European Journal of Information Systems

ISSN: 0960-085X (Print) 1476-9344 (Online) Journal homepage: www.tandfonline.com/journals/tjis20

The insider on the outside: a novel system for the detection of information leakers in social networks

Giuseppe Cascavilla, Mauro Conti, David G. Schwartz & Inbal Yahav

To cite this article: Giuseppe Cascavilla, Mauro Conti, David G. Schwartz & Inbal Yahav (2018) The insider on the outside: a novel system for the detection of information leakers in social networks, European Journal of Information Systems, 27:4, 470-485, DOI: 10.1080/0960085X.2017.1387712

To link to this article: https://doi.org/10.1080/0960085X.2017.1387712

EUROPEAN JOURNAL OF INFORMATION SYSTEMS
2018, VOL. 27, NO. 4, 470–485
https://doi.org/10.1080/0960085X.2017.1387712

THE OPERATIONAL RESEARCH SOCIETY
Taylor & Francis
Taylor & Francis Group

EMPIRICAL RESEARCH

The insider on the outside: a novel system for the detection of information leakers in social networks

Giuseppe Cascavilla[a], Mauro Conti[b], David G. Schwartz[c] and Inbal Yahav[c]

[a]Department of Computer Science, University of Rome, Italy; [b]Department of Mathematics, University of Padua, Italy; [c]Graduate School of Business Administration, Bar Ilan University, Israel

ABSTRACT

Confidential information is all too easily leaked by naive users posting comments. In this paper we introduce DUIL, a system for Detecting Unintentional Information Leakers. The value of DUIL is in its ability to detect those responsible for information leakage that occurs through comments posted on news articles in a public environment, when those articles have withheld material non-public information. DUIL is comprised of several artefacts, each designed to analyse a different aspect of this challenge: the information, the user(s) who posted the information, and the user(s) who may be involved in the dissemination of information. We present a design science analysis of DUIL as an information system artefact comprised of social, information, and technology artefacts. We demonstrate the performance of DUIL on real data crawled from several Facebook news pages spanning two years of news articles.

ARTICLE HISTORY
Received 14 March 2016
Revised 18 June 2017
Accepted 27 August 2017

ACCEPTING EDITOR
Pär Ågerfalk

ASSOCIATE EDITOR
Paul Lowry

KEYWORDS
Cybersecurity; online social networks; information leakers; sensitive information; threat detection; design science research

1. Introduction

In the field of information security, an insider threat is defined as "the organisational member who is a *trusted agent* inside the firewall" (Im & Baskerville, 2005). Information security specialists try to protect against information leakage by detecting and blocking insider threats brought on by actors who are, by definition, organisational insiders. Warkentin and Willison (2009) describe the greatest insider threat as the "employee or other constituent with a valid user-name and password (who) regularly interacts with the information assets of the organization" (p. 102). Bishop and Gates (2008) extend the definition of insider threat to those with access irrespective of an *inside* affiliation through either (a) violation of a security policy using legitimate access; or (b) violation of an access control policy by obtaining unauthorised access.

Rather than focusing on access, the field of finance provides a more information-centric definition of an insider, being *anyone who is privy to information that has not been released to the general public*. This is based on the concept of *information asymmetry* (Huddart & Ke, 2007), where an insider is deemed to be anyone who has an *information advantage* over other market participants. In that context, the goal is to detect or prevent *insider trading* which is the practice of trading in the securities markets by those in possession of *material non-public information* (Karsch, 1984). Following this definition, should a senior manager in a public company share unreleased material information about company performance with his neighbour, and as a result that neighbour trades in the public markets, the neighbour has committed an act of *insider trading* and is treated as a de-facto insider (Strudler & Orts, 1999).

Following the above definition, when an organisation has information that it intends to remain hidden, secret, or censored, anyone who possesses that information, regardless of organisational affiliation or access control, becomes an *insider*. What was once considered an organisational problem focused on identifying threats emanating from those connected to the organisation, has quickly become a broader problem in which any member of society at large may have access to material non-public content generated by users through a social network. We harness user-generated content (UGC, see Agichtein, Castillo, Donato, Gionis, & Mishne, 2008) to identify internal and external information leakers. UGC is generally characterised by (1) a broad and unrestricted user base, (2) user identifying information, (3) user social network and (4) the contributed content. Examples of UGC types include blogs, video channels such as YouTube and micro-blogs such as Twitter.

In this work, we focus on comments to articles containing incomplete information, where the hidden information is held by insiders. Our demonstrations

CONTACT Inbal Yahav inbal.yahav@biu.ac.il

present censored military articles in which the identity of personnel is withheld. Insiders – particularly those *in the know* often outside of the military organisation, can share their private information through comments. Similar forms of identity protection can be found in different domains and cultures, such as non-release of rape victim names in trial reporting or non-disclosure of customer names in corporate announcements as discussed in Cascavilla et al. (2015). We suggest that with the present information environment of social networks, cyber-security against an insider threat must consider the external insider – the insider on the outside.

