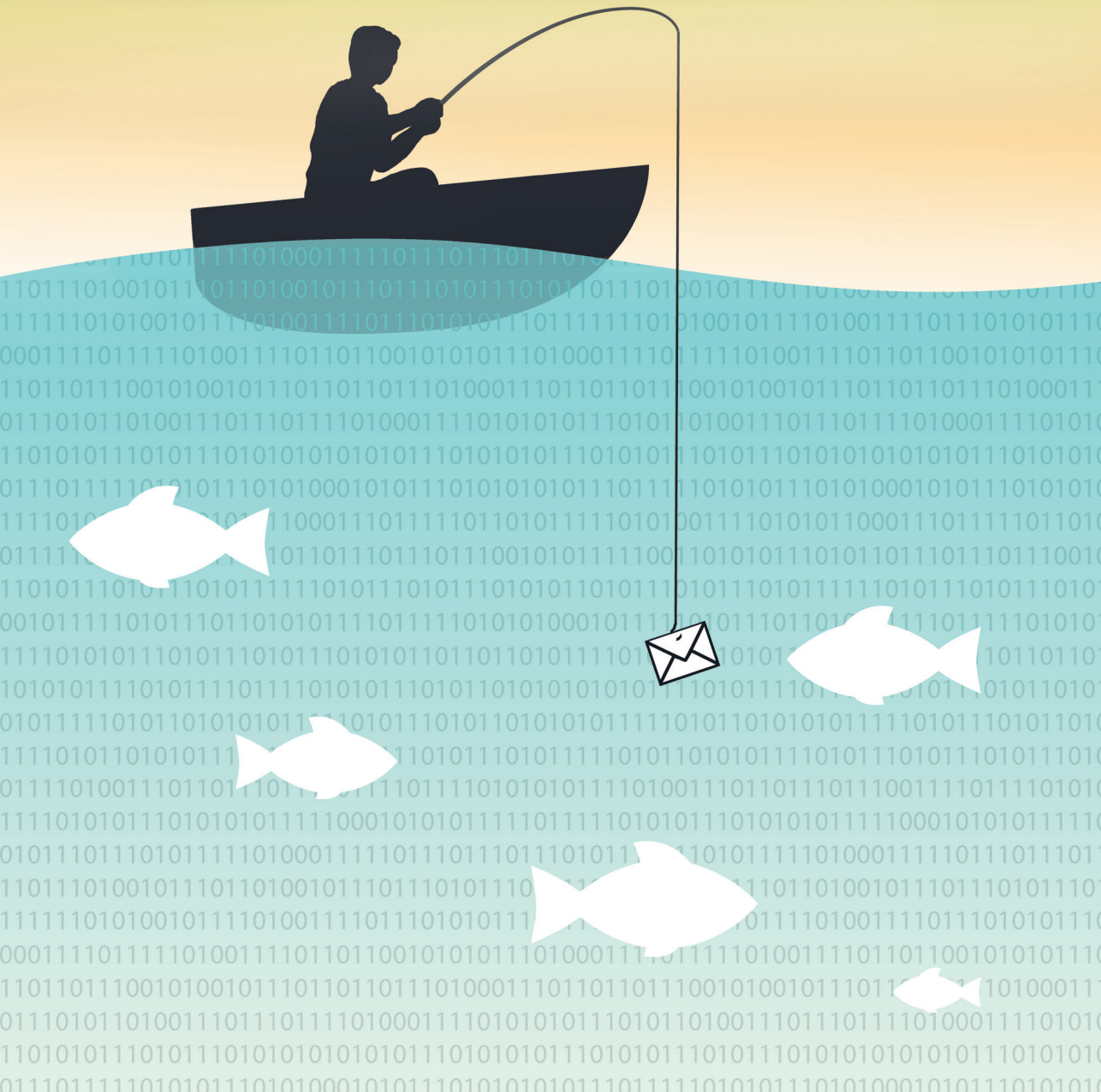


# From Fishing to Phishing

Elmer Lastdrager





# FROM FISHING TO PHISHING

ELMER LASTDRAGER

## PROMOTIECOMMISSIE

Voorzitter	prof.dr. J.N. Kok	Universiteit Twente
Promotoren	prof.dr. P.H. Hartel	Universiteit Twente
	prof.dr. M. Junger	Universiteit Twente
Leden	prof.dr. A. Pras	Universiteit Twente
	prof.dr. M.D.T. de Jong	Universiteit Twente
	prof.dr. F.L. Leeuw	Maastricht University
	dr. H. Borrion	University College London
	dr. Z. Benenson	Friedrich-Alexander Universität

IDS Ph.D. Thesis Series No. 18-455  
Institute on Digital Society  
P.O. Box 217, 7500 AE  
Enschede, The Netherlands



This research was funded through the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement ICT-318003.

ISBN: 978-90-365-4479-5  
ISSN: 1381-3617  
DOI: [10.3990/1.9789036544795](https://doi.org/10.3990/1.9789036544795)

Printed by: Gildeprint Drukkerijen  
Cover design: Remco Wetzels

Copyright © 2018, Elmer Lastdrager, Enschede, the Netherlands.  
All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photography, recording, or any information storage and retrieval system, without prior written permission of the author.



# FROM FISHING TO PHISHING

PROEFSCHRIFT

ter verkrijging van  
de graad van doctor aan de Universiteit Twente,  
op gezag van de rector magnificus,  
prof. dr. T.T.M. Palstra  
volgens besluit van het College voor Promoties  
in het openbaar te verdedigen  
op vrijdag 9 februari 2018 om 16:45 uur

door

ELMER EVERT HENDRIK LASTDRAGER

geboren op 13 februari 1987  
te Groningen, Nederland

Dit proefschrift is goedgekeurd door:

de promotoren

prof.dr. P.H. Hartel

prof.dr. M. Junger

## ABSTRACT

---

Phishing is one of the many types of cybercrime targeting internet users. A phishing message is sent with the aim to obtain information from a potential victim. One of the reasons phishing is popular has to do with the connectivity that the internet provides. A message can be spread to thousands of recipients with little effort and at negligible cost. A successful phishing attack can lead to identity theft and loss of money for the victims. When an organisation is targeted, phishing can lead to, among other things, compromised network security and stolen intellectual property.

Phishing is highly scalable. On the other side of the scalability spectrum are less scalable modus operandi. We categorise less scalable methods as “fishing for information”. In this thesis, we aim to explore the spectrum of scalability. This thesis uses a socio-technical approach by describing both experiments and technical perspectives to “fishing” and phishing.

This thesis starts by exploring definitions of phishing in literature and analysing their concepts. This provides us with a foundation of what constitutes phishing. Following on the definition, we explore two modus operandi that are less scalable than phishing, using USB keys and QR codes. We focus on measuring attack effectiveness on the boundary between the physical (i. e., objects on the floor) and digital world (i. e., getting a computer virus.) By quantifying the effectiveness of an attack using experiments, we investigate the feasibility of less scalable attacks. Then, we investigate the thought patterns that potential victims use in order to assess a phishing email. The thought patterns, or heuristics, determine whether a recipient of phishing becomes a victim or not. Knowledge on people’s thought patterns can be used to improve user training. Subsequently, we created a anti-phishing training to be provided to children. We show that training children is feasible and increases their ability to detect phishing on the short term. Finally, we performed a large-scale analysis of phishing emails in the Netherlands. We discuss patterns in terms of both attacker behaviour as well as recipient behaviour. Our results demonstrate the effectiveness of phishing with different degrees of scalability. Less scalable methods of attack require more effort on the part of the attacker, but provide higher effectiveness. More scalable attacks provide lower success rates, but require less effort than scalable attacks. The contributions in this thesis allow researchers and security professionals to better understand the dynamic nature of phishing.



## SAMENVATTING

---

Phishing is een van de vele soorten cybercrime die zich richt op internetgebruikers. Een phishing bericht wordt verstuurd met als doel om informatie van een slachtoffer te verkrijgen. De goede connectiviteit die het internet met zich meebrengt is een van de redenen dat phishing populair geworden is. Een enkel phishing bericht kan eenvoudig en vrijwel gratis naar duizenden ontvangers tegelijkertijd verstuurd worden. Een succesvolle phishing aanval kan grote gevolgen voor de slachtoffers hebben, bijvoorbeeld door identiteitsdiefstal of diefstal van geld. Echter, wanneer een organisatie doelwit is, kunnen de gevolgen nog veel groter zijn door diefstal van bedrijfsgeheimen of het platleggen van een bedrijfsnetwerk.

Phishing is goed schaalbaar. Aan de andere kant van het schaalbaarheidsspectrum zijn de minder schaalbare modus operandi. Deze minder schaalbare methoden scharen we onder het “vissen (of hengelen) naar informatie”. In dit proefschrift verkennen we dit spectrum van schaalbaarheid. Hiervoor maken we gebruik van een socio-technisch perspectief, waarbij we door middel van zowel experimenten en techniek het “vissen” en “phishen” naar informatie benaderen.

Het proefschrift begint met een onderzoek naar de verschillende definities van phishing in de literatuur. Uit deze definities worden de belangrijkste concepten gehaald. Hiermee bepalen we hoe phishing gezien wordt, iets dat de fundering voor de rest van het onderzoek is. Na de definitie-analyse bekijken we twee niet-zo-schaalbare manieren om een phishing aanval uit te voeren, namelijk door het gebruik van USB sticks en QR codes. Hierbij richten we ons op het meten van de effectiviteit van een aanval die zich bevindt op het raakvlak van de fysieke (een object op de vloer) en digitale wereld (een computervirus). Door middel van experimenten bekijken we de haalbaarheid van aanvallen die minder schaalbaar zijn, bijvoorbeeld een aanval waarbij een USB stick op de grond gelegd wordt. Hierna zoomen we in op phishing door te kijken naar de denkpatronen van potentiële slachtoffers van een phishing e-mail. Denkpatronen (ook heuristieken genoemd) bepalen of de ontvanger van een phishing e-mail slachtoffer wordt, of niet. Kennis over deze denkpatronen kan gebruikt worden om anti-phishing trainingen te verbeteren. Vervolgens kijken we naar een anti-phishing training die speciaal voor kinderen ontwikkeld is. We laten zien dat de training werkt en dat kinderen phishing e-mails beter herkennen na de training. Daarnaast laten we zien hoe lang het duurt voordat deze kennis weer wegzakt, waarna nieuwe trainingen nodig zijn. Als laatste onderdeel van dit proefschrift beschrijven we een analyse op grote aantallen phishing e-mails die door Nederlanders ontvangen zijn.

We beschrijven patronen in zowel het gedrag van de aanvallers (diegene die de phishing e-mails sturen), als in het gedrag van de ontvangers.

De resultaten van dit proefschrift laten de effectiviteit van phishing zien, voor verschillende gradaties van schaalbaarheid. Minder schaalbare methoden van phishing vereisen meer inzet van de aanvaller, bijvoorbeeld door fysieke aanwezigheid, maar bieden relatief hoge effectiviteit. Minder schaalbare methoden van phishing aanvallen zijn minder effectief, maar zijn met minder inzet van de aanvaller uit te voeren. De bijdragen van dit proefschrift stellen onderzoekers en securityspecialisten in staat om de dynamiek achter een phishing-aanval beter te begrijpen.

## ACKNOWLEDGMENTS

---

Summer 2011. I'm working on my master thesis. I'm sitting next to a PhD student, and I see the joy of having a paper accepted, and the frustration of getting rejected. "*Not something I want to do after I finish this.*" And I continue typing.

August 2011. I've graduated from my master studies. On the evening of my graduation, one of my supervisors, Svetla, sends me an email. She suggests me to take a look at two PhD positions in the DIES group. For that, I thank her a lot.

I'm thinking back of my first weeks in the office. Thinking of the meetings with my supervisors, Pieter and Marianne. Pieter, thanks for always being critical and improving my work significantly. Thanks for giving me the opportunity to explore and for believing in the outcome. Marianne, thanks for your endless advice and help on the social-science part of my work. As a computer scientist, I had to learn so much about crime science and experimental research.

I started working in the DIES group, in a 6–8 person office, with only one colleague, with whom I started at the same day. Lorena, thanks for everything. We've met each other in various countries, since our holidays occasionally coincide. What are the odds?

A large office, with empty desks. In the middle of the room a lamp is hanging from the ceiling. I try the light switch, the light does not turn on. When asking for a lightbulb, someone from the facility services shows up and gets angry: which @\$%!& put that lamp there?!?! I did get a new lightbulb though.

I'm thinking about the colleagues that welcomed me to the group. Crazy colleagues. I remember the lunch and coffee breaks. Twelve o'clock lunch. Knocking on all doors. Time to have lunch. Lunch is a lot more fun with a group of people. Three o'clock coffee break, no excuses. Amazing conversations with amazing people. Going out for food and beers. Inspiring suggestions. Thanks for the great time Arjan, Christoph, Jonathan, Eleftheria, Dina, Begül, Jan-Willem, Marco, Michael, and Stefan.

The secretary is the core of the group. Thanks Bertine, and later also Suse. Bertine, thanks for making me feel welcome from day one, for all the talks, and for all the help with everything. You are one of the strongest and friendliest persons I know.

Our group expanded, merged, and eventually became the scs group. Moving offices, and meeting new people. Sharing an office with Jan-Willem and Dan, and traveling with them. Thanks for all experiences. Also, thanks to all other colleagues of the group: Riccardo, Tim, Ali, Steven, Geert Jan, Prince, Alexandr, Thomas, Susanne, Inés, Chris,

Chris, Bence, Roeland, Joao, Carlos, Robson, Glaucia, Hans, Wilbert, Maarten, Andreas, Faiza, Eelco, Frank, Alireza, Roel, Raymond, Ida, Yuxi, Meiru, Didier, Maya, Luuk, Klaas, Luis, and anyone that I might have missed.

Data is very important. Getting access to good data is very difficult. Therefore, I would like to express my gratitude to the Fraudehelpdesk for providing me with the biggest dataset I could ever imagine. A special thanks to John Kellij for the friendly and direct way of working together. Additionally, I want to thank Fleur, Jos, Erwin, and all other employees of the Fraudehelpdesk for their input. I would like to thank Roeland van Zeijst for introducing me to the Fraudehelpdesk, Rob Heijjer for data on phishing reports, and Nicole van der Meulen for statistics on phishing incidents for the large banks.

Throughout my PhD, I've had the honour to supervise many interns and bachelor thesis and master thesis students. This inspired me greatly. Thanks Matthijs, Marjolein, Laura, Henry, Jurgen, Frank, Job, Inés, Lars Nick, Denise, Nolie, Ruben, and all other students that I've worked with.

Life is boring without sports. Throughout my PhD, I've played ice hockey for the Slapping Studs. Thanks to all members and former members for the amazing times. We won some, we lost a lot, but we always had fun.

Without family, nobody gets far. Thanks for your eternal support Edwin and Rita, and for encouraging me to get this far. My gratitude is limitless. My 'bro' Casper, and 'sis' Birgit, I can't imagine better siblings. Thanks for everything. And a big thanks to Tim for all the great times. Finally, Eleftheria. First only as a colleague, later as my partner: thanks for your support during this journey. I'm looking forward to the new journeys to come.

I open my  $\text{\LaTeX}$  editor for the last time. Compiling... *"Please compile without errors."* And it does.



## CONTENTS

---

1	INTRODUCTION	1	
1.1	Scalability	2	
1.2	Modus operandi	4	
1.3	A Model of Phishing	5	
1.4	Research questions	7	
1.5	Contributions and outline	9	
2	TOWARDS A DEFINITION OF PHISHING	12	
2.1	Background	12	
2.2	Method	15	
2.2.1	Selection of Literature	15	
2.2.2	Identification of common words	17	
2.2.3	Identification of concepts	18	
2.2.4	Analysis of concepts	19	
2.3	Results	21	
2.4	Discussion	23	
3	FISHING FOR INFORMATION	28	
3.1	USB Keys	31	
3.1.1	Method	34	
3.1.2	Results	40	
3.1.3	Discussion	43	
3.1.4	Implications	48	
3.2	Phishing With QR Codes	48	
3.2.1	Method	49	
3.2.2	Results	53	
3.2.3	Discussion	55	
3.3	Conclusions	56	
4	HEURISTICS OF PHISHING	59	
4.1	Background	60	
4.1.1	Trust	60	
4.1.2	Characteristics For Victimisation	62	
4.2	Methodology	63	
4.2.1	Subjects	64	
4.2.2	Design	65	
4.2.3	Procedure	66	
4.2.4	Pilot	69	
4.2.5	Analysis	69	
4.2.6	Limitations	73	
4.3	Results	74	
4.3.1	Urgent versus non-urgent	75	
4.3.2	Victimisation	77	
4.3.3	Reading patterns	80	

4.4	Conclusions	82
5	PHISHING EDUCATION FOR CHILDREN	85
5.1	Methodology	87
5.1.1	Design & Concepts	87
5.1.2	Ethics	91
5.1.3	Setting	91
5.1.4	Subjects	92
5.1.5	Analysis	93
5.2	Results	94
5.3	Discussion	101
5.3.1	Limitations	104
5.4	Conclusions	106
6	PATTERNS IN PHISHING	108
6.1	Methodology	109
6.1.1	Email Similarity and Clustering	111
6.1.2	Patterns in Suspicious Emails	115
6.1.3	Behaviour of Targeted Users	118
6.2	Results	119
6.2.1	Context of the data	120
6.2.2	Patterns in Suspicious Emails	123
6.2.3	Behaviour of Targeted Users	127
6.2.4	Impact of APATE	130
6.3	Discussion	131
6.3.1	Future Work	133
6.3.2	Policy Implications	133
7	CONCLUSIONS	135
7.1	Discussion of research questions	135
7.2	Future research directions	138
7.3	Final words	139
A	LIST OF ANALYSED DEFINITIONS OF PHISHING	140
B	PHISHING HEURISTICS QUESTIONS	150
C	PHISHING EDUCATION TEACHING AND TESTING MATERIAL	152
C.1	Statistical Assumptions	152
C.2	Slides of Presentation	154
C.3	Phishing Test	163
	BIBLIOGRAPHY	175

## INTRODUCTION

---

*Phishing* is a scalable act of deception whereby impersonation is used to obtain information from a target ([Chapter 2](#)). Offenders impersonate governmental organisations, financial institutions, but also retailers and service-oriented companies (Anti-Phishing Working Group, [2015b](#)). A typical scenario includes an offender who sends out an email pretending to be from a bank to its customers. Using a fake message, the targets are deceived to perform a certain action, such as clicking on a link, calling a number, or sending a reply with information. Phishing attacks hit the news headlines on a daily basis. The general public receives phishing emails, companies suffer from attacks that started with a phishing email, and even governments are targeted. Generally an offender expects a benefit, or *return of investment*, from committing a crime (Cornish and Clarke, [1986](#), [2014](#)). In the case of phishing, scalability is important to obtain a benefit. The response rate to phishing messages may be low, but due to scalable methods of sending phishing messages, a sufficient number of targets can be reached. Email is such a scalable medium for sending phishing messages. An individual can send thousands of emails per minute using a single computer. Using botnets, a single person can send messages to millions of targets almost simultaneously. A phishing offender can send messages and monetise the obtained information from anywhere in the world. This leads to phishing being a flexible and dynamic type of digital fraud.

Despite countermeasures such as spam filters, blacklists and user training, the general public still receives phishing emails (see [Chapter 6](#)) and continues to ‘bite the hook’. Indicating the prevalence of phishing is difficult. Phishing studies traditionally start by indicating the loss of phishing in terms of money (e.g., Sheng, Kumaraguru et al., [2009](#); Almomani, Gupta et al., [2013](#); Leukfeldt, [2014](#); Hong, [2012](#)). However, such statistics are often biased (Florêncio and Herley, [2011](#); Moore and Clayton, [2010](#)). Furthermore, people do not necessarily know they are a victim. When a victim fills in his information on a phishing website, or replies to a phishing message, he does not necessarily realise the mistake. When the information consists of credentials to a bank website, the loss of money will likely alert the victim about the attack. However, when other information is stolen (consider a copy of a passport), this may not be clear to the victim. The victim may realise what happened only when his information gets misused later, for example, if the information is used for getting a phone subscription and the victim receives the bills. The problem of such misuse of one’s information is known as *identity theft*. According to expert interviews, identity theft is most

often initiated with a phishing attack Paulissen and van Wilsem (2015). A representative survey of Paulissen and van Wilsem (2015) found that 4.6% of the residents of The Netherlands aged over 14 experienced identity theft in the last two years. Statistics Netherlands (2017) found that the number of phishing victims for the period 2012–2014 remained stable at 0.4% of the total population, and went down to 0.3% in 2015–2016. However, only people who are aware of their victimisation from a phishing message are included in that number. Furthermore, victims do not report the phishing attack at all, or report it to institutions other than the police, resulting in under-reporting. A large survey of Statistics Netherlands (2015) on identity theft as a cybercrime (i. e., phishing and skimming) show that only 14% of the victims reported having gone to the police in 2014. In 2016, the number of online identity theft victims reporting to the police went down even more, to 8% (Statistics Netherlands, 2017). In comparison, 80% of the Dutch victims reported their victimisation to a financial institution in 2014 (Statistics Netherlands, 2015). This can be explained by phishing campaigns often targeting banks, and victims being able to get their money back after filing a report. However, it does show that the willingness to report victimisation is low when reporting does not lead to getting back money.

Due to the digital means of communication, cybercrimes are easier to scale than their non-cyber equivalents. With the right knowledge and skills, breaking in to several computers (*hacking*) can be performed with little effort and low risk of being caught. The non-cyber equivalent would be burglary. It is arguably more difficult to break in to ten houses without being caught, than to break in to ten computers without being caught. This is primarily caused by the mandatory physical presence for a burglar. Digital crimes have the advantage of not requiring physical presence. This leads to the ability to target multiple victims and simultaneously victimise them. The ability to target multiple victims and the speed of being able to target subsequent victims, are properties of a crime's scalability. Looking at crimes in terms of their scalability has the advantage of going beyond the exact medium (i. e., cyber or physical) that is used.

### 1.1 SCALABILITY

The concept of scalability can be conceptualised as a dimension, with many gradations. To illustrate this, consider an offender who wants to obtain bank account details from his victim. The least scalable method would be to meet in person and talk to the victim. This requires the offender to come up with a good story and convince the victim to hand out the information. This does not scale: if the offender wants to attack multiple victims, he would need to talk to each of them. Bounded by physical restraints, this requires lots of time and constant concentration.

Furthermore, there is a non-negligible risk of being caught red handed. Therefore, personally talking to the victim is not scalable.

An alternative for verbal communication is writing a letter. Sending a message to someone could be done by writing it down in a letter and sending a messenger to deliver it. An example of such a message is the Nigerian *advance fee fraud* letter (Smith, Holmes and Kaufmann, 1999; Edelson, 2003). Using the postal system, or private messengers, a letter can be delivered to another person. However, there is a fee per letter, and deliveries are often infrequent or delayed. Letters are more scalable than personal contact, because they can be sent to lots of different persons. However, sending large quantities of letters requires a significant investment in terms of time and money. Normal street-side mailboxes would be insufficient and too time consuming to use. Signing a contract with a postal agency to handle so many letters would solve the situation, but makes the offender trivially traceable. Sending letters as *modus operandi* is not scalable, even though it scales better than talking in person.

In the late 18th century, the mechanical telegraph emerged (Standage, 1998). Using semaphore signalling, messages could be transferred at a speed of up to 3 symbols per minute (Encyclopædia Britannica, 2015). With the introduction of the electronic telegraph, it became more efficient to send messages regardless of fog or lack of daylight (Standage, 1998). A message could be transmitted within minutes or hours, compared to days when sending a letter by post. And due to the large scale deployment of the telegraph network, including a transatlantic connection, large numbers of people could be reached. Still there was a high cost per message. From an offender's point of view, this means a high risk investment for running a large-scale fraud. Other ways of cheating were used, taking advantage of the speed at which a telegram arrives. For example, the results of horse races or lotteries could be transmitted by telegram to other parts of the country, where the official results were not known yet and betting was still allowed. The accomplice receiving the telegram could take advantage by betting on the winner or choosing the winning numbers (Standage, 1998).

The introduction of the internet, and more specifically email, was another drastic change in messaging. An email server can process thousands of emails per minute, thereby scaling even better than the telegraph network. Additionally, apart from the need of an email inbox and internet connection, sending and receiving emails is free of charge. The consequences of a large userbase, lack of a central authority and no price per message are significant. Merchants can send advertisements to many potential customers at low cost. As with many new technologies, this simultaneously opened opportunities for offenders as well. In its core protocols, a receiving email server does not authenticate the sender (RFC2821). This allows for unwanted messages and advertisements, called *spam*, to enter the user's email inbox. Currently, many

solutions against spam exist, but are unable to filter all unwanted emails. Therefore, email remains an attractive medium for sending spam (and phishing) messages.

Besides email, other ways of sending phishing messages are being used as well. For example, phishing messages can be distributed using SMS (Castiglione, De Prisco and De Santis, 2009), or by sending prerecorded messages over VOIP (Jakobsson and Myers, 2007). Furthermore, social media platforms like Twitter (Aggarwal, Rajadesingan and Kumaraguru, 2012; Chhabra et al., 2011) and Facebook (Chhabra et al., 2011; Mills, 2009) offer a large number of potential targets. However, whereas it is relatively easy to fake the sender of an email message, this is harder for social media platforms. This, in combination with the mass adoption of email led to the situation where email remains to be the most popular medium for distributing phishing messages.

## 1.2 MODUS OPERANDI

There is no single *modus operandi*, or employed method, for phishing. Instead, offenders choose a subset of the many available options for an attack. Regardless of the methods and tools offenders use, the essence of phishing is simple. At some moment in time, the offender convinces the target to provide information. Information can be almost anything, such as credentials, identity information, or company secrets. The offender uses a medium to send a phishing message to the target. If the target falls for the message, he will return information to the offender. The information does not have to travel on the same medium as the original message. For example, a phishing email could request people to reply by clicking on a link.

In a typical scenario, the offender needs to take three steps: (1) setting up the attack; (2) sending messages and gathering information; and (3) monetising the obtained information. In the setup phase of the phishing scenario, the offender needs to arrange several things. Foremost, he needs to craft a phishing message, typically an email, in which an organisation is impersonated. Typically, banks, package delivery companies and webshops are good candidates. One of the reasons for candidacy is that they are well known and often trusted. On the technical side, the offender needs to obtain lists of email addresses. Furthermore, the offender should get capacity to distribute many emails. Often, this capacity is achieved using *botnets* or hacked servers. Botnets are groups of computers with a virus infection, that are under control of a *botnet herder*. The offender can rent or create such a botnet, and order the infected computers to send out the phishing emails. Alternatively, the offender can break into a web server that runs vulnerable software (Vasek, Wadleigh and Moore, 2015), and use it to distribute emails. Finally, in the typical scenario, the offender needs

to host a phishing website, often called a *landing page*. At this landing page, the victims that fall for the phishing email, are asked to provide information, such as access credentials to the online bank environment. A compromised webserver may be used for this, to avoid linking the attack to the offender.

Once the offender's set-up is ready, it is only a matter of waiting for victims to fall for the attack. Similar to fishing, the offender needs to wait for an inattentive victim to click on the link. Once that happens, the victim will go to the landing page, where the victim is requested to login. When logged in, the credentials are sent to the offender, for example by email. Next, the offender will proceed to the third and final step, which is to monetise the information. Monetising can be done by either selling the information (or credential), or using it. The offender can, for example, log in to the online bank website using the stolen credentials. Then, he will transfer money to the account of a *money mule*, who is an outsider that withdraws the money from his account. What happens after varies a lot. For example, the money mule can send the money via Western Union to the offender (Moore, Clayton and Anderson, 2009), or to an anonymous mailbox, or buy a gift card and email the code of the gift card to the offender.

### 1.3 A MODEL OF PHISHING

Phishing attacks are continuously evolving (Hong, 2012; Jakobsson and Myers, 2007). Countermeasures are implemented to mitigate the newest phishing attacks, only to be followed by a different attack later. This is an ongoing arms race. Offenders choose a *modus operandi*, as well as the accompanying strategy for performing a phishing attack. The chosen *modus operandi* has a certain scalability attached to it. Together, the *modus operandi* and scalability properties lead to a certain effectiveness of an attack.

To clarify this in the present thesis, we want model the relation between scalability and effectiveness for phishing *modus operandi*. The effectiveness is shown as the extent to which an attack is successful, also known as the *success rate*. For example, when an attacker sends 1000 emails, resulting in 50 replies with information, the success rate is 5%. We define the scalability as one of three values: low, medium, or high. For the purpose of our model, we define *low scalability* as the situation where the attacker and the victim have a one-to-one interaction (i. e., one attacker for one victim). Examples of attacks that are low in scalability are face-to-face attacks and phone calls. On the other end of the spectrum is an attack of *high scalability*, where one attacker can have many victims. Highly scalable attacks are one-to-N, for a large N. An example of a highly scalable attack is sending spam emails. In the middle of the spectrum is an attack which has a *medium scalab-*

ility. For an attack to be medium on the scalability spectrum, there should be a one-to-x relation, whereby x is limited by, for example, physical restraints or the need for victim-specific information. For example, sending personalised phishing emails requires the attacker to gather a lot of information for each victim, thereby limiting the potential number of victims. The resulting model is shown in Figure 1, and in the following paragraphs we discuss the data points within the model. Additionally, Figure 1 shows the distinction between *Fishing* for information (i.e., a less scalable attack for information) and *Phishing* (i.e., the scalable version). Methods that have a low scalability can be categorised as social engineering (fishing for information), whereas we consider high scalability methods as phishing.

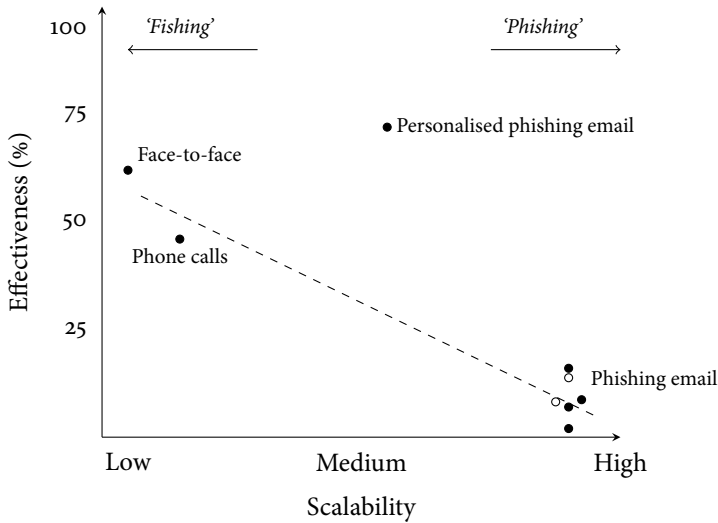


Figure 1: The effectiveness versus the scalability per *modus operandi*. Hollow circles represent real world data.

The model of Figure 1 was filled with data points from the scientific literature. Measurements on the success rate of phishing in the real world are scarce. Most research analyses phishing in a lab setting. There is some data of studies measuring phishing in the wild, or studies performing large-scale measurements on unsuspecting users. Mohebzada et al. (2012) performed two large scale studies ( $N=10,917$ ) to measure the success rate of a phishing email and found success rates of 8.74% and 2.05%. A study of Jakobsson and Ratkiewicz (2006) found success rates of 7% ( $\pm 3\%$ ) when the URL was an IP address and 11% ( $\pm 3\%$ ) when the URL was a domain name ( $N=281$ ). Finally, Jagatic et al. (2007) found a success rate of 16% ( $\pm 7\%$ ) in an experiment with 94 subjects. However, they noted that the relatively high success rate could be due to some contextual information in the email.



Several measurements on phishing in the wild have been performed as well. Notably, these include studies performed by Google and Microsoft on a large user base. Garera et al. (2007) found that 8.24% of the users who view a phishing page will become victim, based on toolbar data from Google. Furthermore, on average 13.78% of all visitors to phishing pages that are hosted on Google Forms submitted data (Bursztein et al., 2014). The numbers suggest a high success percentage. However, one must take into account that these are percentages of people that have already clicked on the link in a phishing email. The actual success rate must therefore be lower. Florêncio and Herley (2007) analysed the browsing behaviour of a 436,000 users by looking at data from the Microsoft Live toolbar, and used these analyses to conclude that 0.4% of the population is victimised by phishing each year. If everybody receives only one phishing email per year, the success rate would be 0.4%. One might argue that the real success rate of an average phishing email must be an order of magnitude lower. For the purpose of modelling phishing, we assume the effectiveness of a general phishing email in terms of success rate to be between 2% and 16% (Mohebzada et al., 2012; Jakobsson and Ratkiewicz, 2006; Jagatic et al., 2007; Garera et al., 2007; Florêncio and Herley, 2007).

In terms of *modus operandi* with medium scalability, Jagatic et al. (2007) harvested information about students and their acquaintances and used this knowledge to perform a personalised phishing attack. The corresponding success rate was 72% ( $\pm 3\%$ ) in an experiment with 487 subjects. Finally, in the low scalability area, we cite two studies related to phishing. Firstly, telephone-based social engineering has a success rate of 46% (Bullee, Montoya et al., 2016) ( $N=118$ ). The second study was a face-to-face social engineering study ( $N=48$ ), with a success rates of 62% (Bullee, Montoya Morales et al., 2015).

Offenders weigh effort and risk against the potential reward (Cornish and Clarke, 2014). Our model shows the combination of the effort (by *modus operandi*) and the potential reward. From the point of view of an attacker, the ideal *modus operandi* consists of a highly scalable attack that has a high effectiveness. However, such an attack may require more effort. In the end, offenders choose a *modus operandi* they consider suitable for getting a return on investment.

#### 1.4 RESEARCH QUESTIONS

The various forms of ‘fishing’ and phishing as a method of obtaining information. As discussed before, one can try to establish the point at which the non-scalable ‘fishing’ stops and the scalable ‘phishing’ starts. However, even though many researchers have published on the topic of phishing, there does not seem to be a central definition of phishing, as further discussed in [Chapter 2](#). Obtaining data on the effectiveness of

various scalable and less scalable attacks would be needed to discuss the scalability and effectiveness properties. One may wonder whether less scalable modus operandi have a better yield than the scalable versions. In other words: how does a physical ‘fishing’ attack compare to a scalable ‘phishing’ attack? This resulted in the following research question:

**RESEARCH QUESTION 1:** How does an attack’s effectiveness relate to the modus operandi’s scalability?

Measuring the effectiveness of an attack is important, as is measuring what influences the effectiveness. When discussing the topic of phishing, one commonly hears the phrase “*I would never fall for a phishing attack.*” However, many internet users become victim of phishing, in the order of 0.3% of the Dutch population (Statistics Netherlands, 2017). When someone receives a phishing email, (s)he will decide at a certain moment whether the email is legitimate or fraudulent. Knowing how this decision process is performed, allows for the creation of better education. This leads to the following research question:

**RESEARCH QUESTION 2:** How do people decide whether or not an email is phishing?

Prevention is important to reduce the number of victims of phishing. Many interventions have been proposed to inform the general public and guide them into making better decisions when receiving a phishing email. Some interventions are targeted towards groups of potential victims, such as university students or employees of a certain company. Children are often not considered potential victims, due to their limited online responsibilities, such as (online) banking. However, they are active online, and therefore a potential target of phishing. Improving their online safety is challenging. Therefore, the fourth research question is:

**RESEARCH QUESTION 3:** How can we reduce the effectiveness of phishing on children?

Providing statistics on the number of phishing attacks, or victims, is difficult due to the lack of an overview. Phishing occurs online and therefore potentially cross-border in the physical world. Victims report to the police, to their financial institutions, to non-profit anti-fraud agencies, or they do not report victimisation at all. Attempts at victimisation are even harder to monitor. However, to describe a phenomenon, or to reduce its impact by prevention, it is important to know the extent of the problem. Therefore, our last research question is:

RESEARCH QUESTION 4: What patterns can be found in phishing campaigns in the Netherlands?

Answering these four research questions leads to a better understanding of phishing, and the answers will hopefully validate our model of phishing.

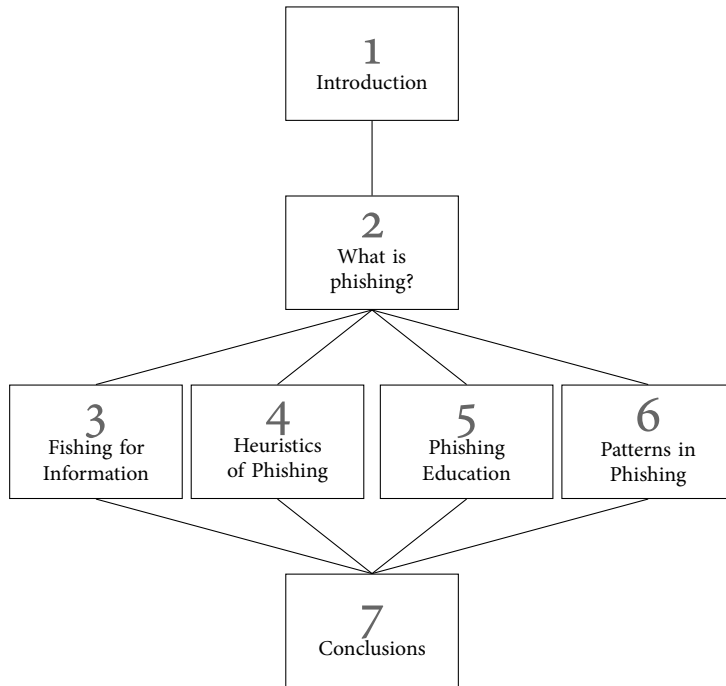


Figure 2: The outline of this thesis.

## 1.5 CONTRIBUTIONS AND OUTLINE

The outline of this thesis is shown in [Figure 2](#). After this introduction, we discuss what phishing is by looking at definitions of phishing. This is followed by four chapters that discuss phishing from various angles. Finally, we provide our conclusions and directions for further research. The thesis is divided into the following chapters:

**INTRODUCTION:** The current chapter provides the motivation for our research, introduces the research questions and provides an overview of the thesis.

**WHAT IS PHISHING?** We discuss phishing as a phenomenon in [Chapter 2](#). Using an extensive literature study, we compare phishing

definitions from existing literature. Core elements of the phenomenon are extracted from literature. Elements lacking a consensus in the literature are discussed in detail. This eventually leads to a uniform and consensual definition of phishing.

**FISHING FOR INFORMATION:** The scalability of an attack's *modus operandi* and its effectiveness influence each other. To measure this, we performed experiments in physical world, as described in [Chapter 3](#). In the experiment, we dropped USB keys and observed the behaviour of the people who found the USB keys. This places the chapter on the boundary between the physical (i. e., objects on the floor) and digital world (i. e., getting a computer virus.) Furthermore, we describe a second experiment on the intersection of the physical and digital worlds. In the second experiment, QR codes pointing to a phishing website were distributed in a hospital. Both experiments explore the risk taking of individuals and quantify the response to fraud cues in the physical world.

**HEURISTICS OF PHISHING:** When someone receives a phishing message but does not become a victim, (s)he will most likely have become suspicious at some moment in time. Whether it is the title, sender or content that alerts the receiver, the person's heuristics have prevented victimisation. At the same time, those heuristics may prove ineffective against a particular message. Since these heuristics depend on the individual, studies are needed to detect these thinking patterns. [Chapter 4](#) describes a lab study where participants have to think out loud while reading a phishing email. The participants' current knowledge about phishing emails can be found by identifying the heuristics that the participants used while reading the email. With this information, training and public awareness campaigns can be fine-tuned.

**PHISHING EDUCATION:** Many interventions against phishing, such as training, TV commercials, or even games, are aimed at adults. This makes sense, since adults have more to lose in terms of money or information. However, due to this focus, the adults of tomorrow are often overlooked. In [Chapter 5](#), the results of a cyber hygiene training tailored to children are shown. With a small intervention, children score significantly better. Additionally, we measured the decay of the training over time.

**PATTERNS IN PHISHING:** In [Chapter 6](#), we describe the prototype of a system that was built to automate the analysis of reported phishing emails. Over 1.4 million emails were reported by the general public to the Dutch Fraud Helpdesk (Fraudehelpdesk, [2016](#)) between 2013 and September 2017. These are emails that found their way to someone's email inbox and were subsequently reported. We show patterns in

the emails in terms of phishing campaigns, as well as patterns in the behaviour of the receivers of these phishing emails.

CONCLUSIONS: Finally, [Chapter 7](#) will conclude with the answers to the research questions and directions for future work. The results of our experiments provide insights in the phishing process from different perspectives.



## 2.1 BACKGROUND

The term *phishing* is currently widely used with thousands of mentions in the scientific literature, lots of media coverage and widespread attention from organisations such as banks and law enforcement agencies. However, this prompts a question: what exactly is phishing? In some publications, the phenomenon of phishing is explicitly defined; in some, it is described by means of an example, while others assume that the reader already knows what phishing is. Many authors propose their own definition of phishing, leading to a large number of different definitions in the scientific literature.

With no scientific consensus, other sources could provide a standard definition. The first point of reference for finding the definition of a word would be a dictionary. Four definitions from prominent English dictionaries are shown in Table 1. Additionally, it lists the definition of the Anti-Phishing Working Group (APWG), a non-profit foundation that keeps track of phishing. The APWG definition is rather lengthy compared to the dictionary definitions. The five definitions vary in the level of detail and the scope of the phenomenon. For example, whereas the American Heritage definition includes phone calls, the others do not. In addition, the goal of phishing differs in the definitions, ranging from financial account details (Collins, APWG) to the more general personal information (Oxford, Merriam-Webster, American Heritage). There is greater consensus about the origin of the term phishing; it was first used around 1995-1996 (Oxford University Press, 2012; Khonji, Iraqi and Jones, 2013; Purkait, 2012; James, 2005) and is a variation on the word 'fishing', something hackers commonly did (Oxford University Press, 2012; Purkait, 2012; James, 2005; McFedries, 2006). In common with fishing, phishing is about setting out 'hooks', hoping to get a 'bite'.

The lack of a standard definition of phishing has been observed previously (Khonji, Iraqi and Jones, 2013; Abu-Nimeh et al., 2007; Al-Hamar, Dawson and Al-Hamar, 2011). This causes several problems for scientists, practitioners and consumers. For scientists, it is difficult to compare research on phishing in a meaningful way. Aggregating research consists of classification (in which attacks are considered phishing), and identification (measuring how often it occurs). Furthermore, countermeasures against phishing cannot be effectively evaluated

<sup>1</sup> This chapter is based on the paper "Achieving a Consensual Definition of Phishing Based on a Systematic Review of the Literature" (Lastdrager, 2014) in *Crime Science*, 3(9), 2014.

Source	Definition
Oxford (UK)	The fraudulent practice of sending emails purporting to be from reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers, online.
Collins (UK)	The practice of using fraudulent e-mails and copies of legitimate websites to extract financial data from computer users for purposes of identity theft.
Merriam-Webster (USA)	A scam by which an e-mail user is duped into revealing personal or confidential information which the scammer can use illicitly.
American Heritage (USA)	To request confidential information over the internet or by telephone under false pretenses in order to fraudulently obtain credit card numbers, passwords, or other personal data.
APWG (USA)	Phishing is a criminal mechanism employing both social engineering and technical subterfuge to steal consumers' personal identity data and financial account credentials. Social engineering schemes use spoofed e-mails purporting to be from legitimate businesses and agencies, designed to lead consumers to counterfeit websites that trick recipients into divulging financial data such as usernames and passwords. Technical subterfuge schemes plant crimeware onto personal computers to steal credentials directly, often using systems to intercept consumers online account user names and passwords – and to corrupt local navigational infrastructures to misdirect consumers to counterfeit websites (or authentic websites through phisher-controlled proxies used to monitor and intercept consumers' keystrokes).

Table 1: Definitions of phishing from four dictionaries and the APWG.

without knowing the extent of the phenomenon. Additionally, having no standard definition is an indication of the immaturity of the field with researchers refining their own definitions over the years (e. g., Kumaraguru, Sheng et al. (2010) and Kumaraguru, Rhee, Acquisti et al. (2007); Moore (2007) and Moran and Moore (2010); and Hong (2012), Xiang and Hong (2009) and Xiang, Hong et al. (2011)). Institutions, such as banks or governments, face problems understanding one another if their definitions of phishing are different. For example, one bank may consider a fraudulent phone call to be phishing, whereas another bank will not, making a comparison of victimisation or countermeasures difficult. Consumers may also experience the downside of a lack of a standard definition. Persons who are less computer literate, for example, may become confused when several awareness campaigns describe phishing differently.



We aim to clarify the definition of the phishing phenomenon by analysing existing definitions, in contrast to most standard definitions, which are developed using expert panels. The resulting definition is based on consensus drawn from literature, and is sufficiently abstract to support future developments. To the best of our knowledge, no previous attempt has been made to synthesise a definition of phishing.

In order to interpret existing definitions of phishing in the right context, one needs a theoretical framework. An initial exploration revealed that phishing contains elements from criminal activities. Crime science theories are used for crime in the physical world, which raises the question of their applicability in the digital world. Previous research supports the idea of applying crime science theories to digital crime (Reyns, Henson and Fisher, 2011; Pratt, Holtfreter and Reisig, 2010; Yar, 2005) and there is limited evidence of its applicability to phishing (Hutchings and Hayes, 2009). Therefore, crime science theories are used to achieve a better understanding of phishing and to provide us with concepts to analyse it. The focus of crime science is on the opportunity for a crime, rather than on the characteristics of the criminal. Three theories on crime opportunity form the foundation of crime science (Clarke, 2009; Felson and Clarke, 1998): the Rational Choice Perspective; Crime Pattern Theory; and the Routine Activity Approach. Each of these theories takes a distinctly different approach to crime (Clarke, 2009). The rational choice perspective offers a view on offender's decision-making, assuming bounded rationality (Cornish and Clarke, 2008). An offender is assumed to make a rational decision and commit a crime if the perceived benefit outweighs the perceived cost. Crime pattern theory (Brantingham and Brantingham, 1993, 2008) focuses on the relation between crime and the physical environment, in particular the crime opportunities that emerge in the daily lives of the offender. According to crime pattern theory, crime is not randomly distributed in time and space. For example, a potential offender may come across opportunities for crime during his regular daily commute. Finally, the routine activity approach (Cohen and Felson, 1979) states that a crime occurs when a likely offender and a suitable target converge in the absence of a capable guardian. Routine activity theory can be interpreted broadly (Reyns, Henson and Fisher, 2011; Pratt, Holtfreter and Reisig, 2010) to include crime without direct contact. For example, in the case of cyber bullying an online chat room can be the location where an offender and victim "meet". The focus on offender decision making within the rational choice perspective makes this theory less suited for reasoning about phishing, since the offender is mostly unknown. Similarly, applying crime pattern theory is difficult for phishing, since it often occurs on the internet. The routine activity approach however, is applicable to phishing (Hutchings and Hayes, 2009) with concepts such as offender and target, especially useful.

To elaborate upon the routine activity approach, crime scripts (Schank and Abelson, 1975; Cornish, 1994) can be used. Crime scripts describe the sequential steps that lead to an offence, much like a film script. Using crime scripts allows for interpretation of definitions of phishing in such a way that the act of phishing is decomposed into several steps. An example of such a step is “Victim receives an email”. To fully understand each definition, we decompose each step into several key concepts. To structure the identification and classification of these concepts, we use the 3A model (El Helou, Li and Gillet, 2010). The 3A model is an activity-centric framework that provides three categories: Actors, Assets and Activities. In the context of phishing, actors are humans (e. g., the offenders) who conduct activities (e. g., send a message) to achieve their goal. The goal itself could be to obtain an asset (e. g., credentials). The routine activity approach together with the tools of crime scripts and the 3A model, are used to identify relevant concepts within each definition.

The goal of the literature search is to find scientific definitions of phishing. We formulated the following research question: *How is phishing defined in the research community?* Three steps are taken to generate a definition. Firstly, relevant literature is selected and definitions of phishing are extracted. Secondly, the concepts of phishing are extracted and scored according to their occurrence. Finally, concepts that are found in most definitions are selected and a standard consensual definition is developed from these concepts.

## 2.2 METHOD

### 2.2.1 Selection of Literature

To obtain data on the existing definitions of phishing, a systematic study of the peer-reviewed scientific literature was performed, following the guidelines of Kitchenham and Charters (2007). Three digital libraries were selected for the search: ACM digital library, IEEExplore and Scopus. The fields relevant to phishing, such as computer science and various social sciences (i. e., psychology or criminology), are covered by these three databases. The literature search (see Figure 3) resulted in 2458 publications up to August 2013 that used the word ‘phishing’ in the title, abstract or keywords. We filtered the publications based on our exclusion criteria: studies had to be written in English to be included in our selection, so that we could run a syntactical analysis on them, and had to be peer-reviewed.

After filtering, the literature set was narrowed down to 312 journal articles and 1774 conference papers. Since it was not feasible to read all publications, we created a subset of the literature to be reviewed manually. Journals generally have less strict review deadlines than conferences,

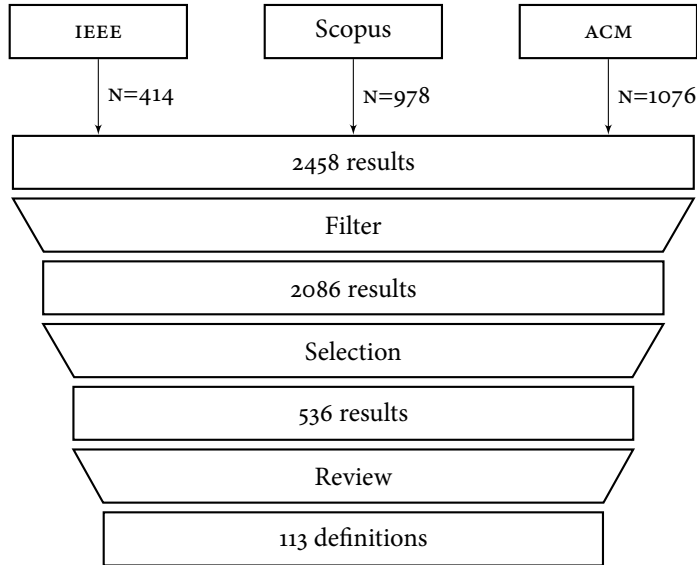


Figure 3: Search, selection and review of the results.

resulting in longer reviews and possibly higher quality. In addition, generally journals have higher limits on the number of pages, resulting in more in-depth articles. Therefore, we included all 312 journal articles in the review. Turning to the 1774 conference papers, we note that in the field of computer science, publishing in conference proceedings is generally favoured (Freyne et al., 2010), whereas journals are preferred in other fields. For the conference papers, we used the number of citations as an indication of quality and based our selection on this criterion. This resulted in the inclusion of 135 conference papers with more than 10 citations each. However, the selection based on citation count may exclude high quality conference publications that have recently been published and thereby have not yet received many citations. Therefore, we included all 69 recent conference papers from 2013 (from January to August) and the 20 newest from 2012.

All 536 eligible publications were manually searched for definitions of phishing by performing a case-insensitive search for the word ‘phish’, so that abbreviations within the paper would also be covered. If a definition was present, it was extracted for further analysis. Studies were excluded if they: (1) did not include a definition, or at least a clear and concise description, of the word phishing; or (2) merely cited a definition of others. If an included paper cited the definition from another peer-reviewed publication (7 occurrences), the cited publication was included in our dataset. The approach involved considering not only explicit definitions but also descriptions of phishing in terms of concepts. Definitions had to be one or two sentences in length, but longer

definitions were included if they were clear and to the point. However, publications giving only a specific example, such as an anecdote, were not included.

Since the search was performed by a single researcher, the extraction of definitions was re-evaluated by a second researcher by randomly selecting 100 publications from the dataset. The second researcher then manually reviewed each publication to identify a definition. The two sets of results were compared and the inter-rater reliability (Cohen's Kappa) was found to be  $\kappa = 0.70$  ( $p < 0.001$ ) with a 95% confidence interval of (0.561, 0.839), indicating substantial agreement and supporting the feasibility of the method.

Careful analysis of the 118 extracted definitions resulted in the exclusion of five of them as non-cited duplicates. Among the duplicate definitions, we selected the definition that had been published the earliest and excluded the others. This reduced our dataset to 113 unique definitions, all of which can be found in the [Appendix A](#).

### 2.2.2 Identification of common words

We initially analysed the definitions in a purely syntactical way (i. e., without context) to obtain an overview of the most commonly used words. The analysis consisted of a simple frequency count of all words to establish which ones occur most often. Although a frequency count removes all contextual information from the individual words, it does give an indication of the relative importance of each word compared to all the others. In addition, words that appear throughout all definitions are probably important to phishing. All definitions were first processed by removing all punctuation, putting all words in singular form and merging different spellings. For example, 'credit-card' became 'credit-card', 'ID theft' became 'identity theft', and 'web page' became 'webpage'. Multiple occurrences of a single word were counted only once per definition to avoid biasing the frequency count. All adverbs were removed, since they give no additional information in a frequency count. Finally, the word *phishing* itself was removed from all definitions, as counting its occurrences would not give any insights. The resulting list of definitions contains normalised words (i. e., singular form, one spelling, no punctuation), which was analysed to get some basic understanding of the concept of phishing. The result of the frequency count was plotted in a 'word cloud' (McNaught and Lam, 2010) as included in [Figure 4](#). In a word cloud, the font size of the words represents the number of occurrences relative to other words, i. e., the word that is mentioned the most, is set in the largest font.



or with more detail about the methods (Forte, 2009). Essentially, these phases are all high-level crime scripts. Using the phases of phishing as a framework, we identified in what way the definitions structure a phishing attack. In each definition, we highlight the words that could relate to a particular phase of phishing, even when the authors do not identify the phases explicitly. For example, Herzberg (2009) defines phishing as '*Password theft via fake websites*', whereas Amin, Ryan and Dorp (2012) state that phishing is '*email soliciting personal information*'. Herzberg focuses on the way passwords are stolen, not on how potential targets are drawn to the websites. Amin, Ryan and Dorp, on the other hand, identify the method of attracting potential targets, but do not explicitly state to whom the personal information is sent, or how this is done. Furthermore, after having highlighted words from the theoretical framework and words relating to the phases of phishing, any remaining words (i. e., nouns, verbs or adjectives) used to define the process of phishing are highlighted as well.

