Technology Arts Sciences
TH Köln

Improving E-commerce fraud investigations in virtual, inter-institutional teams:

Towards an approach based on Semantic Web technologies

MASTER THESIS

by

Andreas Gerlach

submitted to obtain the degree of

MASTER OF SCIENCE (M.SC.)

at

TH KÖLN - UNIVERSITY OF APPLIED SCIENCES
INSTITUTE OF INFORMATICS

Course of Studies
WEB SCIENCE

First supervisor: Prof. Dr. Kristian Fischer
TH Köln - University of Applied Sciences

Second supervisor: Stephan Pavlovic
TH Köln - University of Applied Sciences

Cologne, August 2016

Contact details:

Andreas Gerlach
Wilhelmstr. 78
52070 Aachen
andreas.gerlach@smail.th-koeln.de

Prof. Dr. Kristian Fischer
TH Köln - University of Applied Sciences
Institute of Informatics
Steinmüllerallee 1
51643 Gummersbach
kristian.fischer@th-koeln.de

Stephan Pavlovic
TH Köln - University of Applied Sciences
Institute of Informatics
Steinmüllerallee 1
51643 Gummersbach
stephan@railslove.com

Abstract

There is a dramatic shift in credit card fraud from the offline to the online world. Large online retailers have tried to establish countermeasures and transaction data analysis technologies to lower the rate of fraudulent transactions to a manageable amount. But as retailers will always have to make a trade-off between the *performance* of the transaction processing, the *usability* of the web shop and the overall *security* of it, we can assume that E-commerce fraud will still happen in the future and that retailers have to collaborate with relative parties on the incident to find a common ground on and take coordinated (legal) actions against it.

Combining information from different stakeholders will face issues due to different wordings and data formats of the information, competing incentives of the stakeholders to participate on information sharing as well as possible sharing restrictions, that prevent making the information available to a larger audience. Additionally, as some of the information might be confidential or business-critical to one of the involved parties a *centralized* system (e.g. a service in the cloud) could **not** be used.

This Master thesis is therefore looking into the topic of how far a computer supported collaborative work system based on peer-to-peer communication technologies and shared ontologies can improve the efficiency and effectivity of E-commerce fraud investigations within an inter-institutional team.

Keywords: Peer-To-Peer Communication, Semantic Web, CSCW

Contents

1 Introduction

This introductory section of the Master thesis will first give a section showing the importance and relevance of the topic in the research area of Web Science, followed by a short description of the problem this thesis will focus on as well as an overview of the outline of the thesis.

1.1 Motivation

> *"When it comes to fraud, 2015 is likely among the riskiest season retailers have ever seen, [. . .] it is critical that they prepare for a significant uptick in fraud, particularly within e-commerce channels."*

This statement from Mike Braatz, senior vice president of Payment Risk Management, ACI Worldwide in (Reuters 2015) shows the dramatic shift in credit card fraud from the offline to the online world, that retailers are starting to face nowadays.

In general credit card fraud can occur if a consumer has lost her credit card or if the credit card has been stolen by a criminal. This usually results in an **identity theft** by the criminal, who is using the original credit card to make financial transactions by pretending to be the owner of the card. Additionally, a consumer might hand over her credit card information to an untrustworthy individual, who might use this information for her own benefit. In the real world scenario there is usually a face-to-face interaction between both parties. The consumer, wanting to do business with a merchant or interacting with an employee of a larger business, has to hand over her credit card information explicitly and can deny doing so if she faces a suspicious situation. The criminal on the other hand must get access to the physical credit card first, before she is able to make an illegal copy of it — a process called **skimming**. The devices used to read out and duplicate the credit card information are therefore called skimmers. These can be special terminals, that the criminal uses to make copies of credit cards she gets her hands on, or they can be installed in or attached to terminals the consumer interacts with on her own (Consumer Action 2009). All of these so-called *card-present transaction* scenarios have seen a lot of improvements in security over the last years. Especially the transition from magnetic swipe readers to EMV chip-based credit cards makes it more difficult for criminals to counterfeit them (Lewis 2015).

As of this criminals are turning away from these card-present transaction scenarios in the offline world. Instead they are focusing on transactions in the online and mobile world, in which it is easy to pretend to own a certain credit card. Most online transactions (either E-commerce or M-commerce) rely ***only*** on credit card information like card number, card holder and security code for the card validation process – as of this these interactions are usually called *card-not-present transactions*. This credit card information can be obtained by a criminal in a number of ways. First she might send out **phishing emails** to consumers. These emails mimic the look-and-feel of emails from a merchant or bank, that the consumers are normally interacting with, but instead navigating the consumers to a malicious web site with the intend to capture credit card or other personal information (Consumer Action 2009). Additionally, criminals can **break into the web sites** of large Internet businesses with the goal of getting access to the underlying database of customer information, that in most cases also hold credit card data (Holmes 2015). Additionally, some of the online retailers are not encrypting the transaction information before transmitting them over the Internet; a hacker can easily start a **man-in-the-middle attack** to trace these data packages and get access to credit card and/or personal information in this way (Captain 2015).

Based on this it should come not as a surprise that the growth rate of online fraud has been 163% in 2015 alone (PYMNTS 2016). This results in huge losses for the global economy every year and it is expected that retailers are losing $3.08 for every dollar in fraud incurred in 2014 (incl. the costs for handling fraudulent transactions) (Rampton 2015). These fraudulent transactions also impact the revenue of the online retailers. Here we have seen a growth of 94% in revenue lost in 2015. Overall it is estimated that credit card fault results in $16 billion losses globally in 2014 (PYMNTS 2016) (Business Wire 2015).

While it is possible to prevent fraudulent transactions in the card-present real-world scenario (mainly due to introducing better technology and establishing organizational countermeasures in the recent past), it is more difficult to do so in the card-not-present online- and mobile commerce scenarios, which are lacking face-to-face interactions and enable massive scalability of misusing credit card information in even shorter time frames (Lewis 2015). Large online retailers have tried to establish countermeasures and transaction data analysis technologies to lower the rate of fraudulent transactions to a manageable amount. But this is still an expensive and inefficient solution to integrate into the retailers' business processes, and is largely driven by machine-learning techniques and manual review processes (Brachmann 2015). Additionally, it can be assumed, that the online retailers are getting into a Red Queen race with the criminals here: with every new technology or method introduced they might just be able to safe the status quo. This is largely due to the facts, that there will be no 100% security for such a complex and interconnected system like an E-commerce or M-commerce shop, the criminals will also increase their efforts and technology skills to adapt to new security features and most importantly retailers will always have to make a trade-off between the *performance* of the transaction processing, the *usability* of the web shop and the overall *security* of it.

1.2 Problem Definition

This Master thesis will **not** look into novel techniques and methods to *prevent* credit card fraud in the E-commerce world. This aspect has been seeing a lot of research in the last years.[1] Instead this Master thesis will look into a **concept to optimize the collaboration** between the affected stakeholders in case of an existing credit card fraud in an E-commerce system.

Stakeholders might include **vendors** and other businesses, that the retailer has a long-term business relationship with, **law enforcement agencies**, **payment service providers** like PayPal or Visa, **banks**, and even **competitors**, that are also affected by the Internet fraud. In such a case the merchant usually tries to solve the issue on his own and getting in contact with relative parties by phone or e-mail if necessary. But these communication styles do not fit to the complexity of the task involved, and based on the media-richness model (see Figure 1.1) will result in inefficient and ineffective problem solutions.

Due to the task complexity a **physical face-to-face meeting** with representatives of all involved stakeholders might be a good fit, but arranging such a meeting (same time, same place) with multiple parties, that are globally dispersed, is either economically not feasible or takes a lot of time. But the more time passes for investigating the crime the more difficult it will become to find the criminals and take legal actions against them, which can also reduce the risk of losing the stolen money completely.

[1] please also note the various US patent applications of Google on that matter from 2015, e.g.: "Credit card fraud prevention system and method", "Financial card fraud alert", "Payment card fraud prevention system and method" (Google Patents)

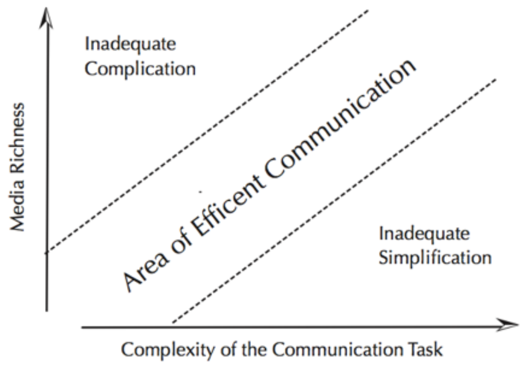

Figure 1.1: The Media Richness Model (Rice 1992)

As of these conditions a **computer-supported collaborative work** (CSCW) system might be an alternative to *cooperate* on an incident of E-commerce fraud (same time, different place). CSCW systems can be categorized by their support for the mode of group interaction as done in the 3C model:

- **communication:** two-way exchange of information between different parties
- **coordination:** management of shared resources like meeting rooms
- **collaboration:** members of a group work together in a shared environment to reach a goal

Based on the level of support for one of these functionalities the various systems can be classified and described (see Figure 1.2) (Koch 2008):

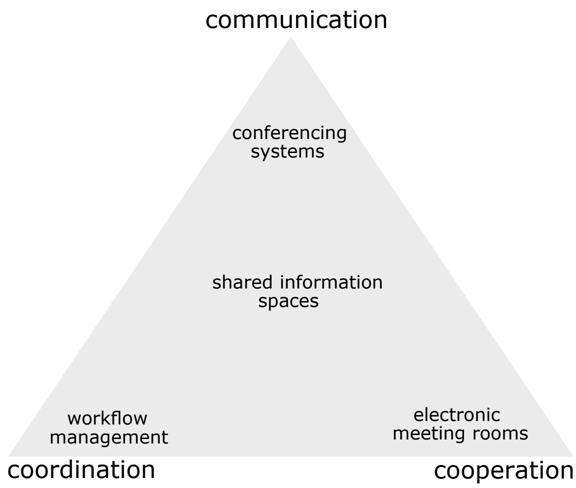

Figure 1.2: The 3C Model (Koch 2008)

A good candidate *could* be a **shared information space**; aka team rooms, cloud storage services or document management systems, that allow to access information at any place, any time and to share information with co-workers — usually with a build in versioning support for artefacts and a workflow component.

However as some of the required information might be confidential or business-critical to one of the involved parties a **centralized system** (e.g. a service in the cloud) could **not** be used in the scenario described here. Another key characteristic of the investigation of an E-commerce fraud is, that it involves information sharing from many different organizations. These different aspects have to be combined into a **shared information space** in a meaningful way to be able to achieve the common group goal on time. Combining information from different stakeholders will face issues due to **different wordings and data formats** of the information, **competing incentives** of the stakeholders to participate on information sharing as well as possible **sharing restrictions**, that prevent making the information available to a larger audience.

Decentralized information sharing architectures, that utilizes **peer-to-peer communication technologies**, are either restricted to a commonly agreed set of data entities and relations (based on an ontology) between all involved parties or are lacking richer semantics for sharing and integrating content between the stakeholders. **Semantic Web technologies** can help lower the barrier to integrate information from various sources into a shared information space, and the advantages of peer-to-peer communication and Semantic Web technologies for information sharing in distributed, inter-organizational settings have been shown in (Staab & Stuckenschmidt 2006).

Still these studies concentrate on making information from different parties searchable and accessible in a distributed, shared information space, which data can be accessed and queried at any time from any participating party. They are not solving the problem of working collaboratively on a common goal in an ad-hoc, loosely-coupled virtual team of disperse organizations by making certain (sometimes sensitive) information available in a shared environment.

Therefore, the **research question** for this Master thesis can be summarized as:

> *In how far can a computer supported collaborative work system based on peer-to-peer communication technologies and shared ontologies improve the efficiency and effectivity of E-commerce fraud investigations within an inter-institutional team?*

1.3 Master Thesis Outline

Before starting with the investigation of E-commerce fraud incidents and their possible examinations the thesis will first analyse related works, that have been looked into during the course of this Master thesis and have had an influence on it.

In the next section, Context Analysis, the thesis will discuss the E-commerce scenario in detail. It starts with a description of the E-commerce shopping process, looks into the stakeholders involved, shows possible kinds of E-commerce fraud incidents and how they are handled today. Based on these discussions the chapter will close with a presentation of the specific scenario selected for further examination within this Master thesis.

After this initial scope setup the thesis will briefly outline the theoretical foundations required for the understanding of the concepts in the solution space. This section starts with a short overview of the important facets of computer-supported collaborative work systems (CSCW), shows the essential specifications of the Semantic Web and ends up with an introduction to the peer-to-peer (P2P) communication techniques and protocols.

Last but not least the conceptualization of a collaborative system, that supports the investigation of E-commerce fraud incidents, takes place. This chapter will lay out and discuss the possibilities for designing and using such a system. The objective is to come up with an approach at the end of this chapter, that might be the best fit for the problem described in the scenario at the beginning.

To conclude the thesis will also show an outcome of the paper work as well as give an outlook, that might be beneficial to decide future progress on this topic.