The cyber environment is becoming increasingly complex. The field of intelligence gathering is concerned with covert operations, attempts to crack and access protected information assets and supporting infrastructures, and the collection and analysis of Open Source INTelligence (OSINT) (Burattin, Cascavilla, & Conti 2015; Kandias, Stavrou, Bozovic, & Gritzalis, 2013). In the intelligence community, the term "open" refers to overt publicly available sources – as opposed to covert or clandestine sources. Hence, OSINT approaches aim at extracting knowledge from publicly available sources (Kandias et al., 2013), which includes on-line comments. Claudio (2009) discusses how both social network analysis and visualisation are fundamental to cyber deterrence strategy, pointing to the growing need to develop advanced detection systems, incorporating linguistic cues (Zhou, Burgoon, Twitchell, Qin, & Nunamaker, 2004) and visualisations to effectively identify OSINT social network threats. Earlier results reported in Schwartz et al. (2017) describe the information leakage problem and detection. Expanding our previous work, the present work describes the full system view and functionality.

An abundance of information is exchanged in the commenting environments of news articles. In this work, we present DUIL, a system for Detecting Unintentional Information Leakers. This novel system for information leaker detection is initially applied to a set of censored news articles. DUIL is comprised of a loosely coupled set of artefacts that implement a multi-stage leaker detection process which can be generalised for the detection of leakers in other information environments by replacing certain artefacts in the system, so long as the UGC characteristics remain. The study and analysis of this type of system naturally points us in the direction of design science research (DSR) (Hevner & Chatterjee, 2010; Hevner, March, & Park, 2004).

Our work presents two **practical contributions**. First, our work raises the important question of leakers, who unintentionally uncover hidden information via UGC, specifically, comments to on-line news in the public sphere. Second, we present a novel end-to-end system that is designed to detect such information leakers, along with their social network. We present a modular architecture system that can be tuned to any (user-) given news context, as long as the data analysed is UGC, as we demonstrate through a collected data-set of censored news articles.

Our work also presents two secondary **theoretical contributions** to design science research that are distinct from the novel information system itself. The first theoretical contribution relates to use of the Lee, Thomas, and Baskerville (2015) IS artefact framework. We show how this framework, when applied to a complex multi-artefact information system, improves expressiveness and clarity in presenting design science research (DSR).

The second theoretical contribution relates to the Gregor and Hevner (2013) DSR Knowledge Contribution Framework. We draw upon the framework of Lee et al. (2015) to present enhancements to the DSR Knowledge Classification framework, extending its applicability to complex multi-artefact information systems and adding expressiveness to the original four-quadrant classification. We believe that the above will contribute to the consistency of DSR reporting.

This introductory section has focused on motivating the research problem, the detection of unintentional information leakers in social networks. In our Literature Review in the next section, we discuss contributions addressing related research problems and elaborate on the research gap. A formal introduction of our chosen methodology in the third section is followed by the fourth section on the Study Scope which system design. We then demonstrate and evaluate in the sixth section the operation and effectiveness of our DUIL system through a series of experiments designed around the current UGC context of news commenting. Our discussion in the final section includes use of the DSR Knowledge Contribution Framework to provide insights into the knowledge contribution of DUIL's design, and reflects upon how the IS artefact framework informs the DSR process.

2. Literature review

2.1. Design science research

Iivari (2015) distinguishes between two different DSR strategies dominating the information systems literature. One strategy, in the tradition of Markus, Majchrzak, and Gasser (2002) and Sein, Henfridsson, Purao, Rossi, and Lindgren (2011) which could be called client-centric or organisation-centric DSR, begins with an attempt to solve a specific client's existing problem and progresses towards a generalisation useful in other contexts. Iivari contrasts this with

a "proof of concept" approach in which a system is constructed to address a general problem and then instantiated as a test of the design theory. We have chosen to follow the "proof of concept" path in the context of a multidisciplinary international collaboration as advocated by Nunamaker, Twyman, Giboney, and Briggs (2017). This is similar to the approach taken by Twyman, Lowry, Burgoon, and Nunamaker (2014) who also address aspects of information leakage in their work. These more recent approaches are combined with the tradition of Hevner et al. (2004), Peffers, Tuunanen, Rothenberger, and Chatterjee (2007), and the formalisations provided in Gregor and Hevner (2013).

