data have been proposed to increase utilization of the vast amounts of collected data across different industries and is believed to serve as a core function in the S^2aaS business model (Mišura & Žagar, 2016). However, the concept is fairly new, and only a few studies have been published on this topic (Kerber, 2017).

The purpose of IIoT data marketplaces is to match buyers and sellers and to facilitate the exchange of a digital good technologically, financially and formally. Given a critical mass of participants, a platform for data exchange is expected to reduce the transaction costs for both parties involved in the trade. This includes the search and information costs, bargaining costs and enforcement costs (Richter & Slowinski, 2019). Based on the works by Mišura and Žagar (2016) and Parera (2017a and b), we have composed a proposed model of possible interactions between agents in IIoT data marketplaces in Figure 2.

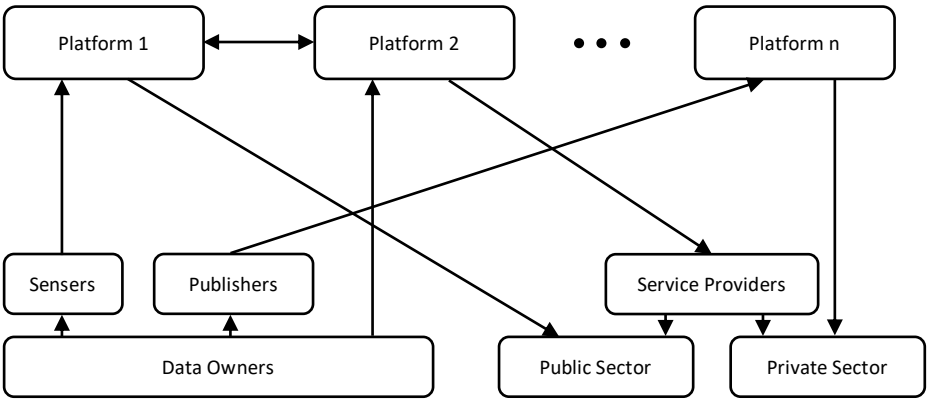

Figure 2 – Potential interactions in IIoT data marketplaces

Data owners are the economic agents who have obtained the property rights of the data stream. As we have learned so far, these property rights are not necessarily well-defined. The data owner can connect any data source in its possession directly to an IIoT data marketplace platform given sufficient compatibility and ease of use. However, in the case where a sensor is not already in place, so-called senser agents may

offer their services in the form of sensor hardware and platform integration installed on the data owner's property or assets. A publisher is an agent who is specializing in integration of existing sensor applications to IIoT data marketplaces. This agent is typically enabling interoperability through development of custom APIs. The platform agents are hosting IIoT data marketplaces and may create strategic alliances and share their data offerings.

In more detail, the marketplace platform can take on responsibilities such as data validation, classification, combination, aggregation, transaction history and payments (Koutroumpis, Leiponen, & Thomas, 2017). Due to the potential volume and variety of data, other important tasks for a marketplace platform is syntactic and semantic vocabulary management and indexing to allow for search-based discovery of data streams (Bröring et al., 2017). That said, the core functionality may turn out to be the creation of trust through reduction of information asymmetry as this is believed to represent a fundamental precondition for sharing data (Richter & Slowinski, 2019).

On the consumer side, we have service providers utilizing the great variety of data streams to enhance their existing services or to develop new ones. These services are subsequently consumed by end-users, but some end-users may also consume raw data streams from the marketplace themselves, such as insurance companies collecting data

from clients' cars (Kerber & Frank, 2017). In other words, there are both direct and indirect consumers of IIoT data streams.

Despite the promising future of the S^2aaS business model and the apparent presence of relevant technology, the current IIoT data marketplaces still seem to be in their infancy (Kerber, 2017; Richter & Slowinski, 2019). A search for active and open marketplaces currently promoting sales of (I)IoT or real-time data underlines the early stages of this business model. All (I)IoT data marketplaces discovered in this study were launched in 2018 and half of the platforms are emphasizing the fact that the platform is still under development. Nonetheless, these marketplaces address what has previously been found to be a lack of providers of dynamic data sources (Liang et al., 2018).

The marketplaces listed in Table 1 were discovered by investigating the top 100 Google search results for keywords "iot data marketplace", "sensing as a service marketplace", "data stream marketplace", "real-time data marketplace" and "buy and sell iot data". The search was performed in April 2019 with Private Browsing activated in Safari on macOS Mojave from South Korea. Industry-specific platforms as well as marketplaces primarily focusing on static data sets are not included.

Table 1 – Current (I)IoT data marketplaces

Name	URL	Status	Launched
IOTA Data Marketplace	data.iota.org	Operational (PoC)	2018
Terbine	terbine.io	Operational (Beta)	2018
databroker dao	databrokerdao.com	Operational	2018
SynchroniCity	iot-data-marketplace.com	Operational (Piloting)	2018
Streamr	streamr.com	Operational	2018

A noteworthy observation from this insight is that most marketplaces offer data that is already assigned to the public domain or that otherwise features public characteristics. Such data streams are typically representing weather data, traffic conditions, public transportation and environmental data such as air pollution and water quality. Hence, the absence of privately held data is prominent. Moreover, three of the discovered marketplaces utilize a cryptocurrency as a medium of exchange and distributed ledgers for transaction records.

Another notable project that is currently under development is BIG IoT which attempts to facilitate (I)IoT platform integration through a common interface and data sharing in a marketplace. BIG IoT aims to overcome the technological barriers to interoperability and unified data management across (I)IoT platforms, ecosystems and other data sources (Bröring et al., 2017).

2.3.1 Use-cases and Value Propositions

In an outlook on the value and pricing of big data, Pantelis and Aija (2013) argue that the decisive factors for exchanging data are quality, price and consumers' willingness to pay – just like with most other intangible goods. However, in addition to featuring negligible marginal cost of production, (I)IoT data is in many cases a byproduct of another process (Pantelis & Aija, 2013). Thus, there are few fixed costs to consider for data producers in IIoT data markets except from any opportunity costs associated with not sharing the data. We will now shift our attention to value propositions from IIoT data sharing and take a closer look at specific use-cases and associated data sources.

Smart devices often include a wide range of sensors to feed control systems with close to real-time information that enables the more intelligent functionality. Machines, products and their users primarily trigger data collection for operational purposes, but this data can also have a secondary value, as we will see in this section.

For further insights on this topic, we will use a connected car as an example. A modern car is equipped with advanced sensors feeding information to internal controls like advanced driver assistance systems as well as the driver itself (Kerber & Frank, 2017). The collected data can contain information on driving behavior, location and navigation, the surroundings of the car, weather data, driving conditions, etc. However,

in line with the motivation behind S^2aaS and IIoT data marketplaces, connected cars can also share this data with third parties. Such stakeholders may be car manufacturers, owners, component suppliers, repair service providers, the government, insurance companies, other financial institutions and more (ibid.).

The imagined use-cases for data harvested from connected cars are many. Some examples are performance analysis, predictive maintenance, auto part supply chain optimization, emergency assistance, traffic monitoring, road condition alerts, road quality reporting, tailor-made insurance pricing, weather forecasting and various smart city applications. A more novel use-case is the collection of data for training neural networks for autonomous vehicles as presented by Tesla Motors.

As explained by Liang et al. (2018), IIoT data streams can be reutilized for either more informed decision-making or the development of new or enhanced technology. The above-mentioned use-cases are spread across both categories. Liang et al. (2018) also cite several studies showing how commercial value is created from related use-cases. Thus, there is little doubt that there are real economic incentives for trading IIoT data.