2 Related Works

2.1 E-commerce Fraud Scenario

- “Fraud in Non-Cash Transactions: Methods, Tendencies and Threats.” (Sobko 2014)
- “Overview of E-Commerce” (Ankhule & Joshy 2015)
- “A Survey on Fraud Detection Techniques in Ecommerce” (Rana & Baria 2015)
- “A Study on E-Commerce Security Issues and Solutions” (Sen et al. 2015)

2.2 Computer Supported Collaborative Work

- “Effects of Sensemaking Translucence on Distributed Collaborative Analysis” (Goyal & Fussell)
- “CSCW and enterprise 2.0 - towards an integrated perspective” (Koch 2008)
- “A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network” (Yang & Chen 2008)
- “Paradox of richness: A cognitive model of media choice” (Robert & Dennis 2005)
- “From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW” (Lee & Paine 2015)

2.3 Peer-To-Peer Communication

- “SWAP: Ontology-based Knowledge Management with Peer-to-Peer Technology.” (Ehrig et al. 2003)
- “RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network” (Cai & Frank 2004)
- “P2P networking: An information-sharing alternative” (Parameswaran et al. 2001)
- “Introduction to XMPP protocol and developing online collaboration applications using open source software and libraries” (Ozturk 2010)
- “Peer-to-peer systems” (Rodrigues & Druschel 2010)
- “Leveraging WebRTC for P2P content distribution in web browsers” (Vogt et al. 2013)
- “Let our browsers socialize: Building user-centric content communities on webrtc”

(Werner et al. 2014)
- “Taking on WebRTC in an enterprise” (Vogt et al. 2013)

2.4 Semantic Web

- “Semantic web technologies for the financial domain” (Lara et al. 2007)
- “Security ontology: Simulating threats to corporate assets” (Ekelhart et al. 2006)
- “Marvin: Distributed reasoning over large-scale Semantic Web data” (Oren et al. 2009)
- “Applying Semantic Technologies to Fight Online Banking Fraud” (Carvalho et al.)
- “The Semantic Web-Based Collaborative Knowledge Management” (Chao et al. 2012)
- “Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations.” (Ashraf et al. 2011)
- “Rule interchange on the web” (Boley et al. 2007)
- “Data linking for the semantic web” (Scharffe et al. 2011)
- “Integrating agents, ontologies, and semantic web services for collaboration on the semantic web” (Stollberg & Strang 2005)
- “GoodRelations Tools and Applications” (Hepp et al. 2009)
- “Drawing Conclusions from Linked Data on the Web: The EYE Reasoner” (Verborgh & De Roo 2015)
- “Schema.org: Evolution of structured data on the web” (Guha et al. 2016)
- “Goodrelations: An ontology for describing products and services offers on the web” (Hepp 2008)
- “A functional semantic web architecture” (Gerber et al. 2008)
- “Towards a financial fraud ontology: A legal modelling approach” (Kingston et al. 2004)
- “Complete query answering over horn ontologies using a triple store” (Zhou et al. 2013)

3 Context Analysis

This chapter will look into the scenario of E-commerce fraud investigation in detail. It will start with an in-depth description of the E-commerce scenario followed by an analysis of the stakeholders involved. It will further describe the kind of information each stakeholder has in her local context and her objectives to take part on the information sharing and collaboration initiative. Based on the analysis of the possible kinds of E-commerce fraud incidents and the current process of their investigation, the chapter will end with a description of the scenario selected for this Master thesis.

3.1 An overview of E-commerce

E-commerce as a term relates to the trading of products or services utilizing a computer network such as the Internet. It is usually categorized into the following four different subfields (Sen et al. 2015):

1. **Business-To-Business (B2B)**: refers to electronic trading between companies with the objective to improve their supply chain processes
2. **Business-To-Consumer (B2C)**: refers to electronic trading between a company and it's consumers (most publicly known example for it is Amazon.com)
3. **Consumer-To-Consumer (C2C)**: refers to electronic trading between consumers (most publicly known example for that is eBay)
4. **Consumer-To-Business (C2B)**: referes to electronic trading between consumers and businesses (most publicly known example for this is TaskRabbit)

This Master thesis will **solely** focus on the **B2C** aspect of E-commerce. In that case a **consumer** is using an **E-commerce shop** of a **merchant** on the Internet to order products or services online. The merchant is offering a catalog of available products or services on the Web, that is available and accessible by the general public and usually has an at least nation-wide if not global reach. The merchant can either run the E-commerce shop software on her own servers (on-premise) or can outsource this additional sales channel to a 3^{rd} party hosting company or **cloud service provider**

(CSP). Also the E-commerce shop software itself can either be developed by the merchant in-house or acquired as a boxed product from an **Independent Software Vendor (ISV)** on the market. For business accounting purposes the merchant also runs a bank account with the **acquirer** (see Figure 3.1).

When placing an order with the merchant online the consumer is normally using a **credit card** for finalizing the transaction. This credit card has originally been handed out by the **issuing bank** to the consumer. Additionally some online shops make it mandatory for the consumer to create an user account with them while others do not. The former is the preferred way when consumers are repetitively buying from that merchant, whereas the latter might be used for one-time or irregular shopping trips online. To be able to connect to the Internet the consumer also relies on a service of an **Internet Service Provider (ISP)**. The whole initial setup for participating on E-commerce activities is found in Figure 3.1.

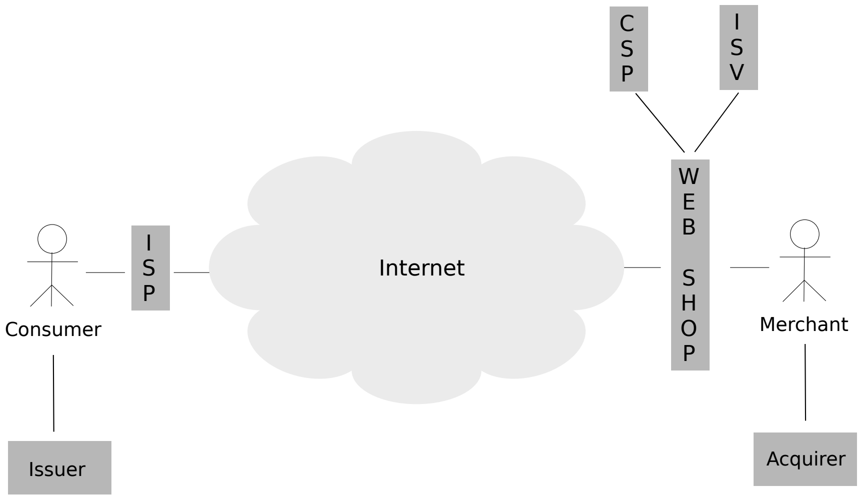

Figure 3.1: E-commerce Fundamentals

When the consumer places the **order** online, the merchant receives at least a **list of products or services** from the current shopping cart of the consumer, the **identification** of the consumer as well as the **delivery address** to ship the physical items to. If the transaction is going to be finalized with a credit card, the consumer will have to provide additional information like her **billing address** and **credit card information** (including the credit card number, the expiry date and the security number of the card).

The merchant usually does not validate the credit card information on her own. For that purpose she is relying on another 3^{rd} party service offered on the Internet by the **Payment Service Provider (PSP)**. These providers are either validating the credit card information themselves based on an user profile the consumer has with the PSP (e.g. a globally available Web service like PayPal) or are connecting to the issuing bank of the card for doing so. For this validation process the merchant is handing over the billing information to the PSP incl. the credit card information given by the consumer.

Either the PSP or the issuing bank is validating the correctness of the information with criterias like:

- is the billing address matching the current consumers' postal address on file?
- is the stated credit card information correct?
- is the credit card still valid?
- is the credit card not marked as being blocked?

The merchant will receive the **status** of the authorization as well as a **payment token** in return. If the authorization was done successfully the merchant will collect the items and send out a shipping request to one of the available **Logistic Service Providers (LSP)** capable to handle the delivery of the order. They will pickup the items at the merchant's facility and ship them to the delivery address stated by the consumer. Usually in parallel the merchant is informing her bank about the order, amount due as well as payment token from the PSP. The acquirer is in charge to withdrawal the amount of the order from the consumer's bank account either via the PSP or directly from the issuing bank, depending on who of them has authorized the initial payment request (a process called clearing) (Visa Europe 2014). The sequence of activities within an E-commerce checkout process is visualized in Figure 3.2.

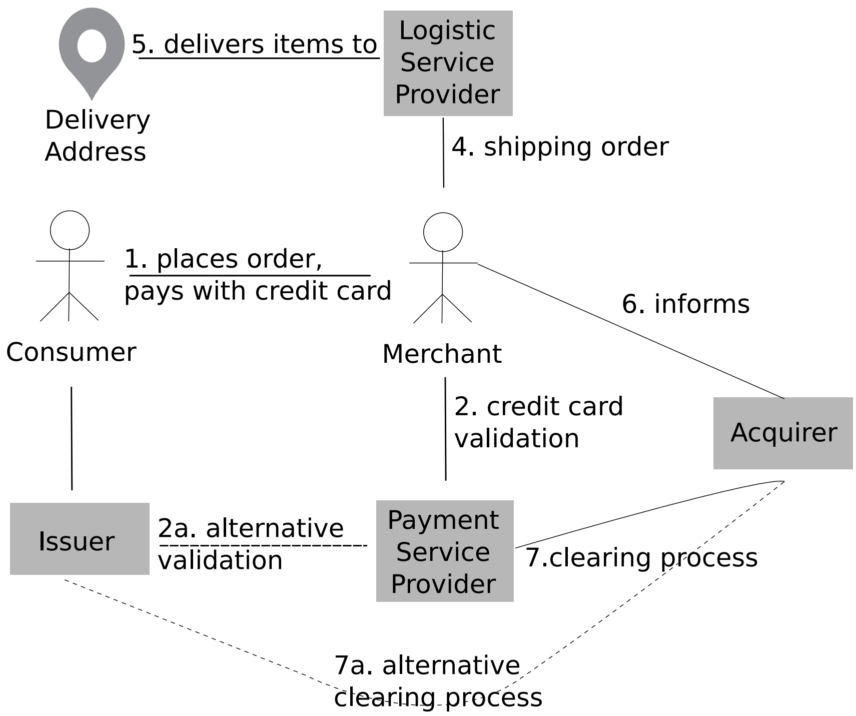

Figure 3.2: E-commerce Checkout Process in detail

3.2 Stakeholders

The following section will look at each stakeholder involved in the E-commerce scenario in detail, lists the kind of information they own or provide to others as well as describes the role of each stakeholder in the E-commerce fraud investigation process (if any).

3.2.1 Consumer

The **consumer** is the initiator of an E-commerce transaction. She is using the shop of a **merchant** on the Internet to order products or services. For doing so she has to know the URL of the Web shop, has to be connected to the Internet via the **ISP** and has to use a standard software called a Web browser on her computer. For the duration of her online browsing session she also owns an unique **IP** address handed out from the ISP.

She might have had a long-term business relationship with the merchant and already owns an **user account** on the Web shop. On the other hand she might be just interested into a one-time shopping trip and might want to order the items without creating an account first — sometimes also called “anonymous shopping” in the E-commerce shops.

The consumer is also having a **bank account** and at least owns a **debit card** with the **issuing bank** to get access to the money on that account. In addition to that she can also hold multipe **credit cards**. A credit card can be issued by the same bank or can be provided by another financial service (e.g. American Express). In any case the organisation that has handed out the credit card to the consumer is called the **issuer**.

If she is going to order items in a Web shop, she will usually browse the product and service offerings of the merchant first and put the items of interest into the **shopping cart**. When finalizing the transaction she has to state the following information to the merchant:

- personal information incl. given name, family name and date of birth
- the address the items should be shipped to
- payment information incl. type of payment and billing address (if different to shipping address)

If she is going to end the transaction with a payment of type **credit card** she will also have to provide specific information of the card that should be used for the payment:

- the owner of the credit card (if it is not belonging to herself)
- the unique credit card number
- the expiry date of the credit card (in format MM/YY)
- the security code of the credit card

The consumer is having a special role in the whole scenario. As the merchant has to deal with the consumer without any face-to-face or real-world interaction online, the consumer is also the least thrustworthy party from the point of view of the merchant. As the Section 3.4 will show the consumer is the main object of question in the case of an E-commerce fraud incident. For the investigation of it the consumer is usually not taking any active part.

3.2.2 Merchant

The **merchant** offers products and services on the Internet to the general public. She might use the Internet as an additional sales channel or rely on it solely for making any business. To provide access to the **Web shop** the merchant has to register a domain name and an **URL** with a local domain name registry. This specific URL refers to a fixed public IP address, that the server, that runs the Web shop software, uses. Normally the merchant does not operate the servers herself, but rely on a service offering from a hosting or **cloud service provider** for that. Also the Web shop software itself is usually not provided by the merchant, but bought from an **ISV** on the market. In any case the merchant has special responsibilities in the Web shop as she has to take care to configure the products, prices, promotions, payment and shipment services available. In addition products can be categorized into departments and subdepartments for easier navigating and searching the offerings in the online shop.

The merchant can decide whether she restricts ordering of products to registered users only, or allows anonymous users too. The main benefit of the former is the possibility to analyse the **shopping behaviour** of individual consumers, whereas the latter will open the business for a wider range of consumers, as it includes also those, that do not want to register with any existing online shop. Nevertheless any consumer activity on the online shop is tracked in the analytic databases of the merchant. This includes not only the items, that have been placed into the shopping cart, but also any product that has been looked at by the consumer during a shopping session. Even if these detailed analytic capabilities are actually a synonym for their usage in target-related advertising, they can also help to decide if a consumer behaves normally or not.