The result of the identification of important words in the sample of 20 definitions is a list of nouns, verbs and adjectives. In several iterations, synonyms and words referring to the same concept are merged. For example, the words 'creditcard numbers', 'credentials' and 'sensitive data' refer to the concept 'information'. In each iteration, we tried to find which words were related in an attempt to merge them into one concept. This resulted in 18 concepts, categorised as 3 actors, 1 asset and 14 activities (see Table 2). All 93 remaining definitions were analysed using these 18 concepts to see whether they can be described as a subset of them. A second rater re-evaluated the extraction of concepts. Since the data are based on the output of the raters, Kappa is not the correct statistic to calculate the level of agreement (Feinstein and Cicchetti, 1990). In this case, the proportion of agreements (agreements divided by non-agreements) was used, which was 0.78. This substantial agreement supports the applicability of the method and indicates the clarity of the theoretical framework for the raters.

The results of the frequency count, as shown in the word cloud, together with the theoretical framework, were used to label the concepts with the most commonly used terminology.

#### 2.2.4 Analysis of concepts

All definitions were scored on the 18 identified concepts that were extracted. Together with the meta-data for each definition (i. e., year of publication, field and country of affiliation of first author), the results were entered into a data file. Frequency analysis was used to determine which concepts were the most important. This frequency analysis consists of establishing whether there is consensus within the set of definitions on whether to include or exclude a concept. For each concept,

we determined whether the definitions agree on either inclusion or exclusion by calculating whether the number of definitions that use the concept differs significantly ( $p < 0.05$ ) from 50% by using Pearson's chi-square test, the results of which can be found in [Table 2](#). This results in three categories: (1) concepts that are used in significantly fewer than 50% of the definitions; (2) concepts where there is no clear consensus; (3) concepts that are mentioned in significantly more than 50% of the definitions. Concepts where there is consensus are either included (category 1) or excluded (category 3). The remaining concepts from category 2, where there is no consensus, are considered in the discussion section.

Finally, we calculate the Pearson's correlation between the year of publication and each concept, to identify evolution of the definitions with respect to the emerging concepts.

### *Validity*

One of the threats to the validity of our study is that the review was conducted by a single researcher. However, subjective decisions are mitigated by following a systematic protocol and discussing this, and the results of the exercise, with senior researchers. Additionally, a second researcher replicated the method. Cases where the second rater disagreed with the initial rater were discussed, which led to the inclusion of six definitions that had previously not been included. For the extraction of concepts, differences were discussed, leading to no changes in the 18 included concepts.

By including peer-reviewed scientific literature only, we were able to search systematically for all publications on phishing in three digital libraries. Due to the goal of this research, i. e., finding out how phishing is defined in the research community, only scientific research was included. Our design suffers from a publication bias, since all included definitions are peer-reviewed. There may be very comprehensive definitions beyond the scientific domain. If this were to be the case, we assume that a large number of research papers would reference this definition.

Although our approach of selecting publications covers a large set of the available literature, there is the possibility of not including a relevant publication. However, we minimise this potential bias by selecting based on citation count (i. e., 10 or more), source (i. e., all journals) and including recent conference papers (i. e., from 2013 and the latest 20 from 2012). If a definition of high importance to the field has been established, it is likely to have been cited by many. In addition, if an included paper cites a definition from another publication, the cited publication is included in our dataset, thereby further decreasing the potential of missing of a key definition. Finally, due to the large number

of definitions, it is unlikely that the results would have been different by including a small number of additional definitions.

The extraction of concepts was based on a sample of the definitions, which could result in certain concepts not being included. We mitigated this by comparing all definitions against the identified concepts, to find out whether any definition had a different concept. Additionally, as mentioned before, another researcher reviewed a random sample of the publications. A consequence of a consensual definition is that it is based on concepts that are used in the majority of the source definitions. We did not conduct any quality assessment of the publications. The quality control was implicitly performed by including all journal articles and highly cited conference papers.

### 2.3 RESULTS

The total sample of selected publications consisted of roughly 22% ( $N=536$ ) of the available peer-reviewed literature. This subset of the literature covers highly cited publications, journal articles and recent publications. The selection covers, in our opinion, most of the important literature on phishing. After review, 113 distinct definitions were extracted from the peer-reviewed literature. The definitions were analysed at the level of words and concepts.

The word cloud (Figure 4) shows the results of the frequency analysis that was used to analyse the words. The five most-used words are *information*, *website*, *user*, *personal* and *email*. From the figure, we can identify the actors, assets and activities. Actors are *user*, *victim*, *attacker*, *bank* and *business*. The assets that were found are *information*, *website*, *email*, *password*, *creditcard*, *username* and *account*. Finally, activities such as an *attack*, *social engineering*, *identity theft* or *spoofing* are most often used.

Eighteen concepts were extracted from the definitions (Table 2). Two of these concepts are common to the routine activity approach: an offender and a target. There is a weak relationship between usage of the concept social engineering in the definition and the year of publication ( $r(105) = .23$ ,  $p = .015$ ). This indicates that recent publications refer to social engineering more often than older publications. The presence of other concepts and the year of publication were not related, giving no evidence of evolution of the definitions with regard to other concepts.

The concepts that are used most frequently in the definitions lead to the following phishing crime script. First, the offender sends a communication to the target, which 62 of the definitions state. Typically, the offender sends the target an email ( $N=30$ ) or sends a message using a method that is not specified ( $N=22$ ), occasionally using other methods such as websites (Olurin, Adams and Logrippo, 2012; Hodgson, 2005; Levy, 2004), social spaces (Piper, 2007), instant messages



(Verma, Shashidhar and Hossain, 2012; Ali and Rajamani, 2012), text messaging (Hinson, 2010) or even letters (Workman, 2008). Then, the target may reply by sending information to the offender, which is mentioned in 64 of the definitions, mostly through the use of a website (N=40). The information that is transmitted, according to 113 definitions, can be categorised as: (1) authentication credentials (N=13); (2) identity information (N=5); (3) sensitive information (N=23); or (4) personal information (N=24). Variations or combinations account for the remaining types of information.

Type	Extracted concept	N	$\chi^2$	p	
Asset	<b>Mentioning information*</b>	105	83.27	.00	} Consensus
Actor	<b>Mentions a target*</b>	87	44.61	.00	
Activity	Phishing is digital*	87	32.93	.00	
Activity	Phishing is internet-based*	84	26.77	.00	
Activity	<b>Using deception*</b>	79	17.92	.00	
Activity	Communication from target to offender	64	1.99	.16	} No consensus
Activity	Communication from offender to target	62	1.07	.30	
Activity	Phishing is a criminal activity	61	0.72	.40	
Activity	<b>Using impersonation</b>	60	0.43	.51	
Activity	Phishing uses websites	56	0.01	.93	
Activity	Phishing uses messages	51	1.07	.30	
Actor	Mentions a trusted third party	50	1.50	.22	
Activity	Phishing is fraud*	43	6.45	.01	} Consensus
Actor	Mentions an offender*	40	9.64	.00	
Activity	Using persuasion*	30	24.86	.00	
Activity	Mentions the later abuse of information*	22	42.13	.00	
Activity	Related to identity theft*	20	47.16	.00	
Activity	Related to social engineering*	19	49.78	.00	

$\chi^2$ -test with df=1. N=113. Boldfaced concepts are included in standard. \* p < 0.05

Table 2: Concepts used in the phishing definitions:  $\chi^2$ -tests are used to determine whether the frequency of use of a concept is significantly more or less than 50% of all definitions.

The results of the analysis of concepts are shown in Table 2. In the literature, there is a consensus that the concepts of deception (N=79), a target (N=87), information (N=105), being digital (N=87) and internet-based (N=84) should be mentioned in a definition. Furthermore, the concepts of fraud (N=43), an offender (N=40), persuasion (N=30), the abuse of information (N=22), identity theft (N=20) and social engineering (N=19) should not be included according to a significant majority of the definitions. There is no consensus for the remaining concepts.

Figure 5 shows the number of publications per year that define phishing, indicating several peaks in the number of definitions within particular years. Partly, this is due to the criteria used in the literature selection. For example, the peak in 2013 is due to the inclusion of all recent conference papers. However, that does not explain the decrease of definitions in 2008, and the increase thereafter. Such changes could indicate emerging consensus about the definition, so that authors start citing earlier definitions they consider useful, or, where there is a rise in the number of definitions, a change in the phenomenon might be developing, requiring redefinition.

The research field and affiliation of the first author show that mostly researchers located in the USA ( $N=53$ ) or in the field of Computer Science ( $N=88$ ) define phishing. Other countries in which the first author is located include the UK ( $N=9$ ), China ( $N=8$ ), India ( $N=7$ ), Canada ( $N=7$ ) and Australia ( $N=6$ ). There is a significant correlation between the year of publication and the first author being affiliated within the USA ( $r(105) = -.46, p < 0.001$ ), indicating that recent definitions originate more often from countries other than the USA. Almost no definitions originate from research fields other than Computer Science, with Psychology ( $N=4$ ) or Law ( $N=3$ ) as largest contributors. For 14 authors, it was not possible to establish the research field (for example, when the first author is a journalist). A possible reason for the large number of computer scientists who produce their own definition of phishing, is that they feel more inclined or capable to define phishing, whereas researchers from other fields would rather use another author's definition, or none at all.

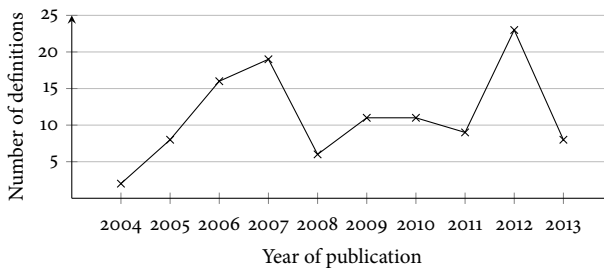


Figure 5: Number of publications with a definition of phishing, till August 2013 ( $N=113$ ).

## 2.4 DISCUSSION

The present study identified concepts of phishing according to the peer-reviewed literature. There is a consensus on most concepts, with seven concepts present in approximately half of the definitions. We discuss each of these concepts and consider whether they should be included

in the definition. However, we first observe that the concept ‘internet-based’ is a subset of the concept ‘digital’ and therefore, one is redundant. As internet-based is the most precise concept, arguably it should be included in the definition. This, however, leads to the discrepancy that instant messaging through an internet-based application on a phone can be phishing, whereas a regular text message on a phone cannot (not internet-based), even though both methods are essentially the same. In our view, phishing was made possible due to the ability to mass-distribute messages. Whereas the internet has served as a catalyst, in facilitating communication cost efficiently, it is not the only way to do so. We propose to replace the concepts of internet-based and digital with *scalability*. Being scalable refers to the ease of scaling from a single occurrence to hundreds, thousands or millions. Whereas digital specifies the encoding used for the channel (in bits, ‘0’ or ‘1’) and internet-based is a specific channel, scalability only requires the channel to support mass-distribution.

We decided to exclude the concept of ‘mentioning a trusted third party’ (included in 50 of the definitions) in favour of impersonation ( $N=60$ ), since deception through impersonation by abusing the target’s trust implies the existence of a trusted third party. The communication between a target and an offender is mentioned in slightly over half of the definitions ( $N=62$  and  $N=64$ ). However, we decided to exclude the explicit mentioning of communication, as this follows from the exchange of information from a target to an offender. Using websites ( $N=56$ ) or messages ( $N=51$ ) as specific channels for phishing were not included since these are absent from a significant majority of the definitions. Phishing as a criminal activity is not included in the list of essential concepts, even though 61 of the definitions mention this, as it is included in deception and furthermore depends on legislation in a particular jurisdiction.

Consequently, the concepts of deception, impersonation, target, information and scalability are the most important aspects of a phishing definition. Therefore, we propose a definition of phishing that comes out of the synthesis of literature and includes all the important concepts that existing definitions have in common:

*Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target.*

A first observation is that our definition provides a high level of abstraction, compared to most alternatives. This derives from the method used. The consequence of this is that there are no details about specific methods (such as email or websites) required to perform a phishing attack. By comparing our definition to those in [Table 1](#), it can be seen that our definition is sufficiently abstract to be compatible with the dictionary and APWG definitions. The Oxford, Collins and Merriam-Webster definitions can be mapped entirely onto our definition, as they are more

specific. For example, our definition does not include the offender's misuse of the obtained information, such as identity theft. The APWG definition is compatible as well, although it is much more specific to what is considered phishing. For example, the APWG definition specifically mentions 'technical subterfuge' schemes that tamper with a target's PC, such as installing a virus, whereas our definition – being broader – states that deception and impersonation are used. Whether or not this is followed by, or consists of, technical subterfuge, is not mentioned. Therefore, we consider the APWG definition to be compatible to ours. Finally, the American Heritage definition is the only one that is not completely compatible, since it mentions the use of a telephone, which does not scale well.

The methods employed in phishing could be used long before the internet became popular. However, the term phishing only arose around 1995-1996 (Oxford University Press, 2012; Khonji, Iraqi and Jones, 2013; Purkait, 2012; James, 2005), indicating that mass-communication is one of the foundations of phishing. Another factor contributing to the success of phishing on the internet is that it is cost-effective for mass-communication (i. e., spreading millions of messages). Although both are potential forms of mass-communication, letters and telegraph messages are more costly to employ on a large scale, whereas sending emails over the internet is cheaper. This contributed to the success of the internet as a channel for phishing. Other channels, such as telegraph messages or text messages, can be scalable, apart from the potentially high costs of sending millions of messages.

Only one indication of the evolution of phishing definitions was found: the tendency to refer to Social Engineering in papers that are more recent. However, there could still have been evolution within the literature on the act of phishing. For example, authors may have identified specific methods of phishing throughout the years, which in our analysis were mapped onto the same concept. Additionally, recent publications that define phishing more often have a first author with an affiliation not in the USA, whereas early definitions originate mainly from the USA. This could indicate that authors from outside the USA feel the need to redefine phishing because of local differences, or indicate more international interest in phishing. However, this could also be a result of the inclusion criteria (i. e., publication in English), or more interest or funding in the United States for phishing research.

## CONCLUSIONS

The goal of this chapter was to identify a consensual definition of phishing from the literature. In the literature search, 113 different definitions were found, indicating that many researchers have thought about a definition of phishing. We identified the core concepts which the re-

search community agrees are part of phishing, resulting in a consensual definition: *'Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target.'*

The principles of phishing were used by offenders long before the advent of the computer and the internet. Before computers became a consumer product, these principles were considered a type of fraud. Digitalisation and mass-communication through networks provide new channels to exploit the same human vulnerabilities on a larger scale. The internet opened many opportunities for new types of fraudulent behaviour, such as phishing. Phishing on particular channels is sometimes named differently, such as smsishing (channel is sms). We consider these types of phishing if they fit the consensual definition that we developed.

The implications for other definitions are mainly caused by the concepts scalable, deception and impersonation. Phishing must use deception by impersonation in order to be called phishing. When no impersonation is used, for example just asking for information, the act cannot be called phishing. Furthermore, it should be easy to scale, implying that one-to-one communication, such as a phone call, is not phishing. Spear phishing, which is phishing with a single target, is possible, as long as the employed method supports scalability.

The main theoretical contribution of this chapter is threefold. Firstly, we validated the findings of Hutchings and Hayes (2009), Reyns, Henson and Fisher (2011) and Pratt, Holtfreter and Reisig (2010) that the routine activity approach, developed for explaining crime in the physical world, can be applied to the digital world. Within the context of phishing, routine activities include, for example, giving one's email address away, time spent on the internet, time spent on email. Such routine activities could lead to more opportunity for victimisation. Additionally, we suggest the notion of crime facilitation to be relevant to cybercrime, and specifically phishing. People can deliberately, negligently or unconsciously facilitate their own victimisation by placing themselves at special risk (Sparks, 1982). The second theoretical contribution of this research is the development of a consensual definition of phishing. Yar (2012) states that networked communications act as a force-multiplier and that the impact is further increased by a space-time compression, whereby actions can occur almost instantly in different locations. Therefore, he argues that new theoretical notions are required for theorising about cybercrime. We believe these notions are manifested in the concept 'scalability' of the consensual definition and therefore constitute the third theoretical contribution.

This research adds a consensual definition of phishing to the body of existing definitions so that others can be weighed against the concepts with consensus within the research community. Research can be aligned by using a common definition, thereby avoiding misinterpretations. Researchers who define phishing differently can relate their

definition to the consensual one, thus positioning better which actions they consider phishing. Furthermore, meta-studies on phishing are better facilitated with our definition. Institutions, such as the police or banks, benefit from a consensual definition as well. Collaboration and data sharing between different organisations is easier if both have a common vocabulary. Organisations labelling phishing incidents according to a consensual definition will find it easier to compare the effectiveness of countermeasures.

Future research could focus on translating and interpreting the consensual definition into other languages. The consensual definition can be related to the definitions that practitioners use, thereby extending this study into the non-scientific domain. Furthermore, a discussion in the research community should establish more clarity on the concepts where there is no consensus at this moment. We believe that the lessons learned in crime science and the theories and tools that crime scientists developed, should be applied to phishing. In particular, we suggest studying the notion of crime facilitation in cybercrime, in addition to crime opportunity. Ultimately, a collaboration of crime science and computer science could help in reducing phishing victimisation and avoid reinventing the wheel.

Having established what phishing is, we now turn our attention to the scalability and effectiveness of phishing in the next chapters.

Obtaining information can be the goal of a cyber attack, or just part of the reconnaissance. When an attack has a specific target, it is important to achieve a high effectiveness. In such cases, scalability is less important than effectiveness. Scalable methods, on the other hand, are distributed on a large scale in order to raise the expected benefits. The effectiveness on a single target can be low, as long as the expected benefits on a large scale are good. Less scalable methods, such as an attack on a specific target, cost more when applied to a single target (or to few targets). However, the expected benefit for such attacks is much higher. In a less scalable attack, the offender is fishing for information, rather than phishing. In this chapter, we discuss attacks that are less scalable, and could be considered ‘fishing’ for information. To illustrate a less scalable attack, we consider the documented attack on Mat Honan as an example (Honan, 2012).

In the Mat Honan attack, two offenders gathered some basic information from him using his personal website and other publicly accessible sources. Then, the offenders called the support of online retailer Amazon to add a (fake) creditcard to the account of Mat Honan, something that can be performed without much validation. In the next step, one of the offenders called Amazon’s support again, this time asking for a password reset. When asked for the last four digits of a creditcard of the account for authentication reasons, the offender provided the digits of the creditcard that he added himself in the previous call. Subsequently, the Amazon support employee reset the password of the account of Mat Honan and informed the caller what the new password was. Now, the offender could login to the web interface of Amazon and see the last four digits of all creditcards: the fake one he added himself, as well as the creditcard of Mat Honan. As the next step in this attack, the offender went on to call technology company Apple, asking for a password reset for his *Apple ID* account. To verify that the caller was really Mat Honan, he had to provide the last four digits of his creditcard. Since the offender obtained the last four digits of the real creditcard of Mat Honan from the Amazon account page, the validation was trivially passed. With help from Apple Support, the offenders got access to the Apple ID of Mat Honan and issued a remote wipe of all his devices.

---

<sup>2</sup> This chapter is an extended version of the published paper “Applying the Lost-Letter Technique to Assess IT Risk Behaviour” (Lastdrager, Montoya et al., 2013), which was published in the Proceedings of the 3rd Workshop on Socio-Technical Aspects in Security and Trust, New Orleans, USA. IEEE Computer Society. Part of this chapter is based upon joint work with Henry Been, Jurgen Kleverwal, Matthijs Gielen and Wouter de Vries.

Such an attack is an example of *social engineering*, which is fishing for information, rather than *phishing*. Social engineering attacks using personal contact are very effective. For example, in an experiment of Bullee, Montoya Morales et al. (2015), employees of a university were asked to give away their office keys to an unknown person, something 62% of the employees did. These kind of attacks are effective, but are less scalable than phishing attacks.

In Figure 1 (from Chapter 1) a relation between the scalability of the modus operandi and the effectiveness was hypothesised. We start by elaborating on this relation with a discussion on the literature of the effectiveness of several modus operandi. The effectiveness of phishing is an often discussed topic. Many methods of quantifying success of phishing can be used. For example, one can measure the monetary loss of the victims, the income of the offender, the number of victims that click on a link, or the number of victims that provide their information. There is no single best measure for phishing, due to the changing modus operandi and ways to monetise the information (Moore and Clayton, 2010). The monetary loss resulting from surveys is often exaggerated due to outliers, especially when the results are extrapolated to the entire population (Florêncio and Herley, 2011). When there is no or little monetary loss (e.g., what is the cost of losing personal information?), the profit gained from a phishing attack is smaller than the damage caused by it (Herley and Florêncio, 2008). The income of the offenders is another measure that could be used to determine their business model, and their return on investment. However, this is both difficult to measure, and, considering the phishing ecosystem, the profitability for an offender is sometimes questioned (Herley and Florêncio, 2008). Phishing effectiveness in terms of monetary gains or losses by itself is therefore not a reliable measure. Other means of measuring the effectiveness are needed.

Instead of measuring the expected benefit to the offender, or the cost for the victim, one could measure the success rate of a phishing message. Consider a typical phishing attack, where victims get 'hooked' by a phishing message (i.e., an email), click on a link and fill in their information on the phishing website. The effectiveness of a phishing message in practice is hard to estimate. However, experiments can establish the success rate of a phishing message. For example, Jagatic et al. (2007) have shown that using personal information in a phishing email, such as sending it from a target's friends, works much better than non-personalised emails. A normal phishing email had a response rate of 16% and a phishing email that pretended to be from a friend lead to a response rate of 72% (Jagatic et al., 2007). In comparison, Kumaraguru, Cranshaw et al. (2009) found that between 41% and 52% of their untrained subjects clicked on a link, and 25-41% provided information to a phishing website.



In all experiments, however, the content of the email is important (Kumaraguru, Sheng et al., 2008). Distributing emails with relevant content, pretending to be from a company that is relevant to the receiver, is much harder on a large scale. Experimental results are likely to overestimate the victimisation rates. At the same time, surveys show that victimisation rates are just a few percent (0.07% up to 2%) of internet users are victimised yearly (see Herley and Florêncio (2008) for an overview and a discussion on the unreliability of these statistics). However, since many emails are caught by technical means of blocking unwanted email, such as spamfilters or blacklists (Purkait, 2012), it is difficult to determine the actual success rate of sending out phishing emails.

Another way of measuring phishing is by looking at the success rates of the phishing websites that users get directed to, after clicking on a link in a phishing email. One way of measuring the success rate of a phishing website is to analyse the log files of the webserver that hosts it (Moore and Clayton, 2007). In the study of Moore and Clayton (2007), the log files of 1695 phishing websites were analysed. A typical phishing website in their sample gets information of about 18 persons per day for the first 24 hours. After that, about 8 persons per day continue providing information to the phishing website. Together with the average time a phishing website is online (61 hours), this suggests that a single phishing website gets about 30 victims (Moore and Clayton, 2007). Victims who fall for a phishing email, click on the link and subsequently go to the phishing website, do not always fill in their information. Bursztein et al. (2014) show that on average 13.7% of the visitors of such as phishing website fill in some kind of information. The worst performing phishing websites still obtained information from 3% of the visitors, whereas the best sites misled 45%. Based on statistics of phishing websites, Herley and Florêncio (2008) estimate that 0.37% of the internet users in the USA provide their credentials to a phishing website annually.

For the other *modus operandi* listed in Figure 1, namely QR code phishing and using USB keys as an attack vector, empirical data on their effectiveness is only sparsely available. Experimentation can provide such data. In this chapter, we describe two experiments that measure effectiveness from a different point of view. Both experiments were performed in the real world (i. e., not in a lab). Firstly, Section 3.1 describes an experiment where USB keys were dropped within a university building. Each dropped USB key was observed, and it was measured how many people would return the picked up USB key to a nearby service desk. Individuals who pick up a USB key and subsequently use it, put their devices at risk of a virus infection. From an attacker point of view, dropped USB keys can be used to obtain access to an organisation by infecting a computer within the network with malware. In the second experiment, QR codes are used to perform a phishing attack on a large organisation, which is described in Section 3.2. QR codes pointing to

phishing websites were included on posters and distributed within the organisation. Finally, [Section 3.3](#) gives concluding remarks.

### 3.1 USB KEYS

Cyber security is an important topic on institutional and personal agendas. To reduce the impact of information security breaches, cost-effective ways to protect against attackers must be first identified. Some risks might be mitigated by implementing information security policies. To test the compliance with such policies, data is required. Within social sciences, many data collection tools which can be adapted to information security are available. Methods to collect data include surveys, interviews, observational research and examining existing materials. Although surveys and expert interviews are often used for obtaining data about information security, there is always the question of the validity of the results. During an interview or in a questionnaire, a person may state to follow the information security policy, but in practice fail to follow it. Therefore, we explore the feasibility of using observational research methods as a tool for collecting data, since in general this will yield more reliable data.

One of the methods of observational research is the lost-letter technique (Merritt and Fowler, 1948; Milgram, Mann and Harter, 1965). It consists of dropping stamped letters in the streets, thus pretending that the letter was lost before it could be posted. Members of the public who see such a letter have the choice of posting the letter, keeping it or not picking it up. The researchers measure the number of letters that are received at the destination address. By varying the addressee's characteristics, one can measure the people's attitude towards certain topics. For example, by addressing letters to different political parties and measuring the return rates of the letters, one can establish popularity of the parties (Shotland, Berger and Forsythe, 1970). It is assumed that supporters of a particular political party will feel more inclined to post the found letter than non-supporters, even if they are aware that they are participants of a lost-letter experiment (Fessler, 2009). In a similar way, the public opinion on various other subjects, such as gay marriage or racism, was measured by changing the addressee (Theodore Montanye, Ronald and Kenneth, 1971; Ahmed, 2010; Forbes, TeVault and Gromoll, 1971; Waugh, Edmund and Rienzi, 2000; Bridges et al., 2002). In other studies, (fake) money was put in the envelope (Simon and Gillen, 1971; Farrington and Knight, 1979, 1980; Gabor and Barker, 1989), the importance of the letter was indicated on the envelope (Simon, 1971; Deaux, 1974) and the influence of the neighbourhood on the return-rate (Holland, Silva and Mace, 2012) was measured. Whereas in the standard lost-letter experiment only the influence of the victim's characteristics on the return rate is measured, researchers

may decide to observe the dropped letter and note the characteristics of the person picking the letter up (Farrington and Knight, 1979, 1980; Gabor and Barker, 1989). The lost-letter technique has been shown to be adaptable to modern techniques, such as the lost-(car)key technique (Forbes, TeVault and Gromoll, 1972), the lost-email technique (Stern and Faber, 1997; Vaes, Paladino and Leyens, 2002; Bushman and Bonacci, 2004; Tykocinski and Bareket-Bojmel, 2009) and the lost-smartphone technique (Symantec, 2012).

We propose the lost USB key technique to measure the attack effectiveness of ‘fishing’. USB keys are important for information security modelling as they may cause issues such as data leaks (The Guardian, 2012; The Canadian Press, 2012) or the infection of a computer network with malware (for example Stuxnet (Falliere, Murchu and Chien, 2011) or malware that is located on USB keys found in public transport (Ducklin, 2011)). Some of these issues can be mitigated by technical means, such as data leaks which can be prevented by requiring users to use encryption. However, as technical solutions do not mitigate all threats, other means are needed to reduce certain risks. People who find and use a lost USB key put their computer at risk of a virus infection (Tetmeyer and Saiedian, 2010) and therefore form a threat to networks. For example, malware infections through USB keys may be prevented by forbidding persons to use untrusted USB keys. These solutions are often implemented as policies within organisations and require compliance of the users. For example, Carnegie Mellon University has a clear policy (Carnegie Mellon University, 2013) on found USB keys: *“Avoid plugging an unknown USB into your computer or a cluster computer. When a USB drive is found unattended, please give it to a cluster consultant, the Computer Services Help Center, a residence assistant (RA) or to Carnegie Mellon campus police.”* The lost USB key technique allows organisations to quantify the user’s compliance with an information security policy. The resulting data may be used as input for modelling users’ behaviour or testing the effectiveness of interventions. Furthermore, the results can influence changes in information security policies, such as disabling USB ports to prevent people to infect a system with malware.

The lost-letter technique and its variations are used to measure altruism (Merritt and Fowler, 1948), but whether or not a person steals a USB key is also influenced by factors other than personality, such as the context. Theories of crime opportunity (Felson and Clarke, 1998) can be used to explain the context of the lost USB key pick up. The Routine Activity Approach (Felson and Clarke, 1998; Felson and Boba, 2010) states that a crime is likely to occur if a likely offender meets a suitable target in absence of a capable guardian. The Routine Activity Approach lists three types of people who can prevent a crime from occurring. First, a handler might convince the offender not to commit a crime. Such a handler may accompany the person picking up the USB key and convince him/her not to steal it but to return the USB key as lost and

found instead. The second type is the aforementioned guardian, who watches the target. A guardian could be the owner of the target or a person close by who watches the situation. The third type is a place manager who is responsible for the setting. An example of a place manager is a receptionist or security guard. Applying this to the lost USB key technique, implies that a subject (i. e., person who picks up the USB key) will converge in space and time with a target (i. e., USB key) in absence of a guardian or place manager and without a handler to hold the subject back. In the lost USB key technique, the target is a USB key that the victim is the alleged owner of.

To investigate whether theft of a lost USB key is related to the victim, subject and situational characteristics, an experiment was performed in a university setting, by dropping USB keys near service desks. We used the methodology of Farrington and Knight (Farrington and Knight, 1979, 1980), who look at the effects of the victim's characteristics, and adapted it to use USB keys instead of letters. This allows comparison of our results to their lost-letter experiments. Farrington and Knight used two groups: a control group consisting of unsealed letters containing no money and an experimental group with unsealed letters containing money. The control group in our experiment consisted of USB keys in their original box and the experimental group consisted of USB keys that were labelled to indicate usage. We hypothesise that USB keys from the control group get stolen more, as they do not contain data and have no risk of a virus, therefore the victim does not lose any data. Alternatively, the resell value might drive theft of brand new USB keys. The ownership of a brand new USB key is not clear, making it a relatively easy target. The USB keys from the experimental group are labelled to indicate the sex of the alleged victim and the importance of the contents. We hypothesise that the victim's sex does not make a significant difference, similar to the observations from Farrington and Knight. We expect USB keys with important content to be returned more (Deaux, 1974). For the subject characteristics, we hypothesise that subjects who are alone, casually dressed, young or put the USB key in their pocket will be likely to steal the USB key and that males are more likely to steal than females (Farrington and Knight, 1979, 1980). Apart from the variables from Farrington and Knight, we note whether the subject was walking in the direction of a service desk prior to picking the USB key up. We hypothesise that subjects who are walking in the direction of a service desk, will be more likely to return the USB key.

In this section, we explore the feasibility of the lost-letter technique to assess risky behaviour in relation to IT security. The contribution is the identification of situational and personal characteristics of the subject and victim that contribute to the theft of a lost USB key. Theft and consequent use of a USB key represent a security threat that organisations are in need of quantifying. Observational research provides a method of objective measurements.

### 3.1.1 *Method*

A field experiment was conducted by using an adapted version of the lost-letter technique that uses USB keys instead of letters. The design was based on the experiments from Farrington and Knight (Farrington and Knight, 1979, 1980), who dropped letters in the streets and observed by whom they were picked up. Teams of researchers dropped USB keys and observed whether they were picked up and, if applicable, by whom.

#### 3.1.1.1 *Design & Concepts*

In the experiment, the concepts of victim and subject are used. The victim is the alleged owner of the USB key and the subject is the person who picks up the USB key. The target is the USB key itself.

The experiment used a  $2 \times 2$  between-subject design. The independent variables were the sex of the victim and the importance of the data on the USB key. The dependent (outcome) variable shows whether or not the USB keys were returned to the service desk. By varying the independent variables, we aim to establish whether the subject's behaviour is influenced by the target's characteristics. In the lost-letter experiment, the recipient's address is listed on the envelope. In the case of a lost USB key, it may not be entirely clear where to return the device. In the lost-key technique (using car keys) by Forbes et al (Forbes, TeVault and Gromoll, 1972), this was solved by attaching a label with name and address information. Similarly, a datafile containing the owner's information could be put on a USB key. In our experiment, we considered USB keys to be stolen if they were not returned to the service desk. The USB keys had labels on both sides to show characteristics of the victim and contents of the USB key. The label on one side showed a male (John) or female (Anna) first name and a surname, whilst the other side showed its importance by labelling its contents to be either academic (thesis, i. e., important) or recreational (music, i. e., not important). Besides the experimental USB keys, a control group consisting of USB keys in their unopened box was used. The person finding a USB key from the control group could directly see that these did not contain any data. Figure 6 shows several of the USB keys that were used in the experiment.

In order to make a comparison of the data, we measured the same variables as Farrington and Knight (Farrington and Knight, 1979, 1980). Additionally, we added the walking direction of the subject relative to the service desk as a variable. A subject can walk to a service desk, away from it or neither (e. g., in parallel). In relation to the continuous data and the comparison with Farrington and Knight, the estimated age was measured as a continuous variable and later categorised, so that our study could be compared to both studies of Farrington and Knight. Farrington and Knight's 1979 study uses a different categorisation com-



(a) Label with name



(b) Label indicating contents (PhD thesis)



(c) Brand new USB key

Figure 6: Examples of the USB keys

pared to their 1980 study (ages 0-30 and above 30 versus 0-20, 21-50 and above 50), but neither justifies why these specific numbers were used. The number of companions was analysed as continuous data and later categorised as alone versus accompanied, to allow comparison with Farrington and Knight. The concept behaviour refers to the actions of the subject directly after picking up the USB key (e. g., whether the subject puts the USB key in his/her pocket or handbag). Clothing was categorised as casual (i. e., jeans and t-shirt), average (i. e., trousers and shirt) and smart (i. e., suit), similar to Farrington and Knight. The measured extraneous and independent variables are listed in [Table 3](#).

### 3.1.1.2 Setting

The USB keys were dropped in nine buildings at three Dutch universities. Each selected building has a lobby containing a service desk with a receptionist, as shown in [Figure 7](#). The USB keys were dropped in or near the lobby area, but not within sight of the receptionist. This was done to prevent people from feeling observed and wanting to please the receptionist by returning the USB key, or from thinking that the receptionist would pick the USB key up and deal with it. In all buildings that were used, the service desk was commonly known to be the first

Characteristic	Explanation	Categories
Time	Time of drop off	Time (i. e., 10:14)
TimeElapsed	Minutes elapsed	0,1,2,...
Type	Experimental group	Control (0), Experimental (1)
Sex of victim	Sex of the victim (label)	Female (0), Male (1)
Contents	Importance label	Recreational (0), Academic (1)
Clothing	Clothing of subject	Casual (0), Average (1), Smart (2)
Age of subject	Estimated age	0,1,2,...
Sex of subject	Sex of the subject	Female (0), Male (1)
Companions	Number of companions	0,1,2,...
Behaviour	Placed in pocket/handbag	No (0), Yes (1)
WalkingDirection	Relative to service desk	Towards (0), Away (1), Other (2)

Table 3: The independent and extraneous variables

point of contact for lost and found items. At the time of the experiment, neither university had a policy about found USB keys.



Figure 7: Example of a service desk at one of the universities.

At each location, USB keys were dropped on three ordinary Wednesdays in September and October 2012, i. e., during term time. In all buildings, the experiment was conducted during three time slots (10am–11am, 1pm–2pm, 3pm–4pm). These time slots were used in an attempt



to reduce the risk of somebody participating twice in the experiment, since finding a similar USB key twice could make people suspicious.

Unused 4 GB USB keys with a retail price of 5 euro were used for this experiment. The USB keys contained no (executable) data. In prior research, Merritt and Fowler (1948) used a fake coin and Simon and Gillen (1971) and Simon (1971) used play-money, which accounts to no economic value, but can, momentarily, lead the subject to believe that the letter contains something of economic value. Farrington and Knight (1979) used real money with values of between 0.20 and 5 GBP.

#### 3.1.1.3 *Subjects*

Subjects were self-selected from the population of people walking through the lobby of one of the buildings. Typically, these include either students or employees of the university, but also contractors (e. g., cleaning staff or construction workers) and visitors. The population of potential subjects of each university is not representative for the population at large. For example, kids or elderly are unlikely to be walking around at the locations of the experiment. In total 106 people picked up a USB key and therefore became subjects in the experiment.

#### 3.1.1.4 *Procedure*

Twenty-seven groups of two or three researchers participated in the experiment. Before starting the experiment, we obtained permission from the faculty's ethical committee (see section 3.1.3.5) and from facility management, which runs the service desks and employs the receptionists. Six weeks before running the experiment, all receptionists were informed about the experiment and who to contact in case of questions. In the morning of the experiments, all receptionists were contacted by phone to make sure that they were aware about the experiment and to ask if they had any questions about the procedure. The receptionists were asked to behave as if they were unaware of the experiment and asked to store the returned USB keys separately from other found items. We considered this procedure essential for running the experiment correctly and for avoiding problems for the receptionist.

The researchers were instructed never to interact with the subjects. They were randomly assigned a location, time and selection of USB keys. Five minutes before the start of the experiment, the students introduced themselves to the receptionist. They would find a suitable location close to the service desk, but not in sight of the receptionist (see Figure 8). One researcher would walk around and pretend to tie his/her shoelaces, look around to see if anybody noticed him/her and drop the USB key before walking away, similar to the procedure used by Farrington and Knight (1980). Another researcher would observe the USB key from a distance of about 20 meters. The researchers pretended



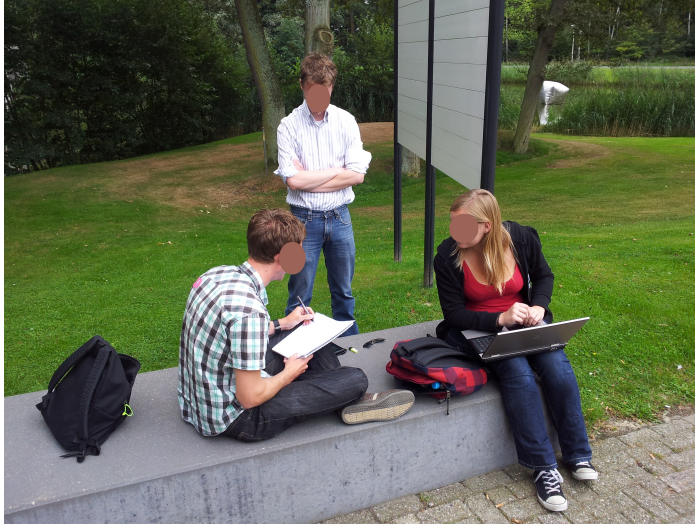


Figure 8: Three researchers in action. To avoid detection, they went outside and placed the USB keys within entrance of the building.

to be working, reading papers or playing with their phones. If somebody picked the USB key up, a form was filled in, taking note of the subject's characteristics and behaviour and of the situation at that moment.

#### 3.1.1.5 Analysis

Fourteen subjects did not look at the labels of the labelled USB keys, or the observers were unsure, and were excluded from the results as they were not fully exposed to the experimental conditions. For similar reasons, Farrington and Knight (1979, 1980) excluded cases in their lost-letter experiment. Subjects that picked a USB key up from the control group (i. e., not labelled) were all included. The exclusion of 14 cases reduced our dataset to 92 cases.

Farrington and Knight presented descriptive statistics and a univariate analysis (i. e., each individual variable in relation to the dependent variable). For comparison, we carried out the same analysis, including the extra variables (WalkingDirection, TimeElapsed and Content) that are specific to the lost USB key experiment. Additionally, several multivariate logistic regression models were developed. We tested whether a multi-level logistic regression was needed to account for similar results within the buildings (i. e., intraclass correlation). We found no significant effect of the individual buildings and therefore for simplicity we present the results of a regular logistic regression. A logistic regression measures the amount of variance in the return rate explained by the predictor (i. e., independent and extraneous) variables. Four models

Type	Characteristic	Category	F&K 1979	F&K 1980	Lost USB key
Experiment	Experiment type	Control	11.1 (N=18)***	10.7 (N=28)**	41.2 (N=17) **
		Experimental	30.1 (N=73)	39.3 (N=112)	12.0 (N=75)
		Male	30.2 (N=43)	50.0 (N=56)	7.9 (N=38)
		Female	30.0 (N=30)	28.6 (N=56)	16.2 (N=37)
Victim	Sex of victim	Academic			14.0 (N=43)
		Recreational			9.4 (N=32)
		Casual	45.8 (N=24)	53.7 (N=54)**	19.6 (N=51)
		Average/smart	22.4 (N=49)	25.9 (N=58)	14.6 (N=41)
Subject	Estimated age	30 or less		48.5 (N=66)*	25.5 (N=51)*
		31 or more		26.1 (N=46)	7.1 (N=42)
	Estimated age	20 or less	63.6 (N=11)*		0.0 (N=10)
		21-50	29.3 (N=41)		20.5 (N=73)
	Sex of subject	51 or more	14.3 (N=21)		11.1 (N=9)
		Male	30.2 (N=43)	41.3 (N=63)	19.1 (N=68)
	Companions	Female	30.0 (N=30)	36.7 (N=49)	12.5 (N=24)
		Alone	34.1 (N=41)	39.4 (N=71)	15.7 (N=51)
	Behaviour	Accompanied	25.0 (N=32)	39.0 (N=41)	19.5 (N=41)
		Placed in pocket	54.2 (N=24)**		75.0 (N=12)***
Walk holding object		19.1 (N=47)		8.9 (N=79)	
Unknown				0.0 (N=1)	
Walking direction	Towards servicedesk	Towards servicedesk			10.4 (N=48)
		Away from servicedesk			25.7 (N=35)
		Other direction			28.6 (N=7)
		Unknown			0.0 (N=2)

Note. N=92. Farrington and Knight (F&K) values from Farrington and Knight (1979, 1980). Significance ( $\chi^2$ ): \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

Table 4: Non-return rates (percent) per characteristic, compared to two lost-letter experiments

were developed: (1) a model based on victim characteristics, (2) a model based on subject characteristics, (3) a combined model based on all characteristics and (4) a compact model, that only uses the best predictor variables. The compact model was developed by narrowing the full model down using the Akaike information criterion (AIC) (Akaike, 1973). The models are reported showing odds ratios between the predictors and the return rate. For example, a predictor in our model with an odds ratio of 2.5 implies that the subject is 2.5 times more likely to steal the USB key if that condition is present. Analysis of the results showed that 'behaviour' is a very good predictor. However, this minimised the odds ratios for the other variables. For clarification, we included two additional models (Suspect model II and Combined model II) that exclude the predictor 'behaviour'. Both models explain less variance, but show more detail for the individual predictors.

Besides the variable age, we included the age squared as predictor in the regression models to compensate for the nonlinearity of the variable, since it is often the case that a given effect increases with age until a certain point and then it decreases. An example of such nonlinearity is criminal activity and age (Felson and Boba, 2010), where criminal activity peaks after childhood and decreases again afterwards. The number of companions (i. e., the size of a group) is not linear either (Klüpfel, 2007), therefore it was squared before including it in the regression. Significance was calculated using Pearson's  $\chi^2$ .

### 3.1.2 Results

The results of a univariate analysis are listed in Table 4 together with the results from both studies of Farrington and Knight (1979, 1980). A significant difference between the control group and the experimental group was found: people return used USB keys more often than brand new USB keys. For the non-return rates of the experimental and control groups, our results are different from the results of Farrington and Knight, where the control group gets stolen significantly less. There is no relation between the time or location of dropping a USB key and the return of the device. The median time before a USB key is picked up is 5 minutes. After 2 minutes and 15 seconds, 25% of the USB keys is picked up and after 10 minutes and 45 seconds 75% is picked up. The fastest time to being picked up was after 10 seconds of being dropped. The maximum time before being picked up was 45 minutes. No relation was found between the elapsed time and the return rate.

#### 3.1.2.1 Victim Characteristics

We did not find any significant results for the victim characteristics, although in our experiment females were victimised more than males. In their 1979 study, Farrington and Knight observed no difference in victim

Characteristic (reference)	Victim model		Subject model		Combined model		Compact model		Subject model II		Combined model II	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Victim's sex (female)												
– Male	0.41*	0.09–1.82			1.40	0.15–12.97					0.57*	0.10–3.05
– Control group <sup>†</sup>	4.96*	1.01–24.30			3.03	0.30–30.41					9.45*	1.33–67.17
Label (recreational)												
– Academic	1.75	0.39–7.81			0.25	0.01–4.42	0.38	0.03–4.57			2.63	0.43–15.90
– Control group <sup>†</sup>							4.96	0.73–33.56				
Clothing (casual)												
– Average			1.00	0.17–5.74	1.19	0.20–7.04			0.55	0.14–2.16	0.63	0.14–2.84
– Smart			2.70	0.16–46.58	2.03	0.09–47.27			1.89	0.13–27.13	2.29	0.11–46.96
TimeElapsed			0.90	0.79–1.02	0.91	0.80–1.04			0.99	0.92–1.06	1.00	0.93–1.08
Age			1.04	0.57–1.89	1.09	0.55–2.14			1.26	0.76–2.09	1.33	0.77–2.32
Age <sup>2</sup>			0.99	0.99–1.00	0.998	0.998–1.008			.996	.998–1.003	0.99	0.99–1.00
Companions			0.51	0.15–1.80	0.48	0.11–2.04			0.78	0.30–2.05	0.87	0.27–2.75
Companions <sup>2</sup>			1.14	0.96–1.34	1.14	0.95–1.37	1.04	0.99–1.10	1.06	0.93–1.21	1.06	0.89–1.28
Behaviour			168.55**	7.66–3710.4	269.15**	7.59–9545.5	69.13***	6.76–706.52				
Subject's sex (female)			1.56	0.21–11.65	2.12	0.25–18.03			1.59	0.37–6.79	1.54	0.31–7.60
WalkingDirection												
(towards service desk)												
– Away			0.91	0.15–5.38	0.84	0.10–6.90			2.82	0.73–10.90	3.55	0.81–15.56
– Other			0.25	0.01–8.89	0.42	0.02–10.78			3.32	0.45–24.21	3.70	0.43–32.18
Constant	0.14**	0.40–0.50	0.16	0.00–3266.1	0.06	0.00–2723.8	0.06***	0.01–0.28	0.01	0.00–21.62	0.00	0.00–6.25
R <sup>2</sup> i.e., variance explained	.103		.418		.474		.412		.134		.240	
Model significance	0.03*		0.00***		0.00***		0.00***		0.33		0.09	

Note. N=92. OR=Odds Ratio. CI=Confidence Interval. <sup>†</sup>The control group was coded with value 2. Due to collinearity, the output of only one control group is included.

Significance ( $\chi^2$ ): \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

Table 5: Predictors of the theft of lost usb keys

sex. However, in the 1980 study of Farrington and Knight males were victimised more than females, although the result was non-significant. Contrary to our hypothesis, the return rate for USB keys labelled as having important contents were returned less than USB keys labelled as containing non-important contents, although the results were not significant.

The interactions sex of subject and sex of victim showed the non-significant result that males stole more from females (20.0%; N=30) than from males (7.7%; N=26), whereas females stole exclusively from males (8.3%; N=12), and never from other females (N=7). Similarly, the interaction of importance with the victim's sex is non-significant, although more USB keys with academic contents of females (21.1%; N=19) were stolen compared to keys with academic content of males (8.3%; N=24).

### 3.1.2.2 *Subject Characteristics*

For the subject's characteristics, two significant differences were found. First, the estimated age of the subject is significant when the categorization of the Farrington and Knight (1980) study is used. People who are younger than 30 years tend to steal more often (25.5%; N=51) than people who are older than 30 (7.1%; N=42). This is in agreement with the 1980 study of Farrington and Knight. The relation between age as a continuous variable and the dependent variable is not significant. The characteristic behaviour is correlated to the non-return of the USB keys. Subjects who put the USB key in their pocket or handbag, steal the device in 75% (N=12) of the cases. Subjects holding the USB key in their hand fail to return the device in only 8.9% (N=79) of the cases.

The other subject characteristics were non-significant. A subject who is alone tends to return the USB keys more often than subjects who are accompanied. This contradicts the results of Farrington and Knight. The results of the other subject characteristics were not significant, but comparable to the lost-letter studies. The characteristic clothing was less important than in the studies of Farrington and Knight; people dressed casually stole in 19.6% of the cases and people dressed average or smart stole in 14.6% of the cases. The sex of the subject was not significantly of influence on the return rate, although males stole slightly more (19.1%; N=68) than females (12.5%; N=24). This is in line with the studies of Farrington and Knight. Interestingly, significantly more men (N=68) than women (N=24) picked up the USB key ( $\chi^2(1) = 21.0$ ;  $p < 0.001$ ). A variable that we introduced in our experiment was the walking direction of the subject, which recorded whether the subject was walking towards the service desk, away from it or in a different direction. Even though the result is not significant, people walking in the direction of a service desk returned the USB key more than people walking in another direction. This is in line with our hypothesis.

### 3.1.2.3 *Models*

The results of the logistic regression models are listed in [Table 5](#). Six models are included. The model with only the victim's characteristics explains around 10.3% of the variance and the model with only the subject's characteristics explains around 41.8% of the variance. The maximum variance we can explain is 47.4%, when all 11 predictor variables are included. The compact model includes only the content, the number of companions squared and the behaviour (whether the key was put in a pocket or handbag) as best predictors and still explains a reasonable 41.2% of the variance. The two models excluding the predictor behaviour explained 13.4% for subject model II and 24.0% for the combined model II, indicating that behaviour is indeed very relevant to predict the return of a USB key.

### 3.1.3 *Discussion*

The current study examined the willingness to return lost USB keys in a university setting and the influence that characteristics of the victim, subject and situation have on the return rate. In case of a lost USB key, the return rate is an indication of risk behaviour, since using a found USB key puts the computer at risk of a virus infection. The results of our univariate analysis ([Table 4](#)) support our hypothesis that USB keys in their original box are stolen more often than USB keys that were used. Furthermore, we found support for the hypothesis that people aged 30 years or younger steal more compared to people who are older than 30. Finally, results show that placing the USB key in a pocket or handbag is a good predictor of theft, which was in line with our expectations. Consequently, the decision to steal is made at the moment of pick up, indicating the feasibility of researching situational and personal characteristics as predictors of risk. No evidence was found to support the other hypotheses.

USB keys from the control group (i. e., in an unopened box) were stolen significantly more than USB keys from the experimental group (i. e., used, with labels). This can be explained by the nature of our experimental set-up. It is likely that subjects estimated the economic value of a used USB key as much lower than the brand new one, therefore the perceived value might have been related to the resell value. The results suggest that subjects who pick a labelled USB key up either perceive its economic value as too low to steal, or have genuinely empathy for the victim, resulting in a higher return rate.

Results showed the elapsed time between dropping a USB key and a subject picking the device up to be low. The implication of this is that a person who loses a USB key containing important content in a public location like a lobby, has only minutes to recover his/her device. The observers indicated that most people who noticed the USB key, picked

it up. However, several observers reported people intentionally kicking away the USB key ( $N=7$ ) or stepping on a USB key and not noticing ( $N=1$ ).

#### 3.1.3.1 *Victim Characteristics*

No evidence was found to support the hypotheses about the characteristics of the victim. The nonsignificant results showed females to be victimised more than males, suggesting further research to establish whether the result is coincidental or contradicting prior lost-letter experiments. The results related to the indication of importance of the contents of the USB key did not yield significant effects, but USB keys labelled as important were stolen more often than the devices containing non-important contents. An interesting interaction is the sex of the victim versus sex of the subject. Results suggest that males steal more from females than from other males and females steal more from males than from other females. However, since these results were not significant, there is no evidence to support statements about a higher likelihood to steal from the opposite sex.

#### 3.1.3.2 *Subject Characteristics*

No evidence was found for using the type of clothing of the subject as predictor of theft. The non-return rates were hardly affected by the clothing, in contrast to results of Farrington and Knight. The estimated age, however, was found to be significant if categorised as 30 years or younger and 31 years or older. Subjects estimated to be 30 years or younger are more likely to keep the USB key than older subjects. This is in line with the results from Farrington and Knight (1980) and age of criminal behaviour in general (Felson and Boba, 2010). When crime is categorised according to the 1979 study of Farrington and Knight, no significant results were found, however, subjects with an estimated age of 20 or younger never stole a USB key in our experiment. The influence of subject's sex on the return rate was marginal, males stole more than females, but there is no significant difference. In prior research, subjects who were alone stole more often, compared to subjects who were accompanied; however, our data show that accompanied subjects steal more often, although these results are not significant. Evidence was found to support that placing the USB key in a pocket or handbag is a very good predictor of theft of the device, suggesting that the decision to steal is taken at the moment when the USB key is picked up. It was hypothesised that, given the opportunity, people would return the USB key. No evidence was found, although subjects walking towards a service desk returned the USB keys more often than subjects walking in another direction, but the result was not significant.