Although this study will not assess current policy evaluations regarding who should be entitled to the ownership of IIoT data, it is worth mentioning that the connected car example is a central scenario in the EU commission's discussions on this topic (Kerber & Frank, 2017). The debatable question in that regard is for instance whether the car manufacturer or car owner should be granted exclusive property rights, and whether any other parties should be granted default access rights for the greater good. Considering that there might be multiple stakeholders involved in both the collection and utilization of data streams, the associated market scenarios could pose complex data governance issues (Kerber & Frank, 2017).

Kerber and Frank (2017) also emphasize that the ideal outcome of the connected car scenario is strongly affected by how the data is made available – that is, if the car is connected to a marketplace platform, to the manufacturer's ecosystem or accessible as a source for multiple stakeholders. This discussion is closely related to the treatment of personal data, but according to the scope of this study we will ignore this fact and assume that all data has been sufficiently anonymized.

For instance, the car manufacturer may have obvious interests in utilizing car data in its service provision and product development, reselling it to other service providers for additional revenue streams, or keeping it undisclosed to protect its competitive edge (Kerber & Frank,

2017). Property rights of this data are typically obtained through a contract between the car owner and manufacturer to clear up any legal uncertainties, but many car manufacturers are today in *de facto* control of the data due to technological barriers (Kerber, 2017). In future IIoT applications, it is envisioned that consumers are explicitly awarded for giving up data. For instance, car owners may expect access to free or improved services, or other financial benefits, as proposed by Perera (2017a), if they choose to share certain data streams.

The open market solution implies that it is the agent being able to create the most value who will eventually obtain control of the data regardless of how property rights are initially distributed (Kerber & Frank, 2017). Although the car manufacturer may technically or contractually prevent the car connecting to an open marketplace, the manufacturers that choose to limit such access would in theory only be better off if that is the most effective solution overall. Otherwise, they may eventually not sell any cars. This is, however, a theoretical scenario that is not very likely unless car buyers are perfectly rational, all cars have perfect substitutes and all agents have perfect information.

That said, Kerber (2017) argues that producer's rights as proposed by the EU Commission would not remedy market failures because the rights are likely to be bargained away in contracts. Thus, unequal bargaining powers would remain, and this could endanger both

competition and innovation through hold-up problems and monopolistic behavior.

2.3.2 Market Structures and Pricing Models

Although IIoT data streams have many theoretical use cases and potential value propositions, we must turn to prior research to investigate how and in which market structures IIoT data streams can be part of commercially viable business models.

In terms of assigning value to data, Fricker and Maksimov (2017) reviewed studies on pricing mechanisms for both static and streamed data and concluded that there is actually no consolidated understanding of how such data products can be priced. Not all reviewed studies considered the open marketplace approach, but interestingly, they were all assuming market structures characterized by monopoly, duopoly or monopsony. Moreover, pricing models surveyed by Fricker and Maksimov (2017) had different aims and varied from profit to social welfare maximization as well as internal consistency and fairness – where internal consistency means that the pricing function is monotonic, and fairness suggests that there is a fair trade-off between price and quality. In terms of pricing attributes, usage, quality, cost, views, and customer profiles were proposed (Fricker & Maksimov, 2017).

As emphasized by many scholars, IIoT data streams can construct big data and can thus be utilized for associated analyses. Therefore, we include the assessment by Pantelis and Aija (2013) on value and pricing of big data in the case of online marketplaces for reference. Pantelis and Aija (2013) apply the mode of data ownership to categorize data from an economic perspective. The authors argue that in a market with open or public data, which would suggest strong competition due to the nearly free availability of such data, neoclassical economic theory suggests that the price will be approaching marginal cost of production. In the case of digital information, we know that marginal cost of (re)production is close to zero. Hence, data markets with open or public data are unlikely to be sustainable (ibid.). Therefore, only privately held, digital goods with a certain degree of differentiation are expected to be commercially viable in an open marketplace. Pantelis and Aija (2013) classify such goods as being mainly private with well-defined and exclusive property rights. In the case of a sustainable market scenario with monopolistic power through data differentiation, Pantelis and Aija (2013) underline that profit maximization is reached through price discrimination with mechanisms such as versioning, market segmentation and personalized pricing.

In a more recent study on data trading, Liang et al. (2018) assessed the full lifecycle of data trading through a comprehensive literature review. The authors concluded that there is no easy way to quantify the value

of data due to the vast amount of use-cases. Thus, the theory of screening will be important as the seller is likely to suffer from imperfect information because the data application is unknown (Stiglitz, 1975).

As for pricing strategies in data markets, Liang et al. (2018) found that data packages, pay-per-use, flat rate, dynamic, two-part tariff and freemium models are all relevant. Liang et al. (2018) separate pricing models into three different categories: economic-based pricing, game theory-based pricing and auctioning. The first category contains typical variants of models like cost-based and consumer perceived value-based pricing, while the second category includes models like non-cooperative, Stackelberg and Bargaining games. Relevant auction models are forward and reverse one-sided, double, sealed-bid and combinatorial auctions. However, the authors emphasize that it is eventually the market structure that will have the greatest influence on price determination. Liang et al. (2018) also found that common techniques for price differentiation of data is versioning through different levels of precision and frequency.

2.4 Digital Rights Management for IIoT

The volatility, regulatory uncertainty and unique economic characteristics surrounding IIoT data streams raise a need for suitable protection mechanisms with respect to obtaining provenance and enforcing property rights. In this regard, the digital age has taught us

the importance of digital rights management to prevent digital goods from being copied, shared and stolen (Liang et al., 2018). This chapter will provide insight into prior works on technological protection mechanisms for data streams.

Central techniques in digital rights management are encryption, watermarks and digital signatures. The portability and reproducibility of digital goods make these measures essential to provide security, access control, usage control, license management and payment fulfillment (Liang et al., 2018). In their literature review on this topic, Liang et al. (2018) categorized digital rights management into three types: software-based, where unauthorized use is typically prevented with encryption and user authentication; multimedia-based, where encryption and watermarking are designed to prevent malicious reproduction; and unstructured data-based management, which is what applies in the context of IIoT data streams (Panah et al., 2016).

As explained by Liang et al. (2018), digital rights management of unstructured data is considered to be more challenging than for software and multimedia because replication and tampering is technically difficult to control. This is because data streams function as carriers of information without a pre-defined application, which makes fulfillment of licensing terms rely heavily on trust and moral obligations. Regardless of any encryption mechanism for secure transportation and

user authentication for access control, the information contained in the data stream must eventually be revealed to the consumer, which in turn enables virtually effortless reproduction. This is a clear distinction from digital media files carrying images, music and video where the perceived value is largely driven by its analogue consumption by humans.

Digital watermarks are often implemented to enforce copyrights of digital media (Cox et al., 2002). For instance, invisible watermarks can be implemented in audio files at frequencies outside the human hearing range, or in middle-frequency parts of images through techniques even sustaining various forms of processing and cropping (Hsu & Wu, 1999). Such watermarks can, for example, contain information about the original buyer. If this fingerprint is carried with any illicit copy of the original work, the malicious agent behind the breach of contract bears a significant risk of being revealed due to the embedded traceability. Thus, digital watermarks can serve as an effective barrier to illegal reproduction and redistribution of digital works (Chen & Wornell, 2001).