Any business transaction that the consumer makes with the merchant is stored in the merchant's databases. A transaction information contains, but is not limited to:

- the personal information of the consumer
- the address the items will be shipped to
- a collection of products with quantities and prices
- the total amount of the order considering promotions, taxes and fees
- the selected payment information

If the consumer pays with credit card the merchant do not handle the payment herself, but relate this action to a **Payment Service Provider**. To initiate the credit card authorization request, the merchant is sending the following information to the Web service endpoint of the PSP:

- consumer's billing address

- given credit card number, expiry date and security number
- identification of the merchant
- final amount of the transaction actually being processed

In return of the payment authorization the merchant will receive and store these payment-related information for the transaction:

- the type of credit card used (e.g. Visa, MasterCard, American Express, ...)
- the name of the credit card owner
- the payment token received by the PSP
- the timestamp and result code of the authorization
- the authority who approved the payment (if the merchant works with multiple Payment Service Providers)

As a merchant will collect a lot of personal and payment-related information over time, she is also one of the major sources of possible data leaks in this scenario. Due to this circumstance the Payment Card Initiative, a group of banks, issuers and PSPs, provides rules and guidelines (aka PCI/DSS standards) for securely handling these kinds of information in an IT system (DSS 2014).

A merchant is one of the main actors in the fraud investigation process. She is interested in figuring out whether the consumer's transaction is valid or not. As in a case of an E-commerce fraud the merchant will mostly have to cover the costs of the incident (see Section 3.4). Also the online merchant's reputation will suffer, if private information from her databases get leacked. If a merchant falls victim to a fraud incident multiple times, the economic damages can finally result in a bankruptcy of the merchant.

3.2.3 Payment Service Provider

The **Payment Service Provider** is offering payment-related services to online merchants. As of this the PSP provides a common Web interface, that the merchant has to communicate with for sending payment authorization requests (see above). The PSP might be able to authorize the payment request on her own or have to route the request to the corresponding **issuer** of the credit card in question. For the former procedure the PSP has to run an own database of registered users with their credit card information (e.g. PayPal). For checking the credit card and authorizing the payment the **merchant** is sending the following information from the transaction:

- credit card owner incl. billing address given
- credit card number
- credit card expiry date
- credit card security number
- identification of the merchant
- current transaction total amount

The PSP has to securely process these information and return the result of the validation to the merchant. The result message also contains an unique payment token, that the merchant can refer to later to initiate the clearing process. As of this the PSP has to persist the credit card and transaction related information in her own backend databases. Following industry standards she should do so according to the PCI/DSS guidelines mentioned above.

The level of activity in the E-commerce fraud investigation process depends on whether the PSP authorizes the payment herself or only acts as routing service between the merchant and the original credit card issuer. In the former case the PSP is more actively involved. In that case she also holds more of the valuable information to analyze the incident. In the latter case she will be able to connect the payment-related request information from the merchant with the related authorization result coming from the issuer.

If the PSP holds sensitive information in her own databases she might also be a source of data leaks. In that case she should put the same precautions in place as an issuing bank will have to do (see next section).

3.2.4 Issuing Bank

The **issuer** is one of the parties in the scenario that knows the owner of the credit card in person. Each individual has to register personally with the issuer to get access to a credit card. This includes providing the following information:

- personal information like given name, family name and date of birth

- the currently registered home address
- the bank account that should be used to settle credit card balances

Even if the two parties do not really meet each other personally, an individual will still have to identify with a valid id card and bank account to receive and activate a new credit card. Beside being the single source of truth about the original credit card owner, the issuer of the card also collects and stores all usages of it. The issuer therefore can provide individual credit card usage patterns, that are not just limited to the online shopping scenario — something a Payment Service Provider can deliver too; but also include transactions the card owner does in the real-world. Needless to say that these are valuable information for the E-commerce fraud investigation.

Still the issuer does not know the details of the transactions, that have been made with the credit card yet. As shown in the section about the PSP above the issuer will just receive an identifier of the merchant, in whose shop the credit card has been used. Based on public available information from a commercial register about the merchant the issuer could at least come up with the retail branch the merchant operates in.

Being the single source of truth about all issued credit cards, their owners and usage patterns the issuer is another high-risk for data leaks. She should as well follow the guidelines from the PCI/DSS standards, should incorporate security standards for her IT systems and processes as well as monitor her backend systems actively with an intrusion detection mechanism.

3.2.5 Acquiring Bank

The **acquirer** holds the bank account of the merchant and is responsible for withdrawing the outstanding amounts of transactions from the accounts of the consumers, or more precisely the issuing bank of each consumer. As of this the acquiring bank is usually not processing any credit card related information from consumers, but refer to the payment tokens that have been given by the PSP or issuer during the authorization process.

Still as a financial institute the acquirer (like the issuer) has to comply with the rules and guidelines of the PCI/DSS and other industry standards to make sure, that her bank accounts and the transaction processing are safe and secure. The detailed analysis of these techniques and procedures as well as possible banking fraud incidents are out of scope of this Master thesis though.

3.2.6 Logistic Service Provider

The **Logistic Service Provider** has two important roles in the E-commerce scenario. First she has access to and control over the items of the merchant for the duration of the transport between the merchant's facility and the consumer's shipping address. And second she holds the information to whom she has handed over the items at the final destination. Although the LSP has nothing to do with any payment-related activities, she is still critical as she will be the last chance for the merchant to stop the delivery of the items (in case a fraud has been detected after initiating of the shipment) or provide information about the person, that has received the items at the shipping address — especially on high-priced goods, which usually require the recipient to show her personal id card and place a signature on the delivery receipt.

For initiating the shipment procedure the merchant is ordering a certain transport service from the LSP and hand over the following information:

- name of the receiver
- delivery address
- list of items to be shipped
- optionally: value of the items if an insurance policy is taken

The LSP at the other hand returns an unique **tracking id** for the shipment. It can be used by the merchant and the consumer to check the status of the shipment online.

As the LSP does not have to deal with the payment-related activities in the E-commerce scenario, she is also not actively involved in the fraud investigation. Still she is of help if an incident is found, as she can stop the delivery or provide useful information about the recipient.

3.2.7 Cloud Service Provider

The **Cloud Service Provider** offers IT services to its customers. These IT services include hardware and software assets, that in the E-commerce scenario a merchant can order to run her Web shop on the Internet. Part of the service level agreement between the merchant and the CSP is a detailed listing of the responsibilities of both parties (who has to take care of what). In most cases the merchant is outsourcing the complete operation of the hardware and software for the Web shop to the CSP; so the CSP will be responsible for making sure the Web shop is available and secure. The

CSP is also constantly monitoring the incoming connections to each public Internet server under her control and can provide information, whether a Web shop of one of the merchants has been compromised or not.

3.2.8 Independent Software Vendor

The **ISV** designs, implements and sells the Web shop software. She has detailed knowledge about the software components and libraries used within the Web shop and checks them regularly for security breaches or vulnerabilities. She also has to verify the software code implemented on her own for vulnerabilities, and has to make sure that the implementation follows industry standards (e.g. PCI/DSS for handling person and payment related information). As of this she can best assert these quality criterias of the Web shop software if needed.

3.2.9 Internet Service Provider

The **ISP** provides a service to the consumer for connecting to the Internet. Each Web request the consumer is doing on her system is routed to the public Internet via the infrastructure of the ISP. Due to regulations and laws the ISP has to store the log files of any Internet session of its customers for a certain amount of time. Especially these log files can be helpful to decide whether a consumer was visiting pages in the dark-side of the Web, or if she falls victim to some phishing attacks (explained later in Section 3.4).

3.3 Data flow for credit card transactions

As the previous chapter shows there are a many stakeholders involved in providing IT hardware, software and services to keep the Web shops on the Internet running. Only a small fraction of those will have to deal with the handling of credit card payments and order fulfillment. These are the critical stakeholders to look at in the case of E-commerce fraud incidents and the actual flow of information between them is displayed in Figure 3.3.

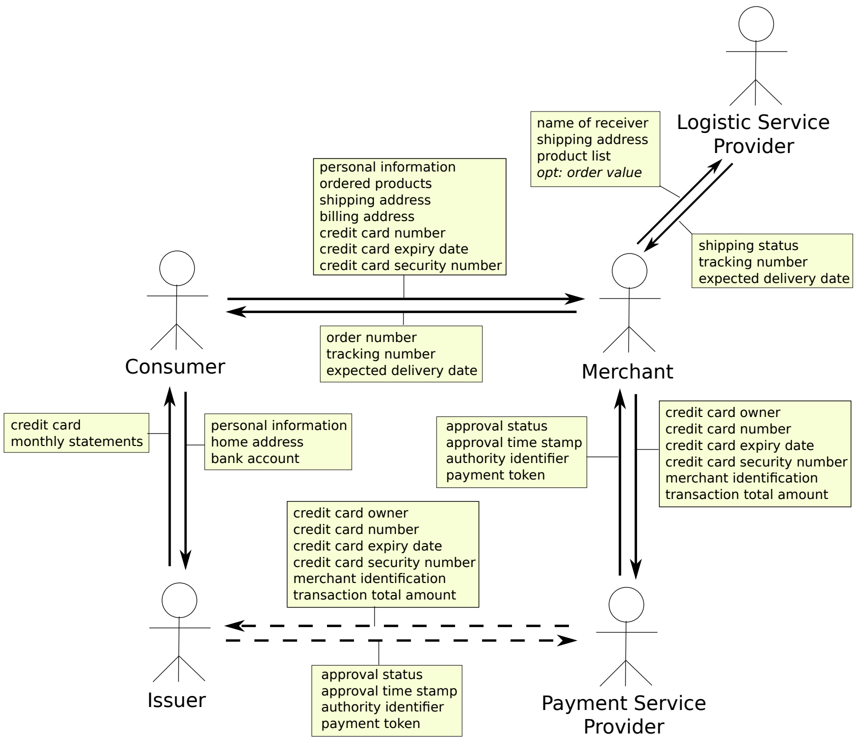

Figure 3.3: Stakeholder and Data Flow in E-commerce scenario

3.4 E-commerce fraud incidents

Based on the previous sections one can come up with strategies a fraudster might use to trick the E-commerce system. To do so the criminal will have to get access to credit card information in the first place. Therefore this section first looks into ways a criminal might get access to credit card and personal information in the E-commerce scenario. After that the section describes possible strategies a fraudster can use to trick the system. The section will end with a discussion of the E-commerce fraud incident handling as it is in place today.

3.4.1 Credit Card data breaches

In the Section 3.3 one could already find the parties, who have access to or store credit card information in the E-commerce scenario, namely:

- the consumer as owner of the credit card
- the issuing bank, who handed out the credit card to the consumer
- the merchant, if the consumer is paying with a credit card

- the Payment Service Provider, if the consumer is paying with a credit card online

The **PSP** does receive the credit card information from the merchant during the authorization of the payment. If the PSP does the authorization herself, she is also the party to store and hold the credit card information in her backend databases. As mentioned earlier the PSP should follow industry standards and guidelines for storing and processing payment related information, especially the PCI/DSS standard. Also she is responsible for monitoring her systems with an intrusion detection program. This will trigger a signal as soon as an hacker got access to the internal databases. In that case the PSP can put the leaked credit card information on an internal blacklist, so that these cards could no longer be used for further payments online. Additionally she will have to send a message to the affected issuers, to which the PSP generally maintains a strong business relationship. The issuer will inform the affected owners and send out new credit cards to them. Due to this procedure in place, one can assume that the safety and security of credit card handling at the PSP can be guaranteed.

The **merchant** receives the credit card information during the checkout process from the consumer. The credit card information is transfered via the public Internet from the consumer to the merchant and could be a victim of a man-in-the-middle attack, in which the hacker is intercepting the communication between the consumer and the merchant with the objective to capture the personal and payment-related information from the data transmission stream. Therefore the merchant should offer the Web shop only via a secure connection using industry standards like TLS to encrypt the information send between both parties. This will make it more difficult for an attacker to get to the plaintext information exchanged between consumer and merchant.

As the merchant is not processing the credit card information directly, she also do not have to store them in her own backend databases. The merchant is asking the PSP or issuer of the card for authorization of the credit card payment and receives an **unique payment token** in response, if the authorization was successful. As stated in the PCI/DSS standard a merchant should never store the whole credit card data, but should use this payment token and shortened credit card information (especially abbreviated credit card numbers) to refer to the single payment later. Due to this one can conclude, that breaking into the systems of a merchant will not result in any leaked credit card information, if the merchant follows these PCI/DSS standards.

The **issuer** is a valuable target for hacking into the backend systems with the objective to leak a massive amount of credit card and owner information. As a financial institute the issuer have to follow a huge set of regulations and safety procedures to be able to participate on the market. It can be assumed that at least the same safety mechanisms are valid as for the PSP are in place. This means constantly monitoring the internal systems with an intrusion detection mechanism and blacklisting any leaked credit card. In addition to the monitoring of all online activities (as also the PSP does) the issuing bank can monitor activities done with the credit card in the offline world too. In case of suspicious activities the credit card can be blocked and a new one will be send out to the credit card owner.

The **consumer** is also a valuable target for eavesdropping on credit card and personal information. She is also the weakest and most unsecure party in the whole E-commerce scenario. As said above a lot of the protection mechanisms of the other parties are relying on implementing industry standards and on constantly monitoring the own systems for malicious activities. This can not be securely said about the computer of the consumer. Whether she is using up-to-date security programs (e.g. an Antivirus tool and a firewall) on her computer or not is out of reach of the other parties to verify. Additionally a consumer can fall victim to a phishing attack, that will send her to a malicious Web site with the intend to get her personal information. In some seldom cases the consumer might cooperate with the fraudster or might be the fraudster itself with the intend to trick the system for her own interest. As of this the E-commerce fraud investigation can not rely on information from the consumer, but instead has to figure out, if the transaction in question was made from the real owner of the credit card or from a frauster.