### 3.1.3.3 *Models*

A logistic regression was used to create six models. A model with only victim characteristics can explain about 10.3% of the variation, whilst a model with only subject characteristics explains about 41.8%, suggesting that subject characteristics are a more important predictor than victim characteristics. The complete model explains 47.4% of the variation in the non-return rate, while our compact model, consisting of three predicting variables, managed to explain 41.2% of the variation. The three best predictors were: the importance of the contents according to the label, the squared number of companions of the subject and whether the subject placed the USB key in his/her pocket or handbag after picking it up. Within our data sample, theft was best predicted based on the situation (accompanied or not, label on USB keys) and person behaviour (placing USB key in pocket).

### 3.1.3.4 *Limitations*

The lost USB key technique inherits several limitations from the lost-letter technique. Similarly to the lost-letter technique (Liggett, Blair and Kennison, 2010), a large sample size is needed to obtain significant results. In the current study, the sample size of 92 is too small to obtain significant results on more variables. Observing the lost items is particularly time consuming, although it can provide insights into the exact behaviour of the subject. Collecting data for this type of research has proven difficult, since the number of locations that are available is limited. Locations can only be used when a service desk, reception or other kind of place manager is active, so that people have the option of returning the USB key to that person.

As far as we could observe, none of the subjects realised that an experiment was being conducted. In the university setting where our experiment was performed, it is common for people to hang around or work in common areas, which is why the observers could remain undetected. At any point in time, there are always people waiting for acquaintances near the entrance and service desk of the buildings, which proved to be an excellent way of hiding the observers. However, one subject reported to the service desk that someone was playing a joke, as he had seen a similarly labelled USB key before. Even though he was aware that something was going on, he did not see the observers. In another situation, a bystander overheard the subject talking to the service desk employee. The bystander mentioned that he had seen such a USB key earlier and that he had inserted it on his/her computer and found that it contained no data. We do not know whether it had been the bystander's intention to find identity information to bring back the USB key. The bystander mentioned to the receptionist that it probably contained a virus and warned the subject about it. However,



this indicates the willingness of people to insert found USB keys in their own computer. This situation points out a limitation of the lost-letter method with a limited set of locations. In our setting, we tried to prevent the subject finding multiple USB keys by spreading the observations over three working days with 2 weeks between the experiments, and by randomly allocating time slots and buildings to groups of researchers.

Another consideration inherent to the use of the lost-letter technique is the self-selection of subjects. We did not take note of the characteristics of people passing by the USB key, so we are unable to make statements about the selection of subjects in relation to the population of potential subjects. Future experiments could consider measuring the number and characteristics of people passing by. This would, however, require more observers. Furthermore, it remains problematic how to reliably measure who sees the USB key, or letter, but decides to not pick it up.

Another issue regarding the validity is the way measurements are recorded. Researchers in groups of two or three record basic properties of the situation and characteristics of the subject. To minimise errors, the researchers were asked to take good care of this. Especially for age estimation this is problematic. Internal discussions within a single team should smooth the age estimation, but unfortunately, we have no measures of inter-rater reliability.

Our interpretation of the return rate is that only USB keys that were brought to a service desk, either immediately or at a later moment, count as being returned. For the control group, this is the only way of returning them. For USB keys from the experimental group, one can think of scenarios in which the subject would try to insert the USB key in his/her computer in an attempt to find identity information of the victim, other than the name on the label. Thus, our non-return rates consist of subjects who stole the USB keys, of subjects who initially took them, but later decided to search for the owner, and subjects who did not consider the service desk as a method of handing lost property in. Twice the USB key got returned to the service desk at a later moment. A construction worker picked up a USB key before going for lunch outside and returned it to the service desk when entering the building again. On a second occasion, a USB key got picked up by a subject when entering the building but initially passed by the service desk, only to return a few minutes later to return the USB key to the service desk. Two USB keys were relocated (i. e., the subject moved the device from one to another location) and for practical reasons, we counted those as not stolen.

The feasibility of the lost USB key methodology depends mostly on the possibility to return the device to somebody who is responsible for the area. Subjects should feel comfortable to return the USB key. If this is not the case, they may prefer to take it home or relocate it at a central location, which would render the method less useful for measuring

altruistic or risky behaviour. In our experiment, the service desk was the most logical – and nearest – location to return the USB key to.

Finally, since we didn't interview the subjects, we are unaware of their motivation for keeping the USB key. Knowing their motivation would be useful information, but it would reveal that an experiment is going on.

### 3.1.3.5 *Ethical Considerations*

As with all lost-letter experiments, there are some ethical considerations (Stern and Faber, 1997). Due to the nature of the lost-letter experiment, informed consent is not feasible, as this would invalidate the experiment. Another option would have been to inform subjects about the experiment afterwards and ask for permission retrospectively in a debriefing. However, this would endanger the rest of the observations, since subjects could tell others about the experiment. Once the rumour spread, people may have been drawn to the lobby to pick up a 'free USB key'. The observers would need to mention to the subject that they would like to interview him/her in connection with the USB key stolen. For these reasons, we decided to observe and not inform the subjects about the experiment. The implication of this is that a subject who stole a USB key kept the device. We did not consider the lack of debriefing or informed consent problematic, as there are no negative consequences for the subjects. However, one of the subjects inserted the labelled USB key in his computer (see [Section 3.1.3.4](#)) and, after observing it was empty, mentioned to the receptionist that it must contain a virus. This could be avoided by putting some files on the device, thereby pretending it is indeed in use.

In an early stage the use of 'call home' software was discussed as a measure of how many people would use the USB key. We considered this unethical in the environment for our experiment, since the buildings of the universities are open for anybody to enter. Students and employees may bring their own device to the university. If any kind of tracking software were to be on the USB keys, there would be negative consequences (i. e., stress) for the subjects if they became aware. However, for organisations that have buildings with proper access control and exclusively use company-owned hardware, the use of a simple tracking tool sending an anonymous 'USB key plugged in'-message may be feasible to collect aggregated information about compliance.

All observers (researchers), research assistants and lecture staff had to sign a nondisclosure agreement regarding the personal identifiable information. During the experiment, observers may recognise a subject, or note information that could be related to a specific person.

### 3.1.4 Implications

The lost USB key methodology provides a method for generating relevant data for IT and facility managers to either design or redesign cyber security policies and test compliance with these policies. Variations of the lost letter experiment may open the field of data generation by providing a method to quantify security issues.

## 3.2 PHISHING WITH QR CODES

Phishing messages can be distributed in many ways. The link, or Uniform Resource Locator (URL), to a phishing website is no exception to that. Within phishing emails, several methods of including URLs exist. For example, URLs can be included as plain sight, or moved to an attachment. However, the offender wants to increase the difficulty of the receiver to see through the deception, and hiding a URL is one of the techniques that can achieve this. Hiding a URL or masquerading its true destination can be performed in many ways, of which we give three examples: (1) link hiding; (2) URL shorteners; and (3) QR codes. Firstly, a URL can be hidden in an <a> tag. The link text may say “Click here to log in”, or <https://bank.example>. However, when the receiver clicks on the link, the link points elsewhere (<http://datathieffs.example>). A second method of masquerading the destination of a link is by using a URL shortening service can be used as an intermediate. In this case, the link points to a ‘short URL’ (such as [bit.ly/1M1wuS1](http://bit.ly/1M1wuS1)), which redirects traffic to the real phishing website. The third example of masquerading the destination of a link, is by representing the link in a way that is not-readable by humans. An example of such a technique is a QR code, such as the one in Figure 9a. By just looking at a QR code, the contents are unknown to the human observer. However, they can be revealed by special software on devices such as smartphones or tablets.



(a) QR code referring to my website



(b) Practical usage

Figure 9: Two examples of QR codes that contain a URL.

Apart from the lack of human readability, QR codes are an attractive method for phishing attacks due to their usage in the physical world (Kieseberg et al., 2010). For example, malicious QR codes (i. e., pointing to a phishing website), can be easily attached to an existing poster using a sticker. Furthermore, existing QR codes can be ‘patched’ by sticking a malicious QR code on top of them. Using both methods, the trust and attractiveness of the original poster is abused by the offender. The victim has to check the URL and make sure it belongs to the brand, just like in a phishing email. The effect of such an attack is local: victims have to be near the QR code to scan it. At the same time, this type of attack is virtually impossible to detect on a large scale. For an organisation running an advertisement campaign, it is not feasible to continuously validate all QR codes on all posters. Or, when no legitimate QR codes are in use, to validate that no malicious QR codes were attached.

The effects of a phishing attack that employs QR codes remain largely unknown. Several experiments in an academic environment were performed, concluding that many people scan QR codes, even when they lead to phishing websites (Been and Kleverwal, 2012; Vidas et al., 2013). Users mostly scan QR codes out of curiosity or for fun (Vidas et al., 2013), indicating that QR codes are not (yet) part of our everyday life. However, there are applications of QR codes, such as paying using the digital currency Bitcoin (Bamert et al., 2013). Additionally, to the best of our knowledge, no practical QR code attacks have been performed. In order to explore whether QR code phishing is an effective method, more research is needed. This leads to the research question: “*Are QR codes an effective attack method to obtain user credentials?*” In order to answer this research question, a real-world phishing attack was performed.

### 3.2.1 Method

An experiment was conducted in The Netherlands within a large organisation with over 3500 employees. The experiment simulated a phishing attack on the employees of the organisation. In the attack, a questionnaire specifically targeted at the employees was designed. The announcement of the questionnaire was printed on posters that contained a QR code with a link to the page of the questionnaire. The topic of the questions were the plans for a new building to be built, a topic that was often discussed amongst employees. This scenario can be considered realistic, since the organisation has just finished another campaign using posters with QR codes. The website of the questionnaire was designed specifically for this experiment.

The design of the poster was based upon an earlier poster of the organisation. An example of the posters in the experimental setting is shown in Figure 10. Apart from a call to action, each poster contained



Figure 10: Example the posters on a poster board. The QR code itself is not visible.

a unique QR code. Each QR code linked to a unique URL, with either a phishing domain or a legitimate domain. Every URL contained an identifier (i. e., ‘posterid=P02’) so that a click could be linked to a physical location. Furthermore, below each QR code, a unique `bit.ly` URL was shown, so that anybody wanting to participate without having a QR code scanner could do so. Nobody used the `bit.ly` links. For half of the posters, the link eventually redirected the user to the phishing domain, the other half redirected to the legitimate domain.

For the experiment, a single website was ran on two domains. The first domain was the legitimate website, running on a domain in the form of `https://topic.mijncompany.example`. The top-level domain `mijncompany.example` was the legitimate intranet website of the organisation, where ‘mijn’ refers to the Dutch word for ‘my’. The sub-domain `topic` referred to the topic of the questionnaire. Thanks to the IT department, we could get a subdomain on the legitimate intranet and obtain a valid SSL certificate for this subdomain. The second domain

was hosting the phishing website. This phishing website was hosted on a domain in the form of `https://topic.mijcompany.example`. An observant visitor would notice the wrong top-level domain, lacking an *n*. Failing to secure the phishing page would result in possible leaking of credentials during the experiment. Therefore, the phishing page was secured using an SSL certificate as well. However, the certificate was requested from a third party using fake name and address information. For the validity of the experiment, this should not matter too much, since a basic SSL certificate can be requested after proving domain ownership. Therefore, phishing websites can use a certificate too<sup>3</sup>.

Both the phishing and legitimate domain showed a login page of the organisation. Visitors were required to login before they could proceed with the questionnaire. Due to the open nature of the organisation's building (i. e., anybody can walk through most of the buildings), having to log in to the questionnaire was not uncommon. After logging in, the employee was presented with the questionnaire. After finishing the questionnaire on either domain, the employees were redirected to the legitimate intranet.

### 3.2.1.1 Concepts

The potential subjects in the experiment were employees of the organisation who were physically in one of the buildings. A subject is an employee who scanned a QR code on one of the posters, and tried to log in to either the legitimate or the phishing website. Potentially, any non-employee could scan the QR code as well, but would not be able to login and therefore not become a subject. However, since the locations of the poster were carefully chosen, they were likely to be only seen by employees.

Five variables were recorded for the subjects: `usernameHash`, `location`, `timestamp`, `useragent`, and `attemptLogin`. The `usernameHash` contained a message digest of the username, obtained through a cryptographic hash function (SHA1). The real username of the subject was not visible to the researchers. The `location` variable referred to the specific poster on which the QR code was scanned. The `timestamp` referred to the day and time of loading the webpage. The `useragent` identification string, i. e., which browser or QR code scanner the subject used, was stored in `useragent`. Finally, the variable `attemptLogin` had three possible values: *valid* if the credentials were correct; *invalid* if the provided credentials were not correct; and *none* when no attempt to login was made. The `usernameHash`, `location` and `timestamp` were used to remove duplicate entries. Duplicates occur not only when someone logs in twice. Some QR code scanning applications automatically open the link when

<sup>3</sup> Additionally, SSL certificates can be obtained without charge at registrars such as Let's Encrypt

scanning a QR code, in order to preview the page. The `usernameHash` was not used for other purposes. The dependent variable of the study is the `attemptLogin`.

### 3.2.1.2 *Subject Selection*

Subjects were self-selected, they participated after scanning the QR code on one of the posters. Due to the self-selection, subjects had to see the posters. Therefore, the location of the posters influences the potential group of subjects. The buildings of the organisation were open to the general public. Clearly, it is important to have the maximum exposure for employees, while limiting the exposure for other people. At the same time, the posters should not be too obtrusive so as to disturb the employees during their work. Furthermore, the experiment should be plausible scenario for a real phishing attack. A real phishing attack should gather some, but not too much attention, so as to not cause employees to warn the security department. To account for all requirements, posters were placed in areas with many employees, such as lunch rooms and coffee corners.

### 3.2.1.3 *Ethics and Risks*

Informing all employees beforehand would invalidate the experiment. Therefore, informed consent was not possible. To make the experiment as risk-free for the subjects as possible, several precautions were taken. Both the phishing website and the legitimate website were only accessible using a secured connection (i. e., using HTTPS). The questionnaire existed on both domains. Therefore, the promised functionality in the form of a questionnaire was delivered to the subjects. Furthermore, the password provided by subjects were not stored by the researchers. Rather, the provided credentials were checked for validity and only a 'valid' or 'not valid' annotation was stored. Additionally, the username of the employee was stored using a cryptographic hash function that produces a message digest. The obtained data was only seen by the researchers conducting the study.

Obtaining the proper permissions for running an experiment is essential. The experimental design, procedures and debriefing were approved by the ethical committee of the Faculty of EEMCS (Electrical Engineering, Mathematics and Computer Science) of the University of Twente. For the present study, the board of directors of the organisation gave permission to conduct the experiment. Several other employees were involved in the preparations as well. Many potential problems had to be taken into account, since employees and subjects may do a number of unexpected actions. For example, an employee may: (1) report the phishing attempt to the IT service desk; (2) inform the local press regarding the phishing attempt; (3) file a take-down request to

the hosting provider; (4) see how the researchers put the posters on the wall and call the security office. To reduce the impact of these scenarios, we informed the manager of the IT service desk, the manager of the project team that managed the new building being constructed, the CERT team, the press officer, the IT administrator responsible for the webmail, and the manager of the physical security. Furthermore, the company hosting the phishing and legitimate websites agreed to suspend the notice-and-takedown procedure for both domains and associated hosting, for the duration of the experiment.

As there was no informed consent, all employees were informed of the experiment afterwards. Subjects were debriefed as a group afterwards, rather than during the experiment, to reduce the risk of them warning other employees. Since employees at key positions within the organisation were informed beforehand, we would be notified of complaints. To the best of our knowledge, no complaints were filed. Even after the debriefing of all staff, we did not receive any questions or complaints.

#### 3.2.1.4 *Running the Experiment*

The posters were distributed by researchers on April 8th, 2013. That day, 42 posters were put on noticeboards and walls throughout the facilities of the organisation. As mentioned before, half of the posters ( $N=21$ ) had a QR code leading to a phishing page, whereas the other half contained a QR code that led to the legitimate page. One week after starting the experiment, the researchers went back to determine whether the posters were still present. Almost half (48%;  $N=20$ ) of the posters were still present, 10 posters (24%) were removed, and the whereabouts of the remaining 12 posters (28%) were unclear<sup>4</sup>.

Three weeks after the introduction of the experiment, it was stopped. At that point, the researchers went to all the poster's locations and physically distributed a survey to all employees they could find. The questions of this survey can be found in [Table 7](#). The researchers went around during the lunch breaks at April 29th and May 7th, 2013. Employees who did not recall having seen the posters, were shown a physical copy of the poster.

#### 3.2.2 *Results*

Only 12 unique visitors were recorded as having seen the website. Due to the lack number of accesses, a meaningful statistical analysis of the results is not possible. However, we do provide descriptives of the results in [Table 6](#). Eight people scanned the QR code on the legitimate poster, and clicked on the link. In comparison, only four people did the same

<sup>4</sup> For these 12 posters, the exact location of the posters was unclear from our notes. Therefore, there is no certainty whether they were removed, or could not be found.



Login	Phishing	Legitimate
No login	50% (N=2)	37.5% (N=3)
Invalid	25% (N=1)	0% (N=0)
Valid	25% (N=1)	62.5% (N=5)
Total	100% (N=4)	100% (N=8)

Table 6: Results of login attempts for the phishing domain and the legitimate domain.

for the phishing poster. Six employees logged in to the questionnaire, one from the phishing domain and five from the legitimate domain. Even though the legitimate poster got more logins, there are too few results to conclude that this is a significant finding.

A further analysis revealed that all subjects obtained the URL through the QR code. The bit.ly links that were shown below the QR code on each poster, were not used at all. Three mobile operating systems were used to click on the links within the QR codes: Android (25%; N=3); ios (67%; N=8); and Windows Mobile (8%; N=1). Posters on 8 out of 42 locations (19%) were followed up on by the subjects.

The posters were distributed at April 8th, which we refer to as day 0. The same day, four hits were registered on the legitimate website of which three logged in. At day 1, two subjects looked at the legitimate domain and one logged in. The phishing domain got its first hit as well, but no login attempts were made there. Two subjects browsed to the phishing domain at day 2, one of them logged in. The first one logged in to the phishing website at 12:29. The second subject scanned the same QR code (i.e., on the same poster) at 12:34, but tried to login using incorrect credentials. Both subjects used a different device (iPhone and HTC Desire Z, respectively). One subject logged in to the legitimate domain at day 3. Finally at day 14, one subject visited the phishing domain without logging in, and another subject visited the legitimate domain, again without logging in.

Half (N=6) of the subjects visited the websites between 2pm and 4pm. Of the remaining six, three subjects visited between noon and 1pm. The remaining three were at 4am (N=1) and between 5pm and 5.30pm (N=2).

The number of subjects for the phishing domain, as well as the legitimate domain, were low. There were no indications of reports to the organisation regarding the phishing domain and/or posters. A separate survey was held to find out why employees did not scan the QR code, the results of which can be found in Table 7. The survey was filled in by 45 employees aged between 19 and 59 ( $M = 36.73$ ;  $SD = 11.9$ ). Four employees refused to fill in their age. Most surveys were filled in by

Question	Answer 'yes'
1 Do you own a smartphone or tablet?	73.3% (N=33)
2 Do you know what a QR code is?	66.7% (N=30)
3 Do you know how to scan a QR code?	37.8% (N= 17)
4 Do you scan QR codes in general?	17.8% (N= 8)
4b If not, why?	
5 Have you seen the poster inside the facilities?	60.0% (N= 27)
6 Did you scan the QR code on this poster?	0.0% (N= 0)
6b If not, why?	
7 Have you seen the URL on this poster?	20.0% (N= 9) <sup>†</sup>
7b If you have seen it, why didn't you use it?	

<sup>†</sup> Two employees (4.4%) did not make a choice for this question.

Table 7: Questions from a survey that was held after the experiment finished. Dichotomous answers. N=45.

women (82.2%; N=37) and only three were filled in by men (6.7%). The remaining five employees did not indicate their sex. The results of the survey show that the majority of the 45 employees own a smartphone or tablet and know what a QR code is. However, only 37% of the employees know how to scan a QR code. Only 17% (N=8) indicates scanning QR codes in practice. The main reason the employees gave for not scanning QR codes is a lack of interest in them, followed by not knowing how to scan them. Since we anticipated this, a URL was provided on the poster. However, only 20% (N=9) indicated having seen the URL. Even though 60% (N=27) employees indicated having seen the poster of the experiment, none of them scanned the QR code. A lack of time was provided most often as reason for not browsing to the URL.

### 3.2.3 Discussion

In the present study, only 12 subjects scanned the QR code and went to either the legitimate or phishing website. One employee entered his/her credentials on the phishing website. Several explanations for the lack of response are likely. According to the results of the survey, a lack of time, as well as a lack of interest were important factors for not scanning a QR code. In the preparation phase of the experiment, the organisations staff that indicated that many people had strong opinions about the new building, which therefore was a good candidate topic

for the experiment. However, the experimental results indicate that the chosen topic of the questionnaire was insufficiently inviting the employees to share their opinion. Using a more controversial topic would likely increase the response rate.

Previous studies performed showed that lots of people scan QR codes in a university environment (Been and Kleverwal, 2012; Vidas et al., 2013). Since the organisation of our study was not a university, this could influence the response rate as well. QR codes appear more in different settings, on commercial posters or in shops. However, they are not commonly used yet, at least not by the employees of the organisation in our experiment. This suggests that QR codes may still be in the early adoption phase.

Even though only one employee was phished, this could be an acceptable result for an attacker that wants access to the internal network of the organisation. With the credentials of a single employee, an offender would be able to enter the network of the organisation. To perform a phishing attack, the offender would need to hang up posters within the organisation. This involves a non-negligible risk of getting caught, due to the requirement of having to put the posters physically. When noticeboards become digitalised, this risk may be reduced. While the researchers were putting the posters on the notice boards, they did receive questions from several employees. The researchers were instructed to reply that they did not know anything about the poster, and that they were paid per hour to just hang up posters. All potentially suspicious employees seemed to accept that explanation. Additionally, there were no reports made to the security department or management on the presence of the phishing posters, or the distribution of them.

### 3.3 CONCLUSIONS

Different *modus operandi* have their own results. In this chapter, we explored two *modus operandi* for obtaining information from people. Firstly, USB keys were dropped to measure how many people would pick it up, and keep it. Depending on the state of the USB keys (used or new), between 12% and 41% was not returned, even though there was ample opportunity to return it. The non return percentage is, therefore, the lower bound of what will be returned in different situations. We expect that more people would pick up and steal a USB key from the floor if there no service desk or authority nearby. Inserting an untrusted USB key in a computer may result in a malware infection (Sood and Enbody, 2013). In the experiment, the USB keys were picked up rapidly. Therefore, it is a very effective method for getting access to an organisational network or person's computer hardware. However, the offender would need to drop an infected USB key close to his target.

The second modus operandi that was explored uses QR codes to perform a phishing attack. Due to the low number of participating subjects, no meaningful statistical results could be found. However, only a few employees of the targeted organisation scanned the QR codes. This leads us to conclude that QR codes are not as commonly accepted yet. Furthermore, the way of attracting people to scan is very important, which may have led to few participants. Looking at the sparse data, four participants browsed to the phishing website through scanning a QR code. One of them (25%) fell victim by providing valid credentials to the phishing website. In contrast, the legitimate website was visited by 8 participants, of which 5 (62.5%) provided valid credentials. However, due to the low number of participants (one victim out of only four scans), further research is required. Finally, performing a successful phishing attack using QR codes depend on the general public being able to scan them. When people scan QR codes only because they are curious (Vidas et al., 2013), other means of phishing are more effective in obtaining information. This may change as QR codes are adopted by the general public.

With these experiments, we have looked at the effectiveness of two modus operandi for performing a phishing or fishing attack. From experiments described in literature, we can find measured success rates of other modus operandi. As explained in [Chapter 1](#), standard phishing attacks have a success rate of between 2% and 16%. In one experiment with personalised phishing, the researchers managed to obtain a success rate of 72% (Jagatic et al., 2007). Social engineering (face-to-face) has success rates of 62% (Bullee, Montoya Morales et al., 2015), and telephone-based social engineering has a success rate of 46% (Bullee, Montoya et al., 2016). The results suggest that scalable attacks are less effective than the non-scalable ones. However, there are outliers (such as personalised phishing). Even though the results of our experiments combined with experiments in the literature suggest a link between the scalability and the modus operandi, more experiments that are specifically targeted to the scalability properties are needed to actually prove a causal relationship. Hence, our experiments alone cannot prove that an attack's effectiveness *follows from* the modus operandi. There may be a confound variable, or non-measured variable, influencing the outcome.

In summary, the modus operandi of attacks that were tested in our experiments had diverse scalability properties. Phishing is very scalable, but less effective compared to spear phishing (Jagatic et al., 2007). Spear phishing using contextual information about the victim is more effective, but scales less, due to the requirement to collect data about the victim. Dropping USB keys is not easily scalable, but people pick up the USB keys within minutes. The requirement of physical presence reduces its scalability. However, a single person can drop many USB keys within a short period of time, and infect the computer of anyone who

picks it up. QR codes are potentially a good candidate for phishing, but require a broader adoption in the general public. They can be scalable if distributed digitally, or less scalable when distributed physically.

Having explored phishing *modus operandi* using experiments, we now turn to the victim's side of phishing by exploring which factors influence the decision to believe a phishing message.



Preventing phishing can be categorised in two parts: (1) technical interventions; and (2) social interventions (Khonji, Iraqi and Jones, 2013). Technical interventions try to hide bad emails (e. g., spam filters), assist in decisions making (e. g., warnings in email clients) or function as a gatekeeper (e. g., blacklists). The second category consists of social interventions, such as education and training of users (Kumaraguru, Cranshaw et al., 2009; Sheng, Magnien et al., 2007; Arachchilage and Love, 2013). Both social and technical interventions are needed to reduce the impact of phishing attacks, and new interventions need to be developed. To develop a new intervention or analyse the effectiveness of an existing one, it is important to know how people read their emails and decide to take action.

The dual-process theory of thinking considers two types of thinking: fast and autonomous (system 1), or slow and controlled (system 2) (Kahneman, 2012). System 1 consists of heuristics that are fast and effortless, whereas system 2 is slow and requires significant mental effort. There is an ongoing debate whether there are two dichotomous types, or rather a scale, or whether there is a single process (Evans and Stanovich, 2013; Kruglanski and Gigerenzer, 2011). However, critics such as Gigerenzer and Todd (1999) do consider the existence of simple heuristics for making decisions. In the context of phishing, system 1 and its heuristics try to figure out whether an email is trustworthy. When a person processes dozens of emails per day, the mental effort of using system 2 for analysing trustworthiness render in-depth analysis by system 2 infeasible. Consequently, the heuristics used by system 1 to consider an email phishing, or activate system 2 to make a decision, are important to a person's digital safety.

Apart from education, such as games or training, email users 'train' themselves every day by processing their email. Users form a risk model and develop heuristics for assessing trustworthiness of communications (Kirlappos and Sasse, 2012). Each user maintains a set of heuristics on which he or she assesses the validity of an email. While these differ per individual, some patterns in decision strategies emerge (Downs, Holbrook and Cranor, 2006). If the set of heuristics is insufficient or incorrect, users may be unable to distinguish phishing from legitimate emails, leading to victimisation. Users train themselves on every spam or phishing message that gets through their spam filters. Thereby, users reconfirm their detection heuristics on emails that they consider

<sup>5</sup> This chapter is based upon joint work with Lars Mol, Hans Heerkens and Marianne Junger

clearly bad. Consequently, users train their heuristics on easy-to-detect phishing emails. Identifying these heuristics is important for improving training and awareness campaigns.

This research aimed to identify the variation in heuristics, or thought patterns, of users. We assumed that reading an email leads to specific thoughts. These thoughts lead to the decision whether or not to take action on a phishing email. We aimed to identify such mechanisms. This lead to the research question of this chapter: *which heuristics do people use when deciding what to do with a phishing email?*

We aimed to establish how users decide whether an email is phishing by performing a think aloud experiment. Establishing how the subjects read emails and which line of reasoning they use for establishing authenticity provides useful theoretical insights (e.g., patterns in decision making) as well as insights that are useful for practitioners (e.g., adjusting training material). We do not aim to test the ability of the subjects to recognise phishing. Therefore, the subjects were not explicitly asked to identify a phishing email, or even to make any decision. Instead, we wanted subjects to process the email as any other email, including any decision making that followed from that.

Important to determining which heuristics people use while reading a phishing email, are patterns in which they read. Such reading patterns consist of the way people read emails, and at what moment they use shortcuts to fasten the processing of an email. For example, a person may only read the sender and title of an email, and when a bank is mentioned, delete the email. We expected that subjects start by reading the sender of the email, followed by the title and the contents of the email. To avoid confusion, we use the word *title* when talking about the subject-header of the email, to avoid confusion with the participants (subjects) of our experiment.

This remainder of this chapter is structured as follows: [Section 4.1.1](#) discusses trust and related work on phishing victimisation. Then, we describe the methodology of the study in [Section 4.2](#), followed by the results in [Section 4.3](#). Finally, we conclude with the implications and how these results might be applied.

## 4.1 BACKGROUND

### 4.1.1 Trust

Between the moments of receiving a phishing message and being victimised, many decisions are made. For example, on a high abstraction level, one needs to open the message, read it and decide what to do with it (i.e. ignoring or taking action such as responding or trashing). A user becomes victimised when trusting a phishing message and consequently revealing personal information. Therefore, trust is important



in the decision making of phishing messages (Kumaraguru, Acquisti and Cranor, 2006). If the phisher manages to create a message that is considered trustworthy by the recipient, heuristics of the reader may fail to raise alert. Some receivers look at the reputation of the sender (Downs, Holbrook and Cranor, 2006), but they may click the link in a phishing email even without a connection between them and the sender (Kumaraguru, Rhee, Acquisti et al., 2007). Having subjects think out aloud may give insights in the decision process. However, a theoretical background of trust is needed in order to classify thoughts of subjects on the (purported) sender of the phishing email. We consider two theories on the trust of phishing messages: (1) a model of trusting the supposed sender (organisation or person); and (2) the truth bias about trusting statements.

The model of trust of Mayer, Davis and Schoorman (1995) establishes three factors that influence the perceived trustworthiness the most: ability; benevolence; and integrity. The ability of a person depends on the influence due to skills, competencies and characteristics of this person within a specific domain. For example, one may not trust the baker to fix a broken car. The benevolence is the perceived will to do good without having an egocentric profit motive. For example, a pedestrian warning pedestrians that the street is slippery does so selflessly. The last factor, integrity, refers to the extent to which the truster believes that the trustee has an acceptable set of principles. For example, two strangers meeting at a conference of a particular political party may feel they share the same set of principles. The extent to which these three factors lead to a feeling of trust is moderated by a person's propensity to trust. The propensity to trust is a factor within each person that determines the likelihood of trusting others. Therefore, the three factors of trust do not influence each individual equally. The resulting trust is weighed against the perceived risk of engaging in a trusting action. The outcome forms the input for future trust decisions. We will apply the elements from Mayer's model of trust to classify trust decisions of subjects.

The model of Mayer, Davis and Schoorman (1995) relates to trust decisions for trusting persons or organisations. In phishing, however, there are trust decisions involving other factors as well. For example, a phishing message may contain many statements on problems (e. g., email inbox ran out of space) and proposed solutions (e. g., click on this link to solve the problem). Therefore, we need to know more about how people trust statements. Furthermore, trust is not an absolute state, nor is it constant in time. This is shown in the *bias to believe or truth bias*, suggesting that people initially make an attempt to believe a statement, only to later make a decision whether or not to disbelieve it (Kahneman, 2012; Levine, Park and McCornack, 1999). Without a truth bias, communicating with others would become too hard (Burgoon and Levine, 2010). For example, consider being present at a birthday party and talking to people you've not met before. It is simply not feasible to

check for evidence or find witnesses to support all of their statements and stories. The truth bias is a heuristic (or mental shortcut) that is employed to save mental effort (Burgoon and Levine, 2010). Applied to phishing messages, this suggests that people first try to believe the phishing message. Only if the contents of a message trigger suspicion, will the receiver disbelieve the message.

#### 4.1.2 *Characteristics For Victimisation*

The relation between the characteristics of the receiver of a phishing email and victimisation shows mixed results in literature. For example, some studies show that males are less prone to phishing emails than females (Jagatic et al., 2007; Kumaraguru, Sheng et al., 2010; Sheng, Holbrook et al., 2010), but a significant relation can not always be found (Leukfeldt, 2014; Alseadoon, 2014; Dhamija, Tygar and Hearst, 2006). Younger adults perform worse than older ones (Sheng, Holbrook et al., 2010; Alseadoon, 2014) and teenagers in particular perform worse than adults (Kumaraguru, Sheng et al., 2010). Again, however, not all studies confirm the age to be related to victimisation (Leukfeldt, 2014; Dhamija, Tygar and Hearst, 2006). In conclusion, the relation between the characteristics of the receiver and the victimisation is subject to debate and is unclear. Therefore, this study will not focus on characteristics of the receiver. Rather, we want to know more about the way users read emails, assess their authenticity and decide whether to take action or not.

In a related study, Downs, Holbrook and Cranor (2006) found that people employ three strategies on deciding how to respond to a particular email: (1) Judge personalisation and professionalism; (2) whether the communication is expected or normal; and (3) the reputability of the sender. Neither of these strategies is sufficient to identify all phishing emails (Downs, Holbrook and Cranor, 2006).

To find heuristics that people may use, one can look at email characteristics that influence the decision to label an email as phishing. If the receiver decides to take action, he needs to perform the requested action, often by filling in information at a website. Several factors that influence the receiver's decision making have been found in prior research. An impersonal salutation (or greeting) is considered to be a warning sign of phishing emails (Downs, Holbrook and Cranor, 2006; Sheng, Holbrook et al., 2010; Dutch Banking Association, 2015; Pfeiffer, Kauer and Röth, 2014). However, its presence or absence does not change the trust of the receiver in the message (Jakobsson and Ratkiewicz, 2006). Lengthy or detailed messages make users evaluate the message by other characteristics than content (Tsow and Jakobsson, 2007), such as design (Pfeiffer, Kauer and Röth, 2014; Tsow and Jakobsson, 2007; Jakobsson, Tsow et al., 2007). Additionally, many users look at linguistic charac-

teristics, such as spelling and grammar (Pfeiffer, Kauer and Röth, 2014; Jakobsson, Tsow et al., 2007; Wang, Herath et al., 2012).

The content of the email is important as well. Phishing messages are perceived as more trustworthy when they link to incidents that got a lot of attention in the media (Tsow and Jakobsson, 2007). Introducing urgency to a phishing email increases the likelihood of responding (Wang, Herath et al., 2012; Vishwanath et al., 2011). Users focus on the urgency cues and pay less attention to other cues that indicate deception, such as spelling or personalisation. Experts recommend to check the location a link points to, before clicking on the link (Downs, Holbrook and Cranor, 2006). However, only some users check the links before clicking (Jakobsson, Tsow et al., 2007), whereas others are not even aware of a way to check the location of a link (Kumaraguru, Rhee, Acquisti et al., 2007). Success rates of phishing vary per experiment, depending on the exact email being used as well as the level of personal information or context that is added. Results range from 7%–17% (Jakobsson and Ratkiewicz, 2006; Jagatic et al., 2007) with a phishing email with no context, to 72%–89% (Jagatic et al., 2007; Egelman, Cranor and Hong, 2008; Ferguson, 2005) when relevant contextual information is present. This indicates that the email user's abilities for assessing authenticity of an email are poor, in particular when the phisher adds context to the email.

Once subjects have decided to take action on a phishing message, they typically are guided to a website to fill in information, which is the second phase of phishing. On the website another trust-assessment will be performed. In the context of websites, users look at the content or only layout (Kumaraguru, Rhee, Acquisti et al., 2007; Kumaraguru, Acquisti and Cranor, 2006; Dhamija, Tygar and Hearst, 2006), or look for non-security related browser information (Kumaraguru, Rhee, Acquisti et al., 2007). This is consistent with the way users judge emails according to the literature.

#### 4.2 METHODOLOGY

A think aloud experiment was conducted in August 2013, following the methodology of van Someren et al. (Someren, Barnard and Sandberg, 1994). Employees from the supporting staff of the University of Twente were asked to participate. By having the subjects verbalising their thinking when reading a phishing email, we aim to establish how they decide whether an email is phishing. Subjects who think aloud will not express all their thoughts, but enough to analyse the structure of their thinking (Ericksson and Simon, 1993). By not informing the subjects of the true purpose of the study, their decision making process will more closely resemble the situation when they are reading emails at home or at work.

Even though the subjects may not be as comprehensive when it comes to describing all security features they might know to exist, the aim of this study is to find the criteria that subjects use in practise. Our design does not include role playing by the subjects, like in Downs, Holbrook and Cranor (2006). Due to the think aloud methodology, users may not express each individual thought (Ericksson and Simon, 1993). Users could have some relevant thoughts that were not recorded in our experiment. However, the method does give insights in the structure of the thought processes.

#### 4.2.1 *Subjects*

The subjects were all employees of non-research departments of the authors's university. They worked in various positions, for example, administration, finance, management or emergency response. This sample of subjects is limited and not representative for the population at large. However, the sample consisted of persons that are able to clearly express themselves, are able to reason about a problem, and know how to get around on the internet.

To get subjects, the departments' management was asked for permission to conduct a study with their employees within working time. Then, the team leaders within each department were contacted to distribute the description of the study with their employees, asking them to participate. A maximum of three employees per department were selected for participating in order to get a more heterogeneous population in terms of background, age and education. Subjects were told that they would participate in a study of marketing and communication through email, and that the researchers were interested in the way people read their emails. Therefore, no references to cybersecurity in general or phishing in particular, were made by the researchers in the recruitment for, and briefing of, the experiment. Consequently, we do not expect the users to be more wary of phishing than during in their regular email use (Parsons et al., 2015).

On the ethical part of the experiment, the subjects were fully informed about the purpose of the experiment, that is, knowing how people read their email. The subjects were told that it was used for a marketing study. After the sessions, the subjects asked not talk to their colleagues about the contents of the email that they read, to prevent informing other subjects. When the experiments were finished, all subjects were provided with an explanation of the study and a summary of the results.

The subject group consisted of 14 men and 10 women ( $N = 24$ ) between the ages of 24 and 63 years ( $M = 47.5$ ,  $SD = 11.0$ ), all of whom had Dutch as their first language. Two thirds of the subjects ( $N = 16$ ) had a degree from a higher education institution (i.e. vocational college

or university Bachelor), of which five had a master's or doctoral degree. When asked how many hours they use a computer (PC or notebook), the subjects' answers ranged from 25 to 80 hours per week ( $M = 41.6$ ,  $SD = 11.8$ ). This includes computer usage professionally and for private use, both at work and at home. Four subjects indicated that they read their email on a daily basis and twenty subjects read their email several times a day.

#### 4.2.2 Design

An experimental protocol was designed following the guidelines of van Someren et al. (Someren, Barnard and Sandberg, 1994). For subjects to think aloud, they need to perform a task that is not too easy, so that they do have to think about it, and not too difficult, so that they have no mental resources left to verbalise their thoughts (Ericksson and Simon, 1993). We consider the task of reading an email appropriately difficult.

To test the influence of urgency on the subjects, two emails were used: a '*normal*' and an '*urgent*' phishing email. The normal email (Figure 11) was taken from the website of a Dutch bank, where it was posted as an example of a recent phishing email. A real-life phishing email was chosen so as to make the experiment as representative as possible. There were no spelling errors and there was a clear call for action, so that subjects had to make a decision. The email contained five paragraphs of text. The email was written in Dutch and starts with "*Dear customer*", after which it claims in the first two paragraphs that a disruption of service has occurred. Then, the email states an apology for the inconvenience and explains what went wrong in the systems of the bank (one large paragraph) and how the bank responded (one paragraph). In the fifth and last paragraph of the email, the sender asks the receiver to log in to the bank's website in order to update their account.

The urgent phishing email was created on the basis of the normal phishing email. The email is shown in Figure 12. It was slightly modified to express more urgency to immediately take action. To achieve this, the title, introduction and call for action were changed to express the urgent need for the reader to click on the link. In the normal email, the receiver was simply asked to log in, whereas the urgent email stated that the receiver's account was blocked and that the receiver is required to login in order to unblock the account. Half of the subjects were shown the normal phishing email and the other half the urgent version. Whether a subject would get the normal or urgent email, was decided randomly prior to the sessions. Each subject was shown only one email: either the normal email, or the urgent one.

The link in both emails, asking to log in to the bank website, pointed to a website on the same machine. No attempt was made to make the

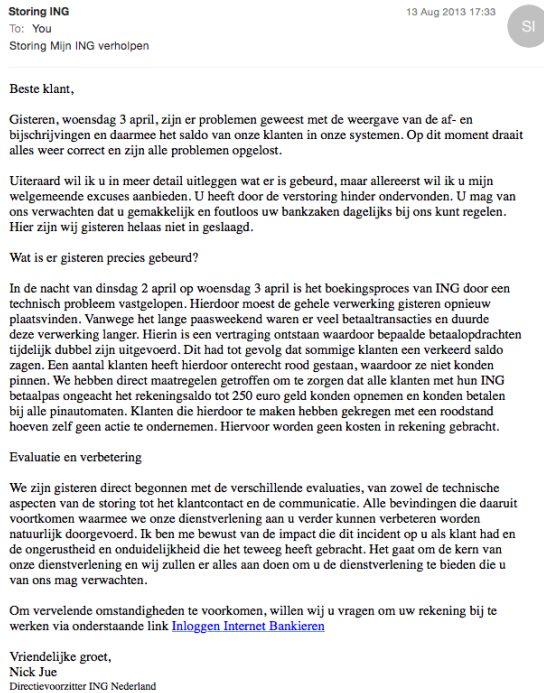


Figure 11: The ‘normal’ version of the phishing email that was used in the experiment.

link look legitimate, except for a filename that included the name of the bank. If subjects clicked on the link, a pop-up warning showing “Clicking hyperlinks can harm your computer.” would display, asking the subject confirm proceeding. When proceeding, the subject would be shown a duplicate website of the login page of the actual banking website. Subject were able to fill in any data, but nothing was stored or processed. When a subject arrived at this stage, we considered the experiment finished.

#### 4.2.3 Procedure

With each subject a 40-minute think-aloud session was held according to a strictly predefined protocol. Before briefing the subjects, they were asked to provide some basic information, such as their job title, education, age, and use of a computer. In the briefing, they were told that the researchers were interested in marketing and communication in email messages from companies. This was done to not raise any awareness in the area of cyber security for the subjects.

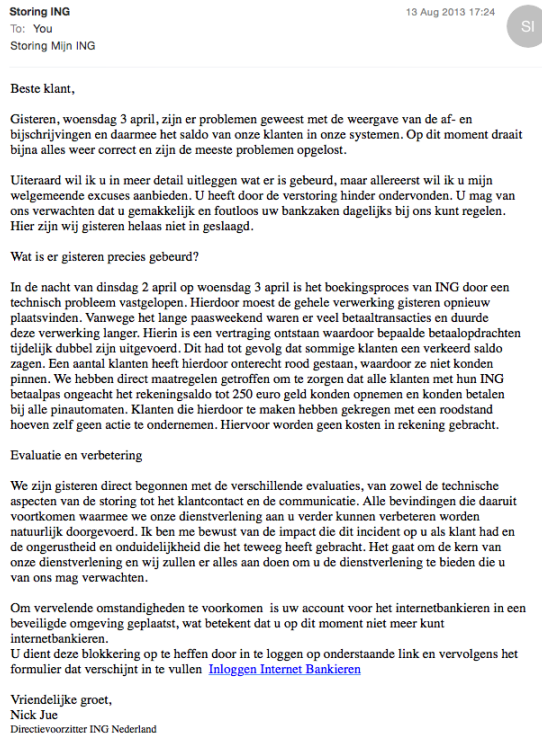


Figure 12: The ‘urgent’ version of the phishing email that was used in the experiment.

The researcher conducting the experiment asked the subjects to verbalise all thoughts and explained what is meant with that. Since this is a difficult task for most people, some exercises were used to practise thinking aloud. Subjects were told to read the email out loud as well. Specifically, subjects were asked to express any feelings they had regarding the emails and indicate when their feelings changed. Subjects were reminded that there are no wrong answers or decisions. It was made clear that the researcher was only present to facilitate the experiment and record the process. Therefore, during the thinking aloud, subjects were told not to address their verbalisation to the researcher. Additionally, the researcher would say nothing other than reminding the subject to keep thinking aloud during silent periods.

The think-aloud warmup exercises consisted of three tasks: a math problem, naming twenty animals and reading aloud an email. In the first task subjects were asked to calculate  $24 \times 36$ . The subjects were interrupted after about a minute, since the task was not to actually compute the result, but to practise thinking out loud. The researcher kept count for the second task, naming twenty animals, so that the subject

was free to focus on thinking out loud. If the subject stopped talking at some point, he was encouraged to keep verbalising his thoughts. Finally, for the third exercise, the subject was given a notebook and instructed to pretend this was his own. The subject was asked to open the email and pretend he received it himself, and read aloud the contents, while verbalising his thoughts. The email pretended to be from a woman cancelling the participation of herself and her husband to the neighbourhood barbecue, due to their emotional response to having to put their pet to sleep (i. e., animal euthanasia). The goal of these exercises is twofold: firstly for the subject to practise thinking aloud; and secondly for the researcher to assess whether the subject sufficiently verbalised. In particular since the researcher is restrained from interacting during the task itself, unless the subject stops talking (Somerén, Barnard and Sandberg, 1994; Ericksson and Simon, 1993).

As equipment, a notebook with Microsoft Outlook 2010 was used for both the warmup exercise as well as the task itself. This software was pre-configured at all workstations provided to employees by the university. All subjects were used to work with it on a daily basis. Both the task and the warmup were recorded using a voice recorder.

The task consists of opening an email in outlook, and reading it. The subjects were asked to process the email and verbalise their thoughts while doing so. This description is purposely open-ended, since we did not want to steer the subjects in a particular direction. A subject who clicks on the link, would be presented a warning message stating that *“hyperlinks can be harmful to your computer and data. (...) Do you want to continue?”* When clicking “yes”, the standard browser (Microsoft Internet Explorer) would be opened, and a phishing website would be shown. This phishing website, as shown in Figure 13, was a duplicate of the real website of the bank. However, instead of logging in with a username and password, the subject was asked to provide his bank account number and the number of his bank card. By itself, these two pieces of information are insufficient to log in to the subject’s online banking. However, such information could be used to impersonate the bank in a social engineering phone call with the victim at a later stage. Any filled in data on the phishing website would be discarded in the client, and no submission was possible. If a subject would fill in their data, the task would be considered complete when they pressed the submit button.

Directly after the subject finished with the task, a short interview was held with the subject. This started with 17 general questions, from prompting subjects about the goal of the email and the task itself, followed by 6 questions in which the subjects could give feedback about the experiment. A complete list of questions can be found in Appendix B. After finishing, the subjects were offered a USB flash drive as a thank-you for participating.





Figure 13: Screenshot of the phishing website that was shown when a subject clicked on the link in the email.

#### 4.2.4 Pilot

Two pilot sessions were conducted in order to identify potential problems with the experimental design. Additionally, it was used to familiarize the researchers with the experimental procedure, test the recording equipment and time the duration of the session. Before starting with the pilot, a coding scheme was devised based on the theories from the related work (i.e., the model of trust and victimisation, see below). In the pilot sessions, the phishing email was shown to the subjects in a Microsoft Word document. It turned out to be problematic for subjects to identify key properties of the email, such as the sender. Therefore, the session of the experiment were held using Microsoft Outlook. During the pilot sessions, several changes were made to the briefing of the subjects to either clarify or remove typographical mistakes.

#### 4.2.5 Analysis

After each session was recorded, the recording was completely transcribed in as much detail as possible. The result is a *protocol* with all spoken sentences in text. The protocols were very detailed. Specifically, it was noted when a subject was silent, mumbling or talking unclearly. During the experiments, no interruptions (e.g., people walking into the room or any sudden sounds) occurred, with the exception of one subject, whose phone rang in the middle of reading the email.

Each protocol was segmented by splitting the spoken sentences in self-contained subsentences called *segments*. These segments were as short as possible while still being meaningful without context. For example, the complex sentence “I do not agree with you since John did not complain” would be split at the marker word “since” to form two segments (one stating a disagreement, one stating an observation). The protocols contains parts of the email that the subjects read out loud. Segments consisting only of parts of the email or website were not included in the reporting of the results. They were, however, used

in the analysis to provide context of the thoughts of the subjects. In the remainder of the paper, the segments that do not consist of text from the email are called *thoughts* or *comments* interchangeably. Any quotes from subjects were translated to English by the researchers. Additionally, excerpts from the phishing email were translated. In order to code the thoughts, a coding scheme was devised.

The coding scheme was based on the related work from [Section 4.1](#) and was operationalised by analysing the two pilot protocols. Based on the pilot study, attributes were included or removed. For example, subjects did not verbalise any thoughts regarding the *integrity* of the organisation that was mentioned in the email, resulting in its removal. Further improvements to the coding scheme were developed throughout the research using an iterative process. For each subject, the corresponding protocol was analysed using the coding scheme, and the coding scheme was checked for each protocol as well. If needed, and after extensive discussion amongst the four researchers, the coding scheme was adapted to fit all protocols.

Category	Attribute	Scale	
		Negative	Positive
Aesthetic	Spelling	Poor	Adequate
	Layout	Bad	Good
	Language	Bad	Good
	Efficiency	Inefficient	Efficient
Content	Believability	Not believable	Believable
Organisation	Ability	Not able	Able
	Benevolence	Bad intentions	Good intentions
	Identity	Scam	Legitimate
Security	Security	Not safe	Safe
Miscellaneous	Choice		
	Other		

Table 8: Attributes in the coding scheme. The Scale columns indicate the typical interpretation of ‘negative’ and ‘positive’ for this variable.

The final coding scheme consists of four categories of attributes: aesthetic; content; organisation; and security. An overview of the categories and attributes can be found in the left two columns of [Table 8](#). Aesthetical attributes denoted comments about the spelling, layout, use of language or efficiency of the email. Spelling mistakes are coded under *spelling*. Comments about the visual appearance of the email, such as a missing logo, are listed under *layout*. The attribute *language* contains thoughts regarding the writing style. The final aesthetic code, *efficiency*, concerns comments on the length or brevity of the email.

The second category of attributes is labelled content, containing comments regarding the *believability* of the emails. For example, a subject may state that he finds it hard to believe that a bank will send such an email. Thirdly, organisational attributes relate to comments about the organisation mentioned in the email. In particular, subjects from the pilot study commented often about the *ability* and *benevolence* of the organisation that appeared to have sent the email. When organisations fulfil the expectations from the subjects, they will be satisfied and comment positively on the organisation's ability. These expectations, however, differ for each subject. The benevolence concerns thoughts of the subjects about the perceived will to do good of the organisation that is mentioned in the email. Specifically, this covers thoughts regarding the will to do good beyond an egocentric profit motive. Additionally, thoughts about the *identity* of the organisation also fall in the category of organisational attributes. These are thoughts of subjects where they are wondering, or stating, whether or not the email is from the organisation. The fourth category is *security* and is used for thoughts on the risk, dangers or safety of the email. For example, a subject verbalising a thought that the hyperlink may be dangerous would be coded as security. Finally, thoughts that did not fit any category were listed as other thoughts. For convenience in analysing, explicit decisions were labelled as choice. All involved researchers discussed and approved the final coding scheme.

During the analysis, each thought was coded with one of the attributes of the four categories and a judgement of the thought: positive (+), neutral (o), or negative (-). *Positive* thoughts are thoughts about an attribute that are in favour of the authenticity of the message. *Neutral* thoughts about an attribute do indicate a judgement on the authenticity of the message. *Negative* thoughts indicate less trust in the authenticity of the message. To give an example, a segment marked as 'believability positive' means that the subject has verbalised a thought indicating believability of the contents, such as "*I heard before that this bank often has service disruptions.*" Furthermore, the segment "*The service disruption did not cause any problems for me.*" was coded as 'believability neutral'. Finally, an example of a segment coded as 'believability negative' is the segment "*That is odd*", which indicates suspiciousness in the contents. For each attribute, a typical interpretation what constitutes as positive, neutral or negative is included in [Table 8](#), and examples are included in [Table 9](#).

Occasionally, labelling a thought as positive, neutral or negative is non-trivial. For example, a positive thought on the use of language of an email could be "*they use typical banking language in this email.*" This can be considered a negative thought regarding the use of language, because it is implied that email is too formally written. However, since the phishing emails were indeed purportedly from a bank, this comment is marked 'positive' for the use of language, since it confirms the

Attribute	Negative	Neutral	Positive
Spelling	Work with a capital W...okay?	Checking the spelling for errors.... <sup>†</sup>	I still see no spelling errors
Layout	The mail is not very neatly made.	I don't see the logo of the bank. <sup>†</sup>	The layout looks good. <sup>†</sup>
Language	First they use plural, than singular.	That is one way of putting it.	They use typical banking language in this email
Efficiency	What a long email.	The email contains five paragraphs. <sup>†</sup>	Yes, i'm curious to read more about this.
Believability	This is a bit weird.	Well, i did not notice it.	Yes, this does worry me.
Ability	They should have sent it directly.	That is not more than logical.	Very good that they've sent this email.
Benevolence	[sighs] Again a disruption.	Anyone can make a mistake.	Good service of them.
Identity	This email does not have to be from the bank.	An email from [bank].	Blegh, indeed, always problems with this bank.
Security	Actually, i don't know where i end up [by clicking].	Let's see where I end up [by clicking].	Clicking never hurts.
Choice		I am not going to do this.*	
Other		Yes...*	

<sup>†</sup> Fictional example: there were no subjects verbalising thoughts labelled with this judgement.