2.2. Information system artefacts

There has been considerable debate around the question of artefacts and their centrality to DSR. Much of DSR work describes information technology (IT) artefacts, which Orlikowski and Iacono (2001), Orlikowski and Iacono (2006) define as "bundles of material and cultural properties packaged in some socially-recognisable form such as hardware and/or software". However, not all agree that the "bundled" IT artefact should be the focal point of information systems research in general and DSR in particular. Lyytinen and King (2004) note that IT artefacts do not deliver value in their own right and must be viewed in the context of a system. Schwartz (2014) advocates the decomposition of IT artefacts into several distinct yet interconnected artefacts. Most recently, Lee et al. (2014) suggest a multi-artefact view when approaching DSR, arguing that the IT artefact is just one element within a broader Information systems artefact, which should be viewed as a construct incorporating information, social, and technology artefacts – and must be addressed as such in design science research.

We have chosen to augment our presentation by addressing the relatively new framework of IS artefacts articulated by Lee et al. (2015). As we will see, this approach is very well suited to the task and in using it to frame our work we believe we contribute to increased understanding and potential use of the framework.

Lee et al. (2015) define a framework, which is comprised of three major elements as follows:

1. A Social artefact – an artefact embodying relationships or interactions among multiple individuals;
2. An Information artefact – an instantiation of information produced by a human participant either directly (as their own creative output) or indirectly (through an individual's invocation of a software program or other automated information production process);
3. A Technology artefact – a human-created tool used to solve a human-defined or perceived problem.

All three interact within a broader systems framework achieving a results that is greater than the sum of its parts, comprising the IS artefact.

3. Methodological approach

This study follows the established approach to design science research as applied by Hevner et al. (2004), Hevner and Chatterjee (2010) and Peffers et al. (2007), resulting in a *proof of concept*. This approach consists of five stages: (1) problem identification, in which we define the scope of our study: detecting information leakers via commenting to on-line news; (2) going through the solution objectives, that is a set of expectations from the system and its design; (3) artefact design, that is the design of the system, followed by (4) demonstration with real data. The fifth stage is the system evaluation according to the objectives set in the second stage.

The DUIL system presented in this study is the result of integrating a series of independently developed and tested artefacts that were adjusted for the purpose of leaker detection. Each distinct artefact addresses a key aspect of the overall system solution. The first three artefacts, including two information artefacts and one social artefact, evolved from a study of the nature of leakage through UGC, specifically, comments. The latter two are technology artefacts which were developed in the context of uncovering hidden network relationships that reveals the potential scope of the leak. The combination and integration of these five artefacts led to the end-to-end leaker detection information system that we present.

In what follows we describe our study scope: disclosing *material non-public information* in the form of comments to news articles in a given context, which occurs as a direct result of social network structures.

4. Study scope

DUIL is designed to detect users who disclose material non-public information through User-Generated Content in social media (Agichtein et al., 2008). UGC in characterised by four main components, (1) a broad and unrestricted user base, (2) user or personally identifying information, (3) user social network, and (4) the contributed content.

The release of material non-public information can occur either maliciously or unintentionally, through discussions in On-line Social Networks (OSN). It is largely common in articles published by news pages, in which information is withheld, hidden, or censored, only to be uncovered by a commenter. The challenges presented by information leaked through commenting on news articles occur in many different contexts including the identity of military personnel, minor victims, minor perpetrators, rape victims, witnesses and others whose identity is considered information to be

withheld from the public, as documented in Cascavilla et al. (2015).

Our solution objectives centre around creating a holistic system to detect information leakers in social networks with an initial focus on Facebook (FB) commenters. While the social network stands at the centre of activities relevant to our work, the types and formats of information that inhabit OSN are myriad. For that reason, *modularity* becomes our first key objective enabling the use of different information artefacts to capture and analyse different OSN information sources, with an initial focus on news articles and comments.

Automation and *accuracy* are two closely linked objectives. The vast quantities of information to be processed and the accuracy of per-module results required to contain a security breach create this necessity.

Our third objective is *visualisation* of leakers' social networks. The detection and identification of information leakers is far from an exact science, and the ability to provide system operators with visualisations of leakers' social network – the extent of the potential leak, is crucial to enabling quick situation assessment and response.

5. System design

The system is designed to identify information leakers through commenting to on-line news, and provide the system user with network visualisation that presents the direct and indirect relationships between the leakers. The architecture of the system resembles a Swiss-Cheese model, a common model in the risk analysis and management field (Reason, 1990), in which at each level non-relevant information is filtered out, and the remaining data are passed to the next module. In our scenario, "relevant information" refers to comments that disclose knowledge and relationships that may lead to uncovering hidden information not released in the news article, and "non-relevant information" refers to comments that do not indicate such disclosure. We denote these comments as *leak enabling* and *non-leak enabling*, respectively. The output of the final module of DUIL is a network visualisation of the relationship between the most relevant commenters.