However, unlike images and audio files, digital watermarks in IIoT data streams would necessarily impose a notable difference in the good. A single number can simply not hold more information than its intrinsic numerical value unless the value itself or its metadata is altered. The difference between more traditional digital content and IIoT data

streams can thus be attributed to the fact that we do not necessarily know the actual use-case or mode of consumption for the latter good. This is why IIoT data streams need to be offered in an open and generic format. In other words, if we did not know that music was to be consumed by human ears, we may not have implemented watermarks at frequencies outside the human hearing range.

In a review of state-of-the-art techniques for digital watermarking, Panah et al. (2016) summarize five popular methods: low-bit encoding, spread spectrum, statistical watermarking, angle coding, and dither modulation. Low-bit encoding techniques like alternation of the least significant bit (LSB) are both simple and common, but often vulnerable to manipulations. In spread spectrum watermarking, information is spread across multiple data samples to better withstand detection and removal. Statistical watermarking embeds information by modifying the statistical characteristics of data. Angle coding is often of complex form and encodes information within the angle between variables. Lastly, dither modulation is the class on which we will focus in this study due to its applicability on data streams. This class contains methods like Quantization Index Modulation (Chen & Wornell, 2001) and applies perturbation over multiple data points to embed a watermark.

All digital watermarking techniques face a trade-off between three conflicting goals: maximizing rate of information imbedding, minimizing

distortion of the original data, and maximizing the robustness against attacks (ibid.). This is in line with the key properties of digital watermarks: invisibility, capacity, robustness, and security (Panah et al., 2016). The two latter properties are sometimes used interchangeably. Cox et al. (2002) assign the security property to the ability to withstand intentional attempts to prevent watermark detection, while robustness covers the ability to survive more innocent operations that may still compromise the watermark. We will use the term robustness to describe both of these properties. When considering robustness in designing digital watermarking techniques, it is common to design based on either security through obscurity or Kerckhoff's principle from cryptography, which assumes that the encryption technology or technique is known, but not the secret key (Panah et al., 2016).

Attacks against digital watermarking can be classified into four main categories (Voloshynovskiy et al., 2001): removal attacks, where information is sufficiently damaged; geometric attacks, targeting the watermark detection mechanism; cryptographic attacks, where the watermark is decoded and then removed or distorted; and protocol attacks, aiming to alter the watermark information.

In terms of watermarking theory and its classification of data types, the IIoT data streams we are assessing in this study are classified as non-media or unstructured in the form of streamed, complex data. Panah et

al. (2016) underline that the unstructured characteristics of IIoT data streams imply that, for example, relational database watermarking techniques are not necessarily applicable. So-called non-media watermarking is a relatively new sub-domain of watermarking and limited studies have been conducted on this topic (Panah et al., 2016). That said, there are more use-cases for watermarking of data generated by sensor networks than we cover in the scope of this study. Some examples are in-network data aggregation and secure query processing, but these applications are mostly relevant before data is made available on a marketplace.

For the purpose of implementing provenance in IIoT data, which is the history of data ownership that is central to the watermarking needs expressed in this study, Panah et al. (2016) highlight two main approaches: embedding through alteration of less significant bits in data samples, or in variable time delays between data samples. However, the latter technique is arguably not an ideal approach because time stamps may in many cases express critical information that should not be modified (ibid.). Therefore, we will focus on bit alteration instead of data point or time stamp alteration in this study. This is based on an assumption that time stamp reliability is considered more valuable than data point accuracy.

Sion, Atallah and Prabhaka (2006) were reputedly the first to propose a technique for watermarking sensor data streams for the purpose of embedding provenance. As a basis for their work, the authors assumed that if value is found in a data stream, it is likely to be tied to the order and accuracy of the data points. Thus, watermarking techniques for data streams should not alter this information in any significant manner. Their study was motivated by the potential threat of malicious agents reselling data streams in secondary markets. The authors also emphasize the unfortunate feature of sensor data in that, despite its scientific usefulness, the provenance is easily disconnected from its content. Sion, Atallah and Prabhaka (2006) developed a technique that implements watermarks through alteration of more significant bits in selected extremes of the data stream. This method proved resilience to transforms such as sampling, summarization, random alterations, and combined transforms.

Chong, Skalka and Vaughan (2010) were the first to propose a self-identifying watermarking technique which utilizes check bits as metadata encoded into insignificant bits of data points, while the actual provenance marks are encoded into more significant bits. This approach achieves some degree of redundancy and shows better robustness to truncation and rounding. This study was the first to show encoding of provenance in streamed data in the form of positive and negative

integers, decimal numbers and low-entropy datasets (Panah et al., 2016).

Except from interpacket delay-based methods as designed by Sultana, Shehab and Bertino (2012), and sequence altering methods as designed by Xiao et al. (2010), forward reference searches with Google Scholar from the original works of Sion, Atallah and Prabhaka (2006) and Chong, Skalka and Vaughan (2010) do not reveal any novel bit altering- or quantization-based techniques designed for IIoT data streams except for those covered in this review. Table 2 summarizes the main IIoT data stream watermarking techniques we have identified and their potential shortcomings with respect to the scope of this study.

Table 2 – Main watermarking techniques for IIoT data

Method	**Weakness**	**Reference**
LSB embedding in selected extremes.	Limited generality and invisibility.	Sion, Atallah and Prabhaka (2006)
Metadata and LSB embedding.	Insignificant bits can be attacked with limited loss of value.	Chong, Skalka and Vaughan (2010)
Data point sequence alternation.	The sequence of data points is considered to be critical for many applications.	Xiao et al. (2010)
Variable interpacket delay.	Interpacket delays are considered to be critical for many applications and are often fixed.	Sultana, Shehab and Bertino (2012)

3 Model

In this chapter, we develop the basis of a digital rights management model for IIoT data streams in order to assess our research objective. The digital rights management model will consist of a mechanism for property rights enforcement in combination with a sustainable pricing scheme for IIoT data.

IIoT data streams are, in general, not protected by copyright, and property rights must therefore be sought through contract law. Unless a data stream can be traced back to the initial buyer, only moral obligations would prevent buyers from breaching the contract that we assume disallows redistribution of data. Moreover, the barrier for malicious agents is weakened by the negligible marginal cost of (re)production of IIoT data streams. And, to make the matter worse, buyers in secondary markets are unlikely to be legally pursued for illicit redistribution due to privity of contract, which will add to the competition in secondary markets and make the price of the data stream approach its negligible marginal cost of (re)production.

Therefore, the property rights enforcement method will be designed to implement traceability in data streams to the point where contract infringement becomes economically unattractive.

3.1 Assumptions

We have seen in our literature review that IIoT data streams are vulnerable to arbitrage due to the challenges involved in enforcing property rights (Mišura & Žagar, 2016). This effect is assumed to harm the data owner and will therefore act as a barrier to data sharing. We will not consider the specific license model in use, but we assume the data is privately held, and that the original owner seeks profit maximization by maintaining its monopolistic power in an IIoT data marketplace. This means that the owner has an interest in preventing buyers from reproducing and distributing the data stream in secondary markets unless those buyers are authorized to do so.