\* For Choice and Other, no positive/neutral/negative judgement was added, therefore only one example is provided.

Table 9: Examples of thoughts for each attribute. Text between brackets was added by the researchers to clarify the thought.

subject's trust in the authenticity of the message. The initial coding was done by one person, after which the results were carefully reviewed and discussed with the other researchers. The coding of segments into attributes is challenging when a segment can be interpreted as being a statement concerning several attributes. These cases were discussed amongst the researchers and the attribute that was considered the most applicable was selected. Segments coded as 'other' or 'choice' were not coded with a judgement (positive, neutral or negative). Furthermore, normative comments such as "*I could expect that to happen*" are excluded from the analysis.

The thoughts of the subjects were the focus of the analysis. Listing the number of thoughts can be biased when some subjects are more verbose in expressing themselves compared to other subjects. In these cases, even a single subject with many verbalised thoughts influences the number of thoughts significantly. Therefore, to allow better inter-

pretation and indicate the importance of an attribute, the number of subjects having thoughts on a particular attribute are reported as well. Furthermore, thoughts marked as miscellaneous (i. e., the Other and Choice attributes) were not coded with a judgement. Instead, those thoughts were used to validate and improve the coding scheme. For example, when thoughts that we considered important to the thoughts processes of the subjects did not fit in the coding scheme, the coding scheme had to be changed. Finally, the answers to the post-task interview were analysed and are reported throughout the results when they provide additional insights.

In discussing the results of the analysis of the data, we establish whether a subject would have been victimised by the phishing attempt. In order to do so, we distinguish three types of the subjects: firstly, *vigilant* subjects explicitly chose to take no action and are not considered victims; *potential victims* choose to take action (i. e., click on the link), but do not provide their personal information; and *victims* choose to take action and provide their personal information to the offender. It is important to remark that this distinction in three types is just a representation of the actions of the subjects within this experiment. In a different setting or at another moment, a subject may choose differently.

#### 4.2.6 Limitations

The aim of this study was to provide an in-depth analysis of the think-aloud protocols. Due to the time-consuming nature of performing and analysing a think aloud study, the sample size is relatively small (24 subjects), resulting in a lack of statistical power. Even though the results cannot be translated to the general population, they do provide insights and show trends in the process of decision making when reading a phishing email, and therefore in the heuristics that people use. The experiment was held in a separate office with only the researcher and the subject, therefore the external validity is limited. However, we used the same equipment (notebook, software) that the subjects used for their daily professional activities. Only two versions of one email was used in the experiment, which makes the results dependent on this specific email. By selecting an email that was actually used for phishing, we aimed to be as close to the real world as possible. Finally, the results from this study are the attributes that our subjects used to read emails and assess authenticity. Other attributes may be found for different emails, since different emails may trigger different heuristics. An large scale experiment (i. e., many emails and many subjects) can be used to give an overview of the popularity of certain heuristics on a population.

Since the coding scheme is partly based on our theoretical perspective, we could have missed attributes that are in the data but are not identified by us. We tried to limit the risk of missing by re-checking the

coding scheme based on the protocols, and by having regular discussions about the results, as well as reviewing the analysis. It is possible that more results would have been found with another theoretical perspective. However, a look at the thoughts coded as miscellaneous revealed no important thoughts, suggesting a good fit of the coding scheme. Due to the iterative approach of re-evaluating the coding scheme regularly while coding the protocols, we aimed to improve the quality of the coding scheme.

When reporting the results, we analysed both the number of thoughts and the number of subjects having thoughts on a particular attribute. The total number of thoughts on a single attribute can be influenced by a small proportion of the subjects. Subjects who feel more comfortable verbalising their thoughts can be overrepresented in the total number of thoughts for particular attributes. However, we wanted to determine which heuristics are used, i. e., the variation in thought patterns. The exact number of occurrences of each attribute were therefore indicative and of lesser importance than the mere usage of each attribute.

#### 4.3 RESULTS

All protocols together resulted in 2288 segments from 24 subjects. Two kinds of miscellaneous attributes were coded: choice and other. In the original email the subjects made 20 choices, whereas 42 choices were made by the subjects who were reading the modified email. Segments marked as *other* were, apart from many hmmm's, remarks like "*I want to search on the internet*" or "*I am a customer of that bank*". The original email had 40 thoughts marked as *other* and the modified email had 154. Excluding these miscellaneous and normative codes, as well as the spoken contents of the emails itself, 553 segments (24%) remained. These 553 segments, which we refer to as thoughts, were further analysed.

An overview of all thoughts of all subjects is given in Table 10, together with their respective judgements, which we labelled as positive, neutral or negative. From the results of the analysis, we make two observations based on the occurrences of attributes. Firstly, out of the nine attributes from the coding scheme, only four attributes were used to label 88% of the thoughts. These four attributes were believability, ability, security and efficiency. The second observation is that the majority of all thoughts are negative (60.2%) in relation to the authenticity of the email, a quarter was neutral and only 15% was positive.

Not often did the subjects use terminology to refer to the phishing email. Only two subjects (8.3%) mentioned the term "phishing", one of them mentioned phishing while reading the introduction, and the other while reading a "How do I recognise phishing?" link on the phishing website. Three subjects (12.5%) called the phishing email "spam".

Attribute	Positive	Neutral	Negative	Total
Believability (content)	45 (8.1%)	84 (15.2%)	106 (19.2%)	235 (42.5%)
Ability <sup>†</sup>	16 (2.9%)	22 (4.0%)	79 (14.3%)	117 (21.2%)
Security	2 (0.4%)	14 (2.5%)	56 (10.1%)	72 (13.0%)
Efficiency	4 (0.7%)	2 (0.4%)	56 (10.1%)	62 (11.2%)
Identity <sup>†</sup>	4 (0.7%)	8 (1.4%)	19 (3.4%)	31 (5.6%)
Benevolence <sup>†</sup>	11 (2.0%)	4 (0.7%)	4 (0.7%)	19 (3.4%)
Language	1 (0.2%)	1 (0.2%)	5 (0.9%)	7 (1.3%)
Layout	0 (0.4%)	0 (0.0%)	6 (1.1%)	6 (1.1%)
Spelling	2 (0.4%)	0 (0.0%)	2 (0.4%)	4 (0.7%)
Total	85 (15.4%)	135 (24.4%)	333 (60.2%)	553 (100%)

<sup>†</sup> Thoughts on the organisation that is mentioned in the email.

Table 10: Level of trust shown by subjects, expressed in number of thoughts.

The remainder of the results section is categorised in three subsections: firstly, an analysis of the two different versions of the phishing emails; secondly, an analysis of the subjects, grouped by their potential victimisation; and thirdly, an overview of reading patterns, indicating specifics that the subjects had thoughts about.

#### 4.3.1 Urgent versus non-urgent

Half of the subjects read a phishing email that was taken from the website of a large Dutch bank. On average, each subject had 19.7 thoughts ( $SD = 18.0$ ) while reading this email. The other half of the subjects received a modified version of the same email, expressing the urgency to take action as soon as possible. The urgent phishing email caused more thoughts for the subjects ( $M = 26.4$ ;  $SD = 21.8$ ) compared to the non-urgent email. Apart from more thoughts, subjects reading the urgent email had roughly the same number of positive thoughts (urgent 14.5%; non-urgent 16.5%), more neutral thoughts (urgent 30.0%; non-urgent 17.0%) and fewer negative thoughts (urgent 55.5%; non-urgent 66.5%). [Figure 14](#) shows the differences between the two versions of the email.

In [Table 11](#), the number of subjects having thoughts on each attribute are listed. Almost all subjects had thoughts on the ability of the bank, as well as the believability of the contents. Introducing urgency cues resulted in fewer negative thoughts on these attributes.

Apart from the general trend that indicates a shift of negative thoughts towards more neutral thoughts, there were exceptions to this on the attribute level. Such exceptions became visible from analysing both [Table 11](#) in combination with [Figure 14](#). For example, introducing urgency led more subjects to consider their security ( $N=9$  instead of  $N=5$ ;

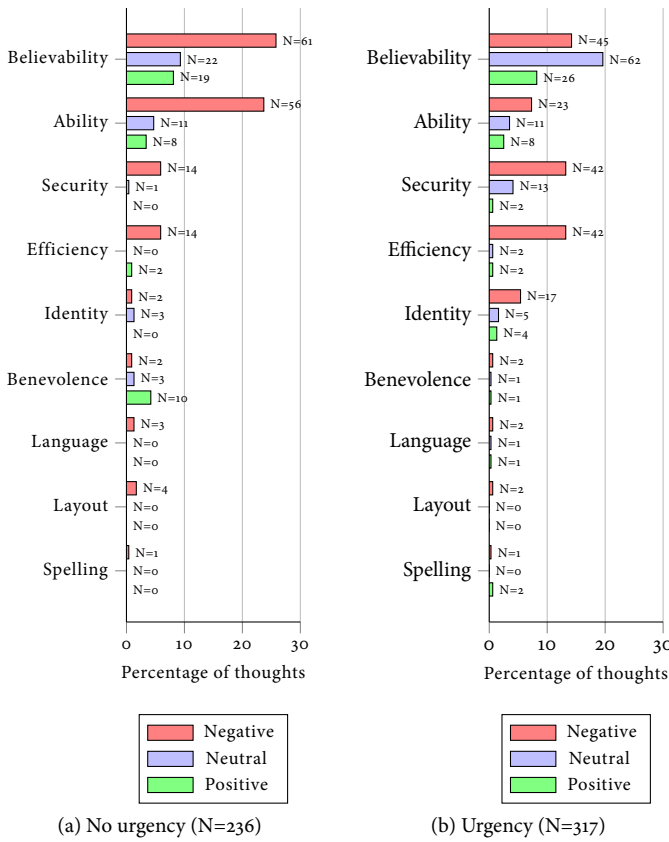


Figure 14: A comparison of thoughts of a phishing email with urgency or without urgency, expressed in number of thoughts per judgement.

in Table 11), and their thoughts were mostly negative ( $N=42$  instead of  $N=14$ ; in Figure 14). Furthermore, more subjects had thoughts on identity of the organisation in the email (i. e., whether the bank was really the sender of this email), and these thoughts were more negative compared to the non-urgent email. Both observations indicate less trust in the authenticity when the email is more urgent. Another difference concerned the efficiency of the email. The urgent email contains 387 words and the non-urgent one has 353 words. When the message is urgent, more subjects mentioned the effectiveness, and their thoughts were more negative. This suggests that the subjects considered the email too lengthy for an important message, such as having one's online banking suspended. Furthermore, introducing urgency reduces thoughts on the benevolence of the organisation. In the non-urgent email, seven subjects had thoughts on the will to do good of the bank from the



email, and these thoughts were mostly positive. In contrast, in the urgent condition, the benevolence was mentioned sparsely by only two subjects.

Attribute	No urgency	Urgency	Combined
Believability	11 (92%)	12 (100%)	23 (96%)
Ability	11 (92%)	11 (92%)	22 (92%)
Efficiency	6 (50%)	9 (75%)	15 (63%)
Security	5 (42%)	9 (75%)	14 (58%)
Identity	4 (33%)	7 (58%)	11 (46%)
Benevolence	7 (58%)	2 (17%)	9 (38%)
Language	2 (17%)	3 (25%)	5 (22%)
Spelling	1 (8%)	3 (25%)	4 (17%)
Layout	2 (17%)	2 (17%)	4 (17%)
Subjects	12 (100%)	12 (100%)	24 (100%)

Table 11: Number of subjects expressing thoughts on each attribute.

#### 4.3.2 Victimisation

Three types of subjects were listed in [Section 4.2.5](#): *vigilant* subjects (no action,  $N=12$ ); *potential victims* (take action, but did not provide their personal information,  $N=8$ ); and *victims* (take action and provided their personal information to the offender,  $N=0$ ). None of the subjects ended up providing their own personal information to the phishing website. Since the subjects did not receive the phishing email on their own account, and due to the lack of role playing, this was expected.

In [Figure 15](#), the decision making moments of the subjects are shown as a flow graph, together with the number of subjects that followed each path. The first step for subjects is to decide whether they want to take action on the email, i. e., clicking on the link in case of the phishing email used in our experiment. Four subjects (17%) did not make a verbalised decision to take action or refrain from doing so. The reason for this is unclear, since neither of them verbalised their thoughts regarding the reasons for not making an explicit decision on pursuing the email. This might be caused by elements from the experimental design, such as the briefing, resulting in the subjects considering clicking out-of-scope of the assignment. In practise, a user reading an email must always make a decision to either take action or not. Since the four subjects did not verbalise their decision, we consider the protocols of those subjects ‘missing values’ with respect to victimisation analysis. The flow graph in [Figure 15](#) does not include these missing values.

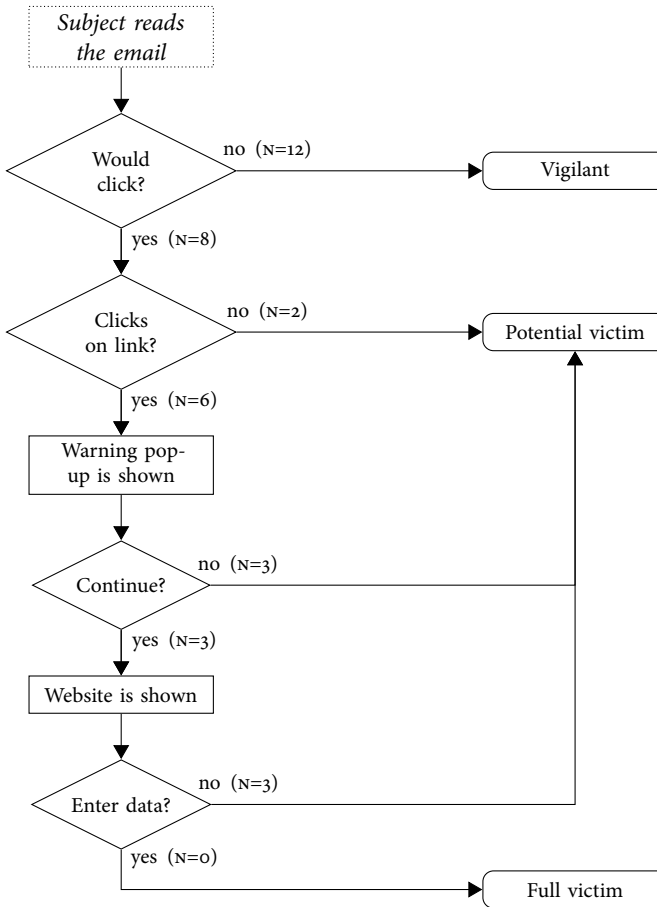


Figure 15: Flow chart showing the victimisation status based on the subject's decisions. Four subjects made no decision and were excluded.  $N=20$ .

Discarding the outliers, 20 subjects (83.3%) remain. The majority of the subjects decided they do not want to take action on the email and are labelled as vigilant ( $N=12$ ; 50%). The remaining 8 subjects (33.3%) indicated they wanted to click on the link. These subjects are considered potential victims. Six subjects (25%) actually clicked on the link, whereas the other two subjects (8%) indicated that they would take action at a later moment. One of these two indicated stated that *"I will not do this now, I will do it later. (...) Good that you sent this email."* The other mentioned not being a customer for this bank, but if his own bank would sent this email, he would click on the link and comply. The two subjects indicating they would click are considered lucky, instead of checking the validity of the request, they decided to not click to external factors (e. g., being in a controller environment). However, in a different

situation, they might fall victim. The six subjects who clicked on the link saw the warning message asking for confirmation to browse to the specified website.

The third category consists of potential victims, a label that is used for 8 subjects ( $N=8$ ; 33.3%). These subjects either clicked on the link and became suspicious due the warning message ( $N=3$ ; 12.5%), went to the website but did not provide their personal information ( $N=3$ ; 12.5%), or did not click, but indicated that they would click if they were customer of this bank ( $N=2$ ; 8.3%).

Attribute	Vigilant				Potential Victims			
	Thoughts		Subjects		Thoughts		Subjects	
Believability	133	45.4%	11	91.7%	85	42.5%	8	100%
Ability	70	23.9%	12	100%	26	13.0%	6	75.0%
Efficiency	36	12.3%	7	58.3%	20	10.0%	5	62.5%
Security	24	8.2%	9	75.0%	48	24.0%	5	62.5%
Identity	17	5.8%	7	58.3%	13	6.5%	3	37.5%
Benevolence	7	2.4%	3	25.0%	2	1.0%	2	25.0%
Layout	2	0.7%	2	16.7%	3	1.5%	1	12.5%
Language	7	0.7%	2	16.7%	2	1.0%	1	12.5%
Spelling	2	0.7%	2	16.7%	1	0.5%	1	12.5%
Total	293	100%	12	100%	200	100%	8	100%

Table 12: Number of thoughts per attribute, grouped by victimisation. None of the subjects provided information, so the ‘victim’ group is not included.

To discover what makes people vigilant, or what causes them to decide to click, an overview of the thoughts of each group is shown in [Table 12](#). In this table, the thoughts of the group of vigilant subjects and the the potential victims are shown. Several small differences between vigilant subjects and potential victims are visible regarding the number of thoughts. Most notably, potential victims have more thoughts on security than vigilant subjects. For example, on close inspection, one subject was verbalising his thoughts very well and clicked on the link. This single subject is responsible for the majority of the security-related thoughts within the potential victims group.

Within the group of six potential victims that clicked on the link, three subjects became suspicious due to the warning message and subsequently clicked ‘cancel’. All three fell for the phishing, in the sense that they did not spot possible harmful intentions until the warning message was shown. After reading the message, they started considering the option that the email was malicious, and therefore did not continue.

For these three subjects, the warning message worked in acknowledging a potentially dangerous situation.

The other three subjects clicked on the link and proceeded despite of the warning message. Two did not read the warning message or did not fully understand it, as indicated by their comments, such as “*I think this screen can go away.*”. We argue that this shows a problem with warnings: simple pop-up messages do not always work as intended, especially when people get used to seeing (and ignoring) warnings. The other subject stated being careful, concluded that just browsing to a website would not hurt, and continued to see what would happen. In his reasoning, the careful subject argued that the message looks reliable, since this particular bank was more often in the news for having service disruptions.

When considering the version of the email (urgent or non-urgent) that the subjects read, an interesting difference between subjects was found. All of the potential victims that continued even after having seen a warning message, had been given the urgency variant of the phishing email. This implies that introducing urgency does indeed work to attract action from the subjects. In contrary, subject reading the non-urgent email either stopped after receiving a warning message, or decided to take action at a later stage. To establish whether urgency indeed urges people to ignore warnings, a future study could investigate this on a larger scale with more subjects. An overview of the crosstable of type of victimisation and type of phishing email is included in [Table 13](#).

		Type of email		
		No urgency	Urgency	Total
Victimisation	Clicked, stopped at warning	3	0	3
	Clicked, continued at warning	0	3	3
	Will click later	2	0	2
	Total	5	3	8

Table 13: Number of subjects for the relationship between urgency and victimisation. Only potential victims ( $N=8$ ) are included.

#### 4.3.3 Reading patterns

Most subjects ( $N=21$ ; 87%) read the email from top to bottom, although the starting point of reading differed. In contrast, three subjects used a different method of reading the email: one started reading the title, continued with the sender address and moved on the the valediction of the letter; the other two read the salutation, proceeded with the sender

address and then continued with the contents of the email. Out of the subjects who read the emails from top to bottom, the most popular point to start reading the email was the salutation ( $N=13$ ; 54%), followed by the sender address ( $N=9$ ; 37%) and the title ( $N=2$ ; 8%). All subjects read the entire email, but this may be due to the task description of ‘reading an email’. However, one subject indicated halfway in the third paragraph of the email that he did not feel like reading the email anymore, and if he would have received it on his own email account, he would stop at that moment. For this subject, the introduction to the call to action was too time consuming.

The fifteen subjects that decided to not comply gave several reasons for becoming alerted. Each subject was either vigilant or a potential victim that got warned, following the labelling from [Section 4.3.2](#). The subjects became alert due to heuristics being triggered, and the reasons they gave for being alerted are related to these heuristics. Three such reasons were given by the subjects: (1) The email is from a bank, therefore it is malicious ( $N=4$ ); (2) A request for taking action is suspicious ( $N=11$ ); and (3) Noticing the link “How to recognise phishing?” on the phishing website ( $N=1$ ). One subject mentioned several reasons. The subject who noticed the link on the website, clicked on it in an attempt to find more information. However, the link was fake and no additional information was shown.

Regarding the salutation of the email (“Dear customer”), it should be mentioned that in Dutch, there are two words that translate to ‘dear’ in English. One way to say “Dear” is to use the word “beste”, which is a more informal way of greeting someone in a written communication. This is the form that was used in the phishing emails. In contrast, the word “Geachte” is a more formal way to say the same thing. One subject commented on the use of the informal salutation in a communication from a bank, stating it is not appropriate. However, the subject kept reading and did not verbalise any negative thoughts on his security or on the believability of the email until the end of the email. In the post-task interview, the subjects were asked whether they remembered the exact wording of the salutation (i. e., formal or informal).

Regardless of whether the subject remembered the salutation correctly, they were asked whether they considered the salutation usual for an email from this organisation? Seven subjects (29%) had no idea what the salutation was. Seven other subjects remembered the informal salutation correctly, and two subjects (8%) found this salutation usual. Six subjects (25%) thought the salutation was the formal version of dear in Dutch, and five (21%) of them considered this a usual salutation. Finally, four subjects (17%) confused the salutation with (parts of) the title of the email (e. g., “Disruption of service”).

In anti-phishing warning campaigns in the Netherlands, such as the ones of the Fraudehelpdesk (2016) and Betaalvereniging (2016), users are advised to check the validity of the sender address, the validity of

any hyperlink appearing in the text, as well as to search for bad spelling or grammar and an impersonal salutation. Several subjects applied one or more of these tactics for checking the validity of an email. For example, two subjects checked the validity of the link in the phishing email. One subject did this by copying the link and manually pasting it into a browser. The second subject put his mouse on the link to see where the link was going to.

Spelling was another heuristic that the subjects applied. Even though both version of the phishing email were free of spelling errors, two subjects reported spotting spelling mistakes. Specifically, one subject saw a word starting with a capital letter in the middle of the sentence. The second subject saw a wrong determiner (*my* versus *mine*) where the correct one was used. Additionally, two other subjects reported positive thoughts regarding the spelling, stating that there were “*thankfully no spelling mistakes*” and “*I still see no spelling mistakes. This could very well be a legitimate email.*”

#### 4.4 CONCLUSIONS

The aim of this chapter was to find which heuristics people use when making a decision to take action upon a phishing email. We analysed heuristics by means of a experiment in which subjects had to verbalise their thoughts. This think aloud experiment was designed in such a way that subject's security awareness was not triggered by the researchers. Therefore, any thoughts on the authenticity of the email, or the security of the subjects themselves, follow from their own level of awareness and not from the experimental design. Out of these nine coded attributes that the subject's had thoughts on, only four attributes were used to label 88% of the thoughts. These four attributes were believability of the contents, ability of the organisation, security and efficiency. Furthermore, the majority of all thoughts showed distrust (60%), a quarter was neutral and only 15% indicated trust.

Half of the subjects were shown a ‘normal’ phishing email, the other half got the same phishing email, but with urgency clues added. The group of subjects reading the urgent email had more thoughts and made more decisions compared to the group reading the non-urgent email. Urgency cues in the phishing email lead subjects to be less negative about the email, and in particular the believability of the contents. Furthermore, the urgency cues triggered three subjects into ignoring the warning message after having clicked on the phishing link. In comparison, subjects from the non-urgent group who received a warning message, stopped and cancelled their actions. The effectiveness of urgency cues supports related literature (Wang, Herath et al., 2012; Vishwanath et al., 2011; Cialdini, 2006), stating that the likelihood of victimisation increases when urgency cues are provided. Additionally, Vishwanath

et al. (2011) stated that subjects would consider other attributes (such as spelling) less often when urgency is introduced. This effect is not visible in our data: subjects do consider most variables, only their feeling (i. e., positive or negative) changes upon the introduction of urgency.

Subjects who realised that the email they were looking at was phishing, did so at different stages. For example, after clicking on the phishing link, the subjects were presented a warning, asking to continue. The more often the subject decides to continue, the less vigilant we consider him/her. This results in a scale, ranging from being vigilant, through potential victim to becoming a victim (see Figure 15). The level of vigilance determines the moment a subject decides to stop. Because it is impossible to be always vigilant, heuristics are used. Therefore, the moment of becoming vigilant is determined by the heuristics that the subjects use.

Already in 2006, Downs, Holbrook and Cranor (2006) reported on an experiment where the decision making of phishing victims was qualified. We extend this study by following a strict thinking out loud protocol, without interaction between the interviewer and the subject. Downs, Holbrook and Cranor (2006) found three strategies that were used by their subjects: (1) Judge personalisation and professionalism; (2) whether the communication is expected or not; and (3) the reputability of the sender. Compared to our results, strategy 1 (personalisation and professionalism) was not commonly used. Only three subjects verbalised thoughts on either layout or spelling, and the ones that did, only sparsely mentioned them. The second strategy, i. e., whether the communication is expected, was not explicitly measured. However, several subjects mentioned that if they were to have an account at the specific bank, they would take action. The third strategy stated that reputable companies will use email. Our subjects had many thoughts on the bank that supposedly sent the phishing email. Furthermore, receiving an email from a bank was not considered unusual by the subjects.

In conclusion, subjects who were reading a phishing email without prior priming had three main thought patterns. Firstly, the subjects assessed the believability of the contents rather than searching for technical evidence of authenticity, such as the link location, sender address or email headers. The second pattern indicated that the subjects related the contents of the email (i. e., the story and accompanying request) to their expectation of how the supposed sender (bank) would write an email. Finally, introducing urgency changed the way the subjects interpreted the outcome of their heuristics. Overall, the subjects' thoughts become less negative of nature, and more subjects ignore warning messages as a consequence of perceived urgency.

Future studies could use the results of this study on a larger scale, with more subjects and a variation of legitimate and phishing emails. Furthermore, it would be interesting to study how a users' profession

relates to the heuristics this person develops. Finally, besides of measuring current heuristics that people employ, measuring a change in heuristics after training or by using software, could be studied in future work.

Having measured phishing heuristics of potential victims, we now turn to intervening with the decision making by means of training.



Fraudsters use phishing to convince victims to give out personal information. Commonly, the fraudsters want credentials that are used to access online services, such as online banking. Even though the impersonated brands that are misused in phishing are predominately financial institutions and payment providers, there has been a recent shift towards retailers and service-oriented companies (Anti-Phishing Working Group, 2015b, 2014c). Several countermeasures are currently in use to prevent phishing victimisation: blocking phishing messages and websites, improving interfaces, and training users (Hong, 2012).

Many training programs have focused on adults (e. g., (Jagatic et al., 2007; Mayhorn and Nyeste, 2012; Blythe, Petrie and Clark, 2011; Alnajim and Munro, 2009)). An often overlooked group of potential victims is children, with data about children only sparsely available (e. g., in Kumaraguru, Sheng et al. (2010)). The current generation of children, sometimes referred to as the *digital generation* or *digital natives*, grew up with the internet. The phrase “digital natives” is being criticised (boyd, 2014), since being a child in this generation does by itself not result in being more digitally capable. Instead, there are lots of opportunities for children, as well as adults, to use technology. Indeed, by the age of nine, many European children have access to the internet (Haddon and Livingstone, 2012). Many of the internet services that adults use, such as social media, email, or online gaming, are used by children as well (Brady, 2010). A quarter of European children aged 9-10 and 73% of 13 to 14-year-olds have at least one profile on a social media website (Haddon and Livingstone, 2012). In the USA, 68% of teenagers aged 13-14 use social media (Lenhart, 2015). Children, and in particular teenagers, are very well represented on the internet, with 92% of American children (13-17 years) (Lenhart, 2015) and 60% of European children (9-16 years) going online daily (Haddon and Livingstone, 2012).

One might wonder why children are at risk. To illustrate why children could be targeted, consider the marketing domain. Marketers know that children have influence over what their parents buy and consequently target children in commercials (Calvert, 2008). In addition to marketing on tv, digital marketing offers even more chances to target children specifically (Calvert, 2008; Montgomery et al., 2012). Phishing is commonly thought to be equivalent to theft of credentials

<sup>6</sup> This chapter is based on the paper “How Effective is Anti-Phishing Training for Children?” (Lastdrager, Carvajal Gallardo et al., 2017) which was published in the Proceedings of the Thirteenth Symposium on Usable Privacy and Security (soups) and won the Distinguished Paper Award.

of financial institutions. Since children often don't participate in online banking, what makes them attractive to a phisher? The online footprint of children on social media, websites, and email can be a target by itself. Obtaining access to email or social media accounts is valuable in order to access to a victim's network of friends and family. A phishing message that is sent by a friend is more likely to be opened than one from a stranger (Jagatic et al., 2007). Subsequently, both children and adults within the victim's network can be approached with personalised phishing messages. Alternatively, influencing a child to provide the personal information of his or her parents provides helpful information for a follow-up call or email, even with simple pieces of information such as a phone number or home address. Training is needed to reduce the risk of initial victimisation. Just like adults, children need to develop the ability to identify fraudulent communication, such as phishing emails.

Anti-phishing training can be administered in various ways. Advice can be given on an individual level, such as parents teaching their child how to ride a bike. Alternatively, one may educate a group at the same time; for example, schools teach skills like arithmetic to entire classes. When possible, educating a group of children can be more efficient. Since most children attend school, they are used to getting information in a class setting. Furthermore, when parents are insufficiently experienced to educate their children in the area of cybersecurity, this topic should be taught at school.

Education tackles only a part of the problem. An important issue is knowledge retention. One of the difficulties with user training is the extent to which the audience remembers the lessons over the long term. Retention indicates the effectiveness of training. Additionally, it is important to know how often to repeat training. This is true for traditional training, as well as alternative methods of creating user awareness, such as training by playing games (Kumaraguru, Sheng et al., 2008; Sheng, Magnien et al., 2007). Studies performed on adults found no significant decay in performance from one week up to one month after the intervention (Kumaraguru, Sheng et al., 2010; Kumaraguru, Cranshaw et al., 2009; Mayhorn and Nyeste, 2012; Alnajim and Munro, 2009; Kumaraguru, Rhee, Sheng et al., 2007). This suggests that improvement of awareness after training is retained in the relatively short term. The question arises whether the same applies to children, as well as, more importantly, whether the improvement in awareness is stable over a longer period.

Children are very active online and can be the target of phishing e-mails. Accordingly, like adults, they should be trained to reduce the risk of victimisation. This raises three questions to be answered. Firstly, what are children's abilities to detect phishing emails and websites? Secondly, what effect does cybersecurity training have on the children's ability to detect phishing? Thirdly, after receiving an awareness training,

how well do children retain this knowledge? To answer these questions, we conducted empirical research.

Our contributions are: (1) to our knowledge, we are the first study to focus on the effect of anti-phishing training on children; (2) the training was based on storytelling and resulted in an improved detection of phishing in the short term and an improved detection of legitimate messages after 2–4 weeks; (3) we show that subjects with more online exposure, as well as older children, score better on a phishing identification test.

## 5.1 METHODOLOGY

An experiment was conducted at six schools in the Netherlands, using a cybersecurity training program that was designed for children aged 9–12. We tested their ability to recognise phishing and measured the effect of an intervention.

### 5.1.1 *Design & Concepts*

The experiment used a 2x2 between-group design. The training intervention was given on a group level (i. e., in a classroom), and we wanted to preserve the anonymity of the pupils. Therefore, no identifying information was recorded on the tests. Consequently, we did not record demographic data other than age and sex. The independent variables were the experimental condition (intervention or control) and the retest duration (measured in number of weeks). The outcome variable is the score on the test, ranging from 0 (no correct answers) to 10 (all answers correct). Five other variables were recorded to identify differences between groups and measure for certain individual differences: sex (male/female); age; possession of email address (yes/no); possession of a Facebook account (yes/no); and whether the subject had received a phishing email before (yes/no/unknown).

We will briefly discuss why these variables were included. Firstly, the subject's sex (male/female) was recorded because several phishing studies found that men are less prone to phishing victimisation than women (Jagatic et al., 2007; Kumaraguru, Sheng et al., 2010; Sheng, Holbrook et al., 2010; Blythe, Petrie and Clark, 2011), though other studies found no relationship (Leukfeldt, 2014; Alseadoon, 2014; Dhamija, Tygar and Hearst, 2006). Age was recorded with the expectation that older subjects would outperform younger ones (Kumaraguru, Sheng et al., 2010; Sheng, Holbrook et al., 2010; Alseadoon, 2014). Finally, the Routine Activity Approach states that for a crime to occur, a target and an offender must converge in the absence of a capable guardian (Cohen and Felson, 1979). Consequently, we expected children who are more active online to be more exposed to phishing. Therefore, subjects

were asked whether they possess their own email address and Facebook account, and whether they have received a phishing email in the past.

In this paper, we use the terms “children” or “pupils” interchangeably to refer to the subjects of the study. “Teacher” refers to the school teacher of the pupils. The trainer is a researcher performing the study (by giving the presentation).

To establish the effectiveness of the cybersecurity training, we formed two types of groups: *intervention* and *control*. The intervention group was made up of school classes that received the cybersecurity training, followed by a capability test. To evaluate the effectiveness of the training, we compared the intervention group with a control group that received training after the study was finished (see [Section 5.1.2](#)).

#### 5.1.1.1 *Training and Procedure*

A cybersecurity training program was developed for this experiment, consisting of an interactive presentation and a test. During the 40-minute presentation, the trainer would introduce and discuss cybersecurity with a class of pupils. The presentation (in Dutch) is included in [Appendix C](#). The trainers were researchers and master’s students specialising in cybersecurity. Asking children for their attention during a presentation can be challenging. Storytelling is an efficient method for non-experts to share in an expert’s knowledge (Rader, Wash and Brooks, 2012). Therefore, the trainer used short stories and examples focussed on children to attract their attention.

The presentation provided the children with the necessary means of recognising cyber misbehaviour and advice on what to do. Topics included cyberbullying, hacking, phishing and identity theft. For phishing, we first explained what phishing is. Then, we showed an educational TV commercial that had been designed by the Dutch banking association (Veilig Bankieren (Dutch Banking Association), 2011). Following the commercial, we asked the children in a group discussion what clues one should look for. Afterwards, we introduced four clues for identifying phishing emails: (1) how to find a URL from a hyperlink and how to assess where a URL leads to; (2) grammar, spelling, and the general type of language used; (3) presence of a sense of urgency or use of threats; and (4) the sender address. Furthermore, we showed two clues for websites: (1) the URL and (2) the need for an HTTPS connection when entering any data. During the training, the children were given ample opportunity to tell about their experiences, which helps the attendees remember the message. This led the children to share their own advice on how to prevent victimisation, along with the advice that was included in the training. The trainer informed the children about the effectiveness of their own advice. Where needed, alternative advice was provided.

During the experiment, researchers went to schools in pairs. There were several practical constraints in time and availability. For example, schools had to book time to receive us, so there was a strict requirement to finish in time. Within classes of the intervention group, the trainers gave a presentation to the pupils. After the presentation, the children were given a paper-based phishing awareness test. Classes in the control group were only given the phishing test. No further explanation was provided, other than that the trainers would be back at a later time. Some pupils asked questions about a particular part of the test. The trainers answered that the pupil should pick the answer that made the most sense to the pupil.

After several weeks, each class was visited again. All pupils were given another paper-based phishing test. Finally, each child was given a one-page debriefing letter that explained and summarised the study. Additionally, all subjects were encouraged to discuss the test with their parents and contact one of the researchers with any questions.

#### 5.1.1.2 *Testing*

Establishing the ability of children to detect phishing was measured using a paper-based phishing test. The participating schools did not have a computer available for each pupil. To allow school participation with the least effort, we chose a paper-based test over a computer-based test. The method of testing phishing ability and the introduction to the test can influence the results. For example, Parsons et al. (2015) have shown that primed study participants are significantly better at discriminating between phishing and non-phishing compared to uninformed participants. To reduce this bias, children were not told that the goal was to discriminate phishing from non-phishing. Rather, the test was introduced as a 'cybersecurity test.'

The phishing test consisted of 10 questions, with six emails and four websites to judge. Both legitimate and phishing emails and websites were included. One correct answer was worth a point, and number of correct answers was the student's score on the test. Answering everything wrong would give a score of 0; answering everything correctly gave a 10. For each email or website in the test, a decision had to be made whether or not to take action. Although it was not stated explicitly, the pupils made a phishing or not phishing decision. Participating pupils were asked to note what kind of action they would take. Subjects' scores can vary depending on the type and origin of emails they have to judge (Parsons et al., 2015). Therefore, diversity in the types of emails and websites is essential to obtain a valuable result. Each question contained a clue as to why it should or should not be trusted. Some clues were explicit, such as a wrong link in an email or an unusual sender address. Others were based on the content, such as expressing urgency and spelling errors. For content-based clues, we

made sure to include several in an email or website. All clues were mentioned in the training. The questions, emails, and websites were tailored to children and included a variety of different companies, such as toy stores, TV programs, game websites, a bank, and social media. The questions were not based on real-life phishing emails, since we are unaware of phishing attacks that target children specifically. However, we used existing legitimate emails and websites and adapted them, just like a phishing offender would do.

The tests were aimed at measuring the ability to identify emails and websites as phishing or legitimate correctly. However, using the same phishing test for the initial measurement as well as the re-test could result in the subjects remembering the questions. To avoid this memory effect, three sets of questions were used to measure the ability of children to detect phishing emails and websites. Three versions of the test were made: A, B, and C. Tests A and B included a front page with questions about the online exposure of the subjects. Test C was used in the pilot phase of the experiment and contains reordered questions from Test A.

Each subject got an overall score, the outcome variable. However, human beings generally assume that a message is truthful, and have great difficulty recognising lies (Levine, Park and McCornack, 1999). This has been called the truth bias (Kahneman, 2012; Levine, Park and McCornack, 1999; Burgoon and Levine, 2010). We need to consider two parts in the subjects' performance: detecting lies (phishing) and detecting truth (legitimate). To do so, we made two equal-sized sets of questions. One set contained phishing, the other contained legitimate communications. By separately grading both sets of questions, we could distinguish between the ability to detect lies versus the ability to detect the truth. The overall score of a subject was calculated as the sum of both sets. For example, if a subject scored 3.0 out of 5 for recognising phishing, and 2.5 out of 5 for recognising a legitimate communication, the overall score would be 5.5 out of 10.

#### 5.1.1.3 *Retention*

To measure knowledge retention, each school class took two phishing tests to test their ability to recognise phishing over time. Classes in the intervention condition received the training, followed by a test. Immediately after groups in the intervention condition finished their tests, the correct answers were discussed in class. This allowed the children to ask questions once more and get feedback on their decisions, thereby increasing the learning effect. After either 2 weeks (14 days), 4 weeks (28 days), or 16 weeks (64 days) a second test was done. Classes in the control condition did one test initially, followed by a re-test after 2 or 4 weeks. For the control condition, the results of the tests were not discussed in class. Unfortunately, classes in the control group that were scheduled for a re-test after 16 weeks were unable to participate

the second time. This makes it impossible to compare the intervention group with a control group at 16 weeks. Therefore, our analysis will focus on the retention between 0 and 4 weeks.

### 5.1.2 *Ethics*

As with any experiment with humans, ethics are important. First of all, the design of this study was approved by the institutional review board of the University of Twente. The study was designed such that the subjects were not hurt or distressed in any way. Furthermore, each participating school was asked for permission to conduct the training and test their pupils. Additionally, we asked each participating school to distribute informed consent letters to the parents of their pupils. Parents were asked to sign and return the informed consent, either to the school or by email to the researchers. The contact information of the researchers was included in the informed consent, in case parents had questions. Several parents contacted the researchers. Only when the parents of a pupil had signed the informed consent and returned this to the school could a child participate as a subject.

After finishing the experiment, each subject was given a debriefing letter. The letter was written for the child and encouraged him or her to discuss the training with his or her parents. Furthermore, the contact details of the researchers were included in the debriefing, in case anyone had questions. After finishing the experiment, nobody contacted the researchers with questions.

From the point of view of the experiment, it was important to separate intervention and control groups. We considered it unethical to deprive subjects in the control group of a cybersecurity training. Therefore, after finishing their second phishing test and concluding their participation as subjects, pupils in the control group received the training too.

### 5.1.3 *Setting*

The experiment was held at six schools in the Netherlands, of which five primary schools and one secondary school. Each participating school gave permission for two sessions for at least one class. Every class received two tests (of 20-30 minutes each), and one intervention (about 40 minutes). Classes were randomly assigned to either an intervention group or a control group, and were additionally assigned a retention period by the researchers. All tests were taken individually by the subjects. The researchers were present to answer questions, but would never give away the correct answer. The subjects were told to answer what they would do if they had received the email or visited the website. [Figure 16](#) shows setting of the experiment.





Figure 16: Children doing the cybersecurity test. *Photo courtesy of Brinda Hampiholi.*

#### 5.1.4 Subjects

The subjects were 353 pupils from six participating schools. All subjects were aged between 8 and 13 ( $M=10.66$ ;  $SD=1.05$ ), and over half (54%) were female. Children could join the training only if their parents had given their written consent before the start of the program (refer to [Section 5.1.2](#) for more information). Children who did not have permission from their parents were temporarily sent to another classroom. If changing rooms was not possible, non-participating children were moved to another part of the same classroom to work on another task. Each child was assigned to an intervention or control group, based on the class they were in. This resulted in 181 children in the intervention group who received training, compared with the control group consisting of 172 children. The re-test was taken by 177 children. We included the week 0 data for several classes that were unable to participate for the re-test. Specifically, the missing classes consisted of all control group classes for the 16-week re-test. This resulted in the exclusion of the 16-week intervention group's re-test, since we could not compare them with their control group counterparts. Therefore, the number of subjects in week 0 is significantly higher compared to those for the re-tests in weeks 2 and 4. The exact number of subjects at each stage in the experiment is listed in [Table 14](#).



Group	Week 0	week 2	week 4
Intervention	181	49	38
Control	172	32	58

Table 14: Number of subjects in each stage of the experiment.

### 5.1.5 Analysis

The three research questions guided the analysis. Descriptives of the control groups provided an answer to the first research question (i. e., what are the children's abilities to detect phishing emails and websites?). Furthermore, we tested whether the subject's characteristics influenced the score. An independent group t-test was used to measure the effect of the subject's sex and possession of an email account. The second research question was: what effect does cybersecurity training have on children's ability to detect phishing? To measure this effect, we compared the intervention group and the control group at 0 weeks. This was done using an independent group t-test, showing the difference between trained children (the intervention group) and untrained children (the control group). The third research question quantified the retention of the training. To answer this question, several linear regression models were developed. Firstly, a multi-level model was tested, measuring whether the school attended by the subject accounted for the results of the pupils. Even though the multi-level model was significant, the intraclass correlation was low (i. e., below 0.025). Therefore, linear regression was used instead. We developed several such models.

Model I uses the type of experiment (i. e., intervention or control), the number of weeks, and the interaction of these two as the predictors. ExperimentType shows the effect of the training on the score. The number of weeks indicates retention over time. Additionally, it is interesting to learn whether the effect of the training increases or decreases over time. For example, teaching someone a skill such as biking results in a higher level of skill over time if the person practices on his or her own. Therefore, the interaction between having participated in the intervention and the number of weeks (ExperimentType  $\times$  Weeks) was taken into account as well. With this interaction, we could analyse whether the intervention resulted in better results as time progressed. A second model including social variables was constructed as Model II. Age and sex were added to the variables from Model I. Age was included since related literature suggested that older subjects score better than younger ones. The literature is inconclusive when it comes to sex and phishing victimisation. Therefore, we added sex as a variable. Finally, Model III combines Models I and II and adds the test version and school, to show their potential influence on the overall score of the subjects. The school

and test version variables were moderately correlated ( $r=0.68$ ), as a consequence of Test c being used only in the pilot of the study. This results in collinearity in the model. Therefore, we omitted Test c from the model. These three models were used to predict the subject's overall scores on the tests.

Using the overall score as a measure of the ability to recognise phishing from legitimate is by itself insufficient. As discussed before, one needs to distinguish the differences in the scores of recognising phishing and recognising legitimate communications. To accommodate this, additional models were developed to distinguish lie detection and truth detection in the analysis. This led to the introduction of six models. Phish-I through Phish-III were based on the previously described models I-III, but used the phishing (lies) score instead of the overall one. Additionally, Legit-I to Legit-III were developed to model the scores of the legitimate (truth) questions.

## 5.2 RESULTS

The first research question concerned the ability of children to detect phishing. This translates to the scores of the control group at the beginning of the experiment, at week 0. The average overall score of this control group is a 6.02 (Table 15) on a scale from 0 to 10. The overall score consisted of two parts: phishing (0–5 points) and legit (0–5 points). When considering only the questions that were related to phishing, the control group scores 3.74 on average, with a 95% confidence interval of [3.62, 3.88]. The mean score for labelling legitimate questions as such was lower: 2.26 (95% CI [2.09, 2.44]). In addition to the average scores of the control group, we also measured the effects of several subject characteristics on the overall score for all subjects. There was no significant effect of sex on the score, indicating a lack of evidence that boys performed differently from girls ( $t(633) = -0.62$ ,  $p=0.53$ ). There was a significant effect of age on the score, with older pupils scoring higher than younger ones ( $F(1,633) = 6.28$ ,  $p=0.01$ ,  $R^2=0.010$ , Adj.  $R^2=0.009$ ). The effect of the school on the subject's score was significant ( $F(5,636)=7.54$ ,  $p<0.001$ ,  $R^2 = 0.056$ ). One school scored significantly lower compared to the others ( $B=-0.80$ ;  $p=0.004$ ). Most of the subjects (80.3%) indicated having their own email address. Having one's own email address significantly influenced the score, with subjects having their own email address performing better than those without ( $t(469)=3.68$ ,  $p<0.001$ ). On the topic of social media, 26.6% of the subjects indicated having their own Facebook profile. Subjects with their own Facebook profile scored significantly higher than those without a Facebook profile ( $z=2.330$ ,  $p=0.02$ ,  $r=0.10$ ). Thirdly, when asked whether they had ever received a phishing message, 8.9% answered 'yes', 37.4% answered 'no' and the remaining 53.7% responded that they did not know. Whether

or not the subjects received a phishing email before was not significantly related to the subject's score ( $F(2, 468) = 0.61, p=0.55$ ). A subject's online exposure did result in higher odds of having received a phishing message before ( $F(2,215) = 6.25, p=0.002, R^2=0.040$ ), whereby having an email address was a significant indicator ( $B=0.16, SE=0.05, p=0.04$ ).

To answer the second research question, the effect of the training was measured. Since three paper-based phishing tests were used in the experiment, we wanted the results to be comparable regardless of the version of the test. The mean overall results of pupils taking different tests were not significantly different from each other: A and B ( $t(470)=1.89; p=0.059$ ); A and C ( $t(307)=0.98; p=0.326$ ); B and C ( $t(451)=-1.214; p=0.225$ ). Figure 17 shows the differences in scores in three box plots. The means and confidence intervals under all experimental conditions are listed in Table 15. The training itself resulted in an improvement in the scores of the participants in the intervention group that was statistically significant compared to the control group ( $t(634)=-10.56, p<.001$ ). The effect size was  $r=.39$ , indicating a medium-sized effect (Cohen, 1992). In comparison, if we include only the first measurement (i. e., week 0), there is a significant difference between the untrained and the trained children as well ( $t(351)=-5.19; p<0.001$ ). The training in week 0 had a small effect size of  $r=.27$ . These results show the effectiveness of adding a simple and short cybersecurity training to the curriculum of schools.

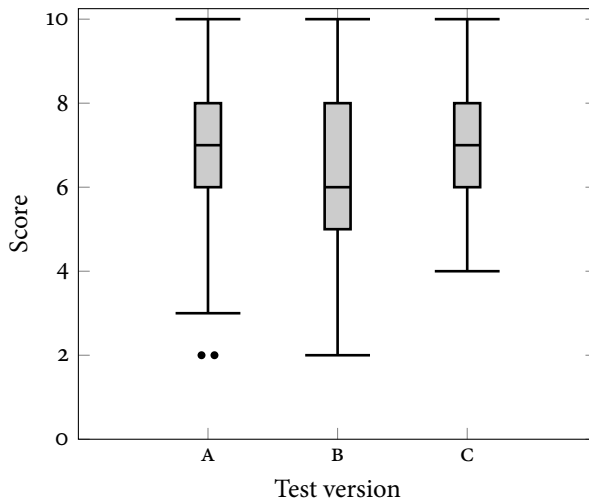


Figure 17: Box plot of three phishing tests of all observations ( $N=636$ ).

To answer the third research question, retention over time was measured. Several linear regression models were constructed, the results of which are included in Table 16. Model 1 shows the influence of the cybersecurity training intervention on the score, as well as the effect

Type	Week	Overall Score		Phishing Score		Legitimate Score	
		Mean	95% CI	Mean	95% CI	Mean	95% CI
Cont	0	6.02	5.79–6.26	3.61	3.45–3.77	2.41	2.20–2.62
Exp	0	6.87	6.65–7.09	4.26	4.15–4.38	2.61	2.41–2.80
Cont	2	5.72	5.21–6.23	4.09	3.74–4.45	1.62	1.17–2.08
Exp	2	7.95	7.58–8.34	4.33	4.12–4.53	3.63	3.28–3.99
Cont	4	6.14	5.75–6.53	3.96	3.70–4.23	2.17	1.79–2.55
Exp	4	8.13	7.67–8.60	4.00	3.73–4.27	4.13	3.81–4.46
Cont	all	6.01	5.82–6.20	3.74	3.62–3.88	2.26	2.09–2.44
Exp	all	7.35	7.19–7.51	4.23	4.15–4.32	3.11	2.97–3.26

Table 15: Mean score and 95% confidence interval per experimental setting.

over time, while controlling for the interaction effect. The resulting Model I is significant and explains 18.6% of the variance ( $F(3,526) = 41.77, p < 0.001$ ). Model II adds social predictors to Model I, resulting in a model that explains 19.8% of the variance ( $F(5,523) = 27.63, p < 0.001$ ). Finally, Model III includes the school as well as the version of the test, as well as the predictors from the other models. Model III is significant and explains 25.7% of the variance ( $F(11,517) = 17.46, p < 0.001$ ). In all three models, the effect of training significantly influenced the score of the subjects throughout the following weeks ( $\beta = 0.23, p < 0.001$ ). Furthermore, the intervention group score significantly higher over time compared to the control group. Figure 18 plots Model III based on the number of weeks passed, split into intervention or control group, to show these effects visually.

To measure the differences in detecting lies from detecting truth, we developed additional models based on Models I, II and III. Instead of using the overall score as the outcome variable, we used the phishing score or the legitimate score, respectively. Since half of the questions were phishing, the scores range from 0 (all wrong answers) to 5 (all correct). Models Phish-I to Phish-III use the score of recognising phishing. The model results can be found in Table 17. Model Phish-I includes the same predictors as the normal Model I, and is significant and explains 8.3% of the variance ( $F(3,526) = 15.36, p < 0.001$ ). Model Phish-II is significant and explains 8.3% of the variance as well ( $F(3,523) = 9.26, p < 0.001$ ). Model Phish-III is significant as well and explains 13.1% of the variance ( $F(11,517) = 9.60, p < 0.001$ ). Compared to the models of the overall scores, different effects emerge. For example, subject age and weeks since intervention in Phish-III are not significant, whereas they are in the overall Model III. The differences are more easily viewed when Model Phish-III is plotted in Figure 19a. At week 0, the intervention group's scores differ significantly from the control group, as shown by

	Model I			Model II			Model III		
	B	SE B	β	B	SE B	β	B	SE B	β
Characteristic (reference)									
ExperimentType (control)	0.92***	0.16	0.28	0.90***	0.16	0.27	1.00***	0.12	0.10
Weeks	0.01	0.06	0.01	0.03	0.06	0.03	0.11	0.12	0.10
Weeks × ExperimentType	0.34***	0.08	0.23	0.36***	0.08	0.24	0.30**	0.08	0.20
Age				0.18**	0.06	0.11	0.19**	0.07	0.12
Sex (female)				0.08	0.13	0.02	0.14	0.14	0.04
Test version (A) <sup>†</sup>									
- Test B							-0.17	0.39	-0.05
School (1)									
- 2							0.89**	0.33	0.16
- 3							0.44	0.31	0.08
- 4							-0.33	0.34	-0.05
- 5							0.30	0.43	0.07
- 6							-0.24	0.47	-0.07
Constant	5.99***	0.11		4.04***	0.69		3.80***	0.85	
R <sup>2</sup>		0.186			0.198			0.257	
Model significance		0.000***			0.000***			0.000***	
N		530			529			529	

Note. Coefficients unstandardised (B) and standardised (β). SE=Standard Error. Significance (X<sup>2</sup>): \* p<0.05; \*\* p<0.01; \*\*\* p<0.001. <sup>†</sup>Due to collinearity, the output of test C was omitted.

Table 16: The linear regression models of the overall score.