The system consists of five loosely coupled modules, corresponding to five artefacts, each responsible for a key part of leak detection and leaker identification:

Module 1 – Articles of Interest (AoI): An information artefact that identifies news articles in a given context. The current implementation detects articles in which personnel names are censored. This results in the creation of an *articles of interest* data-set for further analysis. It should be noted that in real-time systems, in the specific case in which an official censor is releasing news items, this phase can be replaced with expert input such that AoIs are flagged by the page administrator upon posting by the news agency.

Module 2 – Comments of Interest (CoI): This information artefact focuses on comment analysis. The goal of the module is the identification of news article comment discourse in which the commenters exhibit knowledge of sensitive information not released in the article, hence leak enabling. This module results in the creation of a *comments of interest* set, which is passed to the next module. Here too there are multiple approaches to generating this information and the contribution of this artefact to the information system is not in a specific technological approach to comment filtering, but the essential provision of information.

Module 3 – Users of Interest (UoI): This social artefact shifts from comments to commenters and their public user profile. The objective is to filter out users with close-to-zero probability of being leak enabling. The remainder set of commenters and their comments is passed to the next module. As we detail below, social network analysis of the users, their characteristics and interactions are a core part of this artefact's contribution, fitting well the description of a social artefact.

Module 4 – EgoNet: Using additional publicly available information, this technology artefact analyses relevant commenters' egocentric networks to enable the detection of implicit relationships between commenters, who are mutual friends. This network of relationships potentially holds additional information that is related to the hidden content, as well as the extent of the information leakage incident. Created specifically as a technological tool for this purpose moves EgoNet clearly into the category of technology artefact.

Module 5 – Viz: The Viz module is a technology artefact that presents a visualisation of leakers' merged social egocentric networks, received as output in module 4. The visualisation provides the system user with a tool to quickly identify the potential risk level of the leak. As a technology artefact it can be easily repurposed to visualise networks for other types of information systems, but its technological capabilities as a tool remain intact.

This information system design provides for future plans in which different social media sources are analysed to identify the sets of *articles of interest*, *comments of interest* and *users of interest* based on changing criteria and cyber-security needs – necessitating a swap of information and social artefacts.

A detailed description of each module is provided next.

5.1. Module 1: AoI

Given the large set of available FB news pages, the aim of the AoI module is to screen all news posts and generate a database of articles that evolve around a given context: Articles of Interest; along with the full set of comments that follows them, and the users (commenters) who posted them. The set of FB news pages, and screening

query via regular expressions, are defined by the system user. The output of module AoI is a database, composed of three main tables as follows:

Posts table: a set of Articles of Interest. For each AoI we collect the source (NewsPage); the date; and the content (text).

Comments table: a set of comments that follows the articles. For each comment we store the post identification; the commenter identification (the id of the user that wrote the comment); in reply to comment indicator, if the comment was part of a thread; its date; and its content (text).

Commenters table: a set of users who commented on the post. To reduce system complexity, we do not collect information on all users. We later collect information on-the-spot on Users of Interest (UoI), in modules 3 and 4.

5.2. Module 2: CoI

The Comments of Interest (CoI) Module classifies comments into comments of interest (leak enabling comments) or non-interest (non-leak enabling). The module contains two steps: (1) an initialisation step, in which multiple classifiers are trained and tested on previously annotated comments; and (2) a classifying step: where comments are classified into the two classes of using the best performing classifier from step (1).

In the initialisation step, a subset of the data is first split into training and evaluation sets. Domain experts are then asked to label the comments as "leak enabling" or "non leak enabling" classes. Note that manual classification is only done once, yet is essential for the construction of a meaningful and accurate classifier. A data-driven approach is followed to learn expert labelling. Here, multiple classifiers are constructed and trained. Each classifier utilises all or part of the comments' characteristics (e.g. popularity, order, length), and their textual properties such as processed text (Bermingham & Smeaton, 2011) and grammatical parts. Finally, the best classifier is selected based on its performance on the evaluation set. Performance is measured by the C-statistic measure (AKA, Area Under the Curve – AUC), that is, the capacity of the classifier in discriminating "leak enabling" comments from "non leak enabling" comments.

To enhance system performance and avoid missed *leak enabling* comments, we tune the classifier prediction-threshold to minimise False Negatives (type II error). In other words, a comment is classified as *non leak enabling* if the classifier has assigned it with a near-zero probability to be *leak enabling*, and *leak enabling* otherwise.

5.3. Module 3: UoI

Module UoI (Users of Interest) is designed to focus on commenters and their on-line user profile, to create a "leaker profile" of each participant. Users' profiles are collected on demand for the set of comments and commenters received from module UoI. On each user, the following profile information is collected: network size – number of friends and number of followers, and privacy setting (whether the user's profile is kept private of public). Potentially additional input information can be collected on each user to measure her FB engagement and on-line activity.