3.4.2 E-commerce fraud strategies

After a fraudster has got access to leaked credit card information she can come up with the following strategies to trick the E-commerce system:

1. a fraudster owns **one** leaked credit card information and try to use it for ordering products from **multiple** merchants on the Internet
2. a fraudster owns **multiple** leaked credit card information and try to use them for ordering products from **one** merchant on the Internet
3. a combination of the two cases mentioned before, that can also be related to as a series of the first fraud activity

In the **first scenario**, in which the fraudster is trying out a leaked credit card for ordering products on Web shops of various merchants online, each of the merchant only sees the transaction that takes place in her system. It will make it more difficult for the merchant to decide whether this is a fraud transaction or not, as she is not aware of the attempts the fraudster has done on other merchant's Web shops.

As each merchant will rely on a PSP or issuer to verify the credit card payment, it is in the responsibility of these parties to recognize fraud transactions in this scenario. To be able to do this the PSP and also the issuer are monitoring the usage of credit cards and are actively looking for suspicious activities. The fraud prevention mechanisms in place are mostly working on a rule-based, and in some cases also score-based systems running in the internal network of the PSP and issuer. These systems are fed with the information the merchant sends with the payment authorization request and will come up with either:

1. Yes, this looks like a fraudulent transaction and will be blocked
2. No, this seems to be a valid transaction and will be acknowledged
3. Maybe, this transaction might be valid, but there is some uncertainty in the validation of it. These edge cases are routed to a human operator of the PSP or issuer to decide on how to proceed with it

As a recent study shows the success rate of the fraud prevention systems heavily relies on the techniques used to validate the transaction data (Rana & Baria 2015). The outcome is, that ca. 70 to 80 percent of the fraudulent transactions will be recognized as such and blocked successfully. That still means 20 to 30 percent of fraudulent transactions could not be recognized as such. For handling these cases the organisations employ special trained staff, that is operating 24/7 and 365 days a year.

As stated in the introduction of this Master thesis there is a shift from the offline credit card fraud to the online world. This is also resembled in current figures of E-commerce frauds, whose show that it makes up ca. 85 percent of all credit card fraud attempts and have on average a transaction amount of 500 to 600 EURO.

As the PSPs and the issuers do not have any order details, they can only decide on the information given during the authorization request (see Section 3.2). At most they can validate the branch the merchant is operating in, and it might come as no surprise that the fraudsters are regularly using Web shops of merchants, who offer either electronics, clothings, entertainment- or travel-related products. These are also the most commonly used sources of **valid** E-commerce transactions and therefore make any fraudulent transaction very difficult to detect.

At the end it might be the owner of the credit card, who detects suspicious activities on her credit card account and informs the issuing bank about it. Based on current regulations and laws the issuing bank has to rollback the fraudulent transaction on request of the consumer, which means that the merchant will have to cover the costs of the E-commerce fraud (as she is not receiving the money for the products that has been already shipped to the fraudster).

Looking at the **second scenario** of the E-commerce fraud strategies at the beginning of this section, a merchant will receive multiple requests from a fraudster, who is trying out various leaked credit cards for finishing an order process. This kind of E-commerce fraud can be recognized at the systems of the merchant based on the same source IP address of the requests or due to having the same shipping address for orders with different credit cards. Due to that one can conclude, that also merchants must take an active role in the fraud prevention process (if they do not do so already) and try to minimize the amount of fraudulent transactions taken place on their Web shops. As this scenario is likely be manageable with additional fraud prevention mechanisms at the merchant without the need to involve other parties of the E-commerce scenario, this second scenario falls out of scope of this Master thesis.

3.4.3 E-commerce fraud incidents handling

If the fraud prevention systems at the PSP or issuer are detecting a suspecious transaction an operator of a special department within the organisation will be informed about the transaction via a notification on his computer. This operator will have to decide if the transaction looks valid and should be acknowledged or seems to be fraudulent and has to be denied. To be able to decide this she is going to look into the usages of the credit card in question for the near past. Whereas it will be easy to recognize that a credit card, that was just being used in a shop in Germany could not be used in a shop in US or Asia within a short timeframe due to physical constraints in the real world, the same consumer can order products from an US or Asian online retailer with ease within minutes. So these initial geographical contraints, that work well with real world usage patterns of credit cards (a strategy called geo-fencing), will no longer work in the E-commerce scenario.

So the operator has to found her decision on the transaction information at hand. Initially she can check for the amount that should be paid with credit card. One can assume that small amounts will be covered by the PSP or issuer, who will take over the risk for a false authorization. With an increased value of the items ordered the PSP or issuer is putting back the risk to the merchant in case of a customer complaint later. At a second glance she can verify whether the consumer has had any business relationship with the merchant in the past as well as validate the retail branch the merchant operates in. Though these are weak hints for investigating the validity of an E-commerce transaction (see above).

To make a solid decision the operator will have to get in contact with the merchants the credit card has been used with recently and ask for additional information like:

- does the consumer owns an user account with the Web shop?
- does the shipping address matches the billing address for the order?
- if not has the user ordered products to this shipping address in the past?
- what has been ordered, incl. detailed product information like brand, model, product categories, ...

In some cases the PSP or issuer might have had a business relationship with the online merchant in the past and already knows the support personell from the merchant to contact for querying, but in most cases the contact person might not be known to the PSP or issuer resulting in asking the general support via the contact formulars on the merchants Web site.

Executing this procedure with multiple online merchants from different countries will be time consuming and raise a lot of efforts at the PSPs, issuers and merchants. Getting the right information might take time as the correct addressee from the support department of the merchant is unknown, the merchant do not have specialized staff at hand to handle these kind of issues, or there are misunderstandings on handling the case due to language barriers or different incentives. Therefore one can assume that this detailed analysis of any suspecious transaction will not take place today, but most of these transactions will be acknowledged after a first short look at them.

Still the merchants as well as the PSP and issuer have a high incentive for keeping the success rate and numbers of these fraudulent activities low. For the PSP and issuing bank there are regulations stating that at maximum only 1 thousands of the overall transactions (numbers for the EU) should be fraudulent. This keeps the pressure on these financial instituts to invest in fraud prevention techniques for being able to stay in business. For the merchant it is of high interest that a fraudulent transaction can be resolved before the fraudster receives the ordered products. In the worst case scenario of just one successful fraudulent transaction in an E-commerce shop, past experience shows that this will trigger hundreds if not thousands of following attempts from other fraudsters.

3.5 Scope of this Master Thesis

As laid out in the previous section the most interesting E-commerce fraud scenario is the one, in which a fraudster uses a leaked credit card information to order products or services from various merchants on the Internet. This is currently most likely to be successful as there is a lack of information on the side of the merchant as well as the PSP or issuer. Each of the affected merchants just noticed the single transaction that takes place on her Web shop, without knowing about the other attempts the fraudster do on other Web shops. The PSP and issuer will notice the active use of the credit card on different Web shops though, but do not have any information about the transaction details. Therefore they could not correlate these information to check for suspicious activities.

Based on the current strategies of the fraudsters, whose will use the leaked credit card information in commonly used Web shops that deal with electronics, entertainment or travel-related products and services, it is more likely that these fraudulent transactions will not be recognized on time by the existing fraud prevention mechanisms in place.

A simple approach to solve this issue would be to just share more information of the ongoing transaction between the merchants, the PSPs and the issuers. This might fail due to the restrictions and regulations for sharing personal related information, that each party has to follow though. Additionally adapting and harmonizing the communication interfaces between the Web shops from various online merchants and different PSPs and issuers are an enormous undertaking and will likely fail due to different notion of the communication pattern and exchanged data structures between all affected parties.

This is the scenario this Master thesis will look into in detail and try to solve the most important question: is this really a fraudulent transaction?

Looking into the stakeholders, that can provide useful information to solve this question, one will come up with:

1. **merchant**: providing additional information of the E-commerce transaction in question
2. **PSP/issuer**: providing additional information of the credit card usage and owner
3. **ISP**: can give hints whether a consumer has fallen victim to a phishing attack based on Internet access logs
4. **LSP**: provide information if the order has already been shipped or can be still halted

4 Theoretical Foundations

This chapter will lay out the theoretical foundations for the to-be-designed collaborative system. It will start with an investigation of the CSCW system theory followed by a detailed examination of the Semantic Web standards like RDF, OWL and SPARQL and how they can be used within Semantic Web agents. Last but not least the chapter will look into the concepts of P2P communication technologies by looking into various protocols for information sharing in detail — e.g. XMPP and WebRTC.

4.1 Computer-Supported Cooperative Work

4.1.1 Definition

4.1.2 Types

CSCW systems can be differentiated by their support of communication on the two axis place and time:

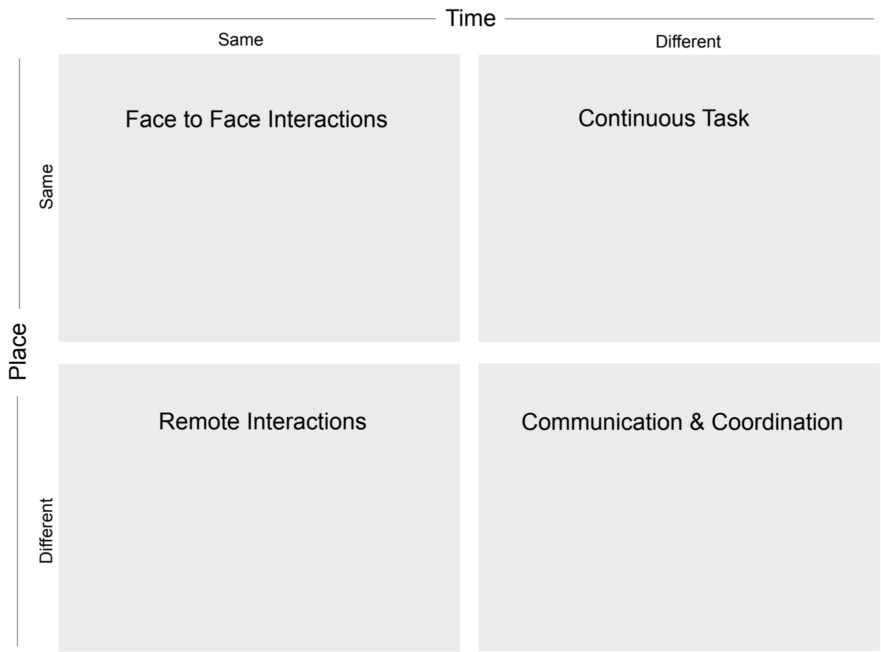

Figure 4.1: CSCW Place/Time Matrix (**?**)

Additionally it is possible to group the CSCW systems based on the 3C model:

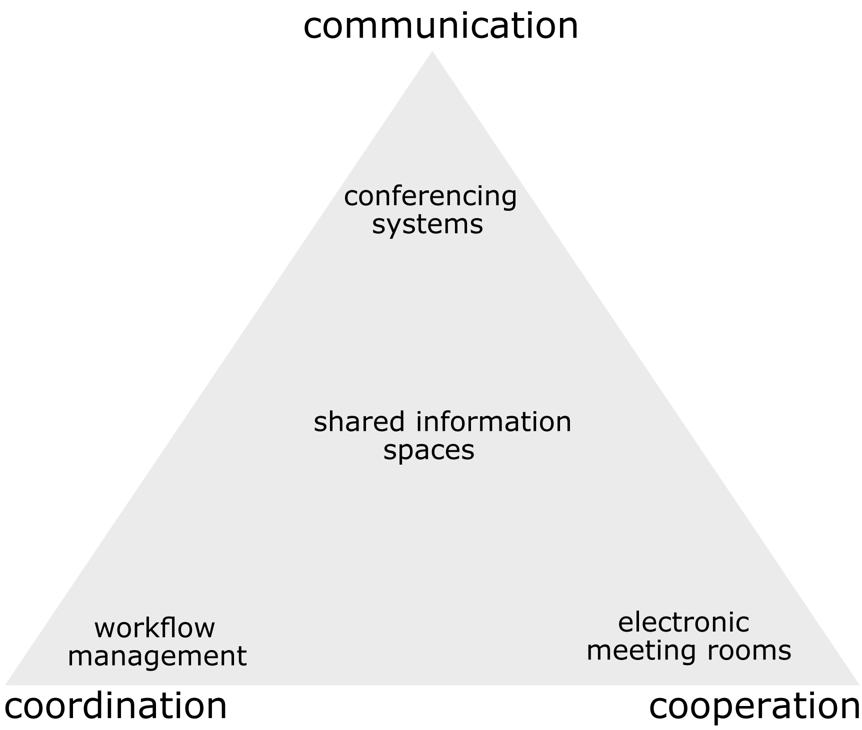

Figure 4.2: The 3C Model (Koch 2008)