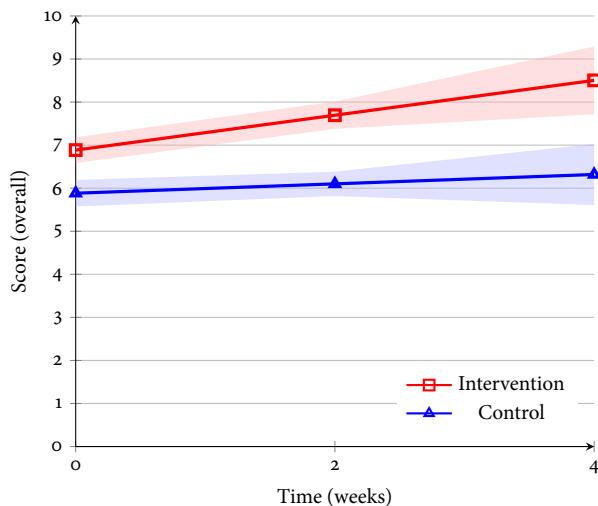


Figure 18: Overall predicted ability scores over time, in number of correct answers (0–10). Shades indicate 95% confidence interval.  $N=529$ .

the confidence intervals. However, in week 4, there is no significant difference between the intervention group and the control group anymore. The control group scored similarly in week 4 compared to week 0. Subjects within the intervention group scored significantly lower in week 4 compared to week 0.

In addition to the three phishing-only models, three legit-only models were constructed. Similarly, three models, Legit-I to Legit-III were constructed based on the overall Models I to III, respectively. The results of these models can be found in Table 18. Model Legit-I was significant and explained 15.1% of the variance ( $F(3,526)=42.57$ ,  $p<0.001$ ). Model Legit-II was significant and explained 16.4% of the variance ( $F(5,523)=29.59$ ,  $p<0.001$ ). Model Legit-III was significant and explained 26.0% of the variance ( $F(11,517)=20.28$ ,  $p<0.001$ ). A graph showing Model Legit-III is included in Figure 19b, with scores ranging from 0 to 5 for all five questions testing legitimacy. There are no significant differences in score at week 0 between the intervention group and the control group for the legitimate scenarios ( $z=-1.17$ ;  $p=0.24$ ). In week 4, however, the scores of the intervention group and control group differ significantly ( $z=-5.85$ ;  $p<0.001$ ). During the experiment, the score of the control group did not change significantly ( $t(228)=1.11$ ;  $p=0.27$ ). In the intervention group, a significant increase in score was observed between week 0 and week 4 ( $z=-6.05$ ;  $p<0.001$ ).

	Model Phish-I			Model Phish-II			Model Phish-III		
	B	SE B	β	B	SE B	β	B	SE B	β
Characteristic (reference)									
ExperimentType (control)	0.65***	0.10	0.34	0.65***	0.10	0.34	0.70***	0.10	0.37
Weeks	0.10**	0.04	0.16	0.10**	0.04	0.16	0.02	0.07	0.04
Weeks × ExperimentType	-0.15**	0.05	-0.18	-0.15**	0.05	-0.17	-0.18**	0.05	-0.22
Age				0.01**	0.04	0.01	0.05	0.04	0.05
Sex (female)				-0.00	0.08	-0.00	0.09	0.08	0.05
Test version (A)†									
- Test B							-0.21	0.24	-0.10
School (1)									
- 2							0.56**	0.21	0.17
- 3							0.08	0.21	0.03
- 4							0.20	0.22	0.06
- 5							0.95**	0.27	0.41
- 6							0.77**	0.29	0.40
Constant	3.63***	0.08		3.50***	0.44		2.59***	0.52	
R <sup>2</sup>		0.083			0.083			0.131	
Model significance		0.000***			0.000***			0.000***	
N		530			529			529	

Note. Coefficients unstandardised (B) and standardised (β). SE=Standard Error. Significance (χ<sup>2</sup>): \* p<0.05; \*\* p<0.01; \*\*\* p<0.001. † Due to collinearity, the output of test C was omitted.

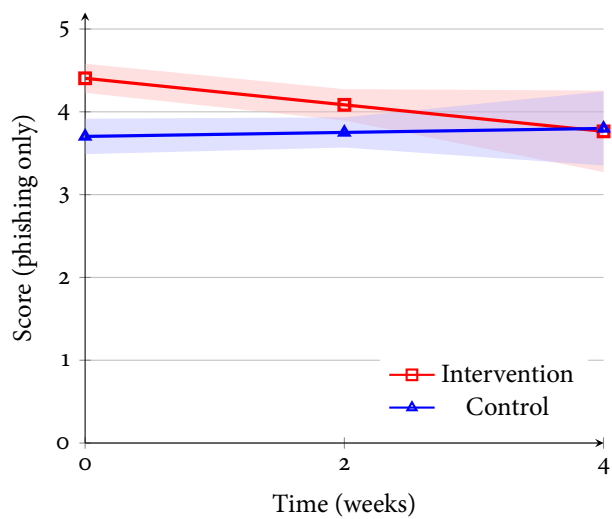
Table 17: The linear regression models of the phishing-only score. The construction of the models is similar to Table 16.

Characteristic (reference)	Model Legit-i			Model Legit-ii			Model Legit-iii		
	B	SE B	β	B	SE B	β	B	SE B	β
ExperimentType (control)	0.27	0.14	0.09	0.25	0.14	0.09	0.30*	0.14	0.10
Weeks	-0.08	0.05	-0.09	-0.07	0.05	-0.07	0.08	0.11	0.09
Weeks × ExperimentType	0.49***	0.07	0.38	0.51***	0.07	0.39	0.48***	0.07	0.37
Age				0.17**	0.06	0.11	0.14*	0.06	0.10
Sex (female)				0.08	0.12	0.03	0.05	0.12	0.02
Test version (A)†									
– Test B							0.04	0.35	0.01
School (1)									
– 2							0.33	0.29	0.07
– 3							0.36	0.26	0.07
– 4							-0.54	0.29	-0.10
– 5							-0.65	0.40	-0.18
– 6							-1.02*	0.41	-0.35
Constant	2.36***	0.11		0.54	0.62		1.21	0.74	
R <sup>2</sup>		0.151			0.164			0.260	
Model significance		0.000***			0.000***			0.000***	
N		530			529			529	

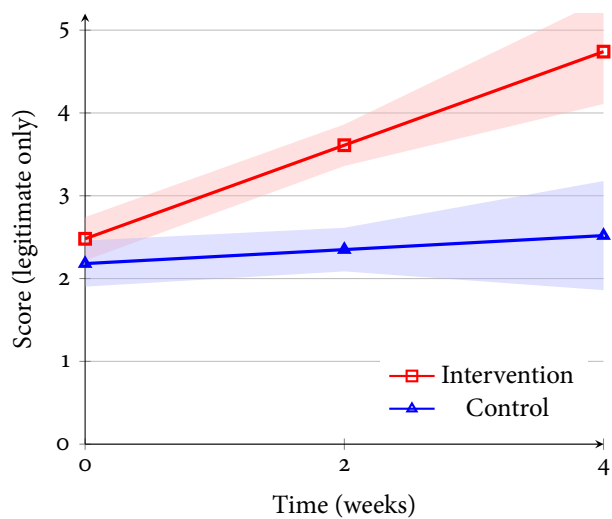
Note. Coefficients unstandardised (B) and standardised (β). SE=Standard Error. Significance (χ<sup>2</sup>): \* p<0.05; \*\* p<0.01; \*\*\* p<0.001. †Due to collinearity, the output of test c was omitted.

Table 18: The linear regression models of the legitimate-only score. The construction of the models is similar to Table 16.





(a) Includes only the phishing questions



(b) Includes only the legitimate questions

Figure 19: Predicted ability score split by phishing and legitimate. Shades indicate 95% confidence interval. N=529.

5.3 DISCUSSION

The concept of testing the ability to detect phishing in an educational setting is challenging (Robila and Ragucci, 2006). Getting the attention of children aged 8–13 to focus on cybersecurity is no less of a challenge. Untrained children are mediocre at discriminating phishing emails and

websites from legitimate ones, scoring 6.02 out of 10 in our experiment. However, subjects trained in a single 40-minute training session and interactive discussion scored 6.87 out of 10, an increase of 14% over their untrained peers. The overall score by itself is not sufficient as a measurement of accuracy, since humans are generally not very good at recognising lies (Levine, Park and McCornack, 1999). Therefore, we distinguished the correctness scores for phishing and legitimate questions.

We found that training improved the ability to recognise phishing directly following the training, but it did not significantly change the ability to identify legitimate emails correctly. This phenomenon has been discussed in the literature. Hauch et al. (2014) have shown in a meta-analysis that training improves both overall accuracy and lie detection, but not truth detection accuracy. This was also the case in our experiment; the subjects did not score significantly better on truth accuracy of legitimate emails and websites on the test directly following the training, compared to the control group. This can be explained by the focus of our training on how to detect phishing. According to Hauch et al. (2014), if the focus of training is on deception detection, the subject's post-training truth accuracy remains unaffected. An alternative explanation would be that the training made the subjects paranoid. However, if that were to be the case, the subjects would have to score lower on recognising legitimate emails, which was not the case.

The overall scores of trained subjects improved significantly over time, indicating a good knowledge retention of the subjects. Within the control group, the overall scores remained stable. When considering only the phishing questions, subjects from the intervention group suffered from a small decay in their ability to recognise phishing. Specifically, after 4 weeks, the ability of the intervention group to recognising phishing matched the level of the control group. Regardless of the decay over time, the scores on the phishing questions were relatively high, with averages of correct answers between 3.7 and 4.4 questions. Since 5 was the maximum, we believe that there is a ceiling effect: many subjects achieved the highest score, and could not improve their scores further. Our test consisted of 10 questions composed of two sub-tests, five legitimate and five phishing. This means that subjects could not receive higher scores than 5 on both sub-tests, which is the maximum on our measures. When many subjects have the maximum score, their scores cannot be distinguished. Figure 19b illustrates this clearly for the intervention group. Therefore, only less-performing subjects could increase their score after training. The subsequent score decay over time shows that the effect of the training, in terms of the ability to recognise phishing emails, fades within a month. To the best of our knowledge, no similar phishing tests have been undertaken with children, making comparisons with other phishing literature difficult. There are studies on phishing interventions with adult subjects, which found no signif-

ant decay of the trained subject's abilities after 7 to 28 days (Kumaraguru, Sheng et al., 2010; Kumaraguru, Cranshaw et al., 2009; Mayhorn and Nyeste, 2012; Alnajim and Munro, 2009; Kumaraguru, Rhee, Sheng et al., 2007). However, there are major methodological differences, since the abovementioned studies use interactive, computer-based methods of training, such as playing games (Kumaraguru, Sheng et al., 2010; Mayhorn and Nyeste, 2012; Kumaraguru, Rhee, Sheng et al., 2007) or roleplaying (Alnajim and Munro, 2009). However, within the field of social engineering, it has been reported that an intervention to increase awareness is subject to significant decay (Bullee, Montoya et al., 2016), showing social engineering awareness returning to pre-intervention levels after two weeks.

While the phishing score decreased slightly over time, the score for legitimate questions followed a rather different pattern. The score over time increased significantly, contrary to our expectations. After two and after four weeks, subjects in the intervention group were able to correctly recognise legitimate scenarios significantly better than subjects from the control group. The cybersecurity training may have triggered the interest of the children, causing them to pay more attention to messages they receive, or to think about the lessons learned. Another possible explanation is that the subjects trained themselves based on emails they received in their daily lives. This may be compared to learning how to ride a bike, where an initial set of skills and knowledge is needed to start biking, and with more practicing, performance increases over time. In other words, training made the children look more closely at the emails they received, after which they were better at identifying legitimate emails.

Further trainings, sometimes called boosters, could be used to increase these abilities and counter decay of the ability to recognise phishing (Purkait, 2012; Kumaraguru, Cranshaw et al., 2009). However, regular training is costly. In the context of children, it may be infeasible for schools to introduce boosters on a regular basis. This is especially the case when the retention of knowledge is short (i. e., a month). Training using different methods, such as letting the subjects play a game (Kumaraguru, Sheng et al., 2010; Kumaraguru, Rhee, Sheng et al., 2007), may be less affected by this disadvantage since the subjects can play the game regularly without supervision. Before introducing additional training, however, better measurements should be used to identify the problem better. One possible fix is an extensive test with more questions and more challenging questions, which could be used to avoid a possible ceiling effect. That way, subjects would be less likely to get the maximum score, and decay or increase effects should be more visible.

Another finding is that older children score better than younger ones. This is in line with similar studies about phishing interventions on adults. In several studies, young adults perform worse than older ones (Sheng, Holbrook et al., 2010; Alseadoon, 2014). In particular, a

large-scale study (Kumaraguru, Sheng et al., 2010; Sheng, Holbrook et al., 2010) found that teenagers between 13 and 17 perform worse than adults in phishing tests. A possible reason for this result is lower education and fewer years of internet experience (Sheng, Holbrook et al., 2010). Furthermore, subjects in this study who have their own email address or a Facebook profile scored significantly higher than other subjects. This suggests that, indeed, internet experience may be an influential factor. Another factor that could influence the subject's score is the training itself. Despite efforts to make all trainings similar, there are group dynamics involved, especially when relying on interaction (e.g., stories) with the subjects.

Other candidate relations did not significantly contribute to the final score of a child. In particular, the sex of the child had no significant influence on the score, when controlling for other variables. Specifically for children, sex differences are not necessarily to be expected at all. For example, boys only begin to take more risks than girls between the ages of 9 and 11 (Slovic, 1966). The lack of differences could be explained by the age groups of the children that participated. Additionally, even for adults and adolescents, the existence of a relation between sex and phishing knowledge is doubtful in existing literature (Leukfeldt, 2014; Alseadoon, 2014; Dhamija, Tygar and Hearst, 2006). The interaction between age and sex did not predict phishing knowledge of children either.

### 5.3.1 *Limitations*

There are several limitations to the results of this study. Even though the intervention condition was given per class, this did not prevent children in one class from talking to their peers in other classes. Since all parents were informed and asked for permission beforehand, they could have discussed the topic of cybersecurity with their children. Unknown external factors may be responsible for the increase over time. For example, the participating children may have seen one of the phishing awareness commercials on tv. Personal experience of the researchers was that indeed one of these three explanations was plausible. One of the colleagues at the University of Twente, who was not involved in the study, had a child in the intervention condition. The colleague mentioned that his children and the other parents were enthusiastic about the intervention and that he had talked about it at home. This example could explain the increase in ability over time that was observed. Moreover, this colleague had other children in the same school. Hence, the intervention could have influenced children in the control condition. However, we do not see indications of that effect in the data.

A possible critique on the study is that the children know that they are being tested. The results, therefore, do not necessarily reflect their ability when receiving an email in the wild. While this is true, we consider the tests an appropriate way to measure the subject's ability to recognise phishing. The subjects' scores are arguably different from how they would respond to a phishing email in their own inbox, since more factors are involved. Factors such as language (an eight-year-old Dutch child receiving an English email) and expectancy (not having a bank account) could increase their real world score. On the other hand, factors like attention (doing other things in parallel) and limited interfaces (not being able to check the link on a tablet computer) could affect resilience in the real world. Furthermore, the subjects received a second test a period of time after the first. This means that they know what to expect when they start the second test.

This study may suffer from an assignment bias. Even though the groups were assigned at random to one of the conditions, the number of schools that participated is limited. Furthermore, all schools are located in two cities in the east of the Netherlands. The results might be affected by factors unknown to the researchers. A nation-wide study on randomly selected schools could counter such biases regarding region and quality of teaching.

A presentation (or lecture) is one way to deliver a message to pupils. Other ways of teaching may be more efficient, such as using games (Domínguez et al., 2013). We chose a traditional presentation-based intervention because it is relatively simple to apply to current primary schools. The pupils do not need to have access to a computer, and a presentation and paper-based test fit in well with the rest of the daily program and activities. Alternatively, game-based anti-phishing solutions (Kumaraguru, Sheng et al., 2008; Sheng, Magnien et al., 2007) may yield better results and could have different retention properties.

Using a paper-based test with images raises questions regarding the representativeness of the resulting score compared to real-world phishing. Whereas using static images or screenshots is not optimal, they have been used before in phishing experiments (Sheng, Holbrook et al., 2010; Tsow and Jakobsson, 2007; Parsons et al., 2015). We believe there is little difference between seeing an image on a screen or seeing one printed on paper. Furthermore, not all subjects may be equally computer literate, and using static images on paper results in a level playing field.

Finally, all students filled in the tests anonymously. Therefore, no repeat measurements were available at an individual level. The analyses could therefore not be performed on repeated-measures samples. Rather, we treated the test results as independent samples. As a consequence, the reported results are conservative and an underestimation, as they miss the power of a repeated-measures test.

#### 5.4 CONCLUSIONS

Children need to understand digital risks to reduce the risk of victimisation on the internet. Understanding digital risks is important for children as well as adults. However, the majority of children are self-taught when it comes to the internet (Brady, 2010), making it unlikely they will systematically learn how to act safely. To learn about the abilities of children in detecting phishing emails and websites, researchers had children aged 8–13 take in a phishing recognition test. Half of the children received training before the test, and the other half did not. Both trained and untrained children were tested for the ability to distinguish phishing emails and websites from legitimate ones. Several schools participated in the study. A first indicator of the practical need for such training arose while performing the experiment. During the training, as more pupils started sharing their stories, they became very enthusiastic and asked lots of questions. In most classes, at least one child knew a phishing victim. These victims were mostly relatives or neighbours. The most common situation in the stories that were told was a victim losing money due to filling in banking credentials on a phishing website. Hearing stories from their peers impacted the children and provided them with a warning message stronger than the presenters could ever give.

Until novel anti-phishing techniques are developed and deployed on a large scale, user training seems to be important. For adults as well as children, that means creating an improved knowledge of the subject for as many individuals as possible. In many countries, all children aged 9 or older attend some form of education. Potentially, this makes it feasible to embed a cybersecurity training in their curriculum, effectively training the entire population of children.

In our experience, both schools and parents are very willing to embed lessons about cybersecurity in the curriculum. Our request to give a training was well received. In particular, incidents with phishing, cyberbullying, and other cyber-threats are often in the news. Teachers and parents reported being worried about those issues. At the same time, teachers at schools where we gave a training, found the course highly informative for themselves as well. Techniques for establishing the validity of an email were unknown to them. Several teachers mentioned that hovering over a hyperlink or checking the sender address were valuable approaches for them. Training teachers should, therefore, be the first step in cybersecurity education. Where needed, universities and practitioners (e. g., IT security firms) could provide help. There are existing initiatives, such as the (ISC)<sup>2</sup> Safe and Secure Online<sup>7</sup> where security professionals visit schools. Such initiatives should be exten-

---

<sup>7</sup> See also <https://iamcybersafe.org/>

ded to more countries and expanded in size, and new ones should be developed.

Training children increased their short-term ability to distinguish phishing from legitimate correctly. Specifically, their ability to recognise phishing increases significantly after an in-class training. However, this increased ability is subject to decay. After four weeks, the ability to recognise phishing for trained children diminished to the level of their non-trained counterparts. This suggests that the training created knowledge, but that this knowledge only lasted through the short term. On the positive side, trained children did continue to perform better in recognising legitimate emails as such. This increases the odds of legitimate communications reaching the end user. Increasing the ability to recognise phishing requires good awareness.

All in all, we believe that researchers and practitioners in the field of cybersecurity should not only focus on adults, but that material for children should be developed in parallel. Phishing, specifically, is too often seen as an adult-only crime. The children of today are the victims of the future.

Having performed user training and measured the outcome, we now focus on measuring phishing in the real world.





Phishing is all about (a lack of) vigilance (Chapter 4). The receiver needs to recognise the risk and decide to not follow up on the phishing message. Reducing phishing victimisation means increase the vigilance of the general population with regards to phishing. Recognising a well-known phishing message is relatively easy for an email user. However, sometimes phishing messages are new and do not look like previous phishing. The receivers could become confused, and in the worst case fail to recognise the message as a scam. A confused phishing message receiver could decide not to act and try to find more information before deciding, or acting upon the phishing message. Publicity campaigns could be used to inform people on this new scam and tell them what to do. In traditional crime such publicity is very common, with media campaigns by the police such as “lock your car door” or “thieves are operating in this area” (Bowers and Johnson, 2005). Another example is a TV show called “The Real Hustle” (Conran and Wilson, 2006), which showed how typical scams were performed, thereby informing the viewers how to prevent victimisation to these scams. Not learning the criminal’s new methods leaves a person more vulnerable and at higher risk of being victimised. The same is true for phishing, where new pretexts are common and different brands get misused for the purpose of victimising targets.

To inform the general public on new types of phishing messages, a good overview of ongoing phishing campaigns is required. This overview should consist of the phishing messages being used, something that continuous monitoring of phishing activity can provide. Furthermore, the impact of these messages on the recipients could be monitored in order to measure the extent of the threat. If potential victims are in doubt, or start asking questions, training or information should be provided to guide them into making the right decisions. Monitoring of phishing trends is therefore important. For traditional crime, monitoring police records can provide such an overview of crime trends. However, since cybercrime is under-reported at the police (Wall, 2010), other means of monitoring need to be used. One could set up a spam trap (honey pot) (Anderson, Fleizach et al., 2007), monitor spam filters (Zhang, Zhu and Yao, 2004), try to infiltrate botnets (John et al., 2009), or collect user reports (Moore and Clayton, 2008). For this chapter, we cooperated with an anti-fraud agency that encourages the general public to report

<sup>8</sup> Parts of this chapter are based the poster and extended abstract “Aplate: Anti-Phishing Analysing and Triaging Environment” (Lastdrager, Hartel and Junger, 2015) which was presented at the 36th IEEE Symposium on Security and Privacy.

phishing emails. This report-based method of monitoring is similar to the reporting of traditional crimes to the police. The advantage of this type of reporting is the influence that potential victims have on what gets reported, as well as any information they can provide on top of the email. For example, people may only have reported phishing emails that appeared in their inbox (i. e., that did not get filtered by their spam filtering). Furthermore, when reporting a *suspicious* email, they may include their thoughts, such as stating that they are unsure about validity of the email. The forwarded emails are the ones that trouble the people who report them. This additional information gives more insights on the impact of phishing than traditional spam traps or filters can provide.

A phishing email is generally not specific to its receiver. Instead, it belongs to a series of emails sent to a number of different email addresses. These emails are part of a phishing *campaign*. At any point in time, many different phishing campaigns are ongoing. In traditional crimes, there is often a non-random distribution in time and space for a crime to occur, as formulated in the crime pattern theory (Brantingham and Brantingham, 1993, 2008). Even though the crime pattern theory was formulated with crime in a physical environment in mind, such patterns occur for online fraud as well (Anti-Phishing Working Group, 2016a). Describing patterns in phishing emails (sent by offenders), as well as the behaviour and reaction of the targets receiving them, leads to a better understanding of phishing, as well as an opportunity for prevention.

A research question was formulated to guide the analysis of the dataset: *what patterns can be found in phishing campaigns in the Netherlands?* We specify our research question further by looking at patterns in terms of the phishing email itself, and patterns in the behaviour of the targets (i. e., the receiver). Specifically, we are interested in how phishing campaigns can be described. Furthermore, we want to know whether any patterns can be used for the prevention of phishing. Due to the nature of our dataset, we have information about the people who received the phishing email. Occasionally, they reported on their feelings or decision strategies, which we analysed.

The remainder of this chapter is organised as follows: [Section 6.1](#) describes the methodology of this chapter's research. This is followed by the results of the research in [Section 6.2](#). Finally, we conclude this chapter in [Section 6.3](#).

## 6.1 METHODOLOGY

The data came from phishing emails that we collected in collaboration with the Dutch Fraud Helpdesk (Fraudehelpdesk, 2016), a non-profit organisation that aims to prevent fraud. Prior to the collaboration, the

Fraud Helpdesk had started asking the general public in the Netherlands to submit phishing emails. To analyse the phishing emails that were submitted, we developed a tool called APATE. APATE is a modular framework written in Python 3, that imports and analyses emails. It was made with scalability in mind, supporting the simultaneous analysis of hundreds of emails per minute on a single server. After analysis, the characteristics of each email are compared and a decision is made whether the email is phishing or contains malware. APATE was deployed in November 2014, after which it was continuously extended with more analyses. At the time of deployment, all submitted phishing emails from January 2013 onwards were analysed.

The submission process of an email goes as follows. A person (*target*) receives a phishing email from an *offender* and decides to submit it to the Fraud Helpdesk. By doing so, the target becomes a *submitter*. APATE imports the submitted phishing email by storing each email on disk and storing the accompanying meta-data in a relational database. This is followed by an in-depth analysis of the email by APATE. Finally, the system makes an assessment and the resulting feedback is automatically provided to the submitter.

Although members of the general public were asked to submit phishing emails, not all emails that were received may be phishing. For example, the submissions can be considered as *suspicious emails*: the submitter thinks they are phishing, but it is unknown whether they are indeed. Therefore, emails need to be grouped and clustered, so as to find emails people consider suspicious. Such emails could be considered to be most likely phishing, or at least spam. Emails in the dataset are not necessarily representative for all existing phishing emails. However, they can be considered to be of ‘good quality phishing’, since they bypassed existing spam filters and ended up in a user’s inbox. Only emails from users who are aware of the existence of the Fraud Helpdesk, and who have the willingness to forward an email, end up in the dataset. There may be a bias in the sense that submitters may be more suspicious than average members of the general public. However, due to the large sample size ( $N=691,876$ ) and large number of people ( $N=135,551$ ) submitting suspicious emails, we are confident that many of the large scale phishing campaigns are present in our dataset.

During the analysis, we have created subsets of the dataset to be used for specific analyses. We have used external data to validate the analyses. Table 19 lists all datasets that were used, as well as the subsets of our dataset, e. g., by only including emails that were attached, or by clustering. For this research, we have used several types of analyses. Sometimes we used the results of our production system by querying the database, and some analyses we performed specifically for this chapter. Furthermore, where possible we automated the processing of the data, but we occasionally had to turn to manual inspection. Table 20 gives an overview of the properties of each analysis.

Data	Size	Source
<i>Internal</i>		
Emails	691,876	Fraud Helpdesk
Emails as attachment	51,708	
Clusters	12,411	
Clusters with at least 5 emails	9265	
<i>External</i>		
Victimisation at Dutch banks	2012–2016	Dutch Banking Association
Statistics on phishing reports	2016	Rabobank
Phishing per industry sector	2012–2016	APWG

Table 19: Overview of the (sub) datasets that were used. The emails were the original internal dataset, which were narrowed down for parts of the research. External datasets were used to validate our results.

Analysis	Type	Usage
<i>General analyses</i>		
Descriptives	Automated	Production
Clustering	Automated	Research
<i>Patterns of emails</i>		
Time and seasonal influences	Automated	Research
Situation	Manual	Research
Persuasion principles	Manual	Research
<i>Patterns of targeted users</i>		
Time of submission	Automated	Production
Speed of submission	Automated	Research
Comments	Manual	Research

Table 20: Overview of the analyses methods and whether each was computed automatically or manually by the researchers, and whether they were performed only once (research) or in production.

### 6.1.1 Email Similarity and Clustering

In a phishing campaign, similar emails are sent to a large number of targets. Often, emails are sent in batches with a unique phishing website or link per batch. Sometimes, each target receives an email with a unique link, to prevent anti-phishing software to efficiently scan links in emails. Additionally, an email may contain some information specific to the target, such as the target's name or address information. A complicating factor is that submitters often forwarded the phishing email as text, resulting in a loss of information compared to submitters forwarding

the email by attaching it as EML file. Forwarding as text may alter the layout and text of the original message, introduce problems between character sets, and the original headers are not preserved. For these reasons, any tool doing similarity detection for forwarded phishing emails needs to account for some variance.

Before doing similarity checking, each email text was cleaned and normalised. HTML was converted to plain text, and URLs and special characters were removed. The similarity checking was performed using simhash, which has been shown to be usable at comparing billions of websites for Google (Manku, Jain and Das Sarma, 2007; Henzinger, 2006). Simhash uses a hash that allow for heuristic near-duplicate emails to be marked similar. Simhash is efficient, but works heuristically, and therefore may not find all similar emails. In experiments with our dataset, simhash did not find all duplicates (more on that later), which is why we added a second analysis based on sentence hashing. Sentence hashing allows for emails that contain the same exact sentences to be efficiently linked. Sentence hashing results in good similarity checking, but fails when sentences contain even the smallest deviation. Having two methods of analysing similarity, we needed to make sure to validate their results. For this, named entity extraction was used, which reveals the organisations or persons that are mentioned in the emails. Named entity extraction works to identify the meaning of an email, but fails to address textual similarity. In conclusion, we used simhash to find similar emails in a fast manner, and sentence hashing to extend these results using a different method. Then, a third method (named entity extraction) was used to analyse whether the topics of the emails were about the same persons or organisations. Before discussing in more detail how we performed the clustering, we first address each of the methods individually.

#### 6.1.1.1 Method 1: *simhash*

Calculating similarity is traditionally done between sets of two texts. However, when working with large datasets of size  $n$ ,  $\frac{n \times (n-1)}{2}$  comparisons are needed. This does not scale. An alternative is to use a more imprecise similarity checking to quickly find possible duplicates. One such method is called *simhash* (Charikar, 2002). Simhash uses fingerprinting to find near-duplicates. A *near-duplicate* is essentially the same text, but differs in a small set of features. In simhash, each text is described by a fingerprint, by default with a length of 64 bits. Two texts are considered equal when their fingerprints differ at most at  $k$  positions, for a predefined Hamming distance  $k$ . The choice of  $k$  is a tradeoff: a low  $k$  misses near-duplicates, a high  $k$  may incorrectly tag pairs as near-duplicates. Manku, Jain and Das Sarma (2007) suggest  $k=3$  serves well in terms of recall ( $\sim 75\%$ ) and precision ( $\sim 75\%$ ). We decided it was more important to include as many as possible candidates, at the cost of

incorrectly tagging near-duplicates. Therefore, we configured simhash with  $k=5$  as suggested by Kumar and Santhi (2012), still having a recall of  $>90\%$  in the experiments of Manku, Jain and Das Sarma (2007). The result of simhashing all emails is that we can perform comparisons on the fingerprints to efficiently find similar emails.

#### 6.1.1.2 *Method 2: sentence hashing*

The second similarity analysis was sentence hashing (also known as sentence-level fingerprints (Wang and Chang, 2009)). For each email, all sentences were extracted. All characters except the latin alphabet (a-z) were removed, any extra spacing was removed, and the sentence was converted to lower case. The resulting sentences were hashed using SHA1 and linked to the email in the database. After the initial processing, finding emails that share sentences is as simple as running an SQL query in the database. Sentence hashing allows us to find emails that contain the same sentences.

#### 6.1.1.3 *Method 3: named entities*

As a third way to check similarity, we extracted named entities from each email. This allowed us to establish the topics of each email. Frog (van den Bosch et al., 2007) was used to extract the named entities. Since Frog is specifically made for Dutch, all non-Dutch emails ( $\sim 30\%$ ) were filtered out before running the analysis. Each word in a sentence was tagged by Frog. However, we included only words that were between 2 and 22 characters long, to prevent non-words from being tagged. In Dutch, 99% of the words are between 2 and 22 characters long (Geloven, 2011), therefore the number of true words not being tagged is expected to be negligible. Finally, only named entities tagged as person or organisation were stored for later analysis.

#### 6.1.1.4 *Clustering based on similarity*

The next step is to combine the knowledge from the three similarity measures and compute clusters of similar emails. For each email, we computed its near-duplicates by comparing its simhash fingerprint to all other emails. This resulted in many clusters of similar emails, based on their simhash fingerprints. Since simhash is not a guarantee, but rather a heuristic, there may be emails not included when they should be (false negatives) or emails that are included when they are not similar (false positives). Emails were mostly forwarded as plain text with transformations, and our primary aim was to reduce the number of false negatives. The clusters that were found using simhash had two potential problems: (1) not all emails were found by the simhash algorithm; and (2) there were several clusters for the same email, due to small transformations and modifications of the email text.

To find more similar emails, we used sentence hashing to extend the clusters. Emails that shared 75% of the sentences to one of the clusters, were added to that cluster (i. e., simhash may not have found these emails). In case of an unequal number of sentences, the smaller email was required to share 75% of the number of sentences of the larger email. We chose for 75% after empirical tests with the dataset (i. e., testing different values and manually analysing the resulting similar emails). Often, emails contained additional text, for example, a disclaimer. The clusters were then assigned a *centroid*, which is the email that is most similar to the other emails within the same cluster.

As discussed before, we observed fragmentation within sets of similar emails, resulting in multiple clusters for emails that were similar. Therefore, the next step was to try to reduce the number of clusters by merging clusters. For each pair of clusters, we compared the simhash fingerprints and sentence similarity of their centroids. When the centroids contained exactly the same sentences, they were merged. Otherwise, if the centroids had a similar simhash fingerprint, they were merged only if one of the two conditions was met: (1) either the clusters had 75% of their members in common; or (2) the clusters had named entities in common. Finally, the resulting clusters were stored in the database for analysis.

#### 6.1.1.5 Validation

To validate the previously mentioned similarity checking and clustering methods, they were tested on samples of the dataset. Each algorithm was tested with different parameters on a sample of the data. For example, after reviewing the literature for good values of  $k$ , we found that using simhash with  $k=5$  worked better for us than  $k=3$ . Similarly, we removed special characters from the emails before doing a text analysis, since it yielded better results compared to the text including special characters. To verify that the clustering worked as expected, we manually inspected 100 randomly selected clusters. For each cluster, we registered the quality (i. e., are all emails alike?) of the cluster, number of wrongly classified emails, and whether the cluster was phishing, spam or something else (e. g., legitimate). Out of the 100 clusters, 94 contained only correct emails, and six contained between one and five emails that should not be in the cluster. The clusters consisted of 8923 emails, leading to less than 0.4% of these emails to be falsely included (false positives), according to the analysed sample. 51 clusters contained phishing emails, 45 clusters were spam and 4 clusters contained something else (e. g., legitimate emails). If the purpose of the email was not clearly phishing, it was marked as spam, which would occur, for example, in emails that claimed the receiver won a prize or get discounts on certain products. While such emails may be considered phishing (i. e., when they ask for personal details and never give a prize or provide

a discount), our classification was conservative and marked those as spam. Moreover, we found that our clustering was not *complete*, i. e., that all similar emails appear in one cluster. Often, several clusters were created due to variations in the emails (e. g., changes made by the offender) or transformations that were made during submission (e. g., character encoding problems). Additionally, we decided to exclude clusters with less than five reports from parts of our analysis, since they may be legitimate emails. As a consequence, phishing campaigns that were not reported by at least five users, were not used in the analysis. However, we considered this an acceptable tradeoff.

When analysing the campaigns, we observed that campaigns would often consist of several *spam runs*. In each run, an email would be sent to the targets. After several weeks, a very similar email would be sent again. To model this behaviour, we analysed all clusters with at least five email reports. For each cluster, we extracted the number of consecutive days at which emails were reported. People may report phishing emails a couple of days late, but at least it provides an overview of the peak moments of a campaign. We then measured the number of *gaps*, i. e., number of days between two spam runs within a particular campaign (or cluster of emails). Mails that were less than 7 days apart, were considered to be from the same spam run. In the same way, mails that were 7 or more days apart, were considered to be different spam runs. This allowed us to calculate the number of spam runs within one campaign.

### 6.1.2 *Patterns in Suspicious Emails*

In the analysis, we focussed on two types of patterns: patterns in the emails (discussed in this section) and patterns in the behaviour of the targets (discussed in the next section). First, we analysed the patterns in the suspicious (forwarded) emails. Applying the broad concepts of the crime pattern theory, we assume that the suspicious emails are distributed non-randomly in time and space. For the time concept, we can extract the time and day at which the emails were sent by the offender. Furthermore, we can try to relate the number of emails in a particular time interval with, for example, seasonal differences or holidays. The concept of space is more ambiguous on the internet, in particular due to their differences with most traditional crime (i. e., the offender and victim can be far apart in terms of physical distance in cyber crime). Therefore, instead of only using space, we look at the broader situation that was created by the offender. Apart from the location of the message (i. e., the email client), the setting of the particular email is important to the success of the attack. We analyse the setting in terms of the impersonated organisation and the persuasion methods that were used.



When analysing suspicious emails, we do not want to include submitted emails that were legitimate (i. e., wrongly classified by submitter as probably phishing). Therefore, we only include emails that were reported at least five times. We use the results of the email clustering and exclude individual emails and clusters with less than 5 members<sup>9</sup>. While this does not guarantee to rule out all legitimate emails, it does filter out individual mistakes. However, for other types of analyses not directly related to the suspicious email being phishing or not (i. e., details on the moment of submission), we used the entire dataset.

#### 6.1.2.1 *Time*

For mails that were forwarded as attachment, we could extract the original date and time at which the email was sent. We looked at the header information of the original email. Occasionally, offenders forge the date header of an email to make their email appear on top, i. e., by claiming the sent date in the future. Therefore, the email *Received* headers listing the email servers that handled the message, were examined. If the date of processing of the last email server (i. e., the one of the submitter) differed more than an hour from the claimed date in the email, the header date was used for the further analysis instead of the claimed date. Finally, we looked for patterns (i. e., positive or negative peaks) in terms of weekend versus weekdays, and time of day.

Additionally, to analyse potential seasonal influences, we manually inspected the ten largest clusters and the distribution of the submissions over time. For each large cluster, the number of reports per week was retrieved and inspected. This data was combined with a list of public holidays in the Netherlands and the number of reports week for each large cluster, over a period of one full year. Finally, two experts from the Fraud Helpdesk were asked about their expectations regarding seasonal changes of phishing emails.

#### 6.1.2.2 *Situation*

Another way to look at the emails is in terms of the situation that was created. The offender drafted an email that seems to originate from a particular organisation. By analysing which organisations were used and from which type of industry they pretend to be, we compare different types of industries and the risk they have to be impersonated. To analyse which types of organisations are abused, we used the results of the named entity extraction. The 1000 most-commonly used named entities were manually inspected and categorised in five industry sectors: (1) financial, (2) retail, (3) internet and telecommunication service providers (ISPs and telcos), (4) government and (5) other. The categories

<sup>9</sup> The threshold of five is used by the Fraud Helpdesk as a minimum number of reports for any type of fraud, before the case is considered.

were inspired by the APWG's quarterly reports (Anti-Phishing Working Group, 2016b), with payment services combined with banks into the financial sector and unreported sectors removed. Each cluster was labelled according to the named entities that were extracted from its centroid or centroids of any clusters that were merged into this cluster. The data was analysed per-sector on an email level, with an email being tagged according to the cluster it was in. The analysis was performed on a quarterly basis, starting from the first quarter of 2014 up to and including the third quarter of 2016.

#### 6.1.2.3 *Persuasion principles*

Furthermore, we include a brief analysis on the methods that the offenders used to persuade the targets to take action. We scored emails on usage of the persuasion principles of Cialdini (2001): reciprocity, consistency, social proof, likeability, authority and scarcity. We will introduce these principles by means of examples. *Reciprocity* is the principle that when someone is given something of value, he or she will feel obliged to do something in return. In our dataset, an example of reciprocity was a promise to give the target a discount or free item, if the he would click on a particular link in the message. *Consistency* (or commitment) is the principle that someone who commits to something small (e. g., signing a petition) will be more likely to commit to something larger afterwards (e. g., donate money). As an example of consistency, some phishing emails mentioned the existence of an appointment or a deal, and pretended it was time for the next step in that agreement. Furthermore, a phishing email can request a small request that is not perceived as dangerous, and request something more after the target has clicked on a link. *Social proof* is a type of conformity. An example of a message that used social proof is "Over 200 others have applied for free better security for your online banking". *Likeability* makes a person more likely to listen to someone that he or she likes. For example, an email where a celebrity recommends the target to perform an action would use to the likeability principle. Alternatively, the offender might include a photo of an attractive person in an attempt to influence the decision of the target. *Authority* is the principle that people will obey requests of authoritative person or organisation. In phishing messages, authority is often used by impersonating an authoritative organisation, such as a bank or a government, or in the title of the supposed sender (e. g., the chairman of a bank). Finally, *scarcity* describes the increase in demand when there is a perceived shortage. Scarcity is most often shown by warning for limited access (e. g., "If you don't click here, your internet banking will be disabled.") or deals that are valid for only a few days (e. g., "request a new bank card within 2 days, otherwise you have to pay €15"). For this specific analysis, we looked at all clusters with at least 5 emails in them (N=9265), to exclude legitimate and small

scale phishing emails. We randomly selected 100 clusters that contained phishing emails. The centroid email of each cluster was analysed for presence of each of the six persuasion principles. A single email can contain zero or more principles.

### 6.1.3 *Behaviour of Targeted Users*

Apart from the patterns in the suspicious emails themselves, we looked at patterns in the behaviour of the targets (the person who submits an email to us) as well. The advantage of our dataset over spam traps or spam filters, is that it contains more information about the combination of suspicious emails and the target. For example, some targets write a message when forwarding a suspicious email. This, together with information such as the time of day or forwarding, reveals a lot about the behaviour of targets. For the target, we analysed two properties that show their email reading behaviour: (1) the time of submission; (2) the difference in time between receiving an email and submitting it. The time of submission shows the time of day at which people process their emails. Therefore, it shows when people are subjected the most to phishing. The time of submission is extracted from the email's Date header.

#### 6.1.3.1 *Submission time*

To find the time that it takes for people to report a phishing email, we included only phishing emails that were forwarded as attachment. For these, we could compare the original arrival date of the email with the arrival date of the submission. For each email that was forwarded as attachment, we extract the receiving time of both the original mail, and the forwarded mail. Then, we calculate the difference between these two times. To avoid erroneous entries, we include only forwarded emails that were sent between 10 seconds and 30 days after the phishing email. This excluded false entries (i.e., forwarded before having receiving the phishing mail), many automatic replies (forwarded within 10 seconds) and extremely slow responses (forwarded after more than 30 days). Additionally, if the date of processing of the last email server that handled the message differed more than an hour from the claimed date in the email, the header date was used for the further analysis instead of the claimed date.

#### 6.1.3.2 *Comments of submitter*

Another way to look at the behaviour of the targets is by reading the comments that they wrote when forwarding a suspicious email. To do this, we randomly selected emails containing comments from 35,075

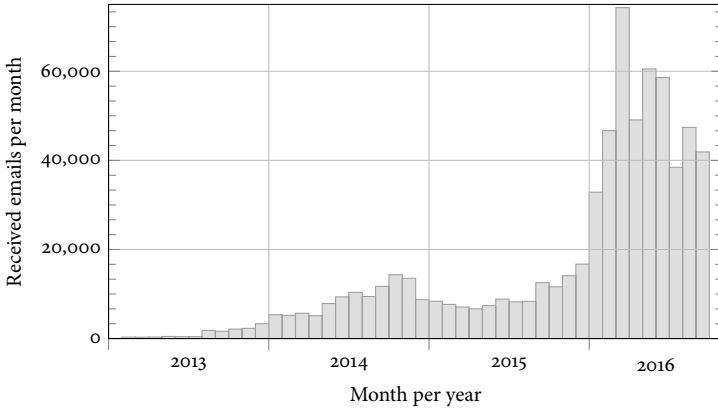


Figure 20: Monthly intake of emails between 2013 and October 2016 ( $N=691,876$ ).

emails that were reported between March 2013 and June 2014<sup>10</sup>. The goal was to end up with 200 *relevant* emails, where relevant emails are ones that contain any form of explanation of the thoughts of the target. To reach 200 relevant emails, batches of 50 emails were randomly selected from the dataset, and analysed for thought patterns. Randomly drawn emails could not be re-drawn later. The selection process for each email started with reading the comment of the submitter. From the comment, it had to be clear that the submitter judged the email to be phishing (or, in general, illegitimate). If that was the case, the comment had to explain why the submitter thought the email was phishing. When the comment stated both the decision (phishing) and the reason for that decision, the corresponding email was included for further analysis. In this way, 3850 emails were scanned to end up with 200 relevant emails that included thought patterns of the targets. Of the 200 relevant emails, we scored the mentioned reasons for finding the email suspicious.

## 6.2 RESULTS

The dataset consisted of 691,876 emails between January 2013 and October 2016. The collection of emails per month is shown in Figure 20. Most emails ( $N=640,168$ ) were forwarded as text, meaning the original headers were lost. Only 7.5% ( $N=51,708$ ) were forwarded as EML attachment, thereby preserving the original headers.

Dutch was the most commonly used language in the submitted suspicious emails (77.9%,  $N=539,147$ ), followed by English (7.0%,  $N=98,659$ ) and German (1.4%,  $N=10,331$ ). Roughly 3.5% ( $N=24,586$ ) could not be parsed correctly, or the language could not be determined. Given the

<sup>10</sup> The analysis of submitter comments was joint work with Nick Grobбен (Grobбен, 2015)

country where the data was collected, it makes sense that most emails were in Dutch. However, at the same time it shows that phishing can be localised. The senders of the phishing emails know in which country the email addresses are used, and therefore in which language they should send the phishing email. Or, in the case of Dutch offenders, emails in Dutch were sent to Dutch email addresses. This could partly be explained by the domain name of the receivers, since this was in 65% (N=88,206) of the cases .nl.

Clustering resulted in 12,411 clusters of similar Dutch emails. Out of these 12,411 clusters, 9,265 contained at least 5 emails. Of all the emails that were written in Dutch (N=539,147), 86.5% (N=466,610) were part of a cluster. Clusters with at least 5 members (N=9265) were analysed to identify spam runs, i. e., waves of emails with no reports in between. For the 9265 analysed clusters, the average number of spam runs per campaign was 3.6. An individual spam run lasted on average 4.5 days, with a median length of 2 days and a maximum of 376 days. 44% of the spam runs lasted only one day, and 90% of the spam runs lasted 9 days or less, including any delays caused by slow forwarding by the submitter. The time between spam runs within a campaign (*gaps*) averaged 49 days, with a median of 21 days.

### 6.2.1 Context of the data

In order to validate our dataset, we looked at three related datasets: (1) the APWG reports; (2) recorded phishing incidents by the Dutch banks; (3) and the number of reported phishing emails at one of the largest Dutch banks.

#### 6.2.1.1 APWG reports

Firstly, we look at the trend analysis of the APWG dataset. The APWG publishes quarterly reports on trends in phishing, and is therefore an excellent dataset to compare our dataset with. Specifically, we looked at the targeted industry sectors in both datasets (see [Figure 25](#)), which are described in more detail in [Section 6.2.2](#). This comparison shows that in terms of targeted industry sectors, our dataset seems to be completely different from the APWG data. Whereas our dataset shows a constantly high proportion of phishing targeting the financial sector, the APWG dataset has relatively more emails from other sectors.

We conclude that the APWG dataset looks different from our dataset in terms of industrial sectors being targeted. Whether received phishing in the Netherlands is different, or whether the Dutch report phishing emails differently, remains unknown. The difference could be explained by the methods used to collect phishing emails (e. g., spam traps or user reports). The APWG has a large body of organisations contributing

phishing emails (Anti-Phishing Working Group, 2016c). Our dataset consists of data from only one organisation, namely the Fraud Helpdesk.

#### 6.2.1.2 *Victimisation at customers of Dutch banks*

The financial sector is the most important targeted industrial sector in our dataset. Therefore, the financial sector could serve as a reference to compare our dataset with. The number of reported phishing emails are not published by the Dutch Banking Association, but they do publish statistics on phishing victimisation in terms of number of victims and amount of money lost. The Dutch Banking Association (2017) publishes aggregated statistics on phishing for the four largest banks (i. e., the domestic systemically important banks: ING, ABN AMRO, Rabobank, SNS Bank (De Nederlandsche Bank, 2014)).

Figure 21 shows the number of successful phishing attacks (i. e., resulting in monetary loss) against the customers of the four aforementioned banks. Only successful attempts that have been reported by the victims to their bank are included in these results. The number of victims is unstable over time, and has been relatively low for the last six quarters. Furthermore, we included the number of reported emails to the Fraud Helpdesk in Figure 21 as well (right axis). We expected an increase in reported phishing emails to match an increase in phishing victims. However, the increase in reported phishing emails for the Fraud Helpdesk seems to be independent of the victimisation of banks. In particular in the year 2016, the number of reports grows by an order of magnitude, but the number of victims remains low. This shows that people are getting better at avoiding victimisation and better at reporting phishing simultaneously.

The average monetary loss per phishing victim for the period of 2012–2016 is shown in Figure 22. These numbers represent the phishing victims of the four largest banks in the Netherlands only. The average loss per subject went down from on average €6,402 in 2012, to an average of €1,334 euro in 2016. At the same time, the combined monetary loss of phishing victims went down from €11.4 million (2012) to €0.7 million (2016) (Dutch Banking Association, 2016). According to the banking association, the reduction in monetary loss is due to measures such as prevention (radio and TV commercials) as well as improved monitoring and detection systems (Dutch Banking Association, 2016).

Looking at the victimisation data should reveal the success of phishing attacks for the financial sector. The data that was published for the banking sector in the Netherlands shows that the success of phishing attacks is reducing over time. The total monetary loss has been decreasing steadily since 2012 (Dutch Banking Association, 2016), and the average monetary loss per victim has been reduced over time as well. A decline in total loss resulting from phishing in the banking sector could follow from an increase in phishing attempts. To quote Herley and

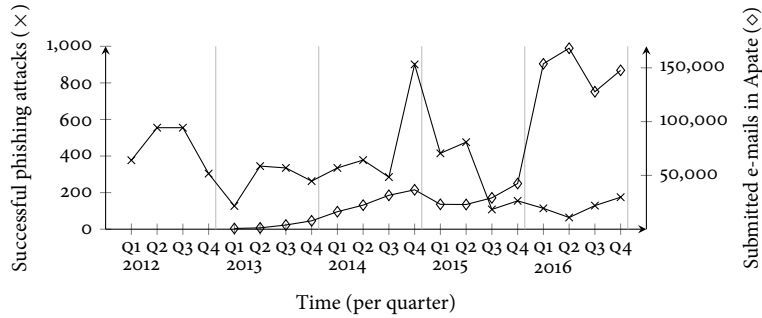


Figure 21: Number of succesful phishing attacks at customers (×) of the largest Dutch banks (Dutch Banking Association, 2017), and the number of reported suspicious emails in the Apaté system (◊).

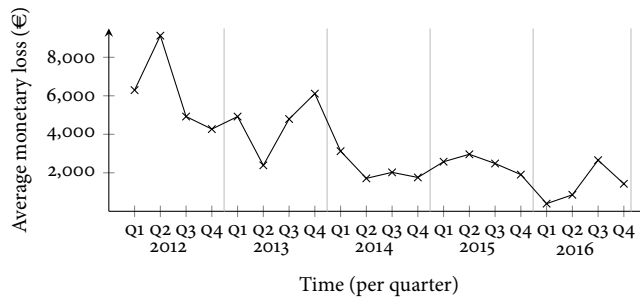


Figure 22: Average monetary loss of a phishing victim at the largest Dutch banks (Dutch Banking Association, 2017)

Florêncio (2008): “as the total phishing effort increases the total phishing revenue declines.” Additionally, people who receive an increased number of phishing emails from a particular industrial sector (e. g., banks), may be more aware for these specific emails, since their awareness is triggered more often and improved. Specifically for our dataset, the 2016 increase in the number of reported emails did not result in a different distribution in industrial sectors (Section 6.2.2). The same holds for the 2015 decrease in number of phishing reports. Therefore, we believe that the fluctuations in the number of reported phishing emails are merely a reflection of the brand awareness of the Fraud Helpdesk. We believe that these changes reflect the willingness of people to report phishing, as well as being a reflection of how acquainted people are with reporting phishing to the Fraud Helpdesk.

### 6.2.1.3 *Forwarded phishing emails at a large bank*

The third (external) perspective to our dataset is the collection of user reports of banks. In the Netherlands, all banks have a specific email address at which people can forward phishing emails that they have received. The Rabobank, the second largest bank of the Netherlands (Kleinnijenhuis, 2015), received on average 24,000 forwarded phishing emails per month as of the months before March 2017 (Rabobank, 2017). About 4% (Apr 2015–Feb 2017) of these emails result in a notice and takedown (NTD) procedure. The low percentage of NTD's is due to the large number of spam or phishing messages targeting other organisations that is received.

If brand awareness is indeed an important factor that determines the number of phishing messages that are reported, as discussed before, the number of reported phishing emails should give an indication of the willingness to report phishing. With on average 24,000 reported phishing emails per month, the Rabobank receives about half the number of emails per month compared to the Fraud Helpdesk average of 2016 ( $N=49,975$ ). In our opinion, this shows that our dataset is rather large in comparison.

## 6.2.2 *Patterns in Suspicious Emails*

In our analysis of the suspicious emails, we looked at patterns over time and in terms of the setting of the emails. To analyse the moment phishing emails were sent, we included only emails that were forwarded as attachment ( $N=51,708$ ). Most emails were sent on weekdays (Monday 16.7%,  $N=8642$ ; Tuesday 17.7%,  $N=9138$ ; Wednesday 16.7%,  $N=8615$ ; Thursday 16.9%,  $N=8728$ ), with a decline on Friday (13.4%,  $N=6913$ ), as is shown in Figure 23. This decline increases further in the weekend (Saturday 9.6%,  $N=4969$ ; Sunday 9.1%,  $N=4703$ ), consistent with literature in the field of phishing (Bursztein et al., 2014; Moore and Clayton, 2007; Ramzan and Wüest, 2007). Furthermore, the ratio of emails per day closely resembles data from a study of Ramzan and Wüest (2007).