Similar to CoI, module UoI has two steps: Initialisation and classification. In the initialisation step, a best data-driven classifier is selected and trained on the previously labelled comments to estimate the probability of a commenter to be *leak enabling*. Here again, the classifier threshold is set to minimise False Negatives. Those identified as UoI are then passed to the module EgoNet.

5.4. Module 4: EgoNet

To rebuild the egocentric network of a UoI we use SocialSpy (Burattin et al., 2015). SocialSpy was developed to retrieve the lists of friends of each UoI, given her publicly available information, such as public friends, pictures, group memberships, and page likes.

The tool implements four strategies, each using a different type of information from the OSN to rebuild the friends list of a given UoI.

The first three strategies are based on liked pages. Statistics show that Facebook *Like* and *Share* buttons are used over 22 billion times a day, on approximately 7.5 million Facebook pages (He, 2013). Furthermore, like and share information is usually available even for users with high privacy settings.

The fourth strategy exploits likes and comments from the picture of a given user. Although pictures for users are only partially available when a user has high privacy settings, a recent survey shows that many users are unaware of Facebook's privacy options (Consumer Reports Magazine 2012), or too lazy or inexperienced to properly modify them (Madejski, Johnson, & Bellovin, 2012). Given that, we expect that this strategy will highlight *strong* relationships with the UoI and both public and private other users (Jones, Settle, Bond, Fariss, Marlow, & Fowler, 2013).

- *Strategy 1* exploits *like* pages of a given profile. Based on the theory of homophily, we can assert that other users who like the same page(s) share

common interests with the user of interest, and hence have higher probability of having friendship relationships. Strategy 1 operates as follows. Using the public Facebook pages of the each UoI, Strategy 1 retrieves the list of liked pages left public by the UoI profile. The strategy then retrieves the list of these pages' fans (users who liked these pages). Next, for each fan Strategy 1 queries (via the *Mutual Content Page (MCP)* Constine, 2010) whether he is a friend of the User of Interest. The output of this strategy is a list of friends tuples of the format $\{UoI_i, friend_j\}$.

- *Strategy 2* is similar to the first strategy, yet differs in the way that probabilities are set: users who share *like* pages with small sets of fans receive higher probability of sharing friendships with the user of interest. The reasoning here is that these pages are likely to target a narrower interest, and therefore, the homophily value of the users who like it is higher.
- *Strategy 3* is the opposite of Strategy 2. That is, the probability of sharing mutual friends is higher for users who like pages with more fans. The idea behind this strategy is that, fetching *like* pages from max-to-min number of fans, results in a bigger user-pool faster, even when crawling for a single page, in which mutual friendships with the user of interest can be found.
- *Strategy 4* exploits public pictures of the given user. The tool then retrieves the list of users who like public pictures of a given UoI, or commented on them. Once the tool obtains the list of users, it checks the friendship between them and the target ID using the MCP.

Among these four strategies, Strategy 4 has proven to be the fastest and with the highest average (37.12%) of friends found. Respectively, Strategy 1 with an average of 17.5% of retrieved friends, Strategy 2 with an average of 17.4% of retrieved friends, and Strategy 3 with an average of 20.8% of retrieved friends (Burattin et al., 2015). Based on these results we follow Strategy 4 in our experiments. A detailed description of our implementation of this strategy is given in Appendix 1.

5.5. Module 5: Viz

Module Viz groups and visualises the egocentric networks obtained in Module EgoNet (visualisation is done via Gephi https://gephi.github.io/). The main goals of this module are to (1) find overlaps between leakers networks, which may provide additional information on the profile of the leakers and the nature of the leak, (2) examine the extent and the potential diffusion of the information leakage, thus the risk associated with the leak, and (3) provide the system user with a network visualisation of these findings.

5.6. Summary of DUIL's design

DUIL is designed to detect information leakers via commenting to on-line news articles with a current focus on FB news pages. To maximise efficiency, the system follows a modular design, in which each module is standalone and can be removed or replaced by a context-specific module designed for different purposes. Our understanding of these modules, their main roles, and replacability within the overall IS artefact is enhanced through their respective characterisations as information, social, and technology artefacts.

Conceptually, the system operates through three phases: First the *search space* is set – the list of articles and comments that may contain information leaks (Module 1). Second, the complexity of the analysis is reduced by filtering out noise and thus, decreasing the size of the search space (Modules 2 & 3). Lastly, the user is presented with a basis for detection and risk assessments of the information leak (Modules 4 & 5). Figure 1 summarise the design of DUIL.

6. Demonstration

In this section, pursuant to design science methodology, we present a series of case studies demonstrating the system. We describe the collection of our test data, and show how we used DUIL to obtain identities of leakers and their social networks.