The IIoT data covered by the scope of this study is machine-generated, non-personal data consisting of time-stamped real numbers. An example of such a data stream can be the speed of a connected car, which is part of core functionality needed to safely operate the vehicle in addition to being of potential interest to others. For instance, developers of autonomous driving systems may utilize this information in combination with other variables in their software development. On the other hand, the same data stream can also be valuable to insurance companies in case of an accident or for personalized pricing schemes based on driving behavior. However, it is natural to assume that the insurance company prefers a lower level of precision compared to the software developer if the data stream is made available at a lower price

through product versioning. Another aspect of this multi-stakeholder scenario is that the perceived value of the data stream will increase if it can be authenticated – that is, a mechanism where the originality of the data stream can be confirmed through its watermark.

We will therefore utilize so-called quality discrimination by applying versioning of the data stream as a measure to maximize profit as supported by prior literature (Liang et al., 2018). Regarding the perceived value of the data stream, we will assume that utility is expressed as a linear function of quality. And because use-cases of the same data stream may vary between different stakeholders as well as often not being known to the seller, we generalize quality to be described as the precision of the data stream. Hence, the perceived value of a data stream can be expressed as a linear function of the number of digits precision per data point in the stream. The precision level is the only factor of quality considered due to the potential elimination of certain use-cases caused by interpacket delay-based watermarking or the complete removal of certain data points. Therefore, we also assume that any attacks on watermarks attempt to maintain as much as possible of the original data stream and thus its quality and value.

As we only consider non-personal, machine-generated sensor data in this study, we assume that such data streams are a byproduct of

operating a product or process, and that there are only marginal additional costs, which are negligible, involved with producing or making the data available for sale – except for costs associated with enforcing property rights. In other words, the profit maximizing objective is aligned with revenue maximization.

3.2 Watermarking Technique

To embed provenance in IIoT data streams, we apply a digital watermarking technique based on alternation of less significant digits. Based on the review by Panah et al. (2016), we argue that this is the most relevant technological protection measure for this purpose. However, as opposed to prior works by Sion, Atallah and Prabhaka (2006) and Chong, Skalka and Vaughan (2010), we put additional emphasis on simplicity and the ability for the watermarking technique to provide product versioning and thus support associated pricing models. These requirements are motivated by our line of argument that there is a need for viable property rights enforcement and pricing models to facilitate increased IIoT data sharing. In addition, the technique is designed with respect to achieving the three main goals of digital watermarking: maximizing rate of information imbedding; minimizing distortion of the original data; and maximizing the robustness against attacks (Chen & Wornell, 2001). The watermarking principles applied in this study are inspired by the concepts of Quantization Index Modulation (ibid.).

3.2.1 Function

We propose a watermarking technique based on rounding operations. Rounding operations are normally applied with the purpose of approximating a fractional decimal number by a number with fewer digits but can also be applied to reduce the accuracy of integers. In computer science, this operation may typically be applied when compressing a data stream from double-precision (64-bit) to single-precision (32-bit). However, in our model, rounding can also be utilized for both watermarking and for reducing precision for the purpose of product versioning.

Tie-breaking rules are needed when rounding a decimal or digit that is exactly halfway between preceding integers. That is, if 9.5 should be rounded to 9 or 10. If it was not for fractions equal to 0.5, all round-off errors would be symmetric by always rounding to the nearest integer. The default rounding mode in the technical standard for floating-point arithmetic IEEE 754 is “round half to even”. This means that a midway floating point will be rounded to the nearest even integer value. In other words, 9.5 is rounded to 10 and 8.5 is rounded to 8. This method has no positive/negative bias and no bias toward/away from zero and will minimize the sum of expected errors. A similar tie-break rounding convention is “round half to odd”. As with “round half to even”, this rule also features the absence of positive/negative bias and bias toward/away from zero.

The proposed watermarking technique is named Deterministic Alternation Between Integer Tie-breaks (DABIT). DABIT implements a seemingly invisible repeating watermark in IIoT data streams consisting of any real number with or without fractions consisting of two or more digits. The method works by altering between “round half to even” and “round half to odd” operations according to a predefined sequence for every encountered tie-break. The embedded watermark can, for example, represent a 64-bit binary code identifying the initial buyer of the data stream. A “round half to even” tie-break operation expresses a binary 1, and a “round half to odd” operation expresses a binary 0. In this way, DABIT enables close to non-biased watermarking of data streams with a negligible loss of precision and accuracy. By comparing a DABIT-encoded data stream with the unwatermarked time series, every case of a tie-break rounding will express a piece of the watermark. Thus, the full data stream does not necessarily need to be kept as reference. The watermark embedding sequence is conceptualized as a Python function in Figure 3, in which the precision of each data point of the data stream *x* is reduced by one digit:

```
# x[] is an unwatermarked data stream
# y[] is the resulting watermarked data stream
# t is the timestamp
# i[] is the 64 bit watermark
# n is the watermarking sequence number (0–63)

if int(repr(x[t])[-1]) == 5:
    if i[n] == 0:
        y[t] = RoundHalfToOdd(x[t])
    else:
        y[t] = RoundHalfToEven(x[t])
    n += 1
else:
    y[t] = round(x[t])
if n == 64:
    n = 0
```

Figure 3 – Generic example of the watermarking process

Note: Python's built-in round() function uses the default rule "round half to even" in accordance with IEEE 754. Thus, round(x[t]) gives the same result as the custom function RoundHalfToEven(x[t]), but the latter is used for clarity in the watermark embedding process.

3.2.2 Example

Table 3 illustrates how an example of the first four bits (1001) of a repeating 64-bit watermark can be embedded in combination with product versioning. The Value column contains the original, unwatermarked data stream with its corresponding time stamps in the

left-most column. For every new pricing tier and thus quality level, the precision of each data point is reduced by one digit. In every instance where the reduction in precision involves a tie-break rounding operation, “round half to even” and “round half to odd” are used deterministically according to the bit sequence of the predefined watermark. When the bit to embed is 1, “round half to even” is applied and vice versa.

Table 3 – Example of watermarking of an IIoT data stream

Time	Value	Tier 1	DABIT	Tier 2	DABIT	Tier 3	DABIT	Tier 4	DABIT
t	43.846	43.85	-	43.8	1	44	-	40	-
t+1	43.525	43.52	1	43.5	-	44	1	40	-
t+2	42.947	42.95	-	42.9	0	43	-	40	-
t+3	42.493	42.49	-	42.5	-	43	0	40	-
t+4	43.065	43.07	0	43.1	-	43	-	40	-
t+5	43.715	43.71	0	43.7	-	44	-	40	-
t+6	44.150	44.15	-	44.1	0	44	-	40	-
t+7	44.519	44.52	-	44.5	-	45	0	40	1
t+8	44.655	44.66	1	44.7	-	45	-	50	0
t+9	45.152	45.15	-	45.2	1	45	-	50	0
t+10	45.301	45.30	-	45.3	-	45	-	40	1
t+11	45.527	45.53	-	45.5	-	46	1	50	-

In this way, the watermark will only be visible to the party having access to two adjacent data streams for comparison.

3.2.3 Robustness

In this chapter, we will discuss the robustness of DABIT against potential geometric, removal, cryptographic, and protocol attacks in scenarios

we find relevant for IIoT data streams. Although a wide range of sophisticated attacks are possible, we do emphasize our assumptions of value being associated with the quality of the data stream. Therefore, any attacks reducing the accuracy or removing proof of authenticity are expected to drastically reduce the value of the data stream.