### 6.2.2.1 *Time*

When looking at the time of day at which suspicious emails were sent, we found that there is a slight tendency for phishing to be sent during the morning (6AM to 12 noon; 28.9%;  $N=14942$ ) and afternoon (12 noon to 6PM; 32.1%;  $N=16623$ ). Phishing activity reduced during the evening (6PM to midnight; 20.0%;  $N=10323$ ) and night (midnight–6AM; 19.0%;  $N=9820$ ). We expected hardly any difference between night and day similar to spam in general (Gomes et al., 2004), but our data shows a peak during the day (i. e., morning and afternoon). It seems that



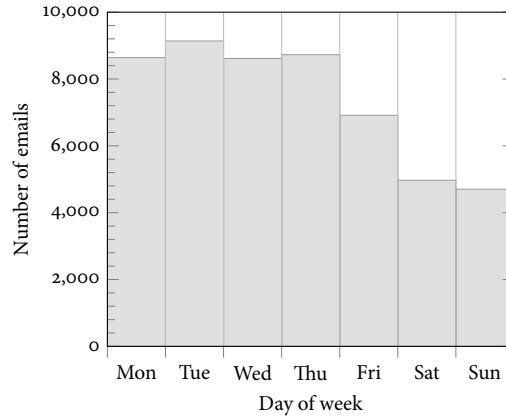


Figure 23: Day of week at which the original email was sent to the target (N=51,708).

phishing offenders are more active during the day (6AM-7PM), as shown in Figure 24. In combination with the higher activity on weekdays, offenders are more active during daytime. In comparison, in a study of email account hijacking (Bursztein et al., 2014), the patterns were clearer and showed that the offenders were working office hours only. All times are in the timezone used in the Netherlands: CET, or CEST when summer time is active. Offenders may be spread in different timezones, accounting for some of the variation.

#### 6.2.2.2 Situation

The ten largest clusters of emails contained phishing campaigns targeting three telecommunications providers and two debt collectors, of which one belongs to the ministry of justice (for collecting traffic fines). Consulting experts led to two hypotheses: (1) that phishing campaigns from debt collectors are related to the end of the month, when employees receive their paychecks; and (2) that phishing using traffic fine as pretext is present during holidays, when many people travel abroad. We found that 42% of the spam runs (N=20; out of 47 total spam runs) in the four debt collector campaigns coincided with the end of the month (i. e., the last week of the month). The other 58% of the spam runs (N=27) for debt collectors peaked at other moments in the month. Moreover, from the ten campaigns that were analysed, we found no evidence of campaigns that were primarily focusing around the holidays. Due to the nature of phishing, a change in the offender's daily routine (i. e., public holiday) may not effect the automated gathering of information.

The email campaigns were spread throughout the year. However, the three ministerial debt collector phishing campaigns peaked around two

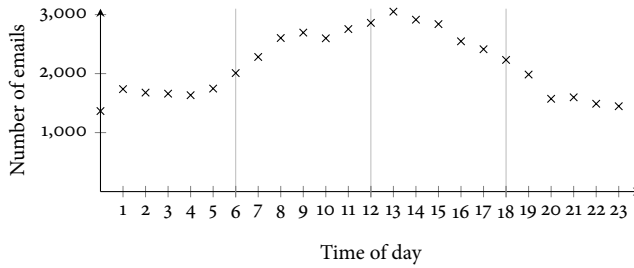


Figure 24: Number of phishing mails sent per hour of the day (N=51,708).

particular holidays (christmas and the spring holidays). None of the telecommunication providers' campaigns showed seasonal differences.

Most of the emails appearing in campaigns abused the name of a financial institution, such as a bank or payment service provider. Between 70% and 83% of the emails claim to be from the financial industry between 2014 up to October 2016, as shown in [Figure 25a](#). Phishing emails pretending to be from a bank are a constant threat. The number of emails in campaigns pretending to be from the government had two peaks (Q2 in 2014 and Q3-4 in 2015), apart from which there were hardly active campaigns. There were several campaigns involving the government, most notably from the tax office (regarding tax returns), ministry of public health (on the topic of health insurance) and the ministry of Justice (traffic fines). Furthermore, the sector ISP and telco slowly increased and is a steady factor in phishing. The largest campaign in this sector informed the targets about an invoice, and clicking on the link or attachment, malware was installed.

Particularly in the financial sector, we found many large clusters of emails that contained the same content with only the name of the bank changed. With one campaign, the offenders targeted multiple banks simultaneously. Also several campaigns from telecommunication providers used multiple versions of the same email.

Compared to the quarterly reports of the APWG (see [Figure 25b](#)), the most notable difference is the large share of government-related phishing emails, and the limited number of emails related to the retail sector in our dataset. For example, in Q2 of 2016 (Anti-Phishing Working Group, [2016b](#)), the APWG reports 43% retail, 29% financial, 12% ISP and only 1% government. For the same period, our data contains 1% retail, 76% financial, 7% ISP and 2% government. This could indicate that phishing in the Netherlands follows different patterns from global phishing. Alternatively, the datasets could be compiled differently, as the APWG gets data from several sources.

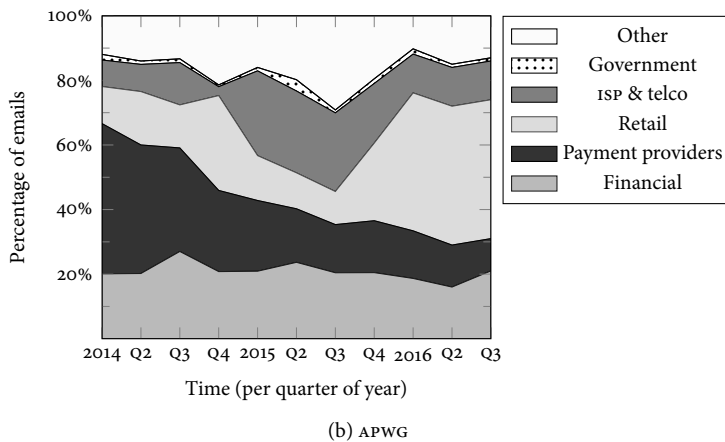
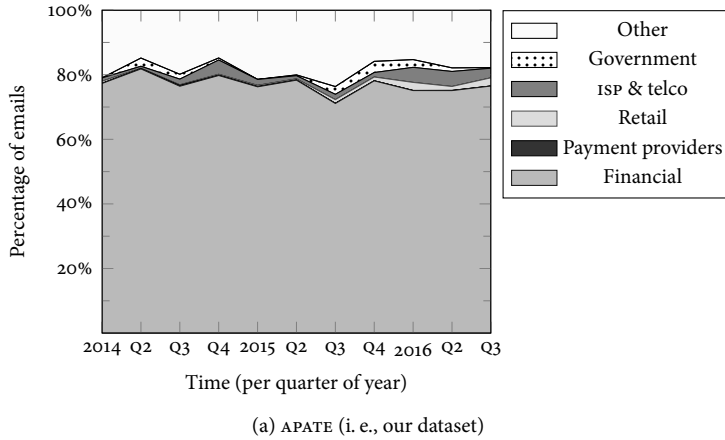


Figure 25: Most targeted industry sectors for our dataset (a) compared to the APWG dataset (Anti-Phishing Working Group, 2015b, 2014c, 2016a,b,c, 2014a,b, 2016d, 2015a).

### 6.2.2.3 Persuasion principles

To analyse the methods that the offenders used, 100 randomly chosen phishing emails representing a cluster of similar emails, were scored for persuasion principles that were used. The results are listed in Table 21. Authority was most the most often used way to persuade the targets to comply. Often, targets are asked on behalf of a bank or governmental organisation to perform an action. Scarcity is the second most-often used method, mostly explained by limited access to an account if the target does not perform the requested action. Reciprocation is used in 28% of the analysed emails, and is seen mostly in emails where the target is promised a free tickets or a coupon for performing an action. The

remaining three principles (social proof, consistency and likeability) were hardly used.

Persuasion Principle	Occurrence
Authority	70% (N=70)
Scarcity	45% (N=45)
Reciprocation	28% (N=28)
Social Proof	6% (N=6)
Consistency	5% (N=5)
Likeability	1% (N=1)

Table 21: Persuasion principles of Cialdini (2001) that were used in phishing emails (N=100). An email can use multiple persuasive principles.

6.2.3 Behaviour of Targeted Users

6.2.3.1 Submission time

The general public submitted phishing emails mostly during office hours, with peaks in the early morning (between 7AM and 10AM). A heatmap of the receiving time of reported suspicious emails is included in Figure 26. In the weekends, there are small peaks in the number of received reports around 10AM. The results show that most people open their emails on working days in the mornings. Particularly on Monday morning, presumably when processing all emails that were received in the weekend, users should be vigilant.

To measure the time it takes for a person to report a suspicious mail, we analysed all emails that contained a forwarded email as attachment (N=52,907). While analysing the emails, we were unable to correctly parse 1,199 emails (i. e., no correct Date headers). This lead to 51,708 emails to be considered for analysis. After excluding replies that were within 10 seconds or replies that were sent longer than 30 days after the original email, 48,279 emails were suitable for further analysis of submitter’s response time. A correction of the claimed date and time of sending the email was applied for 15% (N=7250) of the emails, meaning the claimed date header was inaccurate to due slow mail servers or malicious intent of the sender. A quarter of the emails was forwarded within 1.5 hour, and half of the emails was forwarded within 6 hours. After 24 hours, 80% of the submissions was forwarded and 88% was forwarded after 48 hours. The distribution of emails over the first 48 hours (N=42,654) is shown in Figure 27.

We could find no effect of holidays on the number of phishing emails. That implies that even during their holidays, people still forward phishing emails. The lack of provable seasonal effects can have another ex-

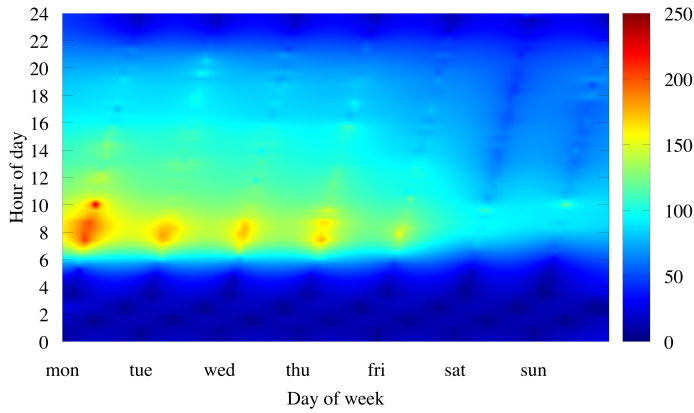


Figure 26: Submission of suspicious emails per minute ( $N=691,835$ ).

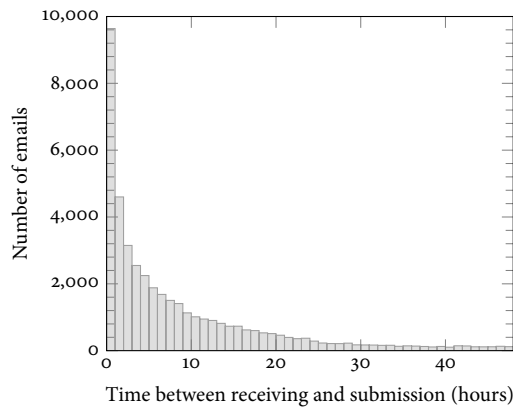


Figure 27: Average time between receiving and submitting a suspicious email ( $N=42,654$ ).

planation. Due to the variation in the number of reports in the dataset caused by other influences (such as media attention), seasonal changes are difficult to find. Every time the Fraud Helpdesk got media attention on the national media, the number of reports went up as a result. These peaks were shown in [Figure 20](#).

People submitting the suspicious emails, did so from 135,551 different email addresses. On average every person submitted 5 suspicious emails. Looking at the domain from which the submitters forwarded the phishing mail (i. e., for *user@hostname.tld*, this would be *hostname.tld*),

we found 38,578 unique domains. By categorising these domains, we found that 39.1% of the reporters used an email address from their ISP. Furthermore, 24.5% used an email address from a large international email provider, such as Google or Microsoft. Due to the nature of these email services, these 63.6% of the reporters could likely be considered *personal* (non-business) email addresses. The remainder consists of either private domains, business addresses or other addresses.

#### 6.2.3.2 *Comments of submitter*

About half of the emails in the dataset contained a *comment*: the person who forwarded the email wrote something with it. We analysed these comments for 200 emails in the dataset. In 7.5% ( $N=15$ ) of the emails, the submitter asks for confirmation on whether or not the email is phishing or not. These people reported a suspicious email hoping to get confirmation on their suspicions. This means that in general, people seem to state little doubt about their decision that the email was correct.

Often, submitters provide personal information together with the phishing email. Provided that phishing is all about gathering information, default signature texts of people provide a lot of information when they forward an email. One in five emails with comments includes more information than just a name. For example, some people included home and work address, phone numbers and other contact information, and sometimes even names of relatives that received similar emails.

The comments were analysed for reasons so as to why people would consider the received email suspicious. The vast majority of the subjects (69%,  $N=138$ ) indicated that they had no relationship with the organisation that the email supposedly was from. The comments would mention not being a customer of the specified bank or not having ordered anything from that web shop. The other properties that were mentioned as being the reasons for being suspicious were far less often mentioned. Ten percent mentioned the sender of the message (name or email address,  $N=20$ ) as being the reason for being alarmed. The reputation of the organisation (8.5%,  $N=17$ ) was listed as a reason too. However, this means that the people don't trust the organisation of which the name was misused by the offender. For example, some submitters mentioned not trusting a particular web shop, whereas the web shop itself was legit. Unusual sentences ( $N=11$ ) and spelling mistakes ( $N=9$ ) were mentioned by only few. Finally, nine comments mentioned having looked at the link, or being alerted because it was requested to click on a link, and 6 comments mentioned that the email was directed to the wrong email address (e. g., when people use a particular email address for all serious emails, and another one for non-important emails).

### 6.2.4 Impact of APATE

'APATE' was initially designed as a system for the employees of the Fraud Helpdesk only. However, it turned out to have significant impact beyond the staff of Fraud Helpdesk. One of the benefits is better prevention. Prior to the introduction of APATE, the Fraud Helpdesk would warn for specific phishing emails at a rate of about one warning per day. Thanks to the fast processing and convenient interface of APATE, a dedicated page with all phishing campaigns could be developed on the website of the Fraud Helpdesk. Between 10 and 30 different phishing emails are uploaded to this phishing page every working day. This page includes the plain text and an image (screenshot) of the phishing email. Visitors can browse and search all phishing campaigns, and filter by targeted company or type of email (e.g., phishing or malware). The page showing phishing emails is rather popular, with 183,224 visits in 2016. A screenshot of the webpage with phishing emails is shown in Figure 28. Specific warnings of high impact phishing campaigns are highlighted on the website, as well as shared through social media.

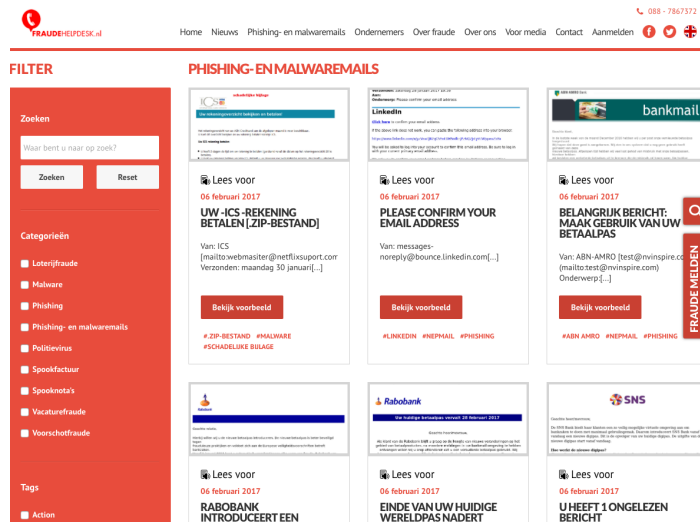


Figure 28: A screenshot of the dedicated phishing page on the website of the Fraud Helpdesk (Fraudehelpdesk, 2017).

A second benefit of APATE is that people who forward a suspicious email will receive an immediate reply from APATE. When we can confirm a suspicious email to be phishing (20-40% of the emails), the reply gives helpful tips to the submitters. For this to work, the phishing email needs to be either known, or contain a blacklisted URL. The same applies for the situation in which a suspicious email contains a malware attachment (<1%). When we are unsure of the exact contents of the email, a

general thank-you message is returned. If the submitter has questions or is unsure whether an email is phishing, (s)he can review existing phishing emails on the phishing page, or call the Fraud Helpdesk for personal assistance.

Pattern
Most phishing emails are sent on Monday to Thursday, and gradually less on Friday, Saturday and Sunday.
Repeated spam runs with the same phishing email are on average 49 days apart (median 21 days).
Between 70% and 83% of the emails were impersonating the financial industry.
Offenders often (70%) use authority to persuade the target to fall for the phish, followed by scarcity (45%).
Targets read phishing emails mostly during office hours, with peaks between 7AM and 10AM.
Half of the emails were forwarded within 6 hours and 80% within 24 hours.
64% of the people reporting phishing, do so from a personal email address.
The main reason (69%) for not trusting an email is not having a relationship with the organisation mentioned in the email.
The impersonated industry sectors in the Netherlands (our dataset) are less diverse than globally (APWG dataset).

Table 22: Overview of the found patterns in the dataset.

Concluding the results, [Table 22](#) provides an overview of the patterns that were found in the dataset.

6.3    DISCUSSION

We collected a large dataset of suspicious emails that were reported by the general public in the Netherlands. By clustering these emails into campaigns, we could analyse phishing on a nation-wide level. The data not only results in insights in the type of phishing campaigns, it also reveals a lot of information about the general public who receives phishing emails.

One of our observations is the repeated use of the same email after several weeks. On average, a phishing campaign consisted of 3.6 different spam runs. The median length between two spam runs in a phishing campaign was 21 days. After these three weeks, there would be another spam run, sending the same email again. There could be several explanations for this behaviour. Offenders need time to gather and monetise the obtained information. They cannot process large quantities at the same time, and are scared that their fraud may be detected before being able to monetise it. After misusing the information from a batch of victims, they are ready to process another batch of victims. Alternatively,



sending a phishing email to millions of subjects requires resources in terms of email servers, either through botnets or by using stolen mail accounts. This, in addition to phishing websites being taken down or blacklisted, results in practical limitations. Sending potential victims to a phishing website that was blacklisted would be a waste. Therefore, using batches of victims, with one phishing website per batch, spreads the risk of a lack of availability, from the offender's point of view. Finally, when alerts are sent out following a phishing campaign, people become more vigilant for a small period of time. However, this increased awareness decreases after 2–4 weeks to lower levels (see [Chapter 5](#) and Bullee, Montoya et al. (2016)). For the same reason an anti-phishing training needs to be repeated, the offenders can use this knowledge as well, by repeating the phishing campaign.

Half of the submitters forwarded the phishing email within 6 hours. This is well below the mean lifetime of a phishing website of 61 hours (Moore and Clayton, 2007). This indicates there is little time to take action, after receiving reports of a phishing email or website. Notice-and-takedown procedures as well as alerts on social media may be too slow in preventing victimisation. This leaves blacklists as the best available method of technical prevention of victimisation. However, blacklisting a phishing website needs to be performed fast if half of the potential victims opens the email within 6 hours.

Fewer reports with phishing emails were received in the weekend compared to weekdays. However, an analysis of the phishing emails showed that fewer were sent in the weekend as well. Furthermore, more phishing mails are sent during office hours in the Netherlands, compare to other moments of the day. Previous research of Bursztein et al. (2014) on account takeover also suggested that the offenders work only during office hours. Our data does show higher activity during daytime, suggesting offenders are likely to be active in the same timezone.

In our dataset, phishing from financial institutions was most prominent, explaining consistently more than 70% of the reported phishing emails. However, their share slowly decreased in 2016 compared to 2015. Having such a large portion of phishing from one particular type of organisation is bad for the vigilance of the general public. They may have trained themselves to recognise phishing mails from financial institutions. When phishing shifts to other industry sectors, the general public may be less aware and therefore be victimised more often. From the ten largest phishing campaigns with a single email, however, three were telecommunication providers. Phishing campaigns misusing financial institutions, even though accounting for a larger share of the total, are less often repeated, resulting in multiple smaller campaigns.

### 6.3.1 *Future Work*

The tool used in this research has been deployed since 2014 and continues to be used by the Fraud Helpdesk to gather information about phishing in the Netherlands. Additionally, the submitted emails are used to inform the general public of the currently active phishing emails.

Future research could include:

- extending the dataset to include malware analysis on the submitted phishing emails (i. e., websites installing malware, or attachments).
- comparing the dataset to existing spam traps and judge the reliability of the dataset, as well as both the proportion of emails being reported and the properties of the reported emails.
- currently, the dataset relies on the general public to report phishing emails. By making it easier to report phishing, more people may be inclined to report phishing, thereby improving the dataset. For example, several large web based email providers include a “Report Phishing” button.

In the application of crime science theories, such as the routine activity theory and the crime pattern theory, to cybercrime, the translation of location or space from the real world to the digital world is non-trivial (Yar, 2005). For example, is the location of a phishing attack the victim’s device, the phishing website, or the offenders device? To avoid such debate, one can measure different variables as a substitute for location, which is difficult to use by itself. For phishing, we propose to use the impersonated organisation as a proxy for the location. This allows modelling victim behaviour regardless of the exact location of the components of a phishing attack (e. g., victim, website, offender). Testing the effectiveness of this approach is subject of future research.

### 6.3.2 *Policy Implications*

The results of this research are important for practitioners and policy makers:

- Organisations should be aware of the behaviour of their employees and take into account that employees will receive phishing emails on their personal email accounts, which they will view during office hours.
- Some moments of the day and in the week are a higher risk when it comes to phishing, especially on Monday morning. Organisations can reduce risk by disallowing employees to open their private email on Monday’s, or all together.

- Blacklists are a common technical method to prevent people from accessing a rogue website. Our study shows that blacklists need to be updated near-realtime in order to stop people from visiting phishing websites.
- Other technical means (such as DMARC, SPF and DKIM) should be promoted or even enforced by policy makers, since they stop the most trivial forms of phishing. When implemented, DMARC disallows the usage of legitimate domains as sender of phishing emails by enforcing SPF or DKIM, which in combination with education, can assist targets in recognising phishing emails. SPF specifies which IP addresses are allowed to send emails for a domain, and DKIM allows email servers to digitally sign emails.
- For organisations allowing abuse reports to be submitted, we recommend to always provide feedback to the submitter. Many of our submitters wanted to be informed about the progress of their report, sometimes even calling the help desk for more information. Providing submitters with detailed and up-to-date information may encourage them to continue reporting abuse.
- Most people indicated that not having a relationship with an organisation is the reason for distrusting an email. Policy makers could target the general public with specific advice on how to recognise illegitimate messages, both offline in the form of letters, and online in the form of spam and phishing emails. Only when people have good heuristics for assessing the legitimacy of an email, can they be sufficiently at moments their vigilance is lower than normal (e. g., due to disturbances or stress).

These recommendations can reduce the risk of phishing attacks, even though phishing will always be present.

Having discussed patterns of phishing in the Netherlands, we now turn to the conclusions of this thesis.



## CONCLUSIONS

---

In this thesis, we researched phishing by performing experiments and measurements at the individual and the national level. We began by giving definitions of phishing that exist in literature and developed a consensual definition: *Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target.* Then, we performed experiments with scalable and less-scalable forms of phishing and analysed decision making of phishing victims in a lab study. We tested a phishing prevention training on children and measured the retention rate. Finally, we presented an overview of phishing in the Netherlands based on the analysis of a large body of phishing emails.

In the remainder of this chapter, we will restate the research questions, discuss our findings and give directions for future work.

### 7.1 DISCUSSION OF RESEARCH QUESTIONS

In this section, we discuss each of the research questions and the corresponding experiments.

**RESEARCH QUESTION 1:** How does an attack's effectiveness relate to the modus operandi's scalability?

For phishing attacks, we conjectured that there is a relationship between the scalability of an attack and the resulting effectiveness. We tested the scalability properties of two forms of attacks that are less-scalable than phishing attacks by email. We performed an experiment to measure what happens when a USB key is dropped on the floor. Picking up and subsequently using a found USB key poses a risk to one's digital security. It turned out that used USB keys were taken in 12% of the cases, versus 41% of the new boxed USB keys. It would be relatively easy for a skilled attacker to seal a USB key with malicious content in a box. In a second experiment, we distributed posters with QR codes targeting employees of a large organisation. Even though the response rate was low, one out of the four people that scanned the QR code fell for the phishing attack. The results of our experiment lead to an updated effectiveness versus scalability figure, as shown in [Figure 29](#).

The ability of an attack to scale easily is related to the personalisation and type of interaction of the attack. One-to-one interaction, such as a face-to-face meeting or a phone call, makes an attack less scalable but more effective. However, the scalability of an attack is viewed mostly

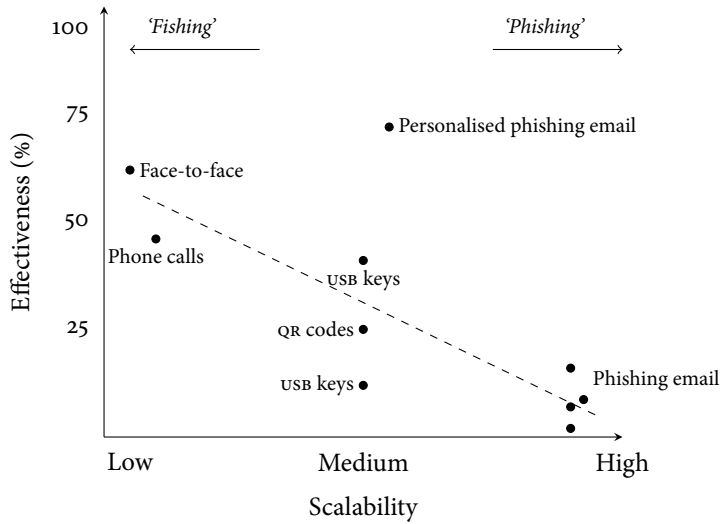


Figure 29: The effectiveness versus the scalability per *modus operandi*, including data from the experiments in this thesis. Hollow circles represent real world data.

from the attacker’s perspective. From the potential victim’s point of view, other factors may influence the effectiveness. One of such factors is the perceived risk by the victim. For example, picking up a USB key may be considered low risk. At the same time, scanning a QR code or clicking on a link in a phishing email, may be considered more risky.

**RESEARCH QUESTION 2:** How do people decide whether or not an email is phishing?

We set up an experiment where participants were asked to think out loud when reading a phishing email. We found three thought patterns of potential victims. Firstly, people assess the believability of the contents of the email and pay little attention to the technical evidence of an email’s authenticity. Secondly, people use their expectations of what an email from the sending organisation should look like, and compare the email with their expectations. Thirdly, when people read an email that introduces urgency, their thoughts become less negative and they are more likely to ignore warning messages. These heuristics show how people become victimised by a phishing email. Following our results, the ‘perfect’ phishing email contains a message that is believable to the receiver, has a writing style matching the expectations of how the supposed sender organisation communicates, and contains urgency cues.

User training should focus on providing users with simple and effective heuristics to recognise those ‘perfect’ phishing emails. For example,

users may feel alerted when they receive an email that contains a sense of urgency. They should be aware that their judgement of such an email may be biased, as has been shown in our experiment. There are other heuristics, such as lack of personalisation, finding spelling mistakes or requests for personal information, but these are easily circumvented with a ‘good’ phishing email. Therefore, we propose that user training consists of two parts: Firstly, simple heuristics to activate their vigilance. Secondly, the know-how to check the technical properties of communication, such as sender address and finding the URL destination for e-mails.

**RESEARCH QUESTION 3:** How can we reduce the effectiveness of phishing on children?

There are many ways of phishing prevention, one of which is user training. We trained children in a school setting on how to distinguish phishing emails from legitimate emails. The trained children scored significantly higher than the untrained children. After four weeks, the subjects’ ability to recognise phishing emails returned to pre-training levels. At the same time, after four weeks there were significant improvements in the subjects’ ability to recognise legitimate emails.

Our study showed that over 80% of the the children (8–13 years) that participated possessed their own email address. This shows the need for cybersecurity education and specifically phishing training that is aimed at children. However, user training in the setting of our experiment results in a lower vulnerability only for a short time. Incorporating cybersecurity as a recurring item in the curriculum, preferably in a more playful way such as educational games, could achieve better results in the longer term.

Our results indicate the feasibility of training children on a complex topic such as phishing emails. The children showed interest in cybersecurity, and learned from the interaction with the researchers. Using story-telling and group discussions worked well when talking about cybersecurity: the children were focussed, shared their own stories and asked questions about the topics.

**RESEARCH QUESTION 4:** What patterns can be found in phishing campaigns in the Netherlands?

To analyse patterns of phishing in the Netherlands, we cooperated with a large anti-fraud agency to collect user reports of phishing emails. This led to a little under 700,000 emails in our dataset to be analysed. After our analysis, the dataset grew to 1.4 million emails. We analysed the presence of phishing campaigns in the dataset and we analysed the timing of the received emails.

Our analyses resulted in two important insights on Dutch phishing. Firstly, after analysing duplicates we found that the offenders use cam-

paings with waves of similar phishing emails. Every campaign consists of on average 3.6 of these waves (or repetitions). Additionally, on average there are three weeks between each consecutive wave of a single email.

The second insight relates to a comparison between our dataset and the dataset of the APWG. A large part of the emails in our dataset (over 70%) claims to be from financial institutions, whereas this number is much lower (around 20%) for the APWG dataset. If we assume that our dataset represents the Dutch situation and the APWG represents either the USA and/or the global state of phishing, this comparison shows that phishing probably is a local activity. This is further supported by increased activity during Dutch daytime.

## 7.2 FUTURE RESEARCH DIRECTIONS

Phishing is a difficult problem to solve, since humans are the weakest link in a system's security. Therefore, technical assistance should support human decision making. This starts with better interfaces. It would be interesting to combine a technical assessment of the email (e. g., with spam filters) with interface design. An email could be shown with a traffic light symbol next to it: green means no threat; orange would indicate an email where not all technical means of validation were successful; and red would mean a likely phishing email. Alternatively, when a recipient is about to be victimised by, for example, clicking on a link in an email, more information can be shown in the screen. This could be in the form of a challenge, such as "out of these five links, which one did you just click on?"; or a simple pop-up showing the exact URL to open.

Another possible line of research is puzzles. Assuming the user is either not fully focussed, or does not even look at the URL, all links in an email would be disabled. To continue opening a link, the user should first solve a small puzzle and be forced to focus on the email. While such interventions are delaying the user, experiments could provide valuable insights that could lead to improved interfaces, and thereby better user decisions.

Email users use a variety of heuristics to decide whether to respond to a phishing email. We have seen that basic training is of limited effectiveness to the ability to recognise phishing. However, users receive phishing emails on a daily basis, and re-affirm their heuristics when they make a correct decision. Future research could investigate the options of showing users well-designed phishing emails on a daily basis, which users can train themselves with.

Finally, email users in the Netherlands are willing to forward phishing emails. It would be interesting when users are able to report phishing to a central authority from within their email client. Email providers



could use the data from such a large body of emails to optimise their spam detection. Using text hashing and pattern matching, suspicious emails with the same contents can be flagged as potentially dangerous.

### 7.3 FINAL WORDS

In this thesis the different aspects of phishing and anti-phishing have been discussed. We collected data on the feasibility of forms of less-scalable phishing, established the heuristics of potential victims, trained children against phishing, and collected and analysed hundreds of thousands of phishing emails. Phishing uses the most powerful bug in computer system: it's owner. No patches and upgrades can remove this vulnerability. However, with the building blocks provided in this thesis, novel social and technical interventions can be built.





## LIST OF ANALYSED DEFINITIONS OF PHISHING

Table 23 lists the 113 definitions of the phenomenon ‘phishing’ that were found with the literature search.

Author	Definition
Adida (2007)	Attackers provide a spoofed web page, where the user is fooled into entering her credentials.
Ahamid, Abawajy and Kim (2013)	Phishing is a type of semantic attack in which victims are sent emails that deceive them into providing sensitive information such as account numbers, passwords, or other personal to phisher.
Al-Hamar, Dawson and Al-Hamar (2011)	(...) a technique of obtaining private information fraudulently and thereafter obtaining money illegally (...)
Ali and Rajamani (2012)	Phishing a fraudulent trick of stealing victim’s personal information by sending spoofed messages, through Instant Messengers via socially engineered messages.
Almomani, Wan et al. (2012)	Such a type of threats, phishing e-mails, is used to steal sensitive and personal data or user’s’ account information from their computers.
Almomani, Gupta et al. (2013)	Phishing is a kind of attack in which criminals use spoofed emails and fraudulent web sites to trick financial organization and customers. Criminals try to lure online users by convincing them to reveal the username, passwords, credit card number and updating account information or fill billing information.
Amin, Ryan and Dorp (2012)	email soliciting personal information
Anderson and Moore (2009)	(...) in which crooks send emails pretending to be from a bank or service provider and inviting its customers to log on at its website.
Bainbridge (2007)	Obtaining information such as a person’s bank account details by sending an e-mail purporting to be from that person’s bank.
Baker, Tedesco and Baker (2006)	(...) the fraudulent and increasingly authentic looking e-mail attempts aimed to lure unsuspecting recipients into sharing sensitive financial and personal information.
Barracclough et al. (2013)	Phishing is an instance of social engineering techniques used to deceive users into giving their sensitive information using an illegitimate website that looks and feels exactly like the target organization website.

*Continued on next page.*

*(continued)*

Author	Definition
Basnet, Mukkamala and Sung (2008)	Phishing is a form of identity theft that occurs when a malicious Web site impersonates a legitimate one in order to acquire sensitive information such as passwords, account details, or credit card numbers.
Beatty et al. (2011)	In a typical phishing scam, the consumer receives an email purportedly from a trusted online vendor (a bank is a typical example). This email contains a call to action, a request to undertake some action that requires the user to disclose their authentication credentials. A hyperlink to the vendor's (supposed) site is provided. Consumers fall prey to this scam when they follow the link and provide their credentials.
Beliakov, Yearwood and Kelarev (2012)	Phishing usually involves acts of social engineering attempting to extract confidential details by sending emails with false explanations urging users to provide private information that will be used for identity theft.
Bergholz et al. (2010)	Phishing emails usually contain a message from a credible looking source requesting a user to click a link to a website where she/he is asked to enter a password or other confidential information.
Biddle, Chiasson and Van Oorschot (2012)	Phishing is a type of social engineering in which users are tricked into entering their credentials at a fraudulent website recording user input.
Brainard et al. (2006)	the fraudulent use of e-mail to capture user passwords (and other information)
Butler (2007)	Phishing represents an online method of identity theft employed by phishers to steal attributes (like passwords or account numbers) used by online consumers.
Cao, Han and Le (2008)	The attacker tricks the user into submitting his/her confidential information (such as password) into a fraudulent web site that has high visual similarities as the genuine one.
Chen et al. (2009)	Phishing is a form of online identity theft associated with both social engineering and technical subterfuge. Specifically, phishers attempt to trick Internet users into revealing sensitive or private information, such as their bank account and credit-card numbers.
Cranor (2008)	Phish e-mails are constructed by con artists to look like legitimate communications, often from familiar and reputable companies, and usually ask victims to take urgent action to avoid a consequence or receive a reward. The desired response typically involves logging in to a Web site or calling a phone number to provide personal information. Sometimes victims need only click on links or open e-mail attachments for their computers to become infected by malicious software –known as malware– that allows phishers to retrieve the data they want or take control of the victim's computer to launch future attacks.

*Continued on next page.*

*(continued)*

Author	Definition
Dhamija and Tygar (2005)	In a phishing attack, the attacker spoofs a website (e.g., a financial services website). The attacker draws a victim to the rogue website, sometimes by embedding a link in email and encouraging the user to click on the link. The rogue website usually looks exactly like a known website, sharing logos and images, but the rogue website serves only to capture the user's personal information. Many phishing attacks seek to gain credit card information, account numbers, usernames and passwords that enable the attacker to perpetrate fraud and identity theft.
Dhamija, Tygar and Hearst (2006)	The practice of directing users to fraudulent web sites.
Dong, Clark and Jacob (2010)	Phishing attacks are well-organised and financially motivated crimes which steal users' confidential information and authentication credentials.
Downs, Holbrook and Cranor (2006)	Phishing emails are semantic attacks that con people into divulging sensitive information using techniques to make the user believe that information is being requested by a legitimate source.
Downs, Ademaj and Schuck (2009)	Attempts to criminally obtain sensitive information (e.g., social security numbers and credit cards) by pretending to be a legitimate businesses.
Drake, Oliver and Koontz (2004)	"Phishing" is an email scam that attempts to defraud people of their personal information including credit card number, bank account information, social security number, and their mother's maiden name.
Egelman, Cranor and Hong (2008)	a scam to collect personal information by mimicking trusted websites
Elmaleh (2007)	This type of unsolicited correspondence has the intention of directing users to a fake web site, facilitating the unauthorised retrieval of personal financial information which can then be used to fraudulently access a user's bank account.
Emm (2006)	It involves tricking computer users into disclosing their personal details (username, password, PIN number or any other access information) and using these details to obtain money under false pretences.
Fernandez et al. (2005)	(...) in which a perpetrator sends an e-mail purporting to be from the victim's Internet service provider, bank, or other company with whom the victim does business. The e-mail asks the victim to update his account information. When the victim complies with the request, he will have unwittingly sent his personal information to a criminal.
Fette, Sadeh and Tomasic (2007)	(...) attacks are launched with the aim of making web users believe that they are communicating with a trusted entity for the purpose of stealing account information, logon credentials, and identity information in general.

*Continued on next page.*

*(continued)*

Author	Definition
Florêncio and Herley (2007)	(...) a victim is lured into submitting her password to a malicious site masquerading as a trusted institution (...)
Forte (2009)	(...) the objective of which is to trick us into revealing sensitive information.
Fumera, Pillai and Roli (2006)	(...) they try to convince them to surrender personal information like passwords and account numbers, through the use of spoof messages which are masqueraded as coming from reputable on-line businesses such as financial institutions.
Garera et al. (2007)	Phishing is form of identity theft that combines social engineering techniques and sophisticated attack vectors to harvest financial information from unsuspecting consumers.
Gastellier-Prevost and Laurent (2011)	By spoofing the identity of a company that proposes financial services, phishing attacks steal confidential information (e.g. login, password, credit card number) to the Internet users
Geer (2005)	(...) phishing, in which e-mails lure unsuspecting victims into giving up user names, passwords, Social Security numbers, and account information after linking to counterfeit bank, credit card, and e-commerce Web sites.
Gouda et al. (2007)	In this type of attack, an attacker sends fraudulent emails to users, pretending to be the system administrator of a benign website such as an online banking website, and fools users to take login actions on a malicious website, which looks very similar to the benign website, but is set up by the attacker. Once a user tries to login on such a malicious website, his user name and password will be recorded and possibly later will be used by the attacker to login on the benign website.
Gross and Rosson (2007)	Phishing involves an attacker, posing as bank, vendor, or other trusted source, who sends an email asking the recipient to "confirm" personally identifying information by entering it on a website. This information is then used in identity theft.
Guan, Wu and Wang (2012)	(...) the malicious mail, which mostly contains a URL to convince the victims to visit a fraudulent website where sensitive information like credit card numbers and passwords are requested.
Gupta and Pieprzyk (2011)	Phishing is the process of covertly and illicitly obtaining user credentials for future gains.
Halevi, Lewis and Memon (2013)	Phishing is an attack that uses fraudulent electronic mail (email) that claims to be from a trustworthy source. The goal of phishing emails is to get personal information from the users, such as user ID and passwords. The attacker can then use this information to impersonate a user and access the user account for financial gain.
Han et al. (2012)	Phishing employs social engineering to trick a user into revealing his or her web digital identities to a fraudulent web site.
He et al. (2011)	Phishing usually takes a form of a fake webpage whose appearance is similar to the page of a real website in order to steal user credentials and identities.

*Continued on next page.*

*(continued)*

Author	Definition
Herzberg (2009)	Password theft via fake websites.
Hinson (2010)	using spam e-mails, targeted e-mails, short message service (SMS) text messages, phone calls, and even leaflets on the windscreen to fool victims into visiting fake websites and disclosing their login credentials or other personal information
Hodgson (2005)	Phishing attacks simulate established and reputable organisation's Web sites and trick the user into providing personal information that is then used by the criminal to either steal from the victim or use the victim's identity to commit further crimes.
Hong (2012)	Phishing is a kind of social-engineering attack in which criminals use spoofed email messages to trick people into sharing sensitive information or installing malware on their computers.
Huber et al. (2011)	An attacker tries to lure victims into entering sensitive information, such as a password or credit-card number, into a fake website that the attacker controls.
Ilchev and Ilchev (2012)	(...) a popular approach used by criminals to acquire sensitive client data such as personal identification numbers (PINs), transaction authentication numbers (TANs), bank account numbers, credit card numbers and passwords.
Jagatic et al. (2007)	Phishing is a form of deception in which an attacker attempts to fraudulently acquire sensitive information from a victim by impersonating a trustworthy entity.
Jahankhani (2009)	This is a technique used to gain personal information for the purposes of identity theft, using fraudulent e-mail messages that appear to come from legitimate businesses. These authentic-looking messages are designed to fool recipients into divulging personal data such as account numbers and passwords, credit card numbers and Social Security numbers.
Jakobsson and Ratkiewicz (2006)	persuades a user to release sensitive personal or financial information, such as login credentials or credit card numbers.
Jakobsson and Stamm (2007)	Phishing combines the deceitful techniques of con artists with the Internet's scalability to commit identity theft by stealing credentials.
Jo, Jung and Yeom (2013)	Phishing is an attack where fraudulent websites impersonate legitimate counterparts to steal users' confidential information.
Khonji, Iraqi and Jones (2013)	Phishing is a type of computer attack that communicates socially engineered messages to humans via electronic communication channels in order to persuade them to perform certain actions for the attacker's benefit.
Khot, Kumaraguru and Srinathan (2012)	(...) attacker tricks the user into divulging the password information through fraudulent websites and emails.

*Continued on next page.*

*(continued)*

Author	Definition
Kim et al. (2012)	(...) attempts to steal confidential user information such as credit card numbers or passwords and social engineering and spoofing techniques are frequently used.
Kirda and Kruegel (2005)	Phishing is a form of online identity theft that aims to steal sensitive information from users such as online banking passwords and credit card information.
Kirlappos and Sasse (2012)	Tricking computer users to disclose personal information, credit card details, user names, and passwords.
Knight (2005)	The practice is known as phishing, and uses social engineering and technical subterfuge to steal consumers' personal data and bank account details.
Kumaraguru, Rhee, Acquisti et al. (2007)	Criminals lure Internet users to websites that impersonate legitimate sites
Kumaraguru, Sheng et al. (2010)	(...) phishing, in which victims get conned by spoofed emails and fraudulent websites.
Larcom and Elbirt (2006)	Phishing is the act of convincing users to provide personal identification information such as credit card numbers, social security numbers and bank account information for explicit illegal use.
Lenton (2005)	(...) rogue emails usually purporting to be from a bank that direct them to a bogus website or attempt to identify their personal details
Levy (2004)	Phishing (the act of conning a person into divulging sensitive information) commonly uses legitimate-looking Web sites that mimic the online interface of the institution the attacker is misrepresenting (usually a bank, merchant, or ISP)
Li, Helenius and Berki (2012)	Phishing is one type of identity theft, where the aim is to steal confidential information, e.g. credit card number, credentials and social security ID numbers, and the list can go on.
Liu, Guanglin et al. (2005)	Phishing is a criminal trick of stealing victims' personal information by sending them spoofed emails urging them to visit a forged webpage that looks like a true one of a legitimate company and asks the recipients to enter personal information such as credit card number, password and etc.
Liu, Qiu and Wenyn (2010)	Phishing is a kind of online attack widely used by phishers to steal users' accounts and passwords, and other personal information for illegal appropriation.
Ludl et al. (2007)	Phishing is a form of electronic identity theft in which a combination of social engineering and web site spoofing techniques are used to trick a user into revealing confidential information with economic value.
Maurer and Höfer (2012)	(...) the act of stealing personal data of Internet users for misuse (...)

*Continued on next page.*



(continued)

Author	Definition
McFedries (2006)	"Phishing" refers to creating a replica of an existing Web page to fool users into submitting personal, financial, or password data to what they think is their bank or a reputable online retailer.
McNealy (2008)	The sender creates e-mails, resembling those from a well-known companies, requesting that the recipient click on a URL provided, which links to a dummy company Web site where the recipient is asked to input personal information. The e-mail sender may then use the information for illegal purposes.
Mills and Byun (2006)	Stealing personal information by requesting it via fraudulent email messages or Web pages
Mohebzada et al. (2012)	Phishing is a type of social engineering where a potential victim is sent a message that impersonates a legitimate source or organization. Phishing attacks typically lure the targets into revealing confidential information such as password, credit card details, bank account numbers, or any other sensitive information.
Moore (2007)	Phishing is the process of enticing people into visiting fraudulent websites and persuading them to enter identity information such as usernames and passwords. This information is then used to impersonate the victim (...)
Moore and Clayton (2007)	Phishing is the process of enticing people into visiting fraudulent websites and persuading them to enter identity information such as usernames, passwords, addresses, social security numbers, personal identification numbers (PINs) and anything else that can be made to appear to be plausible.
Moran and Moore (2010)	Phishing is the criminal activity of enticing people to visit websites that impersonate genuine bank websites and dupe visitors into revealing passwords and other credentials.
Nykodym et al. (2010)	Phishing is a scam to steal valuable information by sending out fake emails, or spam, written to appear as if they have been sent by banks or other reputable organizations with the intent of luring the recipient into revealing sensitive information such as usernames, passwords, social security numbers, account IDs, ATM PIN's or credit card details.
Olurin, Adams and Logrippo (2012)	Fraudsters can create fake websites to lure users for the purpose of collecting their data. (...) Phishing attacks can steal personal identity information such as username, passwords, and credit card details from unsuspecting users by masquerading as trusted entities, such as PayPal sites.
Parno, Kuo and Perrig (2006)	In phishing, an automated form of social engineering, criminals use the Internet to fraudulently extract sensitive information from businesses and individuals, often by impersonating legitimate web sites.
Paulson (2010)	Phishers typically create webpages that look like those belonging to banks, e-commerce operations, or other businesses on which users might enter financial or accountaccess information. When a user enters such data on a fake page, the phisher captures the information and utilizes it to defraud the victim.

*Continued on next page.*

*(continued)*

Author	Definition
Piper (2007)	Phishing is an attempt provided by vendors using email or Internet social spaces such as MySpace to obtain sensitive personal information such as usernames and passwords, social Security Numbers, credit-card numbers, and others.
Ranganayakulu, Kavisankar and Chellappan (2011)	Phishing is the combination of social engineering and technical exploits which has adverse effects aiming at the monetary gain of the attacker (phisher). (...) Phishing attacks use spoofed e-mails and fraudulent websites designed to fool recipients into divulging personal financial data such as credit card numbers, account usernames and passwords, social security numbers, etc.
Ray and Schultz (2007)	Phishing is a technique that many attackers use to trick computer users into revealing personal or financial information through specially worded email messages or websites.
Ross (2006)	(...) in which con artists send e-mails purporting to be from legitimate organizations, such as banks, in order to inveigle recipients into revealing personal information.
Ross (2009)	Phishing e-mails deceive individuals into giving out personal information which may then be utilized for identity theft.
De Ryck et al. (2013)	(...) the process that involves an attacker tricking users into willingly surrendering their credentials (...)
Saberi, Vahidi and Bidgoli (2007)	Phishing attack is a kind of identity theft which tries to steal confidential data like on-line bank account information.
Shahriar and Zulkernine (2012)	Phishing is a web-based attack that allures end users to visit fraudulent websites and give away personal information (e.g., user id, password)
Emilin Shyni and Swamynathan (2013)	A phishing attack is a criminal activity which mimics a certain legitimate webpage using a fake webpage with an intention of luring end-users to visit the fake website thereby stealing their personal information such as usernames, passwords and other personal details such as credit card information.
Sood, Sarje and Singh (2011)	Phishing is an online identity theft that combines social engineering and web site spoofing techniques to cheat the user by redirecting his confidential information to an untrusted destination.
Stabek, Watters and Layton (2010)	(...) which are also synonymous with identity theft and credit/debit card fraud.
Sweeney (2006)	Phishing, which is the act of sending an email message impersonating a respected organization in an attempt to get the reader to click on the provided link and give personal information.
Thiyagarajan, Aghila Prof. and Prasanna Venkatesan (2012)	In this attack, the attacker tries to mimic as legitimate site and gather critical information from the user which in turn will be used to make control of the user's valuable and critical information.

*Continued on next page.*

(continued)

Author	Definition
Vamosi (2009)	Phishing refers to an attempt to collect usernames, passwords, and credit card data by posing as a legitimate, trusted party.
Varshney, Joshi and Sardana (2012)	Phishing is a deception technique used by attackers for gaining personal information from end users, with the help of fraudulent and spoofed emails, Phished Websites and various deception techniques. The aim of the phisher lies in obtaining personal information or credentials from an end user such as bank account numbers their passwords, credit card details etc.
Verma, Shashidhar and Hossain (2012)	Phishing is a social engineering threat aimed at gleaning sensitive information such as user names, passwords and financial information from unsuspecting victims. Attacks are typically carried out via communication channels such as email or instant messaging by attackers masquerading as legitimate and trustworthy entities.
Vitaliev (2010)	fraudulent messages that attempt to withdraw personal and financial information from the reader.
Wang, Herath et al. (2012)	Email-based deception where a perpetrator (phisher) camouflages emails to appear as a legitimate request for personal and sensitive information is known as phishing.
Wenyin et al. (2012)	Phishing is the criminally fraudulent process of attempting to acquire sensitive information such as user names, passwords, and creditcard details from a victim by pretending to be a trustworthy entity in an electronic communication.
Whittaker, Ryner and Nazif (2010)	We define a phishing page as any web page that, without permission, alleges to act on behalf of a third party with the intention of confusing viewers into performing an action with which the viewer would only trust a true agent of the third party.
Workman (2008)	Phishing is a ruse designed to gain sensitive information from an intended victim by way of e-mail and Web pages or letters that appear to be from genuine businesses, that command the potential victim to supply information to prevent an account from being closed, or as part of a promotion or give-away called a gimmie.
Wu, Miller and Garfinkel (2006)	Phishing attacks typically use legitimate-looking but fake emails and websites to deceive users into disclosing personal or financial information to the attacker. Users can also be tricked into downloading and installing hostile software, which searches the user's computer or monitors online activities to steal private information.
Wu, Miller and Little (2006)	Phishing attacks typically use legitimate-looking but fake emails and websites to deceive users into disclosing private information to the attacker.
Xiang and Hong (2009)	Phishing is a form of identity theft, where criminals create fake web sites that masquerade as trustworthy organizations. The goal of phishing is to trick people into giving sensitive information, such as passwords, personal identification numbers, and so on.

*Continued on next page.*

*(continued)*

Author	Definition
Xiang, Hong et al. (2011)	Phishing is a form of identity theft, in which criminals build replicas of target Web sites and lure unsuspecting victims to disclose their sensitive information like passwords, personal identification numbers (PINs), etc.
Yearwood et al. (2009)	Phishing can be defined as a scam by which an email user is duped into revealing personal or confidential information which the scammer can use illicitly. Phishing attacks use both social engineering and technical subterfuge to steal personal identity data and financial account credentials.
Yee and Sitaker (2006)	(...) phishing attacks, in which the user is fooled into entering a password at an imitation site.
Zhang, Hong and Cranor (2007)	A kind of attack in which victims are tricked by spoofed emails and fraudulent web sites into giving up personal information.
Zhang, Wu et al. (2012)	By masquerading as a trustworthy entity, phishing is a criminally fraudulent process of attempting to acquire sensitive information.
Zhou, Leckie and Karunasekera (2009)	Phishing is a form of social engineering attack, which exploits human vulnerabilities rather than software vulnerabilities.

Table 23: An overview of the articles and papers that define phishing (N=113).

The following questions were asked after the subjects performed the task. The questions were originally in Dutch and translated by the researchers into English.

1. Are there thoughts that you did not verbalise during the task itself, but that you want to share now?
2. Can you formulate what the central message of the email was?
3. What was the goal of the email in your opinion?
4. Can you indicate whether you experienced difficulties performing the task at any point during the experiment?
5. Do you remember the salutation (greeting) of the email?
6. Did you consider the salutation usual for an email from this kind of organisation?
7. Do you remember the valediction of the email?
8. Did you consider the valediction usual for an email from this kind of organisation?
9. Do you remember the sender of the email?
10. To what extent do you consider the sender reliable?
11. Certain information is given in the email. To what extent do you consider the given information reliable?
12. At the end of the email, you are requested to perform an action. Do you recall what you were requested to do?
13. Would you have performed the action?
14. Can you indicate at which moments, in your opinion, you had to make a decision?
15. Which decision did you make?
16. Why did you make this decision?
17. Which alternatives did you consider?

After these questions about the contents, we asked six questions regarding the study itself.

18. Did you have sufficient time to perform the task?
19. Did you consider it difficult or annoying to perform the task?
20. Do you think that performing the task would have been easier without the use of a voice recorder?
21. Do you think that performing the task would have been easier without my presence?
22. Do you feel that you have were able to sufficiently express your thoughts and actions?
23. Is there anything else you would like to add?

## PHISHING EDUCATION TEACHING AND TESTING MATERIAL

---

This appendix includes extra material from [Chapter 5](#). Firstly, [Section C.1](#) describes several assumptions in the statistical analysis. Secondly, [Section C.2](#) includes the slides of the presentation that was given to the subjects. Finally, [Section C.3](#) includes one of the phishing tests used in the experiment.