6.1. Experimental design

We design a full experiment based on real data collected from FB. Use of the system is illustrated through the analysis of articles crawled from FB news pages, in which part of the information is kept private. Specifically, we are interested in articles in which the identity of personnel is withheld. Identity-censorship is one straightforward type of "material nonpublic information". Commenters *in the know*, can share their private information through comments. The experiment utilises the three phases of DUIL.

Module AoI is first used to collect a set of identity-censored news articles, followed by a thread of commenters and comments that can potentially exhibit censorship breaches. Then, we initialise modules CoI and UoI on a subset of the articles collected. The initialisation phase provides us with two classifiers that can be used in real-time. We evaluate the classifiers and report their performance. Finally, we ran modules EgoNet and Viz on 14 selected case studies. In each case we depict a subset of leak enabling commenters (UoI) and construct the network between them. The case studies illustrate the additional information that the network provides, which includes capturing the leakers and their social networks, and often though not always, includes identifying the censored personnel.

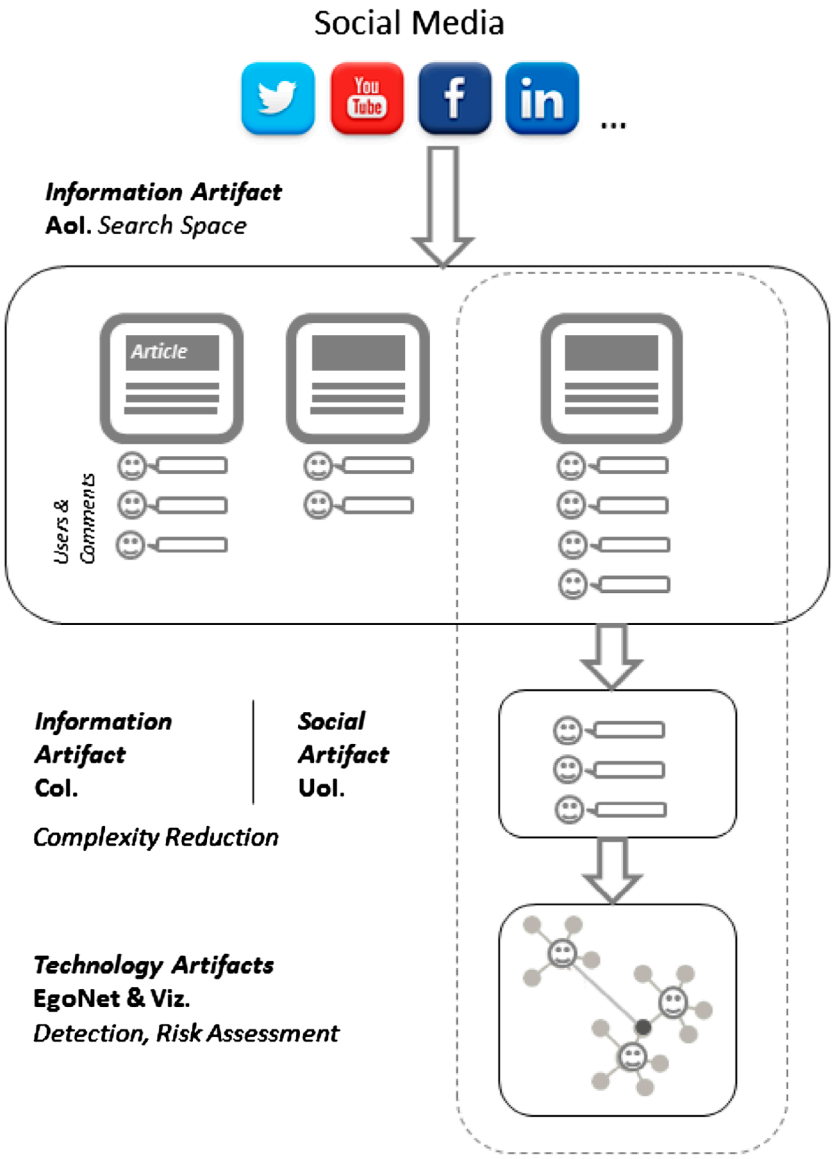

Figure 1. DUIL modules and architecture.

6.2. Description of OSN data collected

We focus on information leakers through comments on Israeli, military-related, news articles published in FB, in which a military personnel name is censored. Censorship in our study is the replacement of a name with a supposedly non-identifying initial (e.g. *Corporal S.*). Information leakage is detected in the comments published by private users, which leads to the identification of the censored person.