4.1.3 Shared Information Spaces

4.1.4 Important aspects of CSCW systems

4.2 Fundamental Web Technologies

4.2.1 The TCP/IP protocol

4.2.2 The HTTP protocol

4.2.3 The XML format

4.3 The Semantic Web

4.3.1 Vision

MKP Chapter 1:
integrate distributed data from various publishers on the Web into smart applications
the Semantic Web delivers the infrastructure for this vision in form of various standard specifications (RDF, RDFS, OWL, SPARQL, ...)
the fundamentals of the World-Wide Web are also supported by the Semantic Web, especially:
- AAA-Slogan: Anyone can say Anything about Any topic
- Open World Assumption: we must always assume that there exist new information unknown to us yet, that can give additional insights
- Non-unique Naming Assumption: different URIs might refer to the same entity or object

as of this any one can extend on existing data entities and contribute her own knowledge / opinions as well as combine existing information in new ways -¿ data wilderness, no common data schema, more of an organic, living system
it heavily depends on the "network effect" and will / might explode with rising number of users / applications
as there will be disagreements on all sorts of topics there is no single ontology for the whole Web, but rather multiple ontologies tat can be integrated and utilised

MIT Chapter 1:
make information on the Web accessible to machines
- allows integration of information across web sites
- is also known as the "Web of Data"

design principles:
1. make structured and semi-structured data available in standardized formats
2. make individual data elements and their relationships accessible on the Web
3. describe the intended semantics of the data in a machine readable format

HTML is just for human consumption and a lot of the structures and semantics of the underlying databases is lost in the transformation process
- use labeled graphs as data model for objects and their relationships (objects == nodes, edges == relationships between them)
- formalize the syntax of the graph in RDF (Resource Description Framework)
- use URIs to identify individual data items and relations
- use ontologies to represent semantics of the data items (either lightweight RDF schema definitions or Web Ontology Language are used for that)

RDFS and OWL are meta-description languages allowing to define new domain-specific knowledge representations
they rely on the basic principles of the Web: supporting distributed, decentralized architectures

some new initiatives for standardizing semantics: schema.org and linkeddata.org
initially it was tried to solve the integration issues with XML, but as it is syntactically more machine- readable it lacks the semantic of the data
- as of this RDF is the basic language of the Semantic Web and describes meta-data as well as content

an ontology formally describe a domain based on terms and their relationships (terms == classes of objects)
hierarchies are supported (even multiple inheritance between objects)
ontologies also include:
- properties
- value restrictions
- disjointness statements
- specifications of logical relationships
goal is to provide a shared understanding of a domain
can help with the necessity to overcome differences in terminology
a mapping for different wordings in an ontology or between ontologies is possible
they can also be useful for generalization or specialization of Web search results

ontologies help with reasoning of objects, they can uncover unexpected relationships and inconsistencies as well as - by utilizing intelligent web agents - make decisions and select course of actions (e.g. "if-then-conclusions" aka Horn logic)
agents can also be used for "validation of proof" of statements of another agent or machine

Semantic Web is a layered approach . . .

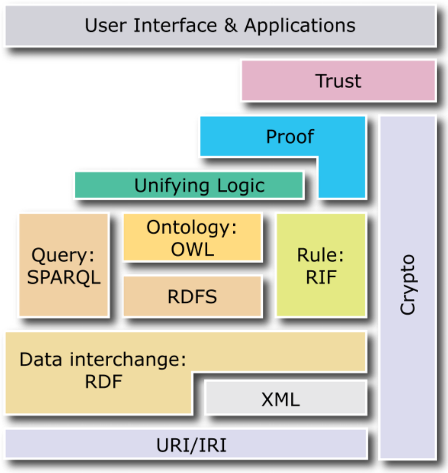

Figure 4.3: The Semantic Web Model (W3C 2013)

4.3.2 Semantic Modelling

MKP Chapter 2:
semantic models
- help people communicate about a fact or situation in the world
- explain and make predictions about the world
- mediate among multiple viewpoints and allow to explore commonalities as well as differences

1. human communication and modelling:
- helps people to coordinate their understanding collaboratively
- knowledge will be gathered, organized, tagged and shared
- when building models in natural human language they are usually open for interpretation of the meaning (e.g. laws)
- interpretation of the text depends on time and context of use -¿ informal model
- the success of informal models can be measured as degree of people supporting the

intended purpose
- tagging systems provide an informal organisation to a large body of heterogenous information
- in addition: models can have different layers with an increasing degree of formality (e.g. in the sector of regulations and laws there are regional, national as well as international laws with different degree of formality)
- informal models might be fitting their purpose in the context of their creation, but might need additional layers of models when their usage get beyond that original context to represent the shared meaning

2. explanations and predictions:
- help individuals to draw their own conclusions based on the information received
- especially useful in “interpretive situations” -¿ something is not set in stone
- explanation plays a crucial role in the “understanding” of a situation; if someone can “explain” it, they usually understood it
- in the Semantic Web explanation might help reuse the whole or parts of an existing model
- prediction is closely related to explanation; if a model offer an explanation for a certain situation, it can also be used to make predictions
- that resembles the fundamental of the scientific method (falsification)
- explanation and prediction require a more formal models than used for human communication (see above)
- usually they are build up from objective statements that are used to describe principles and rules (aka formalism)
- these models can also be used to make predictions
- they allow to evaluate the validity of a model and its applicability to a given situation
- in opposite to human communication formalism doesn’t need extra layers of explanations
- in the Semantic Web there are certain standards (a formalism) for modelling explanations
- these techniques can also be used to validate proofs and make predictions (aka inference)

3. Mediating Variability:
- goes hand in hand with AAA principle of the Semantic Web
- usually one decides for a specific viewpoint based on the information from trusted authorities
- informal approach: let every opinion stay side-by-side and let the consumer choose

which one to follow
- in this scenario the notion depends on the readers interpretation (as is also common in the Web of information)
- can be modelled in an OOP sense with classes and a hierarchy between them (the higher the more general, the lower the more specific)
- works well for known categories of entities (aka taxonomies)
- any model can also be build up from contributions from multiple sources
- usually seen as layers from different sources
- combination of all layers into a complete model
- a simple merge operation on the layers is easy, but might also introduce inconsistencies of viewpoints into the model
- when two or more viewpoints come together on the Semantic Web there will be an overlap of information
- this will result in disagreements and confusions in the beginning before there will be synergy, cooperation and collaboration
- essence of the Semantic Web: provide an infrastructure that supports AAA and help the community to work through the resulting information chaos to come up with a shared meaning

4. Level of expressivity:
- different people contribute information on different levels of expressivity
- each level might be sufficient to answer specific questions while leaving out unnecessary (sometimes confusing and complex) details
- as of this each level has its purpose!
- also on the Semantic Web there are tools for different levels of expressivity, from the least to the most expressive:
1) RDF: foundation for making statements
2) RDFS: basic notion of classes, hierarchies and relationships
3) RDFS+: subset of OWL, more expressive as RDFS, less complex than OWL, but no standard yet. tries to solve some issues with RDFS for industry use
4) OWL: express logic on the Semantic Web like contraints between classes, entities and relationships

- in the context of the Semantic Web modelling is an ongoing process with some well-structured knowledge and some new, unstructured information coming in at the same point in time

4.3.3 Resource Description Language

MIT Chapter 2:
what is needed to exchange information?
1. syntax: how to serialize the data?
2. data model: how to structure and organize the data?
3. semantics: how to interpret the data?

HTML is made for rendering information on screen and for human consumption
RDF brings a flexible data model to the Web:
- basic building block is a **triple** of *entity* - *attribute* - *value* also known as statement (could also be expressed as *subject* - *predicate* - *object*)
RDFS describes the vocabulary that is available

so:
1. syntax: Turtle, RDFa, RDF-XML or JSON-LD
2. data model: RDF
3. semantics: RDFS

foundational elements are:
- resources (aka just a "thing" of interest identified by an URI or URL depending on its accessibility)
- properties (specify the relations between resources, also identified by URIs)
- statements (assign a value to a 'resource-property' relation, value could be another resource or a literal)
- graphs (RDF is a graph-centered data model, could be distributed, Web of Data / Linked Data approaches)

linked data principles:
- use URIs as name for things
- use HTTP URLs so ppl. can look up those things on the Web
- if they do so, provide useful information (HTML and/or RDF, content and/or meta data)
- include links to other URLs so they can discover more/related things

named graph:
- can be used to point to specific statements or (sub-)graphs
- alternative: reification via an auxiliary object

Turtle: Terse RDF triple language
- <subject incl. URI><predicate incl. URI><object incl. URI>.
- literals will be expressed as “value”^^<XML schema data type>and supports *string, integer, decimal, dates, . . .*
- URIs can be prefixed: @prefix: <URI>
- repetition: ‘;’ repeats the subject from previous statement, ‘,’ repeats subject and predicate from previous statement
- named graphs in Turtle via Trig extension:
[...] <predicate incl. URI> [...]

`sample.ttl:`

```
@prefix ns1: <URI>
@prefix ns2: <URI>
@prefix ns3: <URI>

ns1:subject ns2:predicate ns3:object .
```

RDF/XML: RDF represented in XML format
- RDF namespace and root node
- subjects in ‘RDF:description’ node containing ‘RDF:about’ attribute with URI
- predicates and objects are child elements of subject node
- use XML namespeaces for URI of nodes

`sample.xml:`

```
<rdf:Description rdf:about="<subject incl. URI>">
  <ns2:predicate rdf:resource="<object incl. URI>" />
</rdf:Description>
```

RDFa: mixin RDF meta-data into HTML
- ‘about’ attribute on <span>or <div>in HTML
- ‘property’ attribute for literal value assignment
- ‘rel’ and ‘resource’ attributes for non-literals
- use XML namespaces for URI of data nodes
- put ‘[]’ around subject and object notations

`sample.html:`

```
<div about="[ns1:subject]">
  <span rel="ns2:relation" resource="[ns3:object]">
</div>
```

MKP Chapter 3:
- usually data is provided in tables from a database
- if we wanna split those over multiple servers, we can:
1) simply split the tables on a row-basis; the table needs to have the same layout on all servers
2) simply split the tables on a column-basis; the rows in each column need an unique identifier to match up the results
3) break down the whole table into cells and distribute them across all servers
->cells with facts need an unique identifier for the row as well as the column

- therefore RDF uses a triple of subject - predicate - object
- subject and predicate are using an unique identifier based on URI
- the triple can be visualized as directed graph

- data from multiple sources can be combined into a graph, if it can be figured out, which nodes exist in both distributed graphs
- therefore nodes are prefixed with an URI
- this URI should be an URL if the information can be dereferenced on the World-Wide Web
- usually they are used in combination with qnames, which define abbrevations for full-qualified URIs
e.g. qname <URI>
qname:subject predicate qname:object .
- use camel case for identifiers, no spaces are allowed
- W3C defines some qnames themselves:
- rdf: contains identifiers used in RDF
- rdfs: contains identifiers used in RDFS
- owl: contains identifiers used in OWL

- in any case: if you use URLs for your entities at least provide a Web page with the explanation of them

- use rdf:type to specify the type of a subject or object (e.g. geo:Berlin rdf:type geo:City .)

- use rdf:Property to specify an identifier to be used as a predicate (e.g. geo:latitude rdf:type rdf:Property .)
- the references objects could also be literal objects like numbers, dates and strings (they borrow the data type specifications from the XML standard)

- statements can also refer to other statements; this kind of metadata about statements can include:
1) provenance (who has made the statement)
2) likelyhood (what is the probability of this statement)
3) context (the setting in which the statement is valid)
4) timeframe (the time constraints for this statement)

- explicit reification with the predicates rdf:subject, rdf:predicate, rdf:object; e.g.:
q:n1 rdf:subject geo:Berlin
rdf:predicate geo:size
rdf:object geo:MegaCity .

web:Wikipedia m:says q:n1 .
- this sample just qualifies that a source (here: Wikipedia) has made a certain statement (n1); but does say nothing about the statement itself! it is up to the application to decide whether the source (Wikipedia) can be trusted or not!

- RDF triples can be serialized as:
1) N-Triples
2) Turtle
3) RDF/XML
4) RDFa

- blank nodes are commonly used to express unknown or uncertain entities
- they will be described in turtle within []
- an ordered set of items can be represented in turtle as ()

4.3.4 Web Ontologies

Lightweight approach: RDFS
- is about adding semantics to your RDF documents

Start by:
1. specify the **things** to talk about
differentiate between *objects* (real entities) and *classes* (set of entities)
'rdf:type' attribute to assign objects to classes (object = instance of this class)
impose restrictions on the kind of properties used on objects:
- restrictions on values are called 'range' restrictions (object can take values of . . .)
- restrictions on property-object relations are called 'domain' restrictions (this relation applies to objects of . . .)
2. set up relations between classes (inheritance, composition)
3. define properties (registered globally) and the possible hierarchy relationship between them (global properties means you can extend existing RDFS classes with your own properties easily)

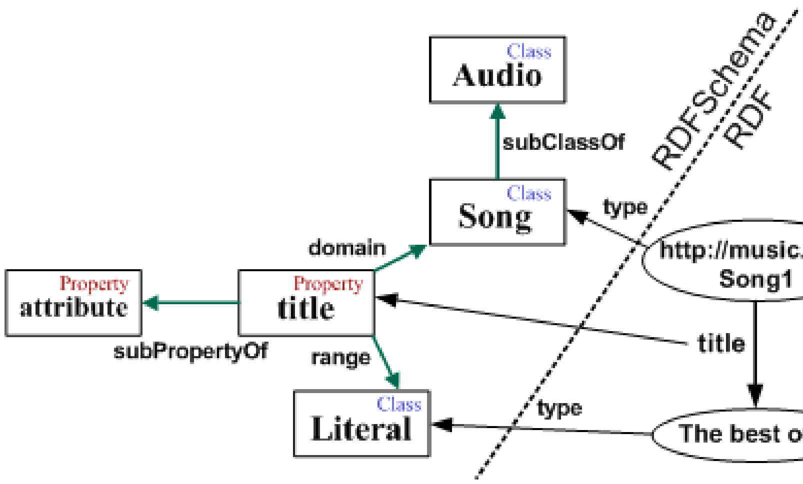

Figure 4.4: RDF Schema sample

RDFS is described in RDF style using:
- core classes like:
- 'rdfs:Resource' (all objects/resources)
- 'rdfs:Class' (all classes)
- 'rdfs:Literal' (all literals)
- 'rdfs:Property' (all properties)
- 'rdfs:Statement' (all reified statements)
- core properties like:
- 'rdfs:type' (specify kind of class)

- ‘rdfs:subClassOf’ (specify inheritance between classes)
- ‘rdfs:subPropertyOf’ (specify inheritance between properties)
- ‘rdfs:domain’ (specify domain restrictions)
- ‘rdfs:range’ (specify range restrictions)
- container classes like:
- ‘rdf:Bag’ (unordered list of entitites)
- ‘rdf:Seq’ (ordered list of entities)
- ‘rdf:Alt’ (list of alternatives/choices)
- ‘rdf:Container’ (superclass for all containers)
- utility classes like:
- ‘rdfs:seeAlso’, ‘rdfs:isDefinedBy’ (links and references to other entities)
- ‘rdfs:Comment’ (comments and notes of entities)
- ‘rdfs:Label’ (human-friendly name of entities)

Missing features in RDFS: . . .