### C.1 STATISTICAL ASSUMPTIONS

Firstly, we checked whether the three phishing tests are equivalent in terms of testing a subject's ability to distinguish phishing from legitimate. Even though the tests were developed in a similar manner, the questions may not be equally difficult. An independent group t-test was used on all combinations of tests: AB; AC; and BC. For the comparisons between tests A and C and tests B and C, the variances were heterogeneous and therefore the Satterthwaite (1946) approximation was used in the t-test.

In a *box plot* (Tukey, 1977), the top of each box represents the 75<sup>th</sup> percentile ( $Q_3$ ), i. e., the median of the upper half of the observations. The bottom represents the 25<sup>th</sup> percentile ( $Q_1$ ), i. e., the median of the bottom half of the observations. The band near the middle of the box is the median of all observations. The lines above and below the box are called *whiskers*. The whiskers represent the top or bottom 25%, i. e., the lower whisker ends at the minimum value and the higher whisker indicates the maximum value. When the data contains outliers, the position of the whiskers can be calculated using the Interquartile Range (IQR):

$$IQR = Q_3 - Q_1 \quad (1)$$

In the presence of outliers, the position of the lower whisker is calculated as  $Q_1 - 1.5 \times IQR$ . The upper whisker is calculated in a similar way:  $Q_3 + 1.5 \times IQR$ . An observation  $v$  is considered an outlier if one of the two following conditions hold:  $v < 1.5 \times Q_3$  or  $v > 1.5 \times Q_3$ . In the box plot, outliers are indicated using dots.

To use a t-test, two assumptions must be met: (1) the data must be normally distributed; and (2) there should be homogeneity of variance. To test whether the data was normally distributed, we used the Shapiro-Wilk test. A non-significant result on the Shapiro-Wilk test indicates that the sample distribution is not significantly different from a normal

distribution. In case of non-normality, the Wilcoxon rank sum test (also known as the Mann–Whitney test) was used instead (Wilcoxon, 1945; Mann and Whitney, 1947). Testing the homogeneity of the variances was done using Levene's test. A non-significant result on Levene's test indicates that the variances are roughly equal, i. e., not significantly different. If the variances were heterogeneous, the Satterthwaite (1946) approximation was used, assuming non equal variances.

Linear regression was used for testing relations with multiple independent variables, or when a variable had more than two possible categories. Several assumptions needed to be checked for each regression (UCLA Statistical Consulting Group, 2016). Firstly, there should be no unusual and influential data. In case of outliers, their effect was measured by performing a regression with and without them. The second assumption is that the residuals are normally distributed, which was tested using Shapiro-Wilk test and a visual inspection of the standardised normal probability plot (UCLA Statistical Consulting Group, 2016). The homoscedasticity was tested using a visual inspection of a plot of the residuals versus the predicted values (Osborne and Waters, 2002). Multicollinearity was tested by using the variance inflation factor (VIF), where a value above the cut-off value of 10 indicates the need for further investigation, even though a higher VIF is not bad per se (O'Brien, 2007). Linearity was visually checked using scatter plots to plot the standardised residuals against each of the predictor variables. Finally, for any two observations, the residual terms should be independent. The scores are not independent, since the same pupils filled in the tests twice. The subjects filled in the tests anonymously, therefore assuming independence will result in conservative estimates regarding the significance and power (see also [Section 5.3.1](#)).

To analyse the variables Sex and HasEmail and HasFacebook, a t-test was used. Both variables were normally distributed and had homogeneity of their variances. The predictor HasFacebook was not normally distributed, and therefore the Wilcoxon signed-rank test was used. For the predictors with several values a linear regression model was calculated. These predictors were: Age (in years); having received a phishing email (yes, no, or don't know); and School. The predictor Age met the assumptions for a linear regression, whereas the ReceivedPhishing predictor and School failed the normality of the residuals check. For the regressions using ReceivedPhishing and School, we used robust standard errors to estimate the standard errors using the Huber-White sandwich estimator (StataCorp, 2013). The questions on whether the subjects have an email address and previously received a phishing message were included in tests A and B only. Therefore, the number of subjects in the analysis varied due to missing values for some subjects.



C.2 SLIDES OF PRESENTATION



## Wat is Cyber Criminaliteit?

---

Gewone criminaliteit, maar:

Op de computer  
Via het internet  
Op een mobiel



UNIVERSITEIT TWENTE.

## In alle soorten en maten

---

**Cyberpesten**

**Phishing**

**Hacken**

**Identiteitsfraude**



UNIVERSITEIT TWENTE.

## Cyberpesten

---

**Video over cyberpesten**

<http://www.youtube.com/watch?v=BWP6Mdnr7is>

UNIVERSITEIT TWENTE.

## Cyberpesten

Wat kun je eraan doen?

---



- Blokkeer de pester
- Praat erover met een ouder/leraar
- Houd je persoonlijke informatie persoonlijk

UNIVERSITEIT TWENTE.

## phishing

---

### Video over phishing

<http://www.youtube.com/watch?v=VcbHo0EOtkA>

UNIVERSITEIT TWENTE.

## phishing

Wat is phishing?



Persoonlijke informatie stelen via:

- Email
- Websites
- Telefoon

Wat voor soort informatie?

Waarom?

UNIVERSITEIT TWENTE.

## Hoe herken je phishing?

Controleer de bestemming van je link:  
links onderin de browser te zien



UNIVERSITEIT TWENTE.

## Hoe herken je phishing?

### Slechte grammatica/spelling

**Beste Klant,**

Er is een Strafrechtelijk onderzoek naar uw gestart met als Dossiernummer 897652

Welkamp wendde zich tot ons betreffende een vordering welke nog niet door u is voldaan, hoewel de betalingstermijn reeds is overschreden. De vordering is aan ons incassobureau overgedragen met het verzoek de schuld te incasseren, eventueel middels een gerechtelijke procedure. Onlangs bent u hierover per brief geïnformeerd. U heeft hierbij geen gehoor gegeven aan de betalingstermijn.

Om verdere kosten te voorkomen bevestigen wij dat wij het verschuldigde bedrag ad. € 695,45 binnen 3 dagen te incasseren van uw rekening.

**KLIK HIER OM BEZWAAR TE MAKEN**



UNIVERSITEIT TWENTE.

## Hoe herken je phishing?

### Dreigende mails

**facebook**

Beste iemand,

We hebben opgemerkt dat je al een tijd niet hebt ingeloopt op je profiel. Als je niet binnen 24 uur inloopt via de onderstaande link inloopt, wordt je profiel verwijderd!

Log nu in:  
<http://privacy.facebook.org/login>

Bedankt voor je medewerking.

---

This message was sent to [jemand@hotmail.com](mailto:jemand@hotmail.com). If you don't want to receive these emails from Facebook in the future, please [unsubscribe](#).  
Facebook, Inc., Attention: Department 415, PO Box 10005, Palo Alto, CA 94303

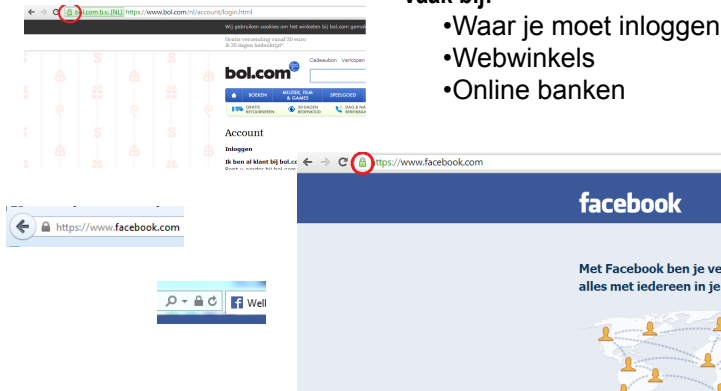
UNIVERSITEIT TWENTE.

## Hoe herken je phishing?

### Veilige websites

Vaak bij:

- Waar je moet inloggen
- Webwinkels
- Online banken



UNIVERSITEIT TWENTE.

## Weet je ze allemaal nog?

Slechte grammatica  
 Verkeerde spelling  
 Dreigende taal  
 Link controleren  
 Eventueel een sleuteltje/https?

Extra tips:  
 Wees voorzichtig met onbekende afzenders  
 Klik niet op elke link

UNIVERSITEIT TWENTE.

## Wat is Hacken?

---

**Computers misbruiken:  
Je cijfers op de schoolpc  
veranderen  
Op een emailadres inbreken  
Een website hacken**



UNIVERSITEIT TWENTE.

## In het nieuws

---

**12-jarige hacker uit Canada  
60 miljoen dollar schade  
Hackte voor videogames**



UNIVERSITEIT TWENTE.



## Identiteitsfraude

---

Facebook profiel onder andere  
naam  
Email account stelen  
Creditcard gegevens stelen



UNIVERSITEIT TWENTE.

## Vragen?

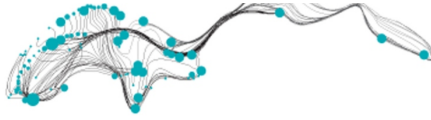
---



UNIVERSITEIT TWENTE.

### C.3 PHISHING TEST

A version of the tests that were printed on paper and distributed amongst the children, is included below. The alternative test was slightly modified to include (partly) different brands in a different order.

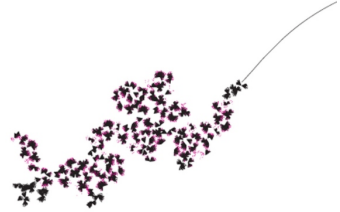


## HOE GA JIJ OM MET HET INTERNET?

In deze vragenlijst krijg je 6 e-mails en 4 websites te zien. Bij elke e-mail of website kun je uitkiezen wat jij er mee zou doen, er zijn 2 antwoorden waaruit je kunt kiezen.

Alle e-mails zijn gestuurd naar het e-mailadres iemand@gmail.com, je mag tijdens deze test doen alsof dit jouw e-mailadres is.

In de vragenlijst zie je soms [Jouw naam] staan. Hier hoeft je je naam niet in te vullen, dit betekent dat de e-mail aan jou gericht is.



Als je een vraag hebt, steek dan je hand op. Er komt dan iemand naar je toe om je te helpen.

Leeftijd:

Jongen ☐ Meisje ☐

Heb je een eigen e-mailadres?

Ja ☐

Nee ☐

Heb je een Facebook account?

Ja ☐

Nee ☐

Heb je wel eens een phishing mail gekregen?

Ja ☐

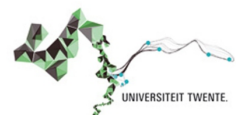
Nee ☐

Weet ik niet ☐

Op de volgende bladzijde begint de test...

Hoe ga jij om met het internet?

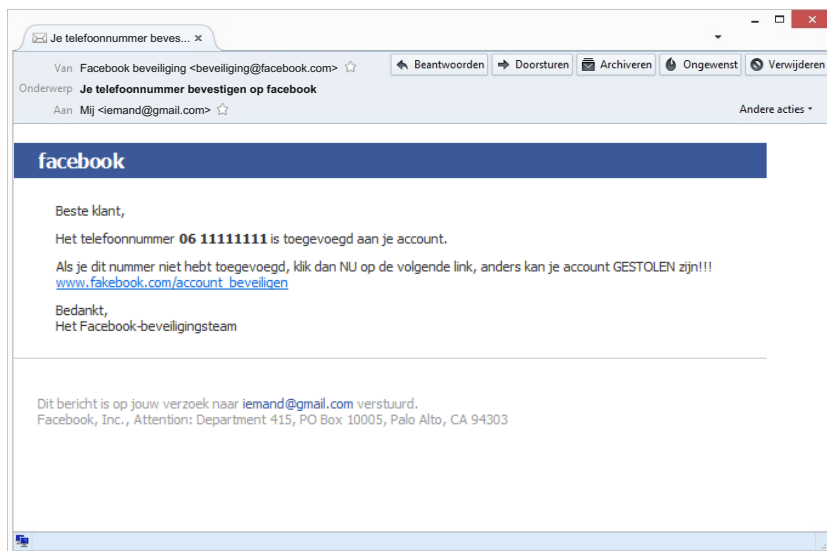
Voor vragen kun je altijd je hand opsteken.





## EEN EMAIL VAN FACEBOOK

Je hebt de volgende email ontvangen:



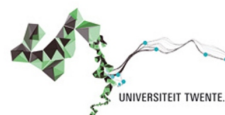
In de e-mail staat dat het telefoonnummer 06 11111111 is toegevoegd aan jouw account, maar dit is niet jouw telefoonnummer. Wat zou jij met deze e-mail doen?

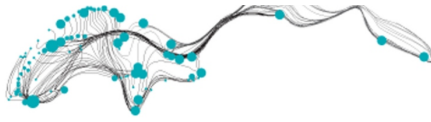
1. Ik verwijder de e-mail, want ik denk niet dat facebook de e-mail gestuurd heeft.
2. Ik klik op de link, om te voorkomen dat mijn account gestolen is.

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

Voor vragen kun je altijd je hand opsteken.





## EEN EMAIL VAN BOL.COM

Je hebt de volgende email ontvangen:



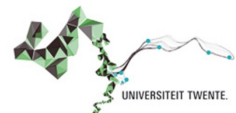
Wat zou je met deze email doen?

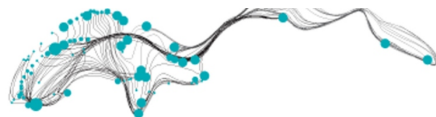
1. Ik verwijder de email, hij zal vast niet echt van Bol.com komen
2. Ik klik op de link om hopelijk een Playstation 4 te winnen

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

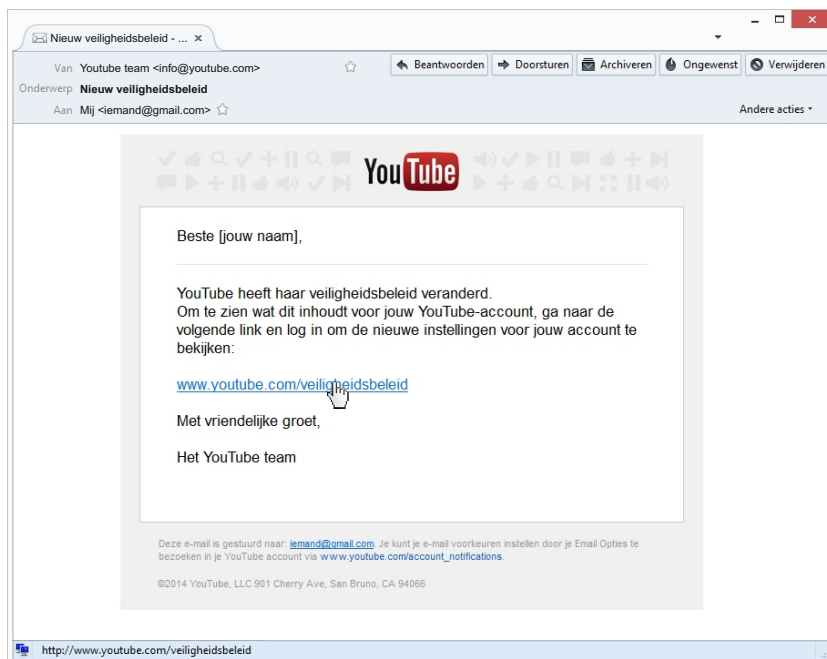
Voor vragen kun je altijd je hand opsteken.





## EEN EMAIL VAN YOUTUBE

Je hebt de volgende email ontvangen:



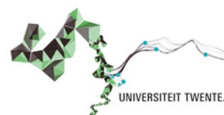
Wat zou je met deze email doen?

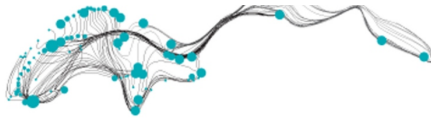
1. Ik klik niet op de link, want ik denk niet dat deze e-mail door Youtube is gestuurd.
2. Ik klik op de link om de nieuwe veiligheidsinstellingen te bekijken.

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

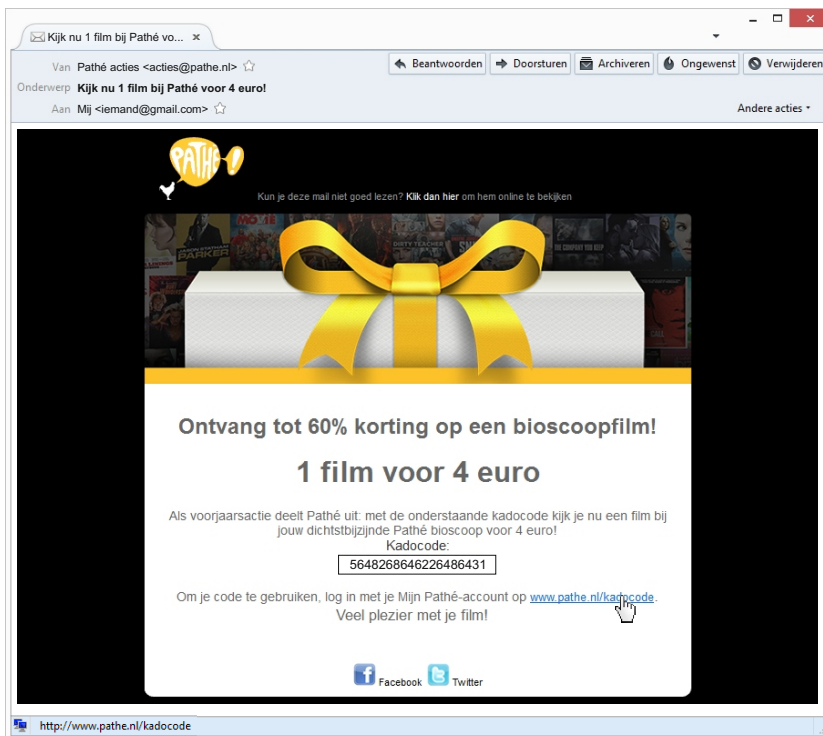
Voor vragen kun je altijd je hand opsteken.





## EEN EMAIL VAN PATHÉ

Je hebt de volgende email ontvangen:



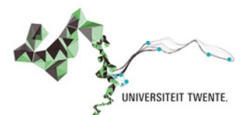
Wat zou je met deze email doen?

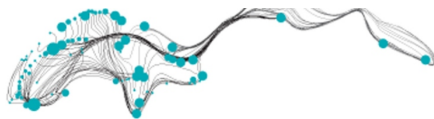
1. Ik klik op de link in de e-mail, om de kadocode te gebruiken.
2. Ik negeer deze e-mail, hij lijkt me niet echt.

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

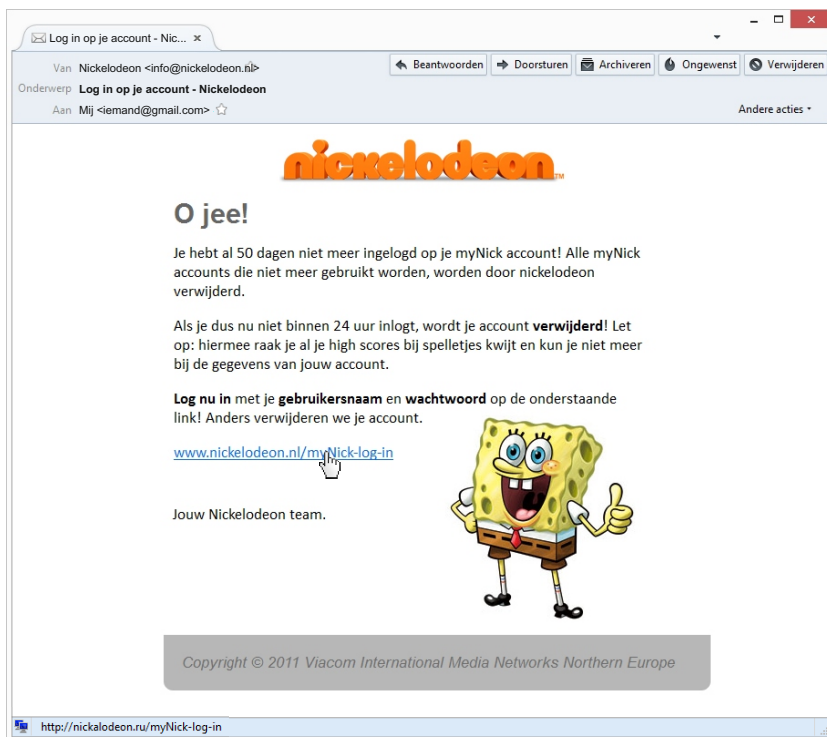
Voor vragen kun je altijd je hand opsteken.





## EEN EMAIL VAN NICKELODEON

Je hebt de volgende email ontvangen:



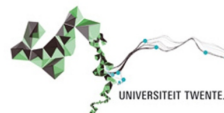
Wat zou je met deze email doen?

1. Ik negeer de mail, hij lijkt me niet echt.
2. Ik klik op de link en log in, want ik wil niet dat mijn account verdwijnt.

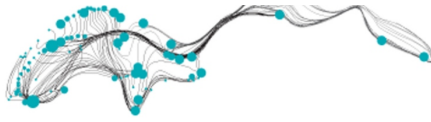
Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

Voor vragen kun je altijd je hand opsteken.

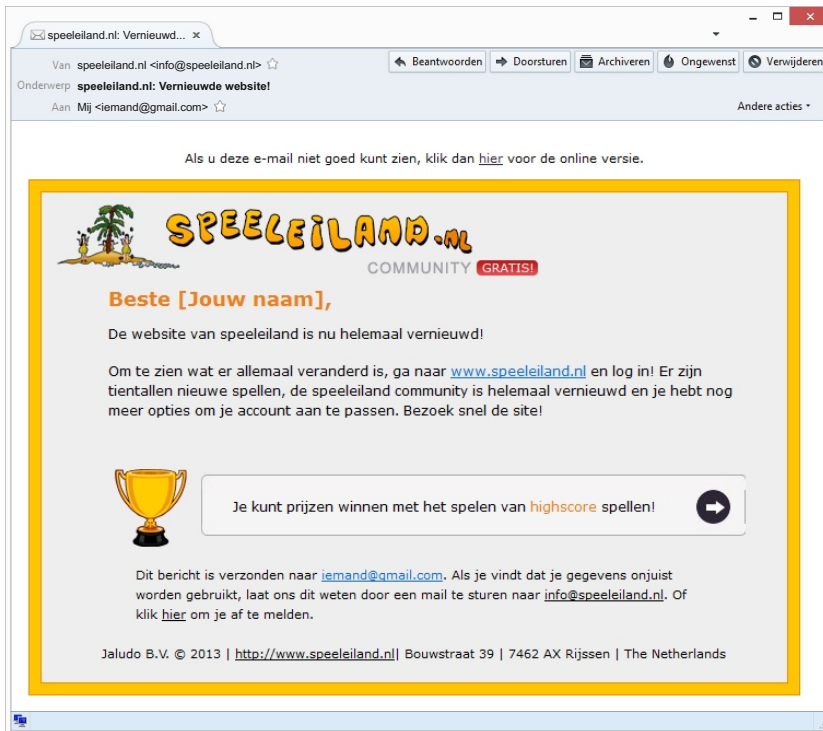






## EEN EMAIL VAN SPEELEILAND

Je hebt de volgende email ontvangen:



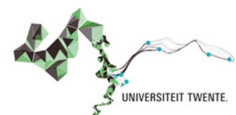
Wat zou je met deze email doen?

1. Ik verwijder deze e-mail, deze e-mail ziet er niet echt uit.
2. Ik klik op de link in de e-mail om de vernieuwde website te bekijken!

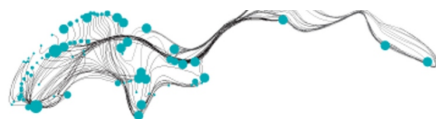
Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

Voor vragen kun je altijd je hand opsteken.



UNIVERSITEIT TWENTE



## WEBSITE 1: LIVE

Je wilt inloggen op je e-mail account bij live.nl. Je ziet de website hieronder.

Aanmelden x

← → ↻ <https://login.live.com/login>

**Outlook**

Microsoft-account Wat is dit?

☐ Aangemeld blijven

**Aanmelden**

Heb je geen toegang tot het account?  
Meld je aan met een code voor eenmalig gebruik

Geen Microsoft-account? [Registreer je nu](#)

---

Microsoft

©2014 Microsoft   Gebruiksrechtovereenkomst   Privacy en cookies   Aanmelden met verbeterde beveiliging (SSL)   Help en ondersteuning   Feedback

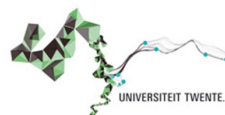
Wat zou jij doen op deze website?

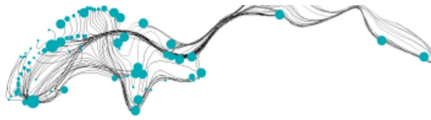
1. Ik log in met mijn e-mailadres en wachtwoord op de website, de website ziet er prima uit!
2. Ik log niet in, want ik denk dat dit een onbetrouwbare website is.

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

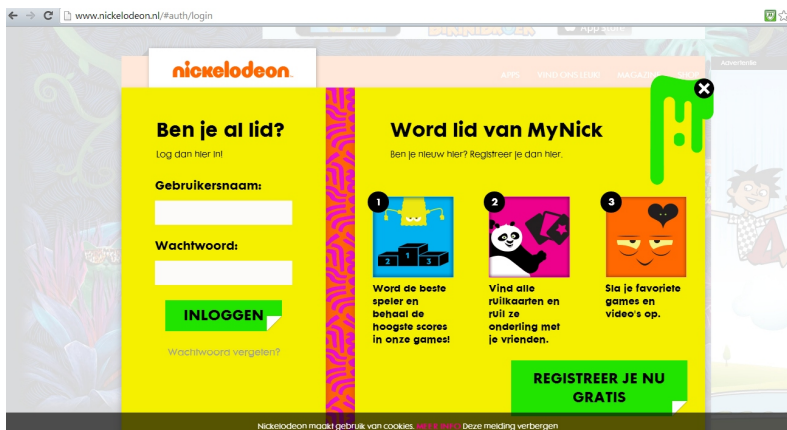
Voor vragen kun je altijd je hand opsteken.





## WEBSITE 2: NICKELODEON

Je hebt een nickelodeon account en wilt inloggen op de site van nickelodeon. Je ziet de onderstaande website.



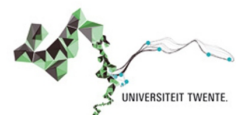
Wat zou jij doen op deze website?

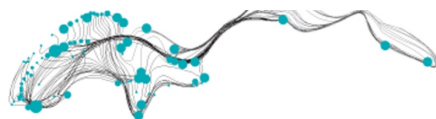
1. Ik log niet in op deze website, want hij ziet er niet zo betrouwbaar uit.
2. Ik zou inloggen op deze website, volgens mij is de echte website van Nickelodeon.

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

Voor vragen kun je altijd je hand opsteken.





### WEBSITE 3: RABOBANK

Je bent samen met je vader/moeder bezig op de computer. Je wilt inloggen op je spaarrekening bij de Rabobank om te zien hoeveel geld er op je rekening staat. Je ziet de onderstaande website.

← → ↻ <http://www.rabobank.nl/klanten/>

Applicaties Plaats voor een snelle navigatie je bladwijzers op deze bladwijzerbalk. [Bladwijzers nu importeren...](#)

**Rabobank**

## Inloggen met de Random Reader

**Log alleen in met de I-toets.**  
**Met de S-toets maakt u geld over.**  
 Ziet u iets ongewoons? Stop en bel 0900 0905.

**Bankpas**

Rekeningnummer/IBAN

☐ Onthouden

Pasnummer

**Random Reader**

- Plaats uw bankpas in de Random Reader
- Druk op **I** (Inloggen)
- Toets uw **pincode** in en druk op **OK**

Vul de toegangscode in die op uw Random Reader verschijnt:

[Inloggen](#) [Annuleren](#) [Help](#)

Ga alleen verder als de adresregel begint met <https://bankieren.rabobank.nl/>...

> Hoe controleert u de veiligheid van uw verbinding?

> Lees meer over veiligheid

**Aanvragen**

Heeft u geen toegang tot Rabo Internetbankieren?

Met Rabo Internetbankieren kunt u altijd via Internet uw rekeningen inzien en transacties uitvoeren.

> Informatie over Rabo Internetbankieren

> Bekijk de demo

**Help**

> Waarom kan ik niet inloggen?

> Waarom krijg ik de melding (942)?

> Waarom krijg ik de melding (947)?

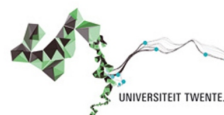
Je weet wat je in moet vullen bij 'Rekeningnummer' en 'Pasnummer'. Wat zou jij doen op deze website?

1. Ik sluit deze website af, want volgens mij is hij onveilig.
2. Ik druk op inloggen en ga kijken wat er op mijn spaarrekening staat.

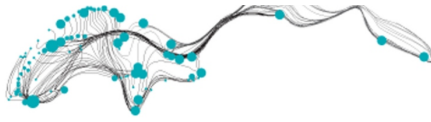
Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

Voor vragen kun je altijd je hand opsteken.



UNIVERSITEIT TWENTE



## WEBSITE 4: YOUTUBE


Je wilt inloggen met jouw e-mail adres iemand@gmail.com op de volgende webiste.

[https://accounts.google.com/ServiceLogin?continue=http%3A%2F%2Fwww.youtube.com%2Fsignin%3Faction\\_handle\\_signin%3Dtrue%26app%3Ddesktop%26featu](https://accounts.google.com/ServiceLogin?continue=http%3A%2F%2Fwww.youtube.com%2Fsignin%3Faction_handle_signin%3Dtrue%26app%3Ddesktop%26featu)



Eén account. Al het beste van Google.

Log in om door te gaan naar YouTube



**iemand**  
iemand@gmail.com

Wachtwoord

Inloggen

[Heeft u hulp nodig?](#)

[Inloggen met een ander account](#)

Eén Google-account voor alles van Google



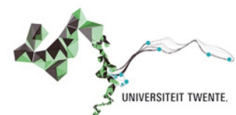
Wat zou jij doen op deze website?

1. Ik zou niet inloggen met mijn wachtwoord, want er klopt iets niet aan deze website.
2. Ik zou wel inloggen met mijn wachtwoord, deze website ziet er prima uit!

Vul hier het getal in van het antwoord dat jij kiest:

Hoe ga jij om met het internet?

Voor vragen kun je altijd je hand opsteken.





## BIBLIOGRAPHY

---

### AUTHOR REFERENCES

- Lastdrager, E. (2014). "Achieving a consensual definition of phishing based on a systematic review of the literature". *Crime Science*, 3(1): p. 9. ISSN: 2193-7680. DOI: [10.1186/s40163-014-0009-y](https://doi.org/10.1186/s40163-014-0009-y).
- Lastdrager, E., L. Montoya, P. Hartel and M. Junger (2013). "Applying the Lost-Letter Technique to Assess IT Risk Behaviour". In: *Third Workshop on Socio-Technical Aspects in Security and Trust (STAST)*. IEEE, pp. 2–9. DOI: [10.1109/STAST.2013.15](https://doi.org/10.1109/STAST.2013.15).
- Lastdrager, E., I. Carvajal Gallardo, P. Hartel and M. Junger (2017). "How Effective is Anti-Phishing Training for Children?" In: *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS)*. Distinguished Paper Award. 12-14 July, Santa Clara, California, USA. URL: <https://www.usenix.org/conference/soups2017/technical-sessions/presentation/lastdrager>.
- Lastdrager, E., P. Hartel and M. Junger (2015). "Apat: Anti-Phishing Analysing and Triaging Environment (Poster, Extended Abstract)". In: *36th IEEE Symposium on Security and Privacy*. San Jose, California, USA. URL: [http://www.ieee-security.org/TC/SP2015/posters/paper\\_58.pdf](http://www.ieee-security.org/TC/SP2015/posters/paper_58.pdf).
- Kernkamp, A., E. Lastdrager, T. Logtens, P. van Slingerland and E. Torreño Dassen (2015). *Cyber Trend Watch Inzicht in trends door het gebruik van Cyber Security Metrics*. Tech. rep. R 11627. TNO.

### OTHER REFERENCES

- Abu-Nimeh, S., D. Nappa, X. Wang and S. Nair (2007). "A comparison of machine learning techniques for phishing detection". In: vol. 269. New York, NY, USA: ACM, pp. 60–69. DOI: [10.1145/1299015.1299021](https://doi.org/10.1145/1299015.1299021).
- Adida, B. (2007). "Beamauth: Two-factor web authentication with a bookmark". In: *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 48–57. DOI: [10.1145/1315245.1315253](https://doi.org/10.1145/1315245.1315253).
- Aggarwal, A., A. Rajadesingan and P. Kumaraguru (2012). "PhishAri: Automatic realtime phishing detection on twitter". In: *2012 eCrime Researchers Summit*. IEEE, pp. 1–12. ISBN: 978-1-4673-2543-1. DOI: [10.1109/eCrime.2012.6489521](https://doi.org/10.1109/eCrime.2012.6489521).

- Ahamid, I., J. Abawajy and T.-H. Kim (2013). "Using feature selection and classification scheme for automating phishing email detection". *Studies in Informatics and Control*, 22(1): pp. 61–70.
- Ahmed, A. (2010). "Muslim Discrimination: Evidence From Two Lost-Letter Experiments". *Journal of Applied Social Psychology*, 40(4): pp. 888–898. DOI: [10.1111/j.1559-1816.2010.00602.x](#).
- Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle". In: *Proceedings of the Second International Symposium on Information Theory*, pp. 267–281.
- Ali, M. and L. Rajamani (2012). "Deceptive phishing detection system: From audio and text messages in instant messengers using data mining approach". In: *Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME)*. IEEE, pp. 458–465. DOI: [10.1109/ICPRIME.2012.6208390](#).
- Almomani, A., T.-C. Wan, A. Altaher, A. Manasrah, E. Almomani, M. Anbar, E. Alomari and S. Ramadass (2012). "Evolving fuzzy neural network for phishing emails detection". *Journal of Computer Science*, 8(7): pp. 1099–1107. DOI: [10.3844/jcssp.2012.1099.1107](#).
- Almomani, A., B. B. Gupta, S. Atawneh, A. Meulenberg and E. Almomani (2013). "A survey of phishing email filtering techniques". *IEEE Communications Surveys and Tutorials*, 15(4): pp. 2070–2090. ISSN: 1553877X. DOI: [10.1109/SURV.2013.030713.00020](#).
- Alnajim, A. and M. Munro (2009). "An Evaluation of Users' Anti-Phishing Knowledge Retention". In: *International Conference on Information Management and Engineering*. ICIME '09. IEEE, pp. 210–214. DOI: [10.1109/ICIME.2009.114](#).
- Alseadoon, I. M. (2014). "The Impact of Users' Characteristics on Their Ability to Detect Phishing Emails". PhD thesis. Queensland University of Technology.
- American Heritage Dictionary (2013). *Phishing*. URL: <http://www.ahdictionary.com/word/search.html?q=phish> (Retrieved 2013-12-20).
- Amin, R., J. Ryan and J. van Dorp (2012). "Detecting Targeted Malicious Email". *IEEE Security Privacy*, 10(3): pp. 64–71. ISSN: 1540-7993. DOI: [10.1109/MSP.2011.154](#).
- Anderson, D. S., C. Fleizach, S. Savage and G. M. Voelker (2007). "Spam-scatter: Characterizing internet scam hosting infrastructure". In: *Processings of the 16th USENIX Security Symposium*, pp. 135–148. URL: [https://www.usenix.org/legacy/event/sec07/tech/full\\_papers/anderson/anderson.html/](https://www.usenix.org/legacy/event/sec07/tech/full_papers/anderson/anderson.html/).
- Anderson, R. and T. Moore (2009). "Information security: Where computer science, economics and psychology meet". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and*



- Engineering Sciences, 367(1898): pp. 2717–2727. DOI: [10.1098/rsta.2009.0027](#).
- Anti-Phishing Working Group (2013). *Phishing Activity Trends Report, 2nd Quarter 2013*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q2\\_2013.pdf](http://docs.apwg.org/reports/apwg_trends_report_q2_2013.pdf) (Retrieved 2013-12-20).
- Anti-Phishing Working Group (2014a). *Phishing Activity Trends Report, 1st Quarter 2014*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q1\\_2014.pdf](http://docs.apwg.org/reports/apwg_trends_report_q1_2014.pdf) (Retrieved 2014-06-23).
- Anti-Phishing Working Group (2014b). *Phishing Activity Trends Report, 2nd Quarter 2014*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q2\\_2014.pdf](http://docs.apwg.org/reports/apwg_trends_report_q2_2014.pdf) (Retrieved 2014-08-28).
- Anti-Phishing Working Group (2014c). *Phishing activity trends report, 3rd Quartile 2014*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q3\\_2014.pdf](http://docs.apwg.org/reports/apwg_trends_report_q3_2014.pdf) (Retrieved 2016-03-22).
- Anti-Phishing Working Group (2015a). *Phishing Activity Trends Report, 1st-3rd Quarter 2015*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q1-q3\\_2015.pdf](http://docs.apwg.org/reports/apwg_trends_report_q1-q3_2015.pdf) (Retrieved 2015-12-23).
- Anti-Phishing Working Group (2015b). *Phishing activity trends report, 4th Quartile 2014*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q4\\_2014.pdf](http://docs.apwg.org/reports/apwg_trends_report_q4_2014.pdf) (Retrieved 2016-03-22).
- Anti-Phishing Working Group (2016a). *Phishing Activity Trends Report, 1st Quarter 2016*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q1\\_2016.pdf](http://docs.apwg.org/reports/apwg_trends_report_q1_2016.pdf) (Retrieved 2016-10-23).
- Anti-Phishing Working Group (2016b). *Phishing Activity Trends Report, 2nd Quarter 2016*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q2\\_2016.pdf](http://docs.apwg.org/reports/apwg_trends_report_q2_2016.pdf) (Retrieved 2016-10-23).
- Anti-Phishing Working Group (2016c). *Phishing Activity Trends Report, 3rd Quarter 2016*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q3\\_2016.pdf](http://docs.apwg.org/reports/apwg_trends_report_q3_2016.pdf) (Retrieved 2016-12-20).
- Anti-Phishing Working Group (2016d). *Phishing Activity Trends Report, 4th Quarter 2015*. URL: [http://docs.apwg.org/reports/apwg\\_trends\\_report\\_q4\\_2015.pdf](http://docs.apwg.org/reports/apwg_trends_report_q4_2015.pdf) (Retrieved 2016-03-22).
- Arachchilage, N. A. G. and S. Love (2013). “A game design framework for avoiding phishing attacks”. *Computers in Human Behavior*, 29(3): pp. 706–714. ISSN: 07475632. DOI: [10.1016/j.chb.2012.12.018](#).
- Bainbridge, D. (2007). “Criminal law tackles computer fraud and misuse”. *Computer Law and Security Report*, 23(3): pp. 276–281. DOI: [10.1016/j.clsr.2007.03.001](#).
- Baker, E., J. Tedesco and W. Baker (2006). “Consumer privacy and trust online: An experimental analysis of anti-phishing promotional effects”. *Journal of Website Promotion*, 2(1-2): pp. 89–113. DOI: [10.1080/15533610802104166](#).
- Bamert, T., C. Decker, L. Elsen, R. Wattenhofer and S. Welten (Sept. 2013). “Have a snack, pay with Bitcoins”. In: *Peer-to-Peer Com-*

- puting (P2P), 2013 *IEEE Thirteenth International Conference on*, pp. 1–5. DOI: [10.1109/P2P.2013.6688717](https://doi.org/10.1109/P2P.2013.6688717).
- Barracough, P., M. Hossain, M. Tahir, G. Sexton and N. Aslam (2013). “Intelligent phishing detection and protection scheme for online transactions”. *Expert Systems with Applications*, 40(11): pp. 4697–4706. DOI: [10.1016/j.eswa.2013.02.009](https://doi.org/10.1016/j.eswa.2013.02.009).
- Basnet, R., S. Mukkamala and A. Sung (2008). “Detection of phishing attacks: A machine learning approach”. *Studies in Fuzziness and Soft Computing*, 226. Ed. by Prasad: pp. 373–383. DOI: [10.1007/978-3-540-77465-5\\_19](https://doi.org/10.1007/978-3-540-77465-5_19).
- Beatty, P., I. Reay, S. Dick and J. Miller (2011). “Consumer trust in e-commerce web sites: A meta-study”. *ACM Computing Surveys*, 43(3): 14:1–14:46. ISSN: 0360-0300. DOI: [10.1145/1922649.1922651](https://doi.org/10.1145/1922649.1922651).
- Been, H. and J. Kleverwal (2012). *Phishing using QR codes: A case study at the University of Twente*. Internal Report. University of Twente.
- Beliakov, G., J. Yearwood and A. Kelarev (2012). “Application of rank correlation, clustering and classification in information security”. *Journal of Networks*, 7(6): pp. 935–945. DOI: [10.4304/jnw.7.6.935-945](https://doi.org/10.4304/jnw.7.6.935-945).
- Bergholz, A., J. De Beer, S. Glahn, M.-F. Moens, G. Paaß and S. Strobel (2010). “New filtering approaches for phishing email”. *Journal of Computer Security*, 18(1): pp. 7–35. DOI: [10.3233/JCS-2010-0371](https://doi.org/10.3233/JCS-2010-0371).
- Betaalvereniging (2016). *Phishing*. URL: <https://www.veiligbankieren.nl/fraude/phishing/> (Retrieved 2016-01-20).
- Biddle, R., S. Chiasson and P. Van Oorschot (2012). “Graphical passwords: Learning from the first twelve years”. *ACM Computing Surveys*, 44(4): 19:1–19:41. ISSN: 0360-0300. DOI: [10.1145/2333112.2333114](https://doi.org/10.1145/2333112.2333114).
- Blythe, M., H. Petrie and J. A. Clark (2011). “F for fake: Four Studies on How We Fall for Phish”. In: *Proceedings of the 2011 annual conference on Human factors in computing systems*. CHI ’11. New York, NY, USA: ACM, pp. 3469–2478. ISBN: 978-1-4503-0228-9. DOI: [10.1145/1978942.1979459](https://doi.org/10.1145/1978942.1979459).
- Bose, I. and A. Leung (2008). “Assessing anti-phishing preparedness: A study of online banks in Hong Kong”. *Decision Support Systems*, 45(4): pp. 897–912. DOI: [10.1016/j.dss.2008.03.001](https://doi.org/10.1016/j.dss.2008.03.001).
- Bowers, K. and S. Johnson (2005). “Using publicity for preventive purposes”. In: *Handbook of Crime Prevention and Community*. London: Willian Publishing, Chap. 13. DOI: [10.4324/9781843926146](https://doi.org/10.4324/9781843926146).
- boyd, d. (2014). *It’s Complicated: The Social Lives of Networked Teens*. Yale University Press. ISBN: 978-0-300-16631-6.
- Brady, C. (2010). *Security Awareness for Children*. Tech. rep. RHUL-MA-2010-05. London: Royal Holloway, pp. 1–10. URL: <http://www.>

- [ma.rhul.ac.uk/static/techrep/2010/RHUL-MA-2010-05.pdf](http://ma.rhul.ac.uk/static/techrep/2010/RHUL-MA-2010-05.pdf).
- Brainard, J., A. Juels, R. L. Rivest, M. Szydlo and M. Yung (2006). "Fourth-factor authentication: somebody you know". In: *Proceedings of the 13th ACM Conference on Computer and Communications Security*, pp. 168–178. ISBN: 1-59593-518-5. DOI: [10.1145/1180405.1180427](https://doi.org/10.1145/1180405.1180427).
- Brantingham, P. and P. Brantingham (1993). "Environment, routine and situation: Toward a pattern theory of crime". In: *Routine Activity and Rational Choice: Advances in Criminological Theory*. Ed. by R. V. Clarke and M. Felson. Vol. 5. Piscataway, NJ, USA: Transaction Press, pp. 259–294.
- Brantingham, P. and P. Brantingham (2008). "Crime Pattern Theory". In: *Environmental criminology and crime analysis*. Ed. by R. Wortley and L. Mazerolle. Devon, UK: Willan Publishing.
- Bridges, F. S., D. A. Anzalone, S. W. Ryan and F. L. Anzalone (2002). "Extensions of the lost letter technique to divisive issues of creationism, Darwinism, sex education, and gay and lesbian affiliations". *Psychological Reports*, 90(2): pp. 391–400. DOI: [10.2466/pr0.2002.90.2.391](https://doi.org/10.2466/pr0.2002.90.2.391).
- Bullee, J. H., A. L. Montoya Morales, W. Pieters, M. Junger and P. H. Hartel (Mar. 2015). "The persuasion and security awareness experiment: reducing the success of social engineering attacks". *Journal of Experimental Criminology*, 11(1): pp. 97–115. ISSN: 1573-3750. DOI: [10.1007/s11292-014-9222-7](https://doi.org/10.1007/s11292-014-9222-7).
- Bullee, J.-W., L. Montoya, M. Junger and P. Hartel (2016). "Telephone-based social engineering attacks : An experiment testing the success and time decay of an intervention". In: *Singapore Cyber Security R&D Conference, SG-CRC 2015*. Singapore: IOS Press, pp. 1–6. DOI: [10.3233/978-1-61499-617-0-107](https://doi.org/10.3233/978-1-61499-617-0-107).
- Burgoon, J. K. and T. R. Levine (2010). "Advances in deception detection". In: *New directions in interpersonal communication research*. Ed. by S. W. Smith and S. R. Wilson. Sage, pp. 201–220.
- Bursztein, E., B. Benko, D. Margolis, T. Pietraszek, A. Archer, A. Aquino, A. Pitsillidis and S. Savage (2014). "Handcrafted Fraud and Extortion: Manual Account Hijacking in the Wild". In: *Proceedings of the 2014 Conference on Internet Measurement Conference - IMC '14*. ISBN: 978-1-4503-3213-2. DOI: [10.1145/2663716.2663749](https://doi.org/10.1145/2663716.2663749).
- Bushman, B. J. and A. M. Bonacci (2004). "You've got mail: Using e-mail to examine the effect of prejudiced attitudes on discrimination against Arabs". *Journal of Experimental Social Psychology*, 40(6): pp. 753–759. ISSN: 0022-1031. DOI: [10.1016/j.jesp.2004.02.001](https://doi.org/10.1016/j.jesp.2004.02.001).
- Butler, R. (2007). "A framework of anti-phishing measures aimed at protecting the online consumer's identity". *Electronic Library*, 25(5): pp. 517–533. DOI: [10.1108/02640470710829514](https://doi.org/10.1108/02640470710829514).

- Calvert, S. L. (2008). "Children as consumers: Advertising and marketing". *Future of Children*, 18(1): pp. 205–234. ISSN: 10548289. DOI: [10.1353/foc.0.0001](https://doi.org/10.1353/foc.0.0001).
- Cao, Y., W. Han and Y. Le (2008). "Anti-phishing based on automated individual white-list". In: *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 51–60. DOI: [10.1145/1456424.1456434](https://doi.org/10.1145/1456424.1456434).
- Carnegie Mellon University (June 2013). *Social Engineering Using a USB Drive*. URL: <http://www.cmu.edu/iso/aware/be-aware/usb.html> (Retrieved 2013-06-15).
- Castiglione, A., R. De Prisco and A. De Santis (2009). "Do You Trust Your Phone?" English. In: *E-Commerce and Web Technologies*. Ed. by T. Di Noia and F. Buccafurri. Vol. 5692. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 50–61. ISBN: 978-3-642-03963-8. DOI: [10.1007/978-3-642-03964-5\\_6](https://doi.org/10.1007/978-3-642-03964-5_6).
- Charikar, M. S. (2002). "Similarity estimation techniques from rounding algorithms". *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing - STOC '02*. DOI: [10.1145/509907.509965](https://doi.org/10.1145/509907.509965).
- Chen, K.-T., J.-Y. Chen, C.-R. Huang and C.-S. Chen (2009). "Fighting phishing with discriminative keypoint features". *IEEE Internet Computing*, 13(3): pp. 56–63. DOI: [10.1109/MIC.2009.59](https://doi.org/10.1109/MIC.2009.59).
- Chhabra, S., A. Aggarwal, F. Benevenuto and P. Kumaraguru (2011). "Phi.sh/\$oCiaL". In: *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*. New York, New York, USA: ACM Press, pp. 92–101. ISBN: 978-1-4503-0788-8. DOI: [10.1145/2030376.2030387](https://doi.org/10.1145/2030376.2030387).
- Cialdini, R. B. (Feb. 2001). "The Science of Persuasion". *Scientific American*, 284(2): pp. 76–81. ISSN: 0036-8733. DOI: [10.1038/scientificamerican0201-76](https://doi.org/10.1038/scientificamerican0201-76).
- Cialdini, R. B. (2006). *Influence: The Psychology of Persuasion*. Harper Business. ISBN: 0-06-124189-X.
- Clarke, R. V. (2009). "Situational Crime Prevention: Theoretical Background and Current Practice". In: *Handbook on Crime and Deviance*. Ed. by M. D. Krohn, A. J. Lizotte and G. P. Hall. Handbooks of Sociology and Social Research. New York, NY: Springer New York. Chap. 14, pp. 259–276. ISBN: 978-1-4419-0244-3. DOI: [10.1007/978-1-4419-0245-0\\_14](https://doi.org/10.1007/978-1-4419-0245-0_14).
- Cohen, J. (1992). "A power primer". *Psychological Bulletin*, 112(1): pp. 155–159. ISSN: 0033-2909. DOI: [10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155).
- Cohen, L. E. and M. Felson (1979). "Social Change and Crime Rate Trends: A Routine Activity Approach". English. *American Sociological Review*, 44(4): pp. 588–608. ISSN: 00031224.

- Collins English Dictionary (2013). *Phishing*. URL: <http://www.collinsdictionary.com/dictionary/english/phishing> (Retrieved 2013-12-20).
- Conran, A. and P. Wilson (2006). *The Real Hustle*. Television series. United Kingdom.
- Cornish, D. and R. Clarke (1986). *The reasoning criminal: Rational choice perspectives on offending*. Springer-Verlag New York. ISBN: 978-0-387-96272-6.
- Cornish, D. and R. Clarke, eds. (2014). *The reasoning criminal: Rational choice perspectives on offending*. Transaction Publishing. ISBN: 978-1-4128-5275-3.
- Cornish, D. (1994). "The procedural analysis of offending and its relevance for situational prevention". *Crime prevention studies*, 3: pp. 151-196.
- Cornish, D. and R. V. Clarke (2008). "The Rational Choice Perspective". In: *Environmental criminology and crime analysis*. Ed. by R. Wortley and L. Mazerolle. Devon, UK: Willan Publishing.
- Cranor, L. (2008). "Can phishing be foiled?" *Scientific American*, 299(6): pp. 104-110. DOI: [10.1038/scientificamerican1208-104](https://doi.org/10.1038/scientificamerican1208-104).
- De Nederlandsche Bank (2014). *Additional buffer requirement enhances resilience of Dutch systemic banks*. URL: <https://www.dnb.nl/en/news/news-and-archive/dnbulletin-2014/dnb306988.jsp> (Retrieved 2017-04-28).
- De Ryck, P., N. Nikiforakis, L. Desmet and W. Joosen (2013). "TabShots: Client-side detection of tabnabbing attacks". In: *Proceedings of the 8th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, pp. 447-455. DOI: [10.1145/2484313.2484371](https://doi.org/10.1145/2484313.2484371).
- Deaux, K. (1974). "Anonymous altruism: extending the lost letter technique". *The Journal of Social Psychology*, 92: pp. 61-66. DOI: [10.1080/00224545.1974.9923072](https://doi.org/10.1080/00224545.1974.9923072).
- Dhamija, R., J. D. Tygar and M. Hearst (2006). "Why phishing works". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*. November 2005. New York, New York, USA: ACM Press, p. 581. ISBN: 1-59593-372-7. DOI: [10.1145/1124772.1124861](https://doi.org/10.1145/1124772.1124861).
- Dhamija, R. and J. D. Tygar (2005). "The battle against phishing: Dynamic Security Skins". In: *Proceedings of the symposium on Usable privacy and security (SOUPS)*. Pittsburgh, Pennsylvania, pp. 77-88. ISBN: 1-59593-178-3. DOI: [10.1145/1073001.1073009](https://doi.org/10.1145/1073001.1073009).
- Domínguez, A., J. Saenz-de-Navarrete, L. De-Marcos, L. Fernández-Sanz, C. Pagés and J.-J. Martínez-Herráiz (Apr. 2013). "Gamifying learning experiences: Practical implications and outcomes". *Computers & Education*, 63: pp. 380-392. ISSN: 03601315. DOI: [10.1016/j.compedu.2012.12.020](https://doi.org/10.1016/j.compedu.2012.12.020).