An example of information leakage of interest is presented in Figure 2. The headline of the news article, as it appears on the FB page of a network news service, is: "Karakal combat soldier *Corporal S.* who eliminated a terrorist in the course of an incident on the Egyptian border awarded an Honourable citation: Everyone who was with us deserves it, and of course Nathaniel who fell in the battle". Military information policy dictates that the identity of officers in key positions, or involved in key operations, must not be released to the public. The motivation for this policy is a desire to protect the officer and his or her family from being identified and potentially targeted by hostile persons or forces. In this case, the obfuscated term "Corporal S." is identified as the censored element of the news item. A particularly verbose comment associated with the news item states: "The brave combatant is the daughter of a good friend of mine. Do you know where the combatant comes from? From Elad of course!". Using DUIL, the readily available identity of that commenter and his FB Friends can lead us to the identity Corporal S. We therefore, treat this comment and other similar comments as "leak enabling" comments. A detailed description of the case is discussed in our previous work (Cascavilla et al., 2015).

6.3. Experimental results

We report the results of each system module throughout the course of our experiment.

Figure 2. Example of data leakage (translated from Hebrew).

6.3.1. Module 1: AoI

AoI screens through a large set of FB news pages, and collects the set of relevant articles. Screening is done via regular expressions defined by the system user.

In our experimental study, we are interested in the set of censored military-related articles, in which the name of a military personnel is censored. For that, we use a list of regular expression expressions that search for a military rank, followed by an initial letter. For example, the expression "`@lieutenant \c\.@`" corresponds to the military rank lieutenant followed by his first initial (e.g. lieutenant D.)

The data-set collected contains 48 articles with censored personnel names, with an aggregate total of 3,538 comments.

6.3.2. Module 2 (CoI) and Module 3 (UoI)

In this section, we present the classifiers constructed for module CoI and UoI, and discuss their performance. We note that the classifiers we chose here are tuned for the data at hand, and might not be optimised to other data-sets. The process of choosing classifiers, however, is data-independent. In the following we repeat the steps of the general process, followed by the specific tuning for our data.

We begin by splitting the set of articles into training, validation, and holdout sets. Independent reviewers are then asked to label comments in the training set and the evaluation set. This labelling is required for system initialisation and evaluation assessment. Ideally, the training and validation sets consist of small samples from the data, as they are labelled manually, yet are big enough to achieve accurate performance.

In our data, for the purpose of performance evaluation, we split the data into training and validation only, each consisting of 50% of the data. Holdout sample may be defined as comment to all future articles, that are not collected in the current time frame. We then asked four reviewers to classify each comment as either *leak enabling* or *non-leak enabling* by reflecting on the following question for each comment: "Based on this comment, do you believe that the commenter knows the identity of the censored person?". We then followed a Delphi procedure to achieve agreement among the reviewers. We further asked them to identify the elements of the comment that caused them to reach their conclusion.

Out of the 3538 comments collected on 48 articles by the AoI module, the reviewers labelled 149 (4.21%) as *leak enabling* comments. Interestingly, these comments are spread out through 75% (36) of the articles. A summary of the comments classification is presented in Table 1.

After the data are labelled, different classification algorithms are trained on the data. Classification algorithms include but not limited to logistic regression, SVM, classification tree, and classification forest. The methods are evaluated based on their ability in capturing the relationship between *leak enabling* comments and features of the comments, using the C-statistic measure (AKA, Area Under the Curve – AUC). The features used in our current system are divided into three families, each used by the classifiers solely and in combination with other families. The first family contains general quantitative characteristics of the com-

Table 1. Frequency (*f*) and proportion (%) of comments' classes.

Comment type	*f*	%
Leakage comments	149	4.21
Non–Leakage comments	3389	95.79
Total comments	3538	100

ments, such as its order, its length, its popularity (as obtained from FB), and the presence of semiotics in the comment (smiles, winks, etc.). The second family of features is the textual features extracted via the comment level sentiment analysis in Bermingham and Smeaton (2011) on both processed text (lemmatised text, after the removal of stop words), and grammatical parts such as lexical part-of-speech, gender, tense, number (singular, plural), and person (first, second, third). The last family contains the main elements mentioned by the reviewers and are data-specific. In our context, this family include the repetition of censored information within the comment, expression of affection, mention of a location not mentioned in the article, or personal experience related to the content of the article.

In our data-set, out of the models we examined, the best performance on the evaluation set, using cross validation, was that given by a logistic regression, using features from the first and the third families of features. The output of the model and the features selected are given in Table 1. The *C*-statistic of the model, that is, its capacity in discriminating *leak enabling* comments from *non-leak enabling* comments, is 81%. The performance on the model is summarised in the Receiver-Operating characteristic (ROC) Curves in Figure 3. The ROC Curve depicts the trade-off between False Positive Rate and False Negative Rate for different classification thresholds: the probability cut-off for classifying comments as leak enabling.

To minimise missed leak-enabling comments, our system next selects the threshold that minimises False Negatives (type II error), assuring that comments are only classified as *non-leak enabling* if the classifier has assigned it with a near-zero probability, and *leak enabling* else-wise.