Complex Ontologies in Web Ontology Language (OWL):
. . .

4.3.5 Query Language

SPARQL requires a **triple store** - a database containing RDF documents
is also referred to as a *Graph Store*
data is inserted via Bulk load operation or via SPARQL update statements
SPARQL consist of SPARQL Queries that are send over the SPARQL protocol
Clients sends the queries to an HTTP endpoint
Stores on the public Web incl. dbpedia.org, ckan.org, wikidata.org
SPARQL also works with RDFS
SPARQL has similarities to SQL: - each element in a triple might be replaced with a variable like ‘?varName’ like so:
`sample.sparql`:

```
PREFIX ns1:<URI>
PREFIX ns2:<URI>
PREFIX ns3:<URI>

```

```
SELECT ?varName
WHERE {
    ns1:subject ns2:predicate ?varName
}
```

- in the WHERE clause it hosts the graph pattern to match (could be cascaded to go down subgraphs)
- variables can occur at any place in the graph pattern (?subj ?pred ?obj) as select with query everything

LIMIT <n>option at the end for limiting the result set
FILTER (?varName <condition>) in graph pattern can restrict results to match some literal values and supports:
- numbers, dates: $<$, $>$, $=$
- strings: $=$, regex()

open world assumption: resources on the Web are described in different schematas with various properties using different vocabularies
- UNION option in graph pattern combines different matches
- OPTIONAL option in graph pattern only returns those entities if they are available (otherwise empty)

ASK query checks for the existence of a given graph pattern
CONSTRUCT can be used to retrieve a subgraph from a larger graph, can also be used to translate between different schemas
`sample2.sparql:`

```
PREFIX ns1:<URI>
PREFIX ns2:<URI>
PREFIX ns3:<URI>

CONSTRUCT {
    ?varA ns2:predicate ?varB .
    ?varA ns3:predicate ?literalA .
}
WHERE {
    ?varA ns1:predicate ?varB
}
FILTER ( ?varB > x )
```

- SPARQL can be used to harmonize graphs from different sources
- is also used for basic reasoning ala "if found this, assume that"
- can ease hierarchical queries with * or + on the predicate (SPARQL 1.1)
- can help resolving issues with different entities referring to the same object (MKP pg. 95)
- Federated Queries can be used to combine information from distinct sources via SPARQL (MKP pg. 110-112)

- inferencing information from existing triples via SPIN (SPARQL Inferencing Notation)
- like in a taxonomy items can be categorized in an hierarchy (MKP pg. 114)
- inference patterns are used in Semantic Web applications (MKP pg. 115)
* subClassOf - type propagation rule
- inferencing could be done at query time or persistently (MKP pg. 120/121)
- inferences can also be helpful when combining information from unknown sources
- inferencing happens on various levels (RDFS, RDFS+, OWL) with an increased set of complex inferencing rules (MKP pg. 122/123)

4.3.6 Agents and Rules

4.4 Peer-to-peer communication

4.4.1 Centralized vs. Decentralized Web Architectures

- in a classical client-server scenario a single server is storing information and distributing it to the clients
- the information is centralized and under control of the provider

- a P2P network considers all nodes equal
- each node can provide information to any other node
- information in a P2P network has to be indexed so that the correct node is queried for it
- the index itself has to be stored somewhere (e.g. on a central server like Napster or in a distributed manner spread over the nodes of the P2P network)

- a P2P system has an high degree of decentralization
- the system is usually self-organizing (adding new or removing members automatically)

- the whole system is usually not controlled by a single organisation and spread over various domains
- it tends to be more resilent to faults and attacks
- can be used for file & data sharing, media streaming, telephony, volunteer computing and much more

- can be categorized by the degree of centralization into:
1) partly centralized P2P systems (have a dedicated controller node that maintains the set of participating nodes and controls the system)
2) decentralized P2P systems (there are no dedicated nodes that are critical for the system operation)

4.4.2 Initiating a communication session

- depends on the structure of the P2P system - in a partly centralized P2P system new nodes join the network by connecting to the central controller (wellknown IP address)
- in a decentralized P2P system new nodes are expected to obtain via a separate channel the IP address to connect to (usually a bootstrap node that helps to set up the new node)

4.4.3 Finding communication peers

- also known as the overlay network in a P2P system
- can be represented as a directed graph containing the nodes and communication links between them
- can be differentiated between unstructured and structured overlays
- unstructured overlay networks have no constraints for the links between nodes; therefore the network has no particular structure
- structured overlay networks assign an unique identifier from a numeric keyspace to each node; these keys are used to assign certain responsibilities to nodes on the network; as of this routing can be handled more efficiently
- in partly centralized P2P systems the controller is responsible for the overlay formation

- in partly centralized P2P system an object is typically stored at the node that in-

serted the object
- the central controller holds the information about which objects exist and which nodes hold them

- in unstructured systems the information is typically stored on the nodes that introduces them
- to locate an object a query request is typically broadcasted through the overlay network
- often the scope of the request (e.g. the maximum number of hops from the querying node forward) is limited to reduce the overhead on the system

- in structured systems a distributed index is maintained in the form of a distributed hash table
- this DHT holds the hash value of the (index) key and the address of the node that stores the value

4.4.4 Transmitting Data

4.4.5 Available Protocols

5 Concept and Design of the System

This chapter will ...

5.1 Objectives of System

Based on the explanations in Chapter 3, and especially the scope definition for this Master thesis in Section 3.5, the system for investigating E-commerce fraud incidents have to answer the central question:

> *Is this really a fraudulent E-commerce transaction?*

The main stakeholders, that need to be involved in the investigation process are:

- the online merchants
- the PSP
- the issuer
- the LSP

Ideally they would make part of their internal information available for external partners to access and query for, so that the party, who has to authorize or validate the credit card payment can analyse all available information, as depicted in the Figure 5.1.

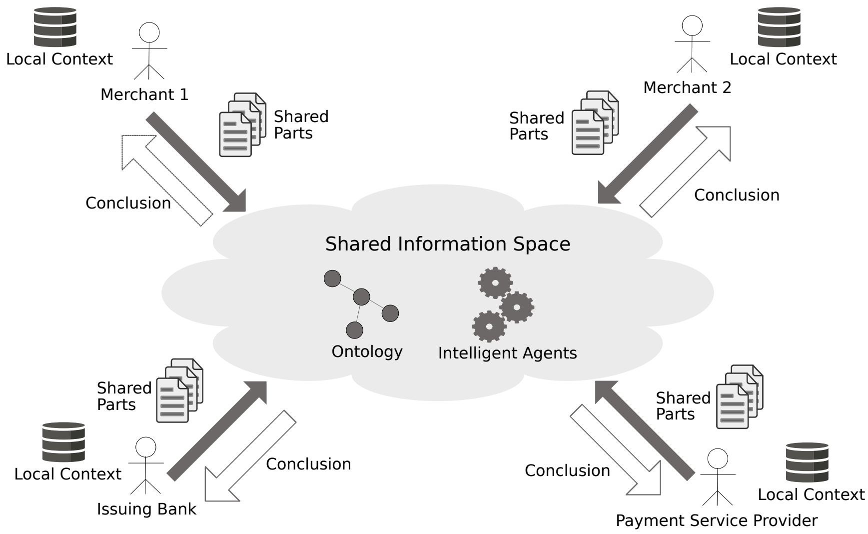

Figure 5.1: System Overview

As the information is coming from various sources there has to be a shared data model that brings all of them together and defines the connection points and references for the parts of each stakeholder. Based on the discussion in Chapter 3 and the analysis of the information each stakeholder has and transmit to others, the following initial schema can be conducted (see Figure 5.2).

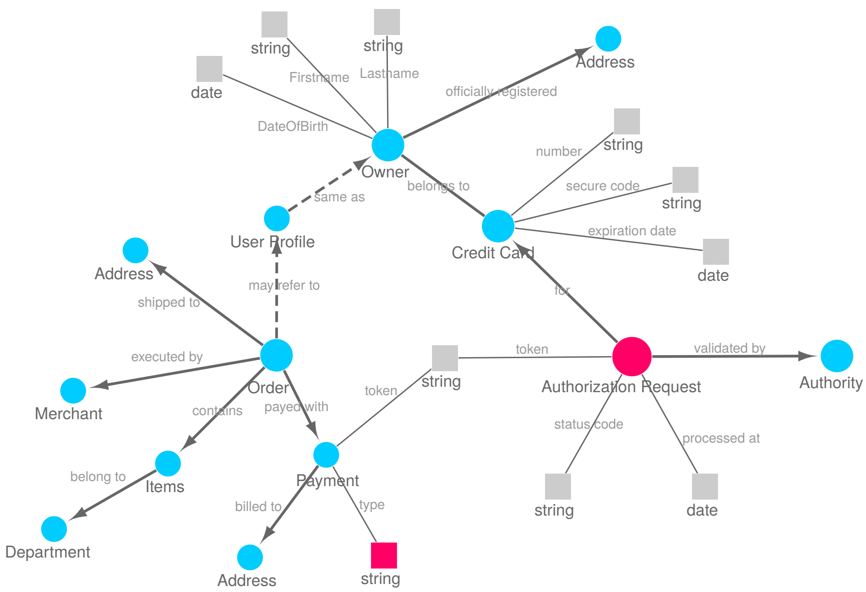

Figure 5.2: Initial Data Model

. . .

5.2 Existing Approaches

When trying to solve issues of information integration between organisations there are already existing solutions, that have to be examined whether they might fit the E-commerce fraud investigation scenario or not.

5.2.1 Web of Services

With the development of the E-commerce scenario there was also a need to integrate business functionality from various service providers operating on the Internet. Good examples for this are the integration of the PSP into the payment as well as the LSP for the shipping process. These approaches resulted in the "Service Oriented Architecture" paradigm, that enables application services provided by different vendors to talk to each other via a public facing programming interface (aka API). The only requirement for such interoperability to work properly is, that each public interface follows some standardised or commonly agreed upon guidelines to be vendor-, platform- as well as language-agnostic. One possible implementation of these concepts are the so-called Web Services, that use the WS* protocols and standards from the W3C with the extensible markup language (aka XML) and the HTTP protocol at their core (Josuttis 2007).

Like the HTML format, that is used to represent Web pages on the Internet, XML is originally based on SGML, but instead of formalising markup tags for structuring and styling textual content it is a meta-language allowing everyone to define his or her own markup languages. In this matter it doesn't dictate what tags are available to structure the information; instead it includes some basic guidelines for creating wellformed and valid documents that uses domain-specific tags, which can be freely defined and structured by the creator of the XML document. Therefore it is better suited in situations where a computer has to parse and evaluate the content of a message (assuming the computer program knows the structure of the message)

In an additional step the author of the API could also specify an XML schema for each message, which describes the structure of the message with all the possible elements, their ordering, nesting level and data types in detail. By doing so the XML parser program can later verify the content of a retrieved message against the XML schema and check if it is a valid document (related to the schema definition). XML schematas are also expressed in XML format and have been standardised by the W3C. Being able to create custom markup languages via XML has a huge benefit for machine-to-machine communication and is the basis for integrating Web Services (via the WS* protocols), but it still has limitations when it comes to figure out the semantics of those XML messages. This is mostly due to the fact that each XML document represents a new markup language and needs a specific XML parser to be understood by the machine; also to distinguish commonly used tag names in an XML document the creator has to place them into specific namespaces (aka XML namespaces). But those XML namespaces further complicate the automatic processing of XML documents and increases the necessity to have custom instances of XML parsers for each XML document (Taylor & Harrison 2008).

Mapping the scenario onto the Web of Services approach would mean, that each merchant has to provide a specific HTTP API endpoint for querying transaction information by the issuer or PSP. Additionally each issuer or PSP has to collect all available information from the affected merchants (based on recent credit card authorization requests) and combine these various transactional data locally to be able to make an analysis of them that might lead to a decision regarding the E-commerce fraud question from above.