- Dong, X., J. Clark and J. Jacob (2010). "Defending the weakest link: Phishing websites detection by analysing user behaviours". *Telecommunication Systems*, 45(2-3): pp. 215–226. DOI: [10.1007/s11235-009-9247-9](https://doi.org/10.1007/s11235-009-9247-9).
- Downs, D., I. Ademaj and A. Schuck (2009). "Internet security: Who is leaving the 'virtual door' open and why?" *First Monday*, 14(1).
- Downs, J. S., M. B. Holbrook and L. F. Cranor (2006). "Decision Strategies and Susceptibility to Phishing". In: *Proceedings of the second symposium on Usable privacy and security*. Pittsburgh, PA, USA: ACM, pp. 79–90. ISBN: 1-4122-6875-3.
- Drake, C., J. Oliver and E. Koontz (2004). "Anatomy of a phishing email". In: *Proceedings of the 2004 Conference on Email and Anti-Spam*.
- Ducklin, P. (Dec. 2011). *Lost USB keys have 66% chance of malware*. URL: <http://nakedsecurity.sophos.com/2011/12/07/lost-usb-keys-have-66-percent-chance-of-malware/> (Retrieved 2011-12-07).
- Dutch Banking Association (2015). *Safe Banking*. URL: <https://www.veiligbankieren.nl> (Retrieved 2015-03-02).
- Dutch Banking Association (June 2016). *Factsheet Veiligheid en fraude*. URL: [https://www.nvb.nl/media/document/000254\\_od15799-nvb-factsheet-veiligheid-en-fraude-06-06.pdf](https://www.nvb.nl/media/document/000254_od15799-nvb-factsheet-veiligheid-en-fraude-06-06.pdf) (Retrieved 2017-03-01).
- Dutch Banking Association (2017). *De Nederlandse Vereniging van Banken*. URL: <https://www.nvb.nl/> (Retrieved 2017-03-01).
- Edelson, E. (June 2003). "The 419 scam: information warfare on the spam front and a proposal for local filtering". *Computers & Security*, 22(5): pp. 392–401. ISSN: 01674048. DOI: [10.1016/S0167-4048\(03\)00505-4](https://doi.org/10.1016/S0167-4048(03)00505-4).
- Egelman, S., L. F. Cranor and J. Hong (2008). "You've been warned: an empirical study of the effectiveness of web browser phishing warnings". In: *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems*. New York, New York, USA: ACM Press, pp. 1065–1074. ISBN: 978-1-60558-011-1. DOI: [10.1145/1357054.1357219](https://doi.org/10.1145/1357054.1357219).
- El Helou, S., N. Li and D. Gillet (2010). "The 3A Interaction Model: Towards Bridging the Gap between Formal and Informal Learning". In: *Proceedings of the Third International Conference on Advances in Computer-Human Interactions (ACHI)*. Saint Maarten: IEEE, pp. 179–184. DOI: [10.1109/ACHI.2010.38](https://doi.org/10.1109/ACHI.2010.38).
- Elmaleh, D. (2007). "Phishing forbidden". *Card Technology Today*, 19(9): pp. 12–13. DOI: [10.1016/S0965-2590\(07\)70137-8](https://doi.org/10.1016/S0965-2590(07)70137-8).
- Emilin Shyni, C. and S. Swamynathan (2013). "Protecting the online user's information against phishing attacks using dynamic encryption techniques". *Journal of Computer Science*, 9(4): pp. 526–533. DOI: [10.3844/jcssp.2013.526.533](https://doi.org/10.3844/jcssp.2013.526.533).

- Emm, D. (2006). "Phishing update, and how to avoid getting hooked". *Network Security*, 2006(8): pp. 13–15. DOI: [10.1016/S1353-4858\(06\)70432-9](#).
- Encyclopædia Britannica (2015). *Telegraph*. URL: <http://www.britannica.com/technology/telegraph> (Retrieved 2015-10-07).
- Ericksson, K. and H. Simon (1993). *Protocol analysis: verbal reports as data*. Revised edition. 2nd. Cambridge, Massachusetts, USA: Bradford Book / MIT Press. ISBN: 978-0-262-55023-9.
- Evans, J. S. B. T. and K. E. Stanovich (May 2013). "Dual-Process Theories of Higher Cognition". *Perspectives on Psychological Science*, 8(3): pp. 223–241. ISSN: 1745-6924. DOI: [10.1177/1745691612460685](#).
- Falliere, N., L. Murchu and E. Chien (Feb. 2011). *W32.Stuxnet Dossier*. Tech. rep. Symantec.
- Farrington, D. P. and B. J. Knight (1979). "Two non-reactive field experiments on stealing from a 'lost' letter". *British Journal of Social and Clinical Psychology*, 18(3): pp. 277–284. ISSN: 00071293. DOI: [10.1111/j.2044-8260.1979.tb00337.x](#).
- Farrington, D. P. and B. J. Knight (1980). "Stealing From a 'Lost' Letter: Effects of Victim Characteristics". *Criminal Justice and Behavior*, 7(4): pp. 423–436. ISSN: 0093-8548. DOI: [10.1177/009385488000700406](#).
- Feinstein, A. R. and D. V. Cicchetti (1990). "High agreement but low kappa: I. The problems of two paradoxes." *Journal of clinical epidemiology*, 43(6): pp. 543–549. ISSN: 0895-4356. DOI: [10.1016/0895-4356\(90\)90158-L](#).
- Felson, M. and R. Boba (2010). *Crime and everyday life*. 4th. SAGE. ISBN: 978-1-4129-3633-0.
- Felson, M. and R. Clarke (1998). "Opportunity Makes the Thief: Practical theory for crime prevention". *Police Research Studies*, 98. Ed. by B. Webb.
- Ferguson, A. (2005). "Fostering e-mail security awareness: The West Point carronade". *EDUCASE Quarterly*, 28(1): pp. 54–57.
- Fernandez, J. D., S. Smith, M. Garcia and D. Kar (2005). "Computer forensics: a critical need in computer science programs". *Journal of Computer Sciences in Colleges*, 20(4): pp. 315–322. ISSN: 1937-4771.
- Fessler, D. M. (2009). "Return of the lost letter: Experimental framing does not enhance altruism in an everyday context". *Journal of Economic Behavior & Organization*, 71(2): pp. 575–578. ISSN: 0167-2681. DOI: [10.1016/j.jebo.2009.03.007](#).
- Fette, I., N. Sadeh and A. Tomasic (2007). "Learning to detect phishing emails". In: *Proceedings of the 16th international conference on World Wide Web (WWW)*, pp. 649–656. DOI: [10.1145/1242572.1242660](#).
- Florêncio, D. and C. Herley (2007). "A large-scale study of web password habits". In: *Proceedings of the 16th international conference on*



- World Wide Web*. Banff, Alberta, Canada: ACM Press, pp. 657–666. ISBN: 978-1-59593-654-7. DOI: [10.1145/1242572.1242661](https://doi.org/10.1145/1242572.1242661).
- Florêncio, D. and C. Herley (2011). *Sex, Lies and Cyber-crime Surveys*. Tech. rep. Redmond: Microsoft Research.
- Forbes, G. B., R. K. TeVault and H. F. Gromoll (1971). “Willingness To Help Strangers As A Function Of Liberal, Conservative Or Catholic Church Membership: A Field Study With The Lost-Letter Technique”. *Psychological Reports*, 28(3): pp. 947–949. ISSN: 0033-2941. DOI: [10.2466/pr0.1971.28.3.947](https://doi.org/10.2466/pr0.1971.28.3.947).
- Forbes, G. B., R. K. TeVault and H. F. Gromoll (1972). “Regional differences in willingness to help strangers: A field experiment with a new unobtrusive measure”. *Social Science Research*, 1(4): pp. 415–419. ISSN: 0049-089X. DOI: [10.1016/0049-089X\(72\)90086-5](https://doi.org/10.1016/0049-089X(72)90086-5).
- Forte, D. (2009). “Anatomy of a phishing attack: A high-level overview”. *Network Security*, 2009(4): pp. 17–19. DOI: [10.1016/S1353-4858\(09\)70042-X](https://doi.org/10.1016/S1353-4858(09)70042-X).
- Fraudehelpdesk (2016). *The Dutch National Anti-Fraud Hotline*. URL: <https://www.fraudehelpdesk.nl/> (Retrieved 2016-01-20).
- Fraudehelpdesk (2017). *Phishing Fraudehelpdesk*. URL: <https://www.fraudehelpdesk.nl/sub-vragen/phishingmails/> (Retrieved 2017-01-06).
- Freyne, J., L. Coyle, B. Smyth and P. Cunningham (2010). “Relative status of journal and conference publications in computer science”. *Communications of the ACM*, 53(11): pp. 124–132. ISSN: 00010782. DOI: [10.1145/1839676.1839701](https://doi.org/10.1145/1839676.1839701).
- Fumera, G., I. Pillai and F. Roli (2006). “Spam Filtering Based On The Analysis Of Text Information Embedded Into Images”. *Journal of Machine Learning Research*, 7: pp. 2699–2720. ISSN: 1532-4435.
- Gabor, T. and T. Barker (1989). “Probing the public’s honesty: A field experiment using the “lost letter” technique”. *Deviant behavior*, 10(4): pp. 387–399. DOI: [10.1080/01639625.1989.9967824](https://doi.org/10.1080/01639625.1989.9967824).
- Garera, S., N. Provos, M. Chew and A. Rubin (2007). “A framework for detection and measurement of phishing attacks”. In: *Proceedings of the 2007 ACM Workshop on Recurring Malcode (WORM)*, pp. 1–8. DOI: [10.1145/1314389.1314391](https://doi.org/10.1145/1314389.1314391).
- Gastellier-Prevost, S. and M. Laurent (2011). “Defeating pharming attacks at the client-side”. In: *Proceedings of the 5th International Conference on Network and System Security (NSS)*, pp. 33–40. DOI: [10.1109/ICNSS.2011.6059957](https://doi.org/10.1109/ICNSS.2011.6059957).
- Geer, D. (2005). “Technology news: Security technologies go phishing”. *Computer*, 38(6): pp. 18–21. DOI: [10.1109/MC.2005.201](https://doi.org/10.1109/MC.2005.201).
- Geloven, S. van (2011). *Woordlengte*. URL: <http://www.opentaal.org/het-laatste-nieuws/153-woordlengte.html> (Retrieved 2016-12-01).
- Gigerenzer, G. and P. Todd (1999). *Simple heuristics that makes us smart*. Oxford University Press. ISBN: 978-0-19-514381-2.



- Gomes, L. H., C. Cazita, J. M. Almeida, V. Almeida and W. Meira (2004). "Characterizing a spam traffic". In: *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, IMC '04*. New York, New York, USA: ACM Press, p. 356. ISBN: 1-58113-821-0. DOI: [10.1145/1028788.1028837](#).
- Gouda, M., A. Liu, L. Leung and M. Alam (2007). "SPP: An anti-phishing single password protocol". *Computer Networks*, 51(13): pp. 3715–3726. DOI: [10.1016/j.comnet.2007.03.007](#).
- Grobbe, N. (2015). *De overeenkomst tussen de anti-phishing campagnes van de NVB en de informatiebehoefte van de burger over phishing*. BA thesis.
- Gross, J. B. and M. B. Rosson (2007). "Looking for trouble: understanding end-user security management". In: *Proceedings of the 2007 symposium on Computer human interaction for the management of information technology*. Cambridge, Massachusetts. ISBN: 978-1-59593-635-6. DOI: [10.1145/1234772.1234786](#).
- Guan, B., Y. Wu and Y. Wang (June 2012). "A Novel Security Scheme for Online Banking Based on Virtual Machine". In: *IEEE Sixth International Conference on Software Security and Reliability Companion (SERE-C)*, pp. 12–17. DOI: [10.1109/SERE-C.2012.28](#).
- Gupta, G. and J. Pieprzyk (2011). "Socio-technological phishing prevention". *Information Security Technical Report*, 16(2): pp. 67–73. DOI: [10.1016/j.istr.2011.09.003](#).
- Haddon, L. and S. Livingstone (2012). *EU Kids Online: national perspectives*. Tech. rep. The London School of Economics and Political Science.
- Halevi, T., J. Lewis and N. Memon (2013). "A pilot study of cyber security and privacy related behavior and personality traits". In: *Proceedings of the 22nd international conference on World Wide Web companion (WWW)*. Rio de Janeiro, Brazil: International World Wide Web Conferences Steering Committee, pp. 737–744. ISBN: 978-1-4503-2038-2. DOI: [10.1145/2487788.2488034](#).
- Al-Hamar, M., R. Dawson and J. Al-Hamar (2011). "The need for education on phishing: A survey comparison of the UK and Qatar". *Campus-Wide Information Systems*, 28(5): pp. 308–319. DOI: [10.1108/10650741111181580](#).
- Han, W., Y. Cao, E. Bertino and J. Yong (2012). "Using automated individual white-list to protect web digital identities". *Expert Systems with Applications*, 39(15): pp. 11861–11869. DOI: [10.1016/j.eswa.2012.02.020](#).
- Hauch, V., S. L. Sporer, S. W. Michael and C. a. Meissner (May 2014). "Does Training Improve the Detection of Deception? A Meta-Analysis". *Communication Research*: pp. 1–61. ISSN: 0093-6502. DOI: [10.1177/0093650214534974](#).
- He, M., S.-J. Horng, P. Fan, M. Khan, R.-S. Run, J.-L. Lai, R.-J. Chen and A. Sutanto (2011). "An efficient phishing webpage detector".

- Expert Systems with Applications, 38(10): pp. 12018–12027. DOI: [10.1016/j.eswa.2011.01.046](https://doi.org/10.1016/j.eswa.2011.01.046).
- Henzinger, M. (2006). “Finding near-duplicate web pages”. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '06*. DOI: [10.1145/1148170.1148222](https://doi.org/10.1145/1148170.1148222).
- Herley, C. and D. Florêncio (2008). “A profitless endeavor: Phishing as a Tragedy of the Commons”. In: *Proceedings of the 2008 workshop on New security paradigms - NSPW '08*. New York, New York, USA: ACM Press, p. 59. ISBN: 978-1-60558-341-9. DOI: [10.1145/1595676.1595686](https://doi.org/10.1145/1595676.1595686).
- Herzberg, A. (2009). “Why Johnny can’t surf (safely)? Attacks and defenses for web users”. *Computers and Security*, 28(1-2): pp. 63–71. DOI: [10.1016/j.cose.2008.09.007](https://doi.org/10.1016/j.cose.2008.09.007).
- Hinson, G. (2010). “There Must be Thirty Ways to Steal Your ID”. *ED-PACS*, 41(5): pp. 1–15. DOI: [10.1080/07366981.2010.495677](https://doi.org/10.1080/07366981.2010.495677).
- Hodgson, P. (2005). “The threat to identity from new and unknown malware”. *BT Technology Journal*, 23(4): pp. 107–112. DOI: [10.1007/s10550-006-0012-2](https://doi.org/10.1007/s10550-006-0012-2).
- Holland, J., A. S. Silva and R. Mace (2012). “Lost letter measure of variation in altruistic behaviour in 20 neighbourhoods.” *PloS one*, 7(8). ISSN: 1932-6203. DOI: [10.1371/journal.pone.0043294](https://doi.org/10.1371/journal.pone.0043294).
- Honan, M. (2012). “How Apple and Amazon Security Flaws Led to My Epic Hacking”. *Wired*. URL: <http://www.wired.com/2012/08/apple-amazon-mat-honan-hacking/> (Retrieved 2013-06-15).
- Hong, J. (2012). “The state of phishing attacks”. *Communications of the ACM*, 55(1): pp. 74–81. ISSN: 00010782. DOI: [10.1145/2063176.2063197](https://doi.org/10.1145/2063176.2063197).
- Huber, M., M. Mulazzani, E. Weippl, G. Kitzler and S. Goluch (2011). “Friend-in-the-middle attacks: Exploiting social networking sites for spam”. *IEEE Internet Computing*, 15(3): pp. 28–34. DOI: [10.1109/MIC.2011.24](https://doi.org/10.1109/MIC.2011.24).
- Hutchings, A. and H. Hayes (2009). “Routine Activity Theory and Who Gets Caught in the 'Net'?” *Current Issues in Criminal Justice*, 20(3): pp. 433–451.
- Ilchev, S. and V. Ilchev (2012). “Modular data hiding for improved web-portal security”. In: *Proceedings of the 13th International Conference on Computer Systems and Technologies*, pp. 187–194. DOI: [10.1145/2383276.2383305](https://doi.org/10.1145/2383276.2383305).
- Jagatic, T. N., N. A. Johnson, M. Jakobsson and F. Menczer (Oct. 2007). “Social phishing”. *Communications of the ACM*, 50(10): pp. 94–100. ISSN: 0001-0782. DOI: [10.1145/1290958.1290968](https://doi.org/10.1145/1290958.1290968).
- Jahankhani, H. (2009). “The behaviour and perceptions of on-line consumers: Risk, risk perception and trust”. *International Journal of Information Science and Management*, 7(1): pp. 79–90.

- Jakobsson, M. and S. Stamm (Nov. 2007). "Web Camouflage: Protecting Your Clients from Browser-Sniffing Attacks". *IEEE Security and Privacy*, 5(6): pp. 16–24. ISSN: 1540-7993. DOI: [10.1109/MSP.2007.182](#).
- Jakobsson, M. and S. Myers (2007). *Phishing and Countermeasures*. John Wiley & Sons, Inc. ISBN: 0-471-78245-9.
- Jakobsson, M. and J. Ratkiewicz (2006). "Designing ethical phishing experiments: A study of (ROT13) rOnl query features". In: *Proceedings of the 15th international conference on World Wide Web - WWW '06*. New York, New York, USA: ACM Press, pp. 513–522. ISBN: 1-59593-323-9. DOI: [10.1145/1135777.1135853](#).
- Jakobsson, M., A. Tsow, A. Shah, E. Blevis and Y.-K. Lim (2007). "What instills trust? a qualitative study of phishing". In: *FC'07/USEC'07: Proceedings of the 11th International Conference on Financial cryptography and 1st International conference on Usable Security*. Berlin, Heidelberg: Springer-Verlag, pp. 356–361. DOI: [10.1007/978-3-540-77366-5\\_32](#).
- James, L. (2005). *Phishing exposed*. Rockland, MA, USA: Syngress Publishing Inc. ISBN: 1-59749-030-X.
- Jo, I., E. Jung and H. Yeom (2013). "Interactive website filter for safe web browsing". *Journal of Information Science and Engineering*, 29(1): pp. 115–131.
- John, J. P., A. Moshchuk, S. D. Gribble and A. Krishnamurthy (2009). "Studying Spamming Botnets Using Botlab". In: *6th USENIX Symposium on Networked Systems Design and Implementation, NSDI '09*. Vol. 9, pp. 291–306.
- Kahneman, D. (2012). *Thinking, Fast and Slow*. Penguin Books UK. ISBN: 978-0-374-53355-7.
- Khonji, M., Y. Iraqi and A. Jones (2013). "Phishing Detection: A Literature Survey". *IEEE Communications Surveys & Tutorials*, 15(4): pp. 2091–2121. ISSN: 1553-877X. DOI: [10.1109/SURV.2013.032213.00009](#).
- Khot, R. A., P. Kumaraguru and K. Srinathan (2012). "WYSWYE: shoulder surfing defense for recognition based graphical passwords". In: *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI)*. Melbourne, Australia, pp. 285–294. ISBN: 978-1-4503-1438-1. DOI: [10.1145/2414536.2414584](#).
- Kieseberg, P., M. Leithner, M. Mulazzani, L. Munroe, S. Schrittwieser, M. Sinha and E. Weippl (2010). "QR code security". In: *Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia - MoMM '10*. New York, New York, USA: ACM Press, p. 430. ISBN: 978-1-4503-0440-5. DOI: [10.1145/1971519.1971593](#).
- Kim, Y.-G., M. Lee, S. Cho and S. Cha (2012). "A quantitative approach to estimate a website security risk using whitelist". *Security and*

- Communication Networks, 5(10): pp. 1181–1192. DOI: [10.1002/sec.420](#).
- Kirda, E. and C. Kruegel (2005). “Protecting users against phishing attacks with AntiPhish”. In: *Proceedings of the International Computer Software and Applications Conference*. Vol. 1, pp. 517–524. DOI: [10.1109/COMPSAC.2005.126](#).
- Kirlappos, I. and M. A. Sasse (2012). “Security Education against Phishing: A Modest Proposal for a Major Rethink”. *IEEE Security & Privacy Magazine*, 10(2): pp. 24–32. ISSN: 1540-7993. DOI: [10.1109/MSP.2011.179](#).
- Kitchenham, B. and S. Charters (2007). *Guidelines for performing systematic literature reviews in software engineering*. Tech. rep. EBSE-2007-01. Software Engineering Group, Keele University.
- Kleinnijenhuis, J. (2015). “Vijf excuses om niet over te stappen van bank”. Trouw. URL: <https://www.trouw.nl/home/vijf-excuses-om-niet-over-te-stappen-van-bank-aa648cff/> (Retrieved 2017-03-01).
- Klensin, J. C. (Apr. 2001). *Simple Mail Transfer Protocol*. RFC 2821. RFC Editor, pp. 1–79. URL: <https://tools.ietf.org/html/rfc2821#section-7.1>.
- Klüpfel, H. (2007). “The simulation of crowd dynamics at very large events – Calibration, empirical data, and validation”. In: *Pedestrian and Evacuation Dynamics 2005*. Springer, pp. 285–296. ISBN: 978-3-540-47064-9. DOI: [10.1007/978-3-540-47064-9\\_25](#).
- Knight, W. (2005). “Caught in the net”. *IEE Review*, 51(7): pp. 26–30. DOI: [10.1049/ir:20050702](#).
- Kruglanski, A. W. and G. Gigerenzer (2011). “Intuitive and deliberate judgments are based on common principles.” *Psychological Review*, 118(1): pp. 97–109. ISSN: 0033-295X. DOI: [10.1037/a0020762](#).
- Kumar, C. V. and G. Santhi (2012). “Optimized near Duplicate Matching scheme for E-mail Spam Detection”. In: *International Conference on Networks and Cyber Security*. Vol. 17, p. 16.
- Kumaraguru, P., A. Acquisti and L. F. Cranor (2006). “Trust modelling for online transactions”. *Proceedings of the 2006 International Conference on Privacy, Security and Trust Bridge the Gap Between PST Technologies and Business Services - PST ’06*, 28: p. 1. ISSN: 03010449. DOI: [10.1145/1501434.1501448](#).
- Kumaraguru, P., J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair and T. Pham (2009). “School of phish: A Real-World Evaluation of Anti-Phishing Training”. In: *Proceedings of the 5th Symposium on Usable Privacy and Security*. New York, New York, USA: ACM Press. ISBN: 978-1-60558-736-3. DOI: [10.1145/1572532.1572536](#).
- Kumaraguru, P., Y. Rhee, A. Acquisti, L. F. Cranor, J. Hong and E. Nunge (2007). “Protecting people from phishing: The Design and Eval-

- uation of an Embedded Training Email System". In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. New York, New York, USA: ACM Press, pp. 905–914. ISBN: 978-1-59593-593-9. DOI: [10.1145/1240624.1240760](#).
- Kumaraguru, P., Y. Rhee, S. Sheng, S. Hasan, A. Acquisti, L. F. Cranor and J. Hong (2007). "Getting Users to Pay Attention to Anti-Phishing Education: Evaluation of Retention and Transfer". In: *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*. eCrime '07. New York, NY, USA: ACM, pp. 70–81. ISBN: 978-1-59593-939-5. DOI: [10.1145/1299015.1299022](#).
- Kumaraguru, P., S. Sheng, A. Acquisti, L. F. Cranor and J. Hong (2008). "Lessons from a real world evaluation of anti-phishing training". In: *2008 eCrime Researchers Summit*. IEEE, pp. 1–12. ISBN: 978-1-4244-2969-1. DOI: [10.1109/ECRIME.2008.4696970](#).
- Kumaraguru, P., S. Sheng, A. Acquisti, L. F. Cranor and J. Hong (2010). "Teaching Johnny not to fall for phish". *ACM Transactions on Internet Technology*, 10(2): pp. 1–31. ISSN: 15335399. DOI: [10.1145/1754393.1754396](#).
- Larcom, G. and A. Elbirt (2006). "Gone phishing". *IEEE Technology and Society Magazine*, 25(3): pp. 52–55. DOI: [10.1109/MTAS.2006.1700023](#).
- Lenhart, A. (2015). *Teens, Social Media and Technology Overview 2015*. Tech. rep. Pew Research Center.
- Lenton, D. (2005). "Bigger phish to fry". *IEE Review*, 51(10): pp. 26–27. DOI: [10.1049/ir:20051001](#).
- Leukfeldt, E. R. (Aug. 2014). "Phishing for Suitable Targets in The Netherlands: Routine Activity Theory and Phishing Victimization". *Cyberpsychology, Behavior, and Social Networking*, 17(8): pp. 551–555. ISSN: 2152-2715. DOI: [10.1089/cyber.2014.0008](#).
- Levine, T. R., H. S. Park and S. a. McCornack (1999). "Accuracy in detecting truths and lies: Documenting the "veracity effect"". *Communication Monographs*, 66(2): pp. 125–144. ISSN: 0363-7751. DOI: [10.1080/03637759909376468](#).
- Levy, E. (2004). "Interface illusions". *IEEE Security and Privacy*, 2(6): pp. 66–69. ISSN: 1540-7993. DOI: [10.1109/MSP.2004.104](#).
- Li, L., M. Helenius and E. Berki (2012). "A usability test of whitelist and blacklist-based anti-phishing application". In: *Proceedings of the 16th International Academic MindTrek Conference*, pp. 195–202. DOI: [10.1145/2393132.2393170](#).
- Liggett, L., C. Blair and S. Kennison (2010). "Measuring Gender Differences in Attitudes Using the Lost-Letter Technique". *Journal of Scientific Psychology*: pp. 16–24.
- Liu, G., B. Qiu and L. Wenying (2010). "Automatic Detection of Phishing Target from Phishing Webpage". In: *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*, pp. 4153–4156.

- Liu, W., H. Guanglin, X. Liu, D. Xiaotie and M. Zhang (2005). "Phishing webpage detection". In: *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pp. 560–564. DOI: [10.1109/ICDAR.2005.190](https://doi.org/10.1109/ICDAR.2005.190).
- Ludl, C., S. McAllister, E. Kirda and C. Kruegel (2007). "On the effectiveness of techniques to detect phishing sites". In: *Proceedings of the 4th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*. Vol. 4579, pp. 20–39.
- Manku, G. S., A. Jain and A. Das Sarma (2007). "Detecting near-duplicates for web crawling". *Proceedings of the 16th international conference on World Wide Web, WWW*. DOI: [10.1145/1242572.1242592](https://doi.org/10.1145/1242572.1242592).
- Mann, H. B. and D. R. Whitney (Mar. 1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". *The Annals of Mathematical Statistics*, 18(1): pp. 50–60. ISSN: 0003-4851. DOI: [10.1214/aoms/1177730491](https://doi.org/10.1214/aoms/1177730491).
- Maurer, M.-E. and L. Höfer (2012). "Sophisticated phishers make more spelling mistakes: Using URL similarity against phishing". In: *Proceedings of the 4th International Symposium on Cyberspace Safety and Security*. Vol. 7672, pp. 414–426. DOI: [10.1007/978-3-642-35362-8\\_31](https://doi.org/10.1007/978-3-642-35362-8_31).
- Mayer, R. C., J. H. Davis and F. D. Schoorman (1995). "An Integrative Model Of Organizational Trust". *The Academy of Management Review*, 20(3): pp. 709–734. ISSN: 0363-7425. DOI: [10.5465/AMR.1995.9508080335](https://doi.org/10.5465/AMR.1995.9508080335).
- Mayhorn, C. B. and P. G. Nyeste (2012). "Training users to counteract phishing". In: *Proceedings of the Human Factors and Ergonomics Society*. Vol. 41, pp. 1956–1960. DOI: [10.3233/WOR-2012-1054-3549](https://doi.org/10.3233/WOR-2012-1054-3549).
- McFedries, P. (2006). "Technically speaking: Gone phishin". *IEEE Spectrum*, 43(4): p. 80. DOI: [10.1109/MSPEC.2006.1611765](https://doi.org/10.1109/MSPEC.2006.1611765).
- McNaught, C. and P. Lam (2010). "Using Wordle as a supplementary research tool". *The qualitative report*, 15(3): pp. 630–643.
- McNealy, J. (2008). "Angling for phishers: Legislative responses to deceptive e-mail". *Communication Law and Policy*, 13(2): pp. 275–300. DOI: [10.1080/10811680801941292](https://doi.org/10.1080/10811680801941292).
- Merriam-Webster (Dec. 2013). *Phishing*. URL: <http://www.merriam-webster.com/dictionary/phishing> (Retrieved 2013-12-20).
- Merritt, C. and R. Fowler (1948). "The pecuniary honesty of the public at large". *The Journal of Abnormal and Social Psychology*, 43(1): pp. 90–93. DOI: [10.1037/h0061846](https://doi.org/10.1037/h0061846).
- Milgram, S., L. Mann and S. Harter (1965). "The Lost-Letter technique: A Tool of Social Research". *Public Opinion Quarterly*, 29(3): pp. 437–438.

- Mills, E. (2009). "Facebook hit by phishing attacks for a second day". CNET News. URL: <http://www.cnet.com/news/facebook-hit-by-phishing-attacks-for-a-second-day/>.
- Mills, J. and S. Byun (July 2006). "Cybercrimes against Consumers: Could Biometric Technology Be the Solution?" IEEE Internet Computing, 10(4): pp. 64–71. ISSN: 1089-7801. DOI: [10.1109/MIC.2006.73](https://doi.org/10.1109/MIC.2006.73).
- Mohebzada, J., A. El Zarka, A. Bhojani and A. Darwish (Mar. 2012). "Phishing in a university community: Two large scale phishing experiments". In: *Proceedings of the International Conference on Innovations in Information Technology (IIT)*, pp. 249–254. DOI: [10.1109/INNOVATIONS.2012.6207742](https://doi.org/10.1109/INNOVATIONS.2012.6207742).
- Montgomery, K. C., J. Chester, S. A. Grier and L. Dorfman (2012). "The New Threat of Digital Marketing". *Pediatric Clinics of North America*, 59(3): pp. 659–675. ISSN: 00313955. DOI: [10.1016/j.pcl.2012.03.022](https://doi.org/10.1016/j.pcl.2012.03.022).
- Moore, T. (2007). "Phishing and the economics of e-crime". *Infosecurity*, 4(6): pp. 34–37. DOI: [10.1016/S1754-4548\(07\)70148-1](https://doi.org/10.1016/S1754-4548(07)70148-1).
- Moore, T. and R. Clayton (2007). "Examining the impact of website take-down on phishing". *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit on - eCrime '07*: pp. 1–13. DOI: [10.1145/1299015.1299016](https://doi.org/10.1145/1299015.1299016).
- Moore, T. and R. Clayton (2008). "Evaluating the Wisdom of Crowds in Assessing Phishing Websites". *12th International Financial Cryptography and Data Security Conference, FC 2008*. LNCS 5143: pp. 16–30. ISSN: 1611-3349. DOI: [10.1007/978-3-540-85230-8\\_2](https://doi.org/10.1007/978-3-540-85230-8_2).
- Moore, T. and R. Clayton (2010). *How hard can it be to measure phishing?* Tech. rep. URL: <https://www.cl.cam.ac.uk/~rnc1/cyberbias.pdf>.
- Moore, T., R. Clayton and R. Anderson (2009). "The Economics of Online Crime". *Journal of Economic Perspectives*, 23(3): pp. 3–20.
- Moran, T. and T. Moore (2010). "The phish-market protocol: Secure sharing between competitors". *IEEE Security and Privacy*, 8(4): pp. 40–45. DOI: [10.1109/MSP.2010.138](https://doi.org/10.1109/MSP.2010.138).
- Nykodym, N., L. Kahle-Piasecki, S. Ariss and T. Toussaint (2010). "Cybercrime and business: How to not get caught by the online phisher". *Journal of International Commercial Law and Technology*, 5(4): pp. 252–259.
- O'Brien, R. M. (Sept. 2007). "A Caution Regarding Rules of Thumb for Variance Inflation Factors". *Quality & Quantity*, 41(5): pp. 673–690. ISSN: 0033-5177. DOI: [10.1007/s11135-006-9018-6](https://doi.org/10.1007/s11135-006-9018-6).
- Olurin, M., C. Adams and L. Logrippo (2012). "Platform for privacy preferences (P3P): Current status and future directions". In: *Tenth Annual International Conference on Privacy, Security and Trust*. IEEE, pp. 217–220. DOI: [10.1109/PST.2012.6297943](https://doi.org/10.1109/PST.2012.6297943).



- Osborne, J. and E. Waters (2002). "Four assumptions of multiple regression that researchers should always test". *Practical Assessment, Research and Evaluation*, 8(2): pp. 1–5. ISSN: 15317714.
- Oxford University Press (Apr. 2010). *Phishing*. URL: <http://english.oxforddictionaries.com/definition/phishing> (Retrieved 2013-12-20).
- Oxford University Press (June 2012). *Phishing, n. OED Online*. URL: <http://www.oed.com/view/Entry/264317> (Retrieved 2013-12-20).
- Parno, B., C. Kuo and A. Perrig (2006). "Phoolproof phishing prevention". In: *Proceedings of the 10th International Conference on Financial Cryptography and Data Security*, pp. 1–19.
- Parsons, K., A. McCormac, M. Pattinson, M. Butavicius and C. Jerram (2015). "The design of phishing studies: Challenges for researchers". *Computers & Security*. In press: pp. 1–13. ISSN: 01674048. DOI: [10.1016/j.cose.2015.02.008](https://doi.org/10.1016/j.cose.2015.02.008).
- Paulissen, L. and J. van Wilsem (2015). *Dat heeft iemand anders gedaan! Een studie naar slachtofferschap en modus operandi van identiteitsfraude in Nederland*. Politie en Wetenschap 82. Apeldoorn: Reed Business, Amsterdam.
- Paulson, L. D. (Apr. 2010). "New Technique Provides Energy Wirelessly". *Computer*, 43(4): pp. 16–19. ISSN: 0018-9162. DOI: [10.1109/MC.2010.110](https://doi.org/10.1109/MC.2010.110).
- Pfeiffer, T., M. Kauer and J. Röth (2014). "'A Bank Would Never Write That!' - A Qualitative Study on E-Mail Trust Decisions". In: *INFORMATIK 2014*. Ed. by E. Plödereder, L. Grunske and E. Schneider. Vol. 232. GI-Edition Lecture Notes in Informatics. Bonn, pp. 2093–2104.
- Piper, P. (2007). "A newer, more profitable aquaculture". *Searcher: Magazine for Database Professionals*, 15(9): pp. 40–47.
- Pratt, T. C., K. Holtfreter and M. D. Reisig (2010). "Routine Online Activity and Internet Fraud Targeting: Extending the Generality of Routine Activity Theory". *Journal of Research in Crime and Delinquency*, 47(3): pp. 267–296. ISSN: 0022-4278. DOI: [10.1177/0022427810365903](https://doi.org/10.1177/0022427810365903).
- Purkait, S. (2012). "Phishing counter measures and their effectiveness – literature review". *Information Management & Computer Security*, 20(5): pp. 382–420. ISSN: 0968-5227. DOI: [10.1108/09685221211286548](https://doi.org/10.1108/09685221211286548).
- Rabobank (2017). Private communication.
- Rader, E., R. Wash and B. Brooks (2012). "Stories As Informal Lessons About Security". In: *Proceedings of the Eighth Symposium on Usable Privacy and Security*. SOUPS '12. New York, NY, USA: ACM, 6:1–6:17. ISBN: 978-1-4503-1532-6. DOI: [10.1145/2335356.2335364](https://doi.org/10.1145/2335356.2335364).
- Ramzan, Z. and C. Wüest (2007). "Phishing attacks: Analyzing trends in 2006". In: *4th Conference on Email and Anti-Spam, CEAS 2007*.



- Mountain View, California, USA. DOI: [10.1038/nprot.2010.202](https://doi.org/10.1038/nprot.2010.202).
- Ranganayakulu, D., L. Kavisankar and C. Chellappan (2011). "Enhanced e-mail authentication against spoofing attacks to mitigate phishing". *European Journal of Scientific Research*, 54(1): pp. 165–175.
- Ray, E. and E. Schultz (2007). "An early look at Windows Vista security". *Computer Fraud and Security*, 2007(1): pp. 4–7. DOI: [10.1016/S1361-3723\(07\)70005-2](https://doi.org/10.1016/S1361-3723(07)70005-2).
- Reyns, B. W., B. Henson and B. S. Fisher (2011). "Being Pursued Online: Applying Cyberlifestyle-Routine Activities Theory to Cyberstalking Victimization". *Criminal Justice and Behavior*, 38(11): pp. 1149–1169. ISSN: 0093-8548. DOI: [10.1177/0093854811421448](https://doi.org/10.1177/0093854811421448).
- Robila, S. A. and J. W. Ragucci (2006). "Don't be a phish: Steps in User Education". In: *Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education*. ITICSE '06. New York, NY, USA: ACM, pp. 237–241. ISBN: 1-59593-055-8. DOI: [10.1145/1140124.1140187](https://doi.org/10.1145/1140124.1140187).
- Ross, D. (2009). "Ars dictaminis perverted: The personal solicitation E-mail as a genre". *Journal of Technical Writing and Communication*, 39(1): pp. 25–41. DOI: [10.2190/TW.39.1.c](https://doi.org/10.2190/TW.39.1.c).
- Ross, P. (Jan. 2006). "Microsoft to spammers: go phish". *IEEE Spectrum*, 43(1): pp. 48–49. ISSN: 0018-9235. DOI: [10.1109/MSPEC.2006.1572356](https://doi.org/10.1109/MSPEC.2006.1572356).
- Saberi, A., M. Vahidi and B. Bidgoli (2007). "Learn to detect phishing scams using learning and ensemble methods". In: *Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 311–314. DOI: [10.1109/WIIATW.2007.4427596](https://doi.org/10.1109/WIIATW.2007.4427596).
- Satterthwaite, F. E. (1946). "An Approximate Distribution of Estimates of Variance Components". *Biometrics Bulletin*, 2(6): p. 110. ISSN: 00994987. DOI: [10.2307/3002019](https://doi.org/10.2307/3002019).
- Schank, R. and R. Abelson (1975). "Scripts, plans, and knowledge". In: *Advance Papers of the Fourth International Joint Conference on Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 151–157.
- Shahriar, H. and M. Zulkernine (2012). "Trustworthiness testing of phishing websites: A behavior model-based approach". *Future Generation Computer Systems*, 28(8): pp. 1258–1271. DOI: [10.1016/j.future.2011.02.001](https://doi.org/10.1016/j.future.2011.02.001).
- Sheng, S., M. Holbrook, P. Kumaraguru, L. F. Cranor and J. S. Downs (2010). "Who falls for phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions". In: *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. New York, New York, USA: ACM Press, pp. 373–382. ISBN: 978-1-60558-929-9. DOI: [10.1145/1753326.1753383](https://doi.org/10.1145/1753326.1753383).

- Sheng, S., P. Kumaraguru, L. Cranor, J. Hong and A. Acquisti (Oct. 2009). "Improving phishing countermeasures: An analysis of expert interviews". In: *2009 eCrime Researchers Summit*. IEEE, pp. 1–15. ISBN: 978-1-4244-4625-4. DOI: [10.1109/ECRIME.2009.5342608](https://doi.org/10.1109/ECRIME.2009.5342608).
- Sheng, S., B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong and E. Nunge (2007). "Anti-Phishing Phil". In: *Proceedings of the 3rd symposium on Usable privacy and security - SOUPS '07*. New York, New York, USA: ACM Press, pp. 88–99. ISBN: 978-1-59593-801-5. DOI: [10.1145/1280680.1280692](https://doi.org/10.1145/1280680.1280692).
- Shotland, R. L., W. G. Berger and R. Forsythe (1970). "A validation of the lost-letter technique". *The Public Opinion Quarterly*, 34(2): pp. 278–281.
- Simon, W. E. (1971). "Return Rates Of "Lost" Letters As A Function Of Whether The Letter Is Stamped And The Apparent Importance Of The Letter". *Psychological Reports*, 29(3): pp. 937–938. ISSN: 0033-2941. DOI: [10.2466/pr0.1971.29.3.937](https://doi.org/10.2466/pr0.1971.29.3.937).
- Simon, W. E. and M. J. Gillen (1971). "Return Rates Of "Lost" Letters As A Function Of Whether The Letter Is Stamped And Amount Of Money Apparently In The Letter". *Psychological Reports*, 29(1): pp. 141–142. ISSN: 0033-2941. DOI: [10.2466/pr0.1971.29.1.141](https://doi.org/10.2466/pr0.1971.29.1.141).
- Slovic, P. (1966). "Risk-Taking in Children: Age and Sex Differences". *Child Development*, 37(1). DOI: [10.2307/1126437](https://doi.org/10.2307/1126437).
- Smith, R., M. Holmes and P. Kaufmann (1999). "Nigerian advance fee fraud". *Trends & Issues in Crime and Criminal Justice*, (121): pp. 1–6.
- Someren, M. W. van, Y. F. Barnard and J. A. C. Sandberg (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press. ISBN: 0-12-714270-3.
- Sood, A. K. and R. J. Enbody (2013). "Targeted cyberattacks: A superset of advanced persistent threats". *IEEE Security and Privacy*, 11(1): pp. 54–61. ISSN: 15407993. DOI: [10.1109/MSP.2012.90](https://doi.org/10.1109/MSP.2012.90).
- Sood, S., A. Sarje and K. Singh (2011). "Dynamic identity-based single password anti-phishing protocol". *Security and Communication Networks*, 4(4): pp. 418–427. DOI: [10.1002/sec.169](https://doi.org/10.1002/sec.169).
- Sparks, R. F. (1982). *Research on victims of crime: accomplishments, issues and new directions*. Crime and delinquency issues. Rockville, USA: National Institute of Mental Health.
- Stabek, A., P. Watters and R. Layton (July 2010). "The Seven Scam Types: Mapping the Terrain of Cybercrime". In: *Proceedings of the Second Cybercrime and Trustworthy Computing Workshop (CTC)*, pp. 41–51. DOI: [10.1109/CTC.2010.14](https://doi.org/10.1109/CTC.2010.14).
- Standage, T. (1998). *The Victorian Internet: The remarkable story of the telegraph and the nineteenth century's online pioneers*. Weidenfeld & Nicolson London. ISBN: 978-1-62040-592-5.

- StataCorp (2013). *Obtaining robust variance estimates*. URL: <http://www.stata.com/manuals13/u20.pdf> (Retrieved 2016-04-25).
- Statistics Netherlands (2015). *Veiligheidsmonitor 2014*. Statistics Netherlands, The Hague. ISBN: 978-90-357-1728-2. URL: <https://www.cbs.nl/nl-nl/publicatie/2015/10/veiligheidsmonitor-2014>.
- Statistics Netherlands (2017). *Veiligheidsmonitor 2016*. Statistics Netherlands, The Hague. ISBN: 978-90-357-2157-9. URL: <https://www.rijksoverheid.nl/documenten/rapporten/2017/03/01/tk-bijlage-veiligheidsmonitor-2016>.
- Stern, S. E. and J. E. Faber (1997). "The lost e-mail method: Milgram's lost-letter technique in the age of the Internet". *Behavior Research Methods, Instruments, & Computers*, 29(2): pp. 260–263. ISSN: 0743-3808. DOI: [10.3758/BF03204823](https://doi.org/10.3758/BF03204823).
- Sweeney, L. (2006). "Protecting job seekers from identity theft". *IEEE Internet Computing*, 10(2): pp. 74–78. DOI: [10.1109/MIC.2006.40](https://doi.org/10.1109/MIC.2006.40).
- Symantec (2012). *The Symantec Smartphone Honey Stick Project*. Symantec. URL: <http://www.symantec.com/content/en/us/about/presskits/b-symantec-smartphone-honey-stick-project.en-us.pdf> (Retrieved 2012-12-28).
- Tetmeyer, A. and H. Saiedian (2010). "Security Threats and Mitigating Risk for USB Devices". *IEEE Technology and Society Magazine*, 29(4): pp. 44–49. DOI: [10.1109/MTS.2010.939228](https://doi.org/10.1109/MTS.2010.939228).
- The Canadian Press (Dec. 2012). *Government USB Key With Personal Info Of Thousands Of Canadians Goes Missing*. Huffington Post. URL: [http://www.huffingtonpost.ca/2012/12/28/government-personal-info-missing-usb-key-canada\\_n-2377503.html](http://www.huffingtonpost.ca/2012/12/28/government-personal-info-missing-usb-key-canada_n-2377503.html) (Retrieved 2012-12-28).
- The Guardian (Feb. 2012). *Nuclear plant data lost by health and safety watchdog employee*. The Guardian. URL: <http://www.guardian.co.uk/environment/2012/feb/17/nuclear-plant-lost-health-safety> (Retrieved 2012-02-17).
- Theodore Montanye, I., F. Ronald and R. Kenneth (1971). "Assessing prejudice toward Negroes at three universities using the lost-letter technique". *Psychological Reports*, 29(2): pp. 531–537. DOI: [10.2466/pr0.1971.29.2.531](https://doi.org/10.2466/pr0.1971.29.2.531).
- Thiyagarajan, P., G. Aghila Prof. and V. Prasanna Venkatesan (2012). "Pixastic: Steganography based anti-phishing browser plug-in". *Journal of Internet Banking and Commerce*, 17(1).
- Tsow, A. and M. Jakobsson (2007). *Deceit and Deception: A Large User Study of Phishing*. Tech. rep. TR649. Indiana University, pp. 1–46.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass.
- Tykocinski, O. and L. Bareket-Bojmel (2009). "The Lost E-Mail Technique: Use of an Implicit Measure to Assess Discriminatory Attitudes Toward Two Minority Groups in Israel". *Journal of Applied*

- Social Psychology, 39(1): pp. 62–81. DOI: [10.1111/j.1559-1816.2008.00429.x](https://doi.org/10.1111/j.1559-1816.2008.00429.x).
- UCLA Statistical Consulting Group (2016). *Regression with Stata: Chapter 2 - Regression Diagnostics*. URL: <http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm> (Retrieved 2016-04-18).
- Vaes, J., M. Paladino and J. Leyens (2002). “The lost e-mail: Prosocial reactions induced by uniquely human emotions”. *British journal of social psychology*, 41(4): pp. 521–534. DOI: [10.1348/014466602321149867](https://doi.org/10.1348/014466602321149867).
- Vamosi, R. (2009). “Security alert: Phishers dangle some brand-new bait”. *PC World*, 27(12): pp. 37–38.
- van den Bosch, A., B. Busser, S. Canisius and W. Daelemans (2007). “An efficient memory-based morphosyntactic tagger and parser for Dutch”. In: *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting*. Ed. by F. van Eynde, P. Dirix, I. Schuurman and V. Vandeghinste. Leuven, Belgium, pp. 99–114.
- Varshney, G., R. Joshi and A. Sardana (2012). “Personal secret information based authentication towards preventing phishing attacks”. *Advances in Intelligent Systems and Computing*, 176: pp. 31–42. DOI: [10.1007/978-3-642-31513-8\\_4](https://doi.org/10.1007/978-3-642-31513-8_4).
- Vasek, M., J. Wadleigh and T. Moore (2015). “Hacking is not random: a case-control study of webserver-compromise risk”. *IEEE Transactions on Dependable and Secure Computing*: to appear. ISSN: 1545-5971. DOI: [10.1109/TDSC.2015.2427847](https://doi.org/10.1109/TDSC.2015.2427847).
- Veilig Bankieren (Dutch Banking Association) (2011). *Nepmail, daar trapt u niet in*. TV commercial (Dutch): <http://youtu.be/VcbHo0E0tKA>. (Retrieved 2017-05-30).
- Verma, R., N. Shashidhar and N. Hossain (2012). “Two-pronged phish snagging”. In: *Proceedings of the 7th International Conference on Availability, Reliability and Security (ARES)*, pp. 174–179. DOI: [10.1109/ARES.2012.51](https://doi.org/10.1109/ARES.2012.51).
- Vidas, T., E. Owusu, S. Wang, C. Zeng, L. F. Cranor and N. Christin (2013). “QRishing: The Susceptibility of Smartphone Users to QR Code Phishing Attacks”. In: *Financial Cryptography Workshops*. Vol. 52-69, pp. 52–69. ISBN: 978-3-642-41319-3. DOI: [10.1007/978-3-642-41320-9\\_4](https://doi.org/10.1007/978-3-642-41320-9_4).
- Vishwanath, A., T. Herath, R. Chen, J. Wang and H. R. Rao (2011). “Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model”. *Decision Support Systems*, 51(3): pp. 576–586. ISSN: 01679236. DOI: [10.1016/j.dss.2011.03.002](https://doi.org/10.1016/j.dss.2011.03.002).
- Vitaliev, D. (Oct. 2010). “The shadownet”. *Engineering Technology*, 5(16): pp. 19–22. ISSN: 1750-9637.
- Wall, D. S. (2010). “The Internet as a Conduit for Criminal Activity”. In: *Information Technology and the Criminal Justice System*. Vol. 15.

- Thousand Oaks, CA, USA: SAGE Publications, Inc., pp. 77–98. DOI: [10.4135/9781452225708.n4](#).
- Wang, J.-H. and H.-C. Chang (2009). “Exploiting Sentence-Level Features for Near-Duplicate Document Detection”. In: *Proceedings of the 5th Asia Information Retrieval Symposium, ARIS*. Vol. 5839. LNCS. Springer, pp. 205–217. DOI: [10.1007/978-3-642-04769-5\\_18](#).
- Wang, J., T. Herath, R. Chen, A. Vishwanath and H. R. Rao (2012). “Phishing susceptibility: An investigation into the processing of a targeted spear phishing email”. *IEEE Transactions on Professional Communication*, 55(4): pp. 345–362. ISSN: 03611434. DOI: [10.1109/TPC.2012.2208392](#).
- Waugh, I., V. Edmund and B. Rienzi (2000). “Assessing attitudes toward gay marriage among selected Christian groups using the lost-letter technique”. *Psychological Reports*, 86(1): pp. 215–218. DOI: [10.2466/pr0.2000.86.1.215](#).
- Wenyin, L., G. Liu, B. Qiu and X. Quan (2012). “Antiphishing through phishing target discovery”. *IEEE Internet Computing*, 16(2): pp. 52–60. DOI: [10.1109/MIC.2011.103](#).
- Whittaker, C., B. Ryner and M. Nazif (2010). “Large-Scale Automatic Classification of Phishing Pages”. In: *Proceedings of the 17th Annual Network and Distributed System Security Symposium (NDSS)*.
- Wilcoxon, F. (1945). “Individual Comparisons by Ranking Methods”. *Biometrics Bulletin*, 1(6): p. 80. DOI: [10.2307/3001968](#).
- Workman, M. (2008). “Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security”. *Journal of the American Society for Information Science and Technology*, 59(4): pp. 662–674. DOI: [10.1002/asi.20779](#).
- Wu, M., R. Miller and S. Garfinkel (2006). “Do security toolbars actually prevent phishing attacks?” In: *Proceedings on the Conference on Human Factors in Computing Systems*, pp. 601–610.
- Wu, M., R. C. Miller and G. Little (2006). “Web wallet: preventing phishing attacks by revealing user intentions”. In: *Proceedings of the second symposium on Usable privacy and security (SOUPS)*. Pittsburgh, Pennsylvania, pp. 102–113. ISBN: 1-59593-448-0. DOI: [10.1145/1143120.1143133](#).
- Xiang, G., J. Hong, C. Rose and L. Cranor (2011). “CANTINA+: A feature-rich machine learning framework for detecting phishing web sites”. *ACM Transactions on Information and System Security*, 14(2). DOI: [10.1145/2019599.2019606](#).
- Xiang, G. and J. Hong (2009). “A hybrid phish detection approach by identity discovery and keywords retrieval”. In: *Proceedings of the 18th International World Wide Web Conference (WWW)*, pp. 571–580. DOI: [10.1145/1526709.1526786](#).

- Yar, M. (2005). "The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory". *European Journal of Criminology*, 2(4): pp. 407–427. DOI: [10.1177/147737080556056](https://doi.org/10.1177/147737080556056).
- Yar, M. (2012). "Sociological and Criminological Theories in the Information Era". In: *Cyber-Safety: An Introduction*. Ed. by W. Stol and R. Leukfeldt. Utrecht: Eleven International Publishing.
- Yearwood, J., D. Webb, L. Ma, P. Vamplew, B. Ofoghi and A. Kelarev (2009). "Applying clustering and ensemble clustering approaches to phishing profiling". In: *Proceedings of the Eighth Australasian Data Mining Conference*, pp. 25–34.
- Yee, K.-P. and K. Sitaker (2006). "Passpet: convenient password management and phishing protection". In: *Proceedings of the 2nd symposium on Usable privacy and security (SOUPS)*. Pittsburgh, PA, pp. 32–43. ISBN: 1-59593-448-0. DOI: [10.1145/1143120.1143126](https://doi.org/10.1145/1143120.1143126).
- Zhang, J., C. Wu, D. Li, Z. Jia, X. Ouyang and Y. Xin (2012). "An empirical analysis of the effectiveness of browser-based antiphishing solutions". *International Journal of Digital Content Technology and its Applications*, 6(7): pp. 216–224. DOI: [10.4156/jdcta.vol6.issue7.27](https://doi.org/10.4156/jdcta.vol6.issue7.27).
- Zhang, L., J. Zhu and T. Yao (2004). "An evaluation of statistical spam filtering techniques". *ACM Transactions on Asian Language Information Processing*, 3(4): pp. 243–269. ISSN: 1530-0226. DOI: [10.1145/1039621.1039625](https://doi.org/10.1145/1039621.1039625).
- Zhang, Y., J. Hong and L. Cranor (2007). "Cantina: A content-based approach to detecting phishing web sites". In: *Proceedings of the 16th International World Wide Web Conference (WWW)*, pp. 639–648. DOI: [10.1145/1242572.1242659](https://doi.org/10.1145/1242572.1242659).
- Zhou, C., C. Leckie and S. Karunasekera (2009). "Collaborative detection of fast flux phishing domains". *Journal of Networks*, 4(1): pp. 75–84.