We argue that the robustness of DABIT against geometric attacks targeting the watermark detection mechanism is relatively strong. This is due to the rigorous relationship between the original data stream and the watermarked version in terms of both time stamps and the order of data points. Hence, in case time stamps were tampered with in a geometric attack, a more advanced watermark detection mechanism may still recognize the order of data points as a reference to where the watermark is embedded. Altering the order of data points will increase the difficulty of watermark detection, but this operation is considered to result in a higher quality loss compared to other attack strategies.

Given that the unwatermarked data stream is not known and the value of the least significant digit is unpredictable, the DABIT watermark is arguably invisible to a malicious agent. However, the technique is potentially vulnerable to cryptographic and subsequent protocol attacks if different agents buy the same data stream with the purpose of averaging them or otherwise compare each data point to detect and possibly alter the watermark. That said, this is arguably the case of all

watermarking methods. A certain level of protection against such attacks can be achieved by utilizing strategic watermarks, but the actual composition of the watermark lies outside the scope of this study.

Lastly, removal attacks are seemingly the most relevant threat against DABIT because this approach can be conducted in a similar manner to the watermark implementation and product versioning itself. Thus, this mode of attack is in line with our assumption of minor noise in data points generally resulting in the lowest perceived quality reduction of IIoT data streams. We will not consider removal attacks such as averaging adjacent data points, which would harm the update frequency, nor the complete removal of selected data points, which would not be a viable approach for attacking an invisible watermark. Instead, we identify relevant removal attacks as being rounding, truncation or adding noise to the data stream. Common for these three removal attacks is their aim to reduce the quality of the data stream in an unbiased manner while maintaining other quality attributes such as frequency and time stamps. To simplify the assessment of these attacks, we combine these three attack modes in a common operation of reducing precision with rounding. We allow for this simplification because the least significant digit of a DABIT watermarked data stream will be equally distorted regardless of whether rounding, truncation or random noise is being applied.

One of the key features of DABIT that helps withstanding the aforementioned removal attacks is that even if the rounding operation is only applied to the least significant digit, the embedded watermark will occasionally affect more significant digits. Given a uniform distribution of least significant digits between 0–9, every 100^{th} rounding operation is expected to impact the second least significant digit. Every $1{,}000^{th}$ rounding operation will impact the third least significant digit and so on. This effect makes the watermark fairly robust against attacks on less significant digits, but it also illustrates the exponentially increasing difficulty to recover the watermark for every decimal being attacked by a malicious agent.

3.3 Economic Reasoning

The economic motivation behind applying digital watermarking is to embed provenance in IIoT data streams. The ability to recover this watermark at a later point is an important tool to fight malicious agents in the marketplace, as well as providing true consumers with a method for verifying authenticity of their data stream.

The perceived risk associated with illicit redistribution of IIoT data streams increases significantly when knowing that a data stream includes traceability back to its initial buyer. In this situation, malicious agents attempting to illicitly redistribute a data stream face two options: to trust their buyer and any subsequent buyers to never reveal the data

stream; or to attack the watermark in an attempt to remove the traces leading back to them. The first option is not considered to be rational due to the portable nature of data streams. Therefore, we will focus on the direct and indirect costs of attacking the watermark in our analysis.

Lastly, product versioning will also help prevent illicit redistribution of data streams in secondary markets. Offering the original data stream at a lower price and precision level will eliminate some incentives for malicious agents attempting to resell attacked, lower-priced versions because they must compete with authentic substitutes. Such authentic, lower-priced versions target consumers that are not willing to pay for the higher-priced options while also supporting the market price through product differentiation.

3.3.1 The Quality Gap

With the introduced quality discriminating approach, it becomes essential to identify the optimal quality gap between two product versions, as illustrated in Figure 4. In this chapter, we will explain how the optimal quality gap can be expressed by producers' and consumers' cost functions.

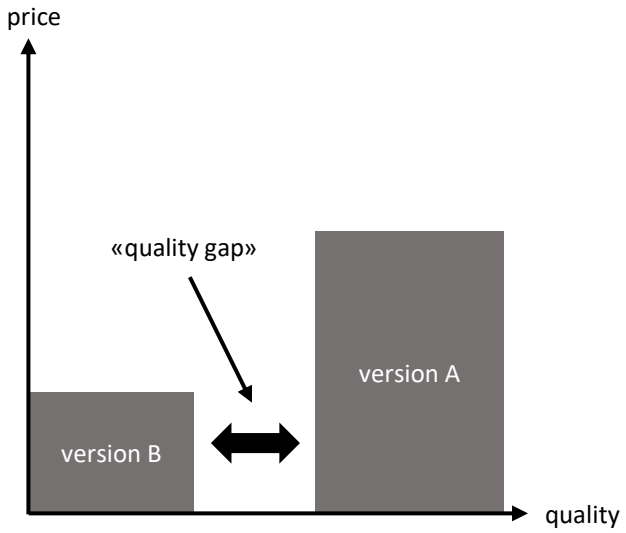

Figure 4 – The quality gap in product versioning

Figure 4 also illustrates the general assumption of consumers' willingness to pay for quality. Consumers requiring high quality will prefer the higher-priced option while consumers with less interest in data precision will resort to the lower-priced option or an attacked version of this in the secondary market. Therefore, the market price difference can express utility as a function of quality. The reason why we do not suggest additional levels of quality is that any levels of quality lower than version B would not be able to compete with attacked versions of A, which would have higher quality in the secondary market.

In terms of identifying the optimal quality gap, we argue that it pays off for the data owner to pursue watermark reconstruction of illegally distributed data streams until the costs of these efforts increase consumers' perceived loss of value due to quality reduction from

watermark attacks. This is because any costs of pursuing malicious agents that exceed consumers' combined willingness to pay for the quality difference between the true and the illicit good cannot be expected to be recovered. When these two cost functions are in equilibrium, the property rights holder is indifferent between pursuing malicious agents through watermark reconstruction and offering an authentic substitute of the illicit good on the marketplace, hence version B in Figure 4. This authentic good can match the precision level of illicit goods and still provide greater value due to its authenticity.

3.3.2 Cost of Watermarking (CoW)

We define cost of watermarking to consist of two main factors: embedding and reconstruction costs. These costs are faced by the property rights holder. Embedding is the cost of implementing the watermark through a rounding operation, which we define as a linear cost function expressed by a cost parameter ρ and the number of rounding operations x per data point.

Reconstruction costs are expressed as an exponential function of cost parameter a and the number of digits x that have been attacked by a malicious agent – in other words, the number of rounding operations applied per data point by the attacker. The exponential property is attributed to the exponentially increasing size of data that needs to be

available and collected in order to detect traces of the watermark for every digit of precision that has been attacked.

$$CoW = \rho\, x \times a\, 10^{x}$$

where:

x is the number of digits processed

ρ is the per digit cost of watermarking (rounding) operation

a is the per data point cost of watermark reconstruction

3.3.3 Cost of Attacking (CoA)

Malicious agents are facing two main factors in their cost function: the actual attack operation; and the resulting loss of value of the data stream due to the reduced quality.