This will also mean that each merchant has to provide a separate entrypoint for issuers and PSPs to give access to selected information from the internal backend databases. These endpoints have to make safe and secure, so that only allowed parties can access these records. In addition the information from the various internal databases of the merchants have to be converted into a format suitable for external consumption — in this case a structured XML file.

Issuers and PSPs on the other hand have to maintain a list of HTTP API endpoints and their respective credentials to access the data from each merchant. Additionally they will have to provide a mapping of each merchant's specific data from the XML file to the internal analytics database of the issuer or PSP. They also have to trigger internal backend services in case of a suspecious E-commerce transaction that will query the HTTP API endpoint of the corresponding merchant for further information.

As conclusion one could easily see that there are huge efforts on all participating parties to provide information to or integrate data from each other. As there is no common way to access the information at the merchant side (beside the lower level HTTP protocol and XML data format), there have to be a lot of collaboration between each combination of stakeholders for deciding on access patterns and needed data structures in the beginning. Most importantly as each HTTP API endpoint will be publicly accessible via the Internet it will also opens up a new opportunity for hackers to get access to personal or payment related information.

5.2.2 Web of Data

"The Web is full of intelligent applications, with new innovations coming every day" (Allemang & Hendler 2011). But each of those intelligent Web applications is driven by the data available to them. Data that is likely coming from different places in the global information space — accessible usually via a custom API on the server hosting those resources (see Section 5.2.1). The more consistent the data available to the smart Web application is the better the service and its result will be. But to support an integration of the data from various Web services the semantics of the information delivered by each service has to be available — and there has to be a generalised, formalised way to express the semantic of that data. The focus on a standard that allows Web services to express the semantics of the data they provide also allows for global scalability, openness and decentralisation, which are the key principles of the World-Wide Web. The Semantic Web tries to give a solution for this problem by providing the Resource Description Framework (aka RDF) and related technologies (e.g. RDF schema, SPARQL, OWL, . . .) for describing, linking and querying the data that a Web service delivers. But it doesn't reinvent the wheel; instead the Semantic Web builds upon existing, proven technologies like XML, XML namespaces, XML schemata and the URI to uniquely address resources on the Web (Allemang & Hendler 2011).

A huge benefit of the Web of Data approach is, that the resources delivered are self-describing. They do not only have a consitent and meaningful syntax, but are also semantically self-contained. As of this each merchant has to provide a semantically description of the resources used in a transaction in a standard way — e.g. by using W3C standards like RDF, RDFa or JSON-LD. Each merchant also have to provide a HTTP API endpoint to access and query for the resources, utilizing a query language like SPARQL.

Each issuer or PSP can access these HTTP API endpoints with her credentials and query for specific information from the public "information database" from a merchant. The results of each query can be easily combined into an existing database based on the merging capabilities of RDF. The resulting analytic database can be used by the issuer or PSP to run queries against or use them with intelligent reasoning tools from the Semantic Web standards for investigation of an E-commerce transaction.

The resulting issues and problem are mostly the same as with the Web of Services approach — beside that the Web of Data offer an unique and integrated way to describe the structure and semantic of the data received from another party. The initial efforts for the implementation of this scenario is also quite high, even if it is lower than with the Web of Services approach. This is mostly due to the fact that there are already some industry-wide and commonly agreed upon ontologies and taxonomies, that are able to describe most of the resources in an E-commerce transaction (e.g. GoodRelations Ontology, Schema.org). As it is more likely that merchants do already use them to encode at least some of the data in their backend databases for machine-to-machine communication, it will also descrease the effort on merchant side to provide them for the issuers and PSPs. Still these parties have to define the kind of queries and reasoners that might be useful to investigate an E-commerce transaction with the objective to figure out if it is fraudulent or not and have to implement them into their own backend systems.

5.3 Design Proposal

5.3.1 Schema.org Vocabulary

The usage of the vocabulary of Schema.org is preferred as . . .

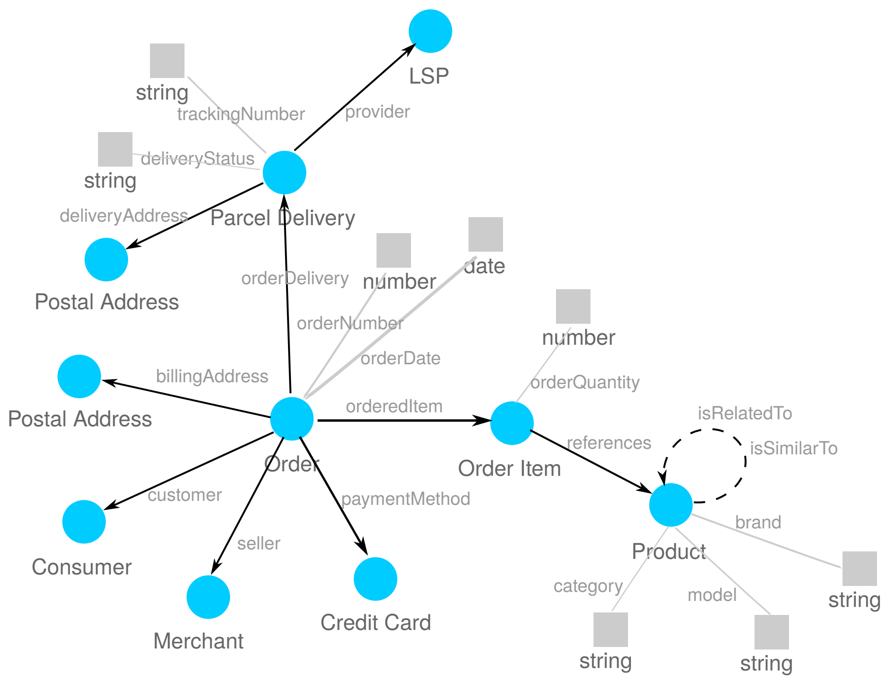

Figure 5.3: Schema.org Mapping

5.3.2 Information encoding

As the merchants will provide semantic meta data for SEO already, one can re-use part of these information for the E-commerce fraud investigation. The data is likely encoded in Microdata, RDFa or JSON-LD.

For the communication the WebRTC is a good approach as it integrates well with existing enterprise IT infrastructures.

5.3.3 Partially centralized Peer-To-Peer System

The issuing bank is the trusted party in this system setup. It will initiate a data sharing session with the other required stakeholders based on the past usage of the credit card in question. During the P2P session the merchants, payment and logistic service providers will share the required information with the issuer. In this process the data from the other stakeholders will be replicated to the issuer, who will build up a graph based on the Schema.org schema mapping. The analysis of the data will be done on top of this graph by the issuer and can also be handled after the initial P2P data sharing session has been ended. If there are any new conclusions drawn from the data, the issuer is in charge to inform the stakeholders afterwards. So the main work will be on the issuer side, who is the major driving party in this scenario, as depicted in Figure 5.4.

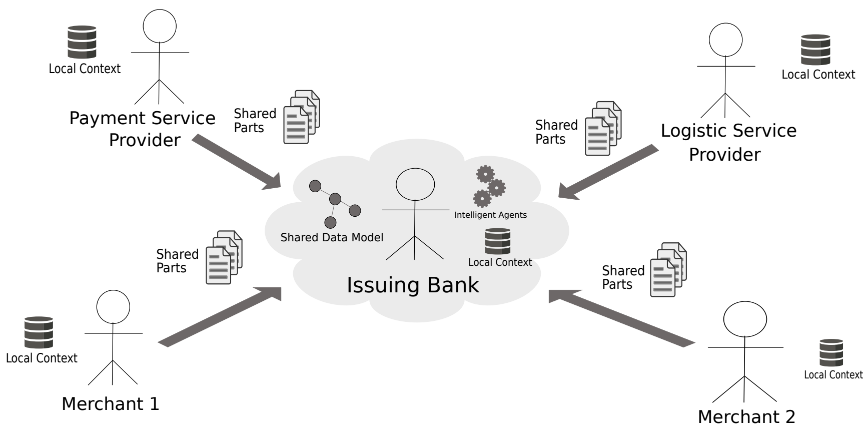

Figure 5.4: Partially centralized P2P system architecture

Main issue with the above mentioned system architecture is, that the merchants, payment and logistic service providers have to hand over their information to the issuing bank of the credit card for analysis. This might be either problematic due to distrust of the objectives of the issuing bank, or not possible at all due to local data sharing restrictions and regulations.

5.3.4 Decentralized Peer-To-Peer System

In the decentralized P2P system architecture each node is equal and keeps their local data ready for analysis if the node is online. If the issuer will have to figure out, whether a transaction is fraudulent or not, she is going to send out various queries to all the available nodes in the P2P cluster asking for certain information that help investigating the case. The other nodes, whose reside on each stakeholder involved, will answering the queries based on the common Schema.org data mapping shown above and send back the results to the issuing bank. The issuer will collect all the results from the various parties and combine them to be able to analyse the issue and come up with a conclusion. The main benefit of this architecture is, that there is no need to duplicate the data from the other stakeholders to the issuing bank. Due to this it can also be a better suited solution if data sharing faces restrictions due to law or regulations. On the other hand this architecture will depend on the nodes being online all the time so the issuer can query for information at any time. So this works only in synchronous communication mode. Additionally there are efforts spread around all the stakeholders to set up and maintain a system for secure data querying functionality, please see Figure 5.5.

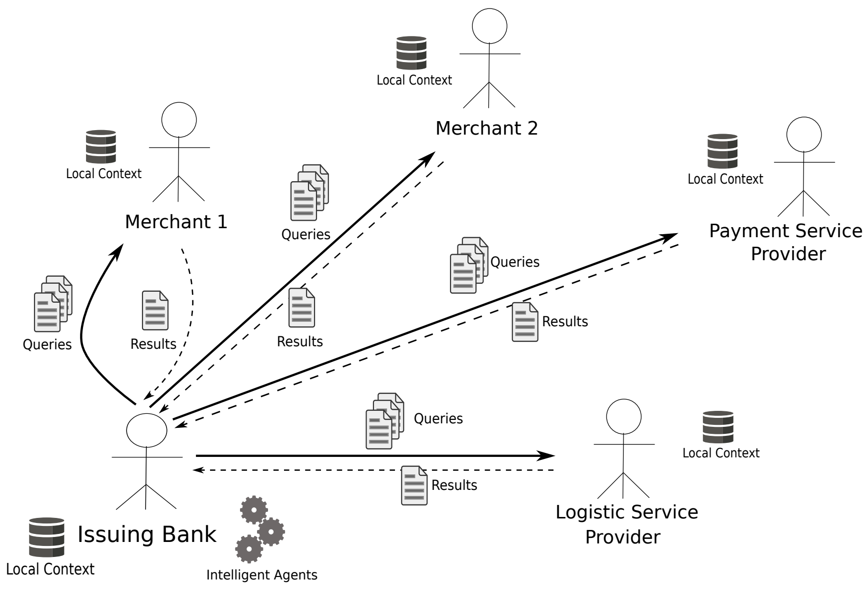

Figure 5.5: Decentralized P2P system architecture

6 Conclusion and Future Work

List of Figures

List of Tables

Glossary

API	Application Programming Interface.
B2B	Business-To-Business.
B2C	Business-To-Consumer.
C2B	Consumer-To-Business.
C2C	Consumer-To-Consumer.
CSCW	computer-supported cooperative work.
CSP	Cloud Service Provider / Hosting Service.
HTML	HyperText Markup Language.
HTTP	HyperText Transfer Protocol.
IP	Internet Protocol.
ISP	Internet Service Provider.
ISV	Independent Software Vendor.
JSON-LD	JavaScript Object Notation for Linked Data.
LSP	Logistic Service Provider.
OWL	Web Ontology Language.
P2P	Peer-To-Peer.
PSP	Payment Service Provider.
RDF	Resource Description Framework.
RDFa	Resource Description Framework in Attributes.
SGML	Standard Generalized Markup Language.
SPARQL	SPARQL Protocol and RDF Query Language.
URI	Uniform Resource Identifier.
URL	Uniform Resource Locator.
W3C	World-Wide Web Consortium.
WebRTC	Web Real-Time Communication.

XML	Extensible Markup Language.
XMPP	Extensible Messaging and Presence Protocol.

Bibliography

Allemang & Hendler 2011
Allemang, Dean; Hendler, James: *Semantic web for the working ontologist: effective modeling in RDFS and OWL.* Elsevier, 2011

Ankhule & Joshy 2015
Ankhule, Gayatri R.; Joshy, MR: Overview of E-Commerce. In: *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)* (2015), pages 196

Ashraf et al. 2011
Ashraf, Jamshaid; Cyganiak, Richard; O'Riain, Seán; Hadzic, Maja: *Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations.* In: *LDOW*, 2011

Boley et al. 2007
Boley, Harold; Kifer, Michael; Pătrânjan, Paula-Lavinia; Polleres, Axel: Rule interchange on the web. In: *Reasoning Web.* Springer, 2007, pages 269–309

Brachmann 2015
Brachmann, Steve: *In the face of growing e-commerce fraud, many merchants not prepared for holidays - IPWatchdog.com |patents & patent law.* http://www.ipwatchdog.com/2015/11/22/growing-e-commerce-fraud-merchants-not-prepared-for-holidays/id=63271/. Version: 11 2015

Business Wire 2015
Business Wire: Global card fraud losses reach $16.31 Billion — will exceed $35 Billion in 2020 according to the Nilson report. In: *Business Wire* (2015), 08. http://www.marketwatch.com/story/global-card-fraud-losses-reach-1631-billion-will-exceed-35-billion-in-2020-accor

Cai & Frank 2004
Cai, Min; Frank, Martin: *RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network.* In: *Proceedings of the 13th international conference on World Wide Web* ACM, 2004, pages 650–657

Captain 2015
Captain, Sean: These are the mobile sites leaking credit card data for up to 500, 000 people A day. In: *Fast Company* (2015), 12. http://www.fastcompany.com/3054411/these-are-the-faulty-apps-leaking-credit-card-data-for-up-to-500000-people-a-day

Carvalho et al.
CARVALHO, Rodrigo; GOLDSMITH, Michael; CREESE, Sadie; POLICE, Brazilian F.: Applying Semantic Technologies to Fight Online Banking Fraud.

