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

ca. 10 pages

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 description of the problem this thesis will focus on as well as an analysis of related works and 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**, **acquirers** like PayPal or Visa, 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)

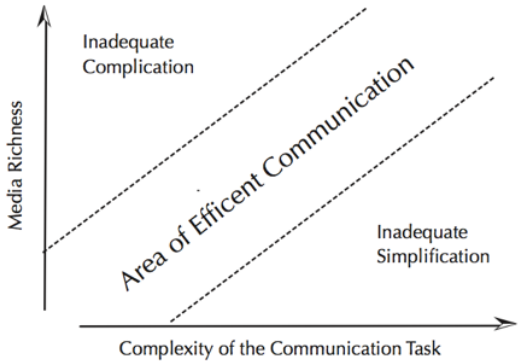

Figure 1.1: The Media Richness Model (Rice 1992)

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):

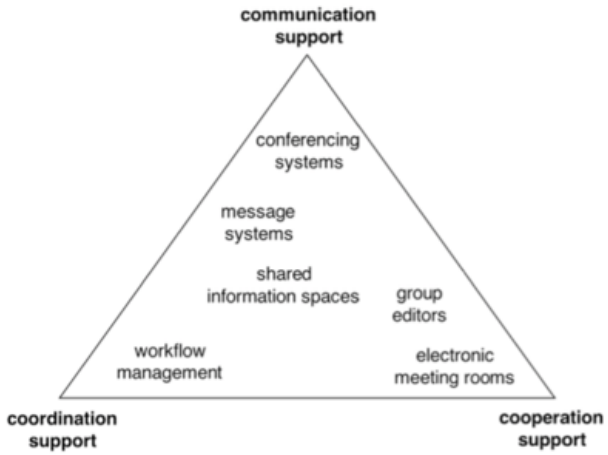

Figure 1.2: The 3C Model (Koch 2008)

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 Related Works

1.4 Master Thesis Outline

2 Context Analysis

ca. 15 pages

This chapter will look into the scenario of e-commerce fraud investigation in detail. It will start with an in-depth scenario description followed by an analysis of all involved stakeholders. 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. The chapter ends with a description of the scope this Master thesis will focus on.

2.1 Scenario Description

2.2 Stakeholder Analysis

2.3 Stakeholder Objectives

2.4 Scope of this Master Thesis

3 Theoretical Foundations

ca. 25 pages

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, WebRTC as well as less known ones like BitTorrent and BitMessage.

3.1 Computer-Supported Cooperative Work

3.1.1 Definition

3.1.2 Types

3.1.3 Shared Information Spaces

3.1.4 Important aspects of CSCW systems

3.2 The Semantic Web

3.2.1 Vision

3.2.2 Resource Description Language

3.2.3 Web Ontologies

3.2.4 Query Language

3.2.5 Agents and Rules

3.3 Peer-to-peer communication

3.3.1 Centralized vs. Decentralized Web Architectures

3.3.2 Extensible Messaging and Presence Protocol

3.3.3 Web Real-Time Communication Protocol

3.3.4 BitTorrent Protocol

3.3.5 Bitmessage Protocol

4 Concept and Design of the System

ca. 30 pages

5 Conclusion and Future Work

ca. 10 pages

List of Figures

List of Tables

Glossary

CSCW	computer-supported cooperative work.
OWL	Web Ontology Language.
P2P	Peer-To-Peer.
RDF	Resource Description Framework.
SPARQL	SPARQL Protocol and RDF Query Language.
WebRTC	Web Real-Time Communication.
XMPP	Extensible Messaging and Presence Protocol.

Bibliography

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

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

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

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

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

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

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

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

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

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

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