Costs associated with performing the attack are considered to be equivalent to the initial watermark embedding process because we assume this to be performed as a rounding operation that reduces the precision level of the data stream. The cost of quality reduction is the consumer-perceived loss of quality of the data stream caused by an attack of the watermark. Due to the invisibility of the watermark, the attack is assumed to obscure or remove one digit per level of strength, *x*. According to our general assumption of a linear utility function of quality, this cost function is also linear.

$$CoA = \rho\, x \times b\, x$$

where:

x is the number of digits processed

ρ is the per digit cost of attacking (rounding) operation

b is the perceived loss of value per reduced precision level

4 Analytical Analysis

4.1 Equilibrium Between CoW and CoA

To identify the optimal quality gap and thus versioning strategy for the described scenario, we determine the equilibrium between the cost of watermarking and the cost of attacking. This equilibrium will identify the quality level at which product version B should be introduced to support the market value of product version A. The graph below indicates where this equilibrium can be found. The x-axis expresses the number of digits attacked by malicious agents in order to remove the watermark and thus the number of digits that drive the cost of the watermark reconstruction process. The cost of attack (CoA) increases linearly with the wasted quality and associated loss of consumers' willingness to pay per attacked digit. The cost of watermarking (CoW), which is largely driven by the reconstruction of watermarks in pursuit of malicious agents, increases exponentially with the number of attacked digits due to the nature of the watermark technique as described for its robustness properties. As argued in our economic reasoning, these two cost functions are equal at the optimal "quality gap" between product version A and B.

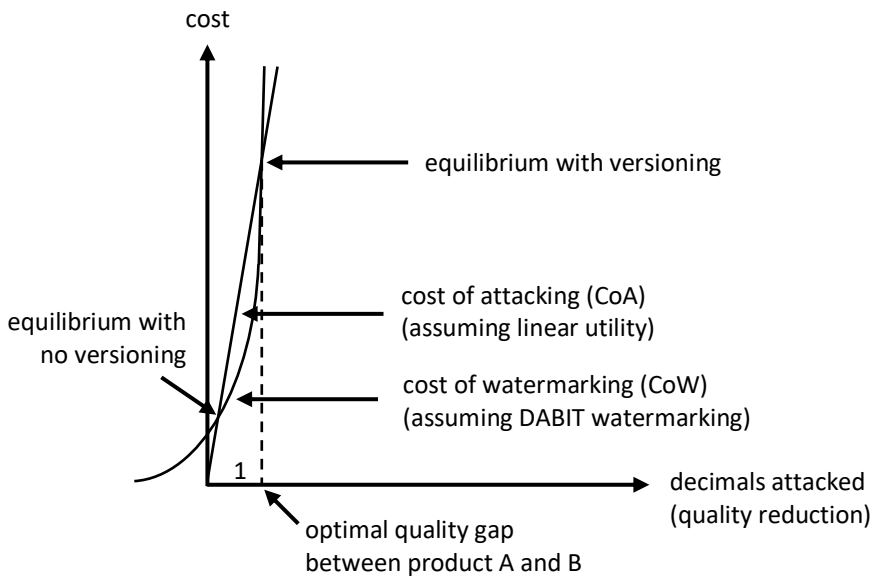

Figure 5 – Equilibrium between CoA and CoW

As illustrated in Figure 5, one equilibrium can occur before there is a full-digit "quality gap", thus when no DABIT watermark should be added. A second equilibrium between the two cost functions CoW and CoA can occur at a greater optimal "quality gap" between product version A and B.

The following equation expresses all equilibria between the two cost functions and thus the ideal product versioning strategy for the described scenario:

$$CoW = CoA$$

$$\rho\, x \times a\, 10^{x} = \rho\, x \times b\, x$$

As we assume the costs of rounding operations for the property holder and the malicious agent to be the same, these two factors cancel each other out and we are left with the costs of pursuing reconstruction of watermarks in attacked primary products, and the perceived loss of value due to reduced quality from watermark attacks:

$$a\, 10^{x} = b\, x$$

4.2 Determining the Optimal Quality Gap

The resulting function *y* below describes the ideal ratio between cost parameters *a* and *b* at different quality gaps *x* between product versions A and B. We name this function the quality gap function:

$$y = \frac{a}{b} = \frac{x}{10^{x}}$$

If the cost parameters are known, the optimal quality gap between the two product versions can be determined with the quality gap curve plotted in Figure 6.

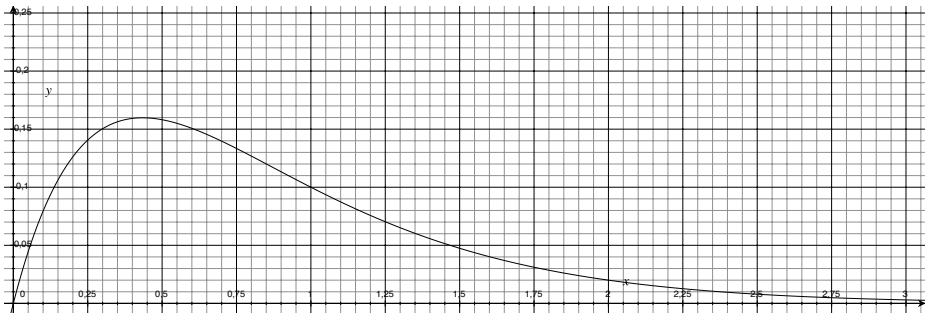

Figure 6 – The quality gap curve

We can see from the quality gap curve that the cost ratio *y* must be lower than 0.1 for the ideal quality gap to be higher than one digit. This is discovered by tracing a horizontal line from any given cost ratio *y* and observe for which values of *x* the line intersects with the quality gap curve. Below is an example of how the ideal quality gap is located if the cost ratio *y* = 0.02:

$$y = \frac{a}{b} = \frac{x}{10^x} = 0.02$$

where

a is the per data point cost of watermark reconstruction

b is the perceived loss of value per reduced precision level

x is the quality gap between product versions A and B

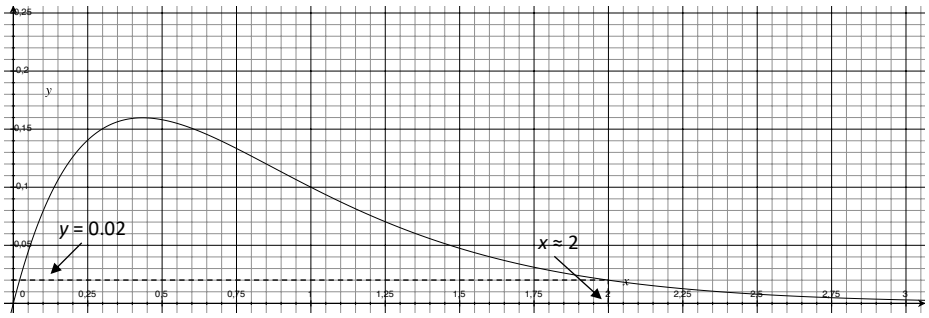

Figure 7 – Practical use of the quality gap curve

The quality gap curve illustrates the two equilibria between cost functions CoW and CoA, but we ignore the equilibrium occurring at less than a one-digit optimal quality gap between product versions. The second optimal quality gap for the suggested cost ratio occurs at approximately $x = 2$, hence at a two-digit quality gap.