Chao et al. 2012
CHAO, Lemen; XING, Chunxiao; ZHANG, Yong: *The Semantic Web-Based Collaborative Knowledge Management*. INTECH Open Access Publisher, 2012

Consumer Action 2009
CONSUMER ACTION: Questions and answers about credit card fraud A Q & consumer aCtion A consumer action publication. Version: 2009. `http://www.consumer-action.org/downloads/english/Chase_CC_Fraud_Leaders.pdf`. http://www.consumer-action.org/downloads/english/Chase_CC_Fraud_Leaders.pdf, 2009. – Forschungsbericht

DSS 2014
DSS, PCI: *Payment Card Industry Data Security Standards*. 2014

Ehrig et al. 2003
EHRIG, Marc; TEMPICH, Christoph; BROEKSTRA, Jeen; VAN HARMELEN, Frank; SABOU, Marta; SIEBES, Ronny; STAAB, Steffen; STUCKENSCHMIDT, Heiner: *SWAP: Ontology-based Knowledge Management with Peer-to-Peer Technology*. In: *Wissensmanagement*, 2003, pages 17–20

Ekelhart et al. 2006
EKELHART, Andreas; FENZ, Stefan; KLEMEN, Markus D.; WEIPPL, Edgar R.: *Security ontology: Simulating threats to corporate assets*. Springer, 2006

Gerber et al. 2008
GERBER, Aurona; MERWE, Alta Van d.; BARNARD, Andries: *A functional semantic web architecture*. Springer, 2008

Google Patents
`https://patents.google.com/?q=credit+card+fraud+prevention&after=20150101`

Goyal & Fussell
GOYAL, Nitesh; FUSSELL, Susan R.: Effects of Sensemaking Translucence on Distributed Collaborative Analysis.

Guha et al. 2016
GUHA, RV; BRICKLEY, Dan; MACBETH, Steve: Schema. org: Evolution of structured data on the web. In: *Communications of the ACM* 59 (2016), Nr. 2, pages 44–51

Hepp 2008
HEPP, Martin: Goodrelations: An ontology for describing products and services offers on the web. In: *Knowledge Engineering: Practice and Patterns*. Springer, 2008, pages 329–346

Hepp et al. 2009
Hepp, Martin; Radinger, Andreas; Wechselberger, Andreas; Stolz, Alex; Bingel, Daniel; Irmscher, Thomas; Mattern, Mark; Ostheim, Tobias: *GoodRelations Tools and Applications.* In: *Poster and Demo Proceedings of the 8th International Semantic Web Conference (ISWC 2009), Washington, DC, USA*, 2009

Holmes 2015
Holmes, Tamara E.: *Credit card fraud and ID theft statistics.* `http://www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php`. Version: 09 2015

Josuttis 2007
Josuttis, Nicolai M.: *SOA in practice: the art of distributed system design.* " O'Reilly Media, Inc.", 2007

Kingston et al. 2004
Kingston, John; Schafer, Burkhard; Vandenberghe, Wim: Towards a financial fraud ontology: A legal modelling approach. In: *Artificial Intelligence and Law* 12 (2004), Nr. 4, pages 419–446

Koch 2008
Koch, Michael: *CSCW and enterprise 2.0 - towards an integrated perspective.* In: *BLED 2008 Proceedings*, 2008

Lara et al. 2007
Lara, Rubén; Cantador, Iván; Castells, Pablo: Semantic web technologies for the financial domain. In: *The Semantic Web.* Springer, 2007, pages 41–74

Lee & Paine 2015
Lee, Charlotte P.; Paine, Drew: *From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW.* In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* ACM, 2015, pages 179–194

Lewis 2015
Lewis, Len: *More vulnerable than ever?* `https://nrf.com/news/more-vulnerable-ever`. Version: 12 2015

Oren et al. 2009
Oren, Eyal; Kotoulas, Spyros; Anadiotis, George; Siebes, Ronny; Teije, Annette ten; Harmelen, Frank van: Marvin: Distributed reasoning over large-scale Semantic Web data. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 7 (2009), Nr. 4, pages 305–316

Ozturk 2010
Ozturk, Ozgur: *Introduction to XMPP protocol and developing online collaboration applications using open source software and libraries.* In: *Collaborative*

Technologies and Systems (CTS), 2010 International Symposium on IEEE, 2010, pages 21–25

Parameswaran et al. 2001
PARAMESWARAN, Manoj; SUSARLA, Anjana; WHINSTON, Andrew B.: P2P networking: An information-sharing alternative. In: *Computer* (2001), Nr. 7, pages 31–38

PYMNTS 2016
PYMNTS: *Hackers and their fraud attack methods.* http://www.pymnts.com/fraud-prevention/2016/benchmarking-hackers-and-their-attack-methods. Version: 02 2016

Rampton 2015
RAMPTON, John: How online fraud is a growing trend. In: *Forbes* (2015), 04. http://www.forbes.com/sites/johnrampton/2015/04/14/how-online-fraud-is-a-growing-trend/#16ffc0ec349f

Rana & Baria 2015
RANA, Priya J.; BARIA, Jwalant: A Survey on Fraud Detection Techniques in Ecommerce. In: *International Journal of Computer Applications* 113 (2015), Nr. 14

Reuters 2015
REUTERS: *Fraud rates on online transactions seen up during holidays: Study.* http://www.reuters.com/article/us-retail-fraud-idUSKCN0T611T20151117?feedType=RSS&feedName=technologyNews. Version: 11 2015

Rice 1992
RICE, Ronald E.: Task Analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. In: *Organization Science* 3 (1992), 11, Nr. 4, pages 475–500. http://dx.doi.org/10.1287/orsc.3.4.475. – DOI 10.1287/orsc.3.4.475. – ISSN 1047–7039

Robert & Dennis 2005
ROBERT, Lionel P.; DENNIS, Alan R.: Paradox of richness: A cognitive model of media choice. In: *Professional Communication, IEEE Transactions on* 48 (2005), Nr. 1, pages 10–21

Rodrigues & Druschel 2010
RODRIGUES, Rodrigo; DRUSCHEL, Peter: Peer-to-peer systems. In: *Communications of the ACM* 53 (2010), Nr. 10, pages 72–82

Scharffe et al. 2011
SCHARFFE, François; FERRARA, Alfio; NIKOLOV, Andriy: Data linking for the semantic web. In: *International Journal on Semantic Web and Information Systems* 7 (2011), Nr. 3, pages 46–76

Sen et al. 2015
Sen, Pritikana; Ahmed, Rustam A.; Islam, Md R.: A Study on E-Commerce Security Issues and Solutions. (2015)

Sobko 2014
Sobko, Oleg V.: Fraud in Non-Cash Transactions: Methods, Tendencies and Threats. In: *World Applied Sciences Journal* 29 (2014), Nr. 6, pages 774–778

Staab & Stuckenschmidt 2006
Staab, Steffen (Hrsg.); Stuckenschmidt, Heiner (Hrsg.): *Semantic web and peer-to-peer*. Springer Science + Business Media, 2006. http://dx.doi.org/10.1007/3-540-28347-1. http://dx.doi.org/10.1007/3-540-28347-1. – ISBN 9783540283461

Stollberg & Strang 2005
Stollberg, Michael; Strang, Thomas: *Integrating agents, ontologies, and semantic web services for collaboration on the semantic web*. In: *Proc. of the First International Symposium on Agents and the Semantic Web, AAAI Fall Symposium Series Arlington, Virginia*, 2005

Taylor & Harrison 2008
Taylor, Ian J.; Harrison, Andrew: *From P2P and grids to services on the web: evolving distributed communities*. Springer Science & Business Media, 2008

Verborgh & De Roo 2015
Verborgh, Ruben; De Roo, Jos: Drawing Conclusions from Linked Data on the Web: The EYE Reasoner. In: *IEEE Software* (2015), Nr. 3, pages 23–27

Visa Europe 2014
Visa Europe: *Processing e-commerce payments*. https://www.visaeurope.com/media/images/processing%20e-commerce%20payments%20guide-73-17337.pdf. Version: 08 2014

Vogt et al. 2013
Vogt, Christian; Werner, Max J.; Schmidt, Thomas C.: *Leveraging WebRTC for P2P content distribution in web browsers*. In: *Network Protocols (ICNP), 2013 21st IEEE International Conference on* IEEE, 2013, pages 1–2

W3C 2013
W3C: *W3C semantic web activity*. https://www.w3.org/2001/sw/. Version: 06 2013

Werner et al. 2014
Werner, Max J.; Vogt, Christian; Schmidt, Thomas C.: *Let our browsers socialize: Building user-centric content communities on webrtc*. In: *2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW)* IEEE, 2014, pages 37–44

Yang & Chen 2008
YANG, Stephen J.; CHEN, Irene Y.: A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network. In: *International Journal of Human-Computer Studies* 66 (2008), Nr. 1, pages 36–50

Zhou et al. 2013
ZHOU, Yujiao; NENOV, Yavor; GRAU, Bernardo C.; HORROCKS, Ian: Complete query answering over horn ontologies using a triple store. In: *The Semantic Web–ISWC 2013*. Springer, 2013, pages 720–736

Declaration in lieu of oath

I hereby declare that this master thesis was independently composed and authored by myself.

All content and ideas drawn directly or indirectly from external sources are indicated as such. All sources and materials that have been used are referred to in this thesis.

The thesis has not been submitted to any other examining body and has not been published.

Place, date and signature of student
Andreas Gerlach

Appendix

Based on chapter 4.1 we can conclude:
1. Face-to-Face Meetings: out-of-scope of this thesis
2. Distance Meetings: lack of collaboration support
3. Continuous Tasks: collaboration in teams, but only works when everyone is online
4. Communicate & Collaborate: allows to work on it in a disconnected mode, but increases communication and coordination efforts as well as might lead to synchronisation issues over time

This either leaves us with two options:
1. build a distributed, synchronous collaboration system, in that ppl. can share and work on content at the same time
2. build a distributed, asynchronous collaboration and communication system, in that ppl. can work on things for themselves and get connected together at a certain point in time for synchronising their findings and develop new insights

In the first variant it can be assumed that:
- stakeholders will initiate a collaborative session for a certain case, the collaboration and information sharing efforts end with finishing the case.
- each stakeholder might just work on his part of expertise in the whole knowledge graph (e.g. named subgraphs per stakeholder). these parts could be easily mirrored on the stakeholders environment (no discrepancies with informations from others)
- the whole knowledge graph is only available during the p2p collaboration session, nevertheless results and findings (per stakeholder?) can be synchronized into the named graph of the stakeholder and be analysed offline
- . . .

In the second variant it can be assumed that:
- every stakeholder holds different parts of the whole knowledge graph, even might hold the whole graph on his machine.
- stakeholders can fill out the information offline, they might get together at irregular intervals to synchronise their efforts and come up with new knowledge graph entries based on the work of the others
- during the synchronisation process there might come up discrepancies due to the different understandings of the stakeholders for a certain aspect of the knowledge graph
- there might also be different findings or result, even contradictory statements, based on the different progress of each stakeholder on the knowledge graph
- . . .

Based on chapter 4.3.2 we might come up with a design of the shared information space that looks like this:

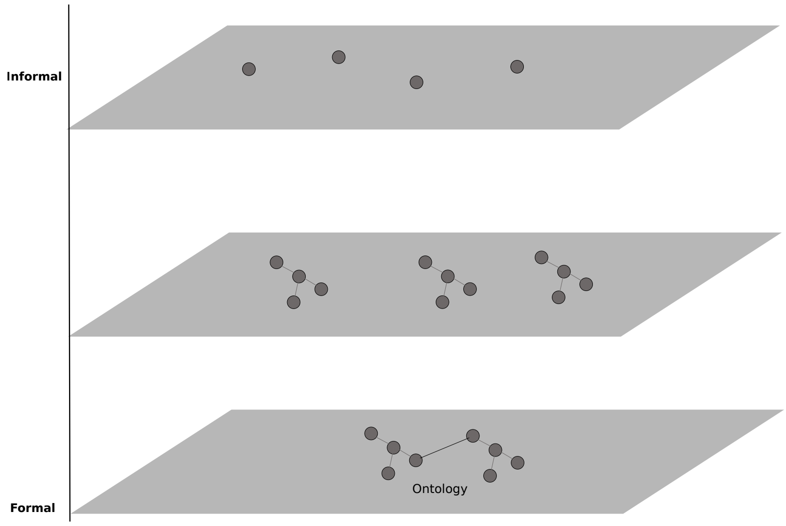

Figure .1: Shared Information Space

- it uses different layers to connect domain specific knowledge
- these layers are ordered based on formalism and models available
- the more informal a layer is the more collaboration is required to come up with a shared understanding / the common sense and connect it with the layer below
- layers can be easily turned on / off in the application to focus on a specific aspect of the investigation