4.3 Applicability of the Quality Gap Function

To further investigate the applicability of the quality gap function for the scenario covered by this study, we calculate the first derivative of y with respect to x:

$$\frac{dy}{dx}$$

$$= \frac{d}{dx}\left[\frac{x}{10^x}\right]$$

$$= \frac{\frac{d}{dx}[x] \cdot 10^x - x \cdot \frac{d}{dx}[10^x]}{(10^x)^2}$$

$$= \frac{10^x - \ln(10) \cdot x \cdot 10^x}{10^{2x}}$$

$$= \frac{1}{10^x} - \frac{\ln(10) \cdot x}{10^x}$$

$$= -\frac{\ln(10) \cdot x - 1}{10^x}$$

We then find the critical point:

$$\frac{dy}{dx} = -\frac{\ln(10) \cdot x - 1}{10^x} = 0$$

$$x = \frac{1}{\ln(10)} \approx 0.43$$

From Figure 6 as well as the mathematical properties of the quality gap function, we see that x ≈ 0.43 is a global maximum. Thus, the quality gap function is applicable for cost factor ratios in the interval (0, 0.43]. We also appreciate that the limit of the quality gap function *y* as *x* approaches infinity is 0. In other words, as consumers' willingness to pay for precision (*b*) grows relatively larger than the per data point cost of watermark reconstruction (*a*), the ideal quality gap approaches infinity. That said, the actual quality gap is practically limited by the number of digits per data point of the data stream.

5 Conclusion

There is little doubt that the use-cases for shared data streams are many and that the resulting advancements in operational efficiency and productivity are likely to be in the public interest. However, it seems like our private interests and lack of trust are currently preventing us from taking full advantage of this opportunity. It is natural to assume that our past experience with digital media piracy may be a decisive factor in this regard. This is why we proclaim that digital rights management models have become more relevant than ever before.

5.1 Summary

Our economic analysis shows how a simple technique for implementing provenance through digital watermarking in data streams can create a basis for more commercially viable IIoT data marketplaces. The key takeaway of this study is the relationship between perceived value of data streams and the efforts associated with enforcing property rights, and how this relationship can be utilized in profit-maximizing pricing strategies.

5.2 Discussion

The combined property rights enforcement and pricing model shows that there is in fact an equilibrium on the quality scale between these opposing forces, which is also the ideal entry point for the product

versioning strategy. However, this equilibrium may not exist for all watermarking techniques and is dependent upon the cost functions of embedding and reconstruction of watermarks. This emphasizes that there might be more considerations to be made when assessing watermarking techniques for IIoT data streams than just technological features alone.

Although the practical use-cases of this model will require further research and more rigorous testing, we have answered our research question and shown how digital watermarking and product versioning can contribute to economically viable IIoT data trading. The proposed model can also increase trust and ensure data owners a greater part of the revenue stream from their data despite the presence of malicious agents in the marketplace.

Moreover, we believe an online data stream authentication service can provide additional value to the proposed digital rights management model. We have mainly focused on the use of watermarks to legally pursue malicious agents when illicit data streams are discovered in secondary markets, but another relevant value-proposition is for consumers who would like to verify the authenticity and quality level of their data streams.

6 Limitations and Future Research

As a complete digital watermarking mechanism requires more sophisticated functionality, which lies outside the scope of this study, the main purpose of the described technology is to illustrate how property enforcement can be combined with pricing strategies in the pursuit of economically viable IIoT data sharing.

As for the quality gap function, we assume that the willingness to pay for quality is known in order to identify the ideal product versioning strategy. However, due to the wide spectrum of use-cases for data, it can be a challenging task to reveal the consumer's willingness to pay.

Moreover, the watermarking technique DABIT features some obvious shortcomings in that it requires a minimum quality loss of one-digit precision in order to be implemented, and each data point must carry more digits than the optimal quality gap. Moreover, all watermarked data points need to be kept for reference for watermark detection, and the technique is not robust against time stamp manipulation, aggregation, and the averaging of multiple data streams. Lastly, DABIT requires careful consideration during distribution in order for data streams that are adjacent in quality to not reveal in which data points the watermark has been embedded.

Although we argue that DABIT addresses shortcomings of alternative watermarking techniques, a common denominator for this study and prior works is the general approach and lack of specific use-cases. Thus, future research should attempt to combine successful features of existing models into novel watermarking frameworks designed for more specific scenarios that are relevant for IIoT data sharing. Such frameworks should also be considered and compared in combination with different pricing strategies.

References

Ashton, K. (2009). That 'internet of things' thing. *RFID journal, 22*(7), 97-114.

Bohannan, C. (2007). Copyright Preemption of Contracts. *Md. L. Rev., 67*, 616.

Bröring, A., Schmid, S., Schindhelm, C. K., Khelil, A., Käbisch, S., Kramer, D. & Teniente, E. (2017). Enabling IoT ecosystems through platform interoperability. *IEEE software, 34*(1), 54-61.

Chen, B., & Wornell, G. W. (2001). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory, 47*(4), 1423-1443.

Chong, S., Skalka, C., & Vaughan, J. A. (2010). Self-identifying sensor data. *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 82-93.

Coase, R. H. (1960). The problem of social cost. *Classic papers in natural resource economics*, 87-137.

Cox, I. J., Miller, M. L., Bloom, J. A., & Honsinger, C. (2002). *Digital watermarking* (Vol. 53). San Francisco: Morgan Kaufmann.

Crabtree, A., & Mortier, R. (2015). Human Data Interaction: Historical Lessons from Social Studies and CSCW. *Proceedings of the 14th European Conference on Computer Supported Cooperative Work*, 3-21.

De Cristofaro, E., Ding, X., & Tsudik, G. (2009). Privacy-preserving querying in sensor networks. *Proceedings of 18th International Conference on Computer Communications and Networks*, 1-6.

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review, 34*(2), 193-203.

Duan, Y., Fu, G., Zhou, N., Sun, X., Narendra, N. C., & Hu, B. (2015). Everything as a service (XaaS) on the cloud: origins, current and future trends. *IEEE 8th International Conference on Cloud Computing,* 621-628.

EU (1996). Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

EU (2006). Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights.

EU (2016). Directive 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.

EU (2017), Digital Transformation Monitor, Germany: Industrie 4.0, January 2017.

Fricker, S. A., & Maksimov, Y. V. (2017). Pricing of data products in data marketplaces. *2017 International Conference of Software Business*, 49-66.

Golchha, N. (2015). Big data-the information revolution. *Int. J. Adv. Res*, 1(12), 791-794.

Governatori, G., Rotolo, A., Villata, S., & Gandon, F. (2013). One license to compose them all. *International semantic web conference*, 151-166.

Guth, J., Breitenbücher, U., Falkenthal, M., Leymann, F., & Reinfurt, L. (2016). Comparison of IoT platform architectures: A field study based on a reference architecture. *Cloudification of the Internet of Things (CIoT)*, 1-6.

Heller, M. A. (1998). The tragedy of the anticommons: property in the transition from Marx to markets. *Harvard law review*, 621-688.

Hsu, C. T., & Wu, J. L. (1999). Hidden digital watermarks in images. *IEEE Transactions on image processing, 8*(1), 58-68.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research, 2*(2), 59-64.

Kerber, W. (2017). Rights on Data: The EU Communication 'Building a European Data Economy' From an Economic Perspective.

Kerber, W., & Frank, J. (2017). Data Governance Regimes in the Digital Economy: The Example of Connected Cars. Available at SSRN 3064794.

Kerber, W., & Schweitzer, H. (2017). Interoperability in the digital economy. *J. Intell. Prop. Info. Tech. & Elec. Com. L., 8*, 39.

Khan, M., Wu, X., Xu, X., & Dou, W. (2017). Big data challenges and opportunities in the hype of Industry 4.0. *2017 IEEE International Conference on Communications (ICC)*, 1-6.

Kim, Y., & Chang, H. (2014). The industrial security management model for SMBs in smart work. *Journal of Intelligent Manufacturing, 25*(2), 319-327.

Koutroumpis, P., Leiponen, A., & Thomas, L. (2017). The (unfulfilled) potential of data marketplaces. *The Research Institute of the Finnish Economy, 53*.

Lewis, M. (2014). *Flash boys: a Wall Street revolt.* WW Norton & Company.

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading and protection. *IEEE Access, 6*, 15132-15154.

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, *6*, 1-10.

Luong, N. C., Hoang, D. T., Wang, P., Niyato, D., Kim, D. I., & Han, Z. (2016). Data collection and wireless communication in Internet of Things (IoT) using economic analysis and pricing models: A survey. *IEEE Communications Surveys & Tutorials, 18*(4), 2546-2590.

Ma, Z. (2017). Digital rights management: model, technology and application. *China Communications, 14*(6), 156-167.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: architecture, opportunities, and open research challenges. *IEEE Access, 5*, 5247-5261.

Milham, M., Craddock, C., Fleischmann, M., Son, J., Clucas, J., Xu, H., ... & Colcombe, S. (2017). Assessment of the impact of shared data on the scientific literature. *BioRxiv, 183814*.

Missier, P., Bajoudah, S., Capossele, A., Gaglione, A., & Nati, M. (2017). Mind My Value: a decentralized infrastructure for fair and trusted IoT data trading. *Proceedings of the Seventh International Conference on the Internet of Things,* 15.

Mišura, K., & Žagar, M. (2016). Data marketplace for Internet of Things. *2016 International Conference on Smart Systems and Technologies (SST),* 255-260.

Mital, R., Coughlin, J., & Canaday, M. (2015). Using big data technologies and analytics to predict sensor anomalies. *Advanced Maui Optical and Space Surveillance Technologies Conference.*

Mueller, M. (1999). Digital convergence and its consequences. *Javnost-the public, 6*(3), 11-27.

Murray, G., Johnstone, M. N., & Valli, C. (2017). The convergence of IT and OT in critical infrastructure.

Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2012). Pricing approaches for data markets. *International workshop on business intelligence for the real-time enterprise,* 129-144.

Nadah, N., de Rosnay, M. D., & Bachimont, B. (2007). Licensing digital content with a generic ontology: escaping from the jungle of rights expression languages. *Proceedings of the 11th international conference on Artificial intelligence and law*, 65-69).

Niyato, D., Alsheikh, M. A., Wang, P., Kim, D. I., & Han, Z. (2016). Market model and optimal pricing scheme of big data and Internet of Things (IoT). *2016 IEEE International Conference on Communications (ICC)*, 1-6.

Panah, A. S., Van Schyndel, R., Sellis, T., & Bertino, E. (2016). On the properties of non-media digital watermarking: a review of state of the art techniques. *IEEE Access, 4*, 2670-2704.

Pantelis, K., & Aija, L. (2013). Understanding the value of (big) data. *2013 IEEE International Conference on Big Data,* 38-42.

Perera, C. (2017a). *Sensing as a service for internet of things: A roadmap.* Lulu.com.

Perera, C. (2017b). Sensing as a service (S2aaS): Buying and selling IoT data. *arXiv preprint arXiv:1702.02380*.

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Sensing as a service model for smart cities supported by internet

of things. *Transactions on Emerging Telecommunications Technologies, 25*(1), 81-93.

Peters, M. (2006). The Challenge of Copyright in the Digital Age. *Revista La Propiedad Inmaterial,* (9), 59-68.

Rafaeli, S., & Raban, D. R. (2005). Information sharing online: a research challenge. *SSRN 999993*.

Richter, H., & Slowinski, P. R. (2019). The Data Sharing Economy: On the Emergence of New Intermediaries. *IIC-International Review of Intellectual Property and Competition Law, 50*(1), 4-29.

Ringel, M., Zablit, H., Grassl, F., Manly, J. and Möller, C. (2018) The Most Innovative Companies 2018, *The Boston Consulting Group*, January 2018.

Shapiro, C., Carl, S., & Varian, H. R. (1998). *Information rules: a strategic guide to the network economy.* Harvard Business Press.

Sheng, X., Tang, J., Xiao, X., & Xue, G. (2013). Sensing as a service: Challenges, solutions and future directions. *IEEE Sensors journal, 13*(10), 3733-3741.

Sion, R., Atallah, M., & Prabhaka, S. (2006). Rights protection for discrete numeric streams. *IEEE Transactions on Knowledge and Data Engineering, 18*(5), 699-714.

Spence, M. (1978). *Uncertainty in Economics.* Academic Press.

Stigler, G. J. (1961). The economics of information. *Journal of political economy, 69*(3), 213-225.

Stiglitz, J. E. (1975). The theory of “screening”, education, and the distribution of income. *The American economic review, 65*(3), 283-300.

Sultana, S., Shehab, M., & Bertino, E. (2012). Secure provenance transmission for streaming data. *IEEE Transactions on Knowledge and Data Engineering, 25*(8), 1890-1903.

Tang, R., Wu, H., Bao, Z., Bressan, S., & Valduriez, P. (2013). The price is right. *2013 International Conference on Database and Expert Systems Applications*, 380-394.

Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J. J., & Su, J. K. (2001). Attacks on digital watermarks: classification, estimation based attacks, and benchmarks. *IEEE communications Magazine, 39*(8), 118-126.

Wiebe, A. (2016). Protection of industrial data–a new property right for the digital economy?. *Journal of Intellectual Property Law & Practice, 12*(1), 62-71.

Xiao, X., Sun, X., Li, F., Wang, B., Xia, Z., & Liang, W. (2010). Watermarking-based intellectual property protection for sensor streaming data. *International Journal of Computer Applications in Technology, 39*(4), 213-223.

Yan, Z., Zhang, P., & Vasilakos, A. V. (2014). A survey on trust management for Internet of Things. *Journal of network and computer applications, 42*, 120-134.

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2013). Sensing as a service and big data. *arXiv preprint arXiv:1301.0159.*

Zech, H. (2017). Building a European data economy. *IIC - International Review of Intellectual Property and Competition Law, 48*(5), 501-503.

Ølnes, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing.

Özyilmaz, K. R., Doğan, M., & Yurdakul, A. (2018). IDMoB: IoT Data Marketplace on Blockchain. *2018 Crypto Valley Conference on Blockchain Technology (CVCBT),* 11-19.

Abstract (Korean)

초 록

산업용 사물 인터넷 (IIoT) 데이터가 제품과 서비스를 위한 중요한 고급 데이터 소스로 여겨지고 있지만, 여전히 수 많은 기업들은 불충분한 산업용 사물 인터넷 데이터 공유 시스템으로 인하여 고충을 겪고 있다. 방대한 분량의 산업용 데이터가 제대로 거래되지 못하고 있으며, 이는 데이터의 커다란 가치 손실로 이어지고 있다. 본 연구에서는 서비스로서의 센싱 (Sensing as a Service) 비지니스 모델이 한정적으로 적용되고 있는 원인이 해당 정보의 경제적, 기술적 특징들을 반영하는 디지털 권리 시스템의 부재에 기인한다고 보고 있다. 따라서 본 연구에서는 산업용 사물 인터넷 데이터에 대한 지적재산권 집행 시스템과 데이터 가격산정 모델을 제안하여 산업용 사물 인터넷 데이터 공유 인센티브 문제를 해결하고자 한다.

주요어 : 4 차산업혁명, 산업용 사물인터넷, 서비스로서의 센싱, 산업용 사물인터넷 데이터 거래, 디지털 권리 관리, 디지털 워터마크

학 번 : 2017-21456