For our data, as observed by the ROC curve, this threshold equals 0.75. Under this threshold approximately 26% of the comments can be ruled out as non *leak enabling*, without significantly increasing the model's False Negatives (less than 5% error). This threshold is marked with dash line in Figure 3.

A similar process is carried for the UoI module. For our data, the best classifier achieved for this module is the logistic regression given in Table 3. The *C*-statistic of the model is 60%. Given the fairly low C-statistic, the low model coefficients and their (in)significance, we conclude that UoI module in our case study does not provide additional information on top of the CoI module. Reasons for this can be attributed to data size

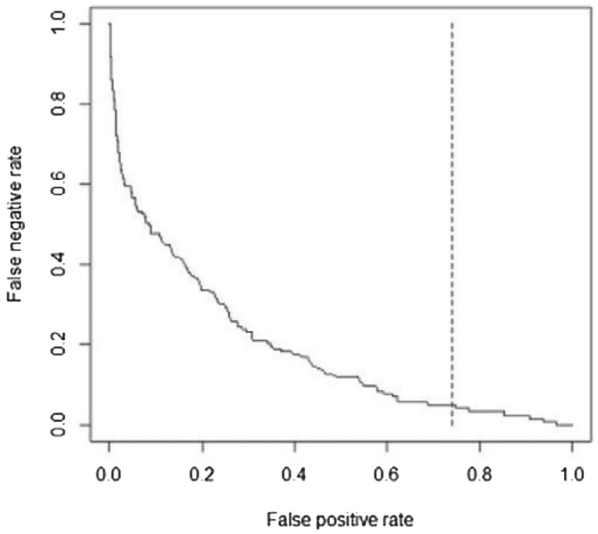

Figure 3. ROC curves of the logistic models.

and information available for each profile. Note that this result only holds for this specific set of articles. However, UoI might be useful for other data-sets or detection purposes.

6.3.3. Module 4 (EgoNet) and Module 5 (Viz)

We run modules EgoNet and Viz on 14 articles randomly selected from our output of AoI. For each article EgoNet and Viz are run on all users screened through module UoI. EgoNet and Viz (re)build and visualise the friends ego networks of the list of UoI, including mutual friends.

Figure 4 illustrate two of the more interesting cases. The black nodes in the figure are the UoI. UoI are surrounded by their friends, some of which are common between them.

In the first case (Figure 4(a)) we can see that four UoI are not FB-friends: there is no direct connection between them. They share a single friend, which we later found to be the censored person from the article. In the second example, plotted in Figure 4(b), it is observed that all five UoI are strongly connected, and share multiple friends.

Due to the nature of the articles and the data chosen, that is, *identity*- censored article, the ego networks may provide us with additional useful information: the *identity* of the censored personnel. This information will become immediately available when the leakers are FB-friends with this person. In the two examples we present, this is the case. In each panel of Figure 4, a single white circular node was manually confirmed to

Table 2. CoI logistic model.

Predictor	Estimate	*p*-value
(Intercept)	−3.77	∼0
Comment Popularity	0.00	∼0
Repetition of censored information	2.28	∼0
Mentions of location	0.84	0.15
Mentions of personal experiences	1.41	∼0
Semiotics (smiles, winks, etc.)	0.66	∼0
Expressions of affection	2.40	∼0

Table 3. UoI logistic model.

Predictor	Estimate	*p*-value
(Intercept)	−0.05	∼0
Network size	0.0007	0.008
Followers	0.006	0.42
Privacy setting	−0.51	0.42

be the censored personnel. Confirming the censored person identity was done thanks to finding the blurred picture from the article, unblurred in the users' profile.

Out of the 14 experiments, we were able to identify the censored personnel in four cases. In each of the other 10 cases, a network was constructed, and mutual friends of UoI were found in eight of the cases. However, we could not verify nor refute their link to the censored personnel.

7. Discussion and conclusion

DUIL is a new type of system that can be considered as a member of the superclass of Social OSINT systems, a form of cyber-threat intelligence system, which are growing in importance (Bowman, 2016; Casanovas, 2017; Jasper, 2017; Nunamaker et al., 2017).

The combination of comment and user mining with risk analysis, and of social network visualisation for risk signalling, is the main system-based contribution. This produces synergies in terms of new analytical capabilities. Such analytical requirements are a moving target which makes the decisions to divide components all the more important.

To provide a more granular discussion we frame DUIL as an information systems artefact as presented by Lee et al. (2015). There are few salient examples of related work that has taken a design science approach focused on the IS artefact rather than the IT artefact. Huhtämki, Russell, and Sill (2016) perform ecosystem analytics by integrating a technology artefact with other artefacts to perform visual network analytics. Both Spagnoletti, Resca, and Sæbø (2015) and Wakefield and Wakefield (2016) tackle social media technologies as a three-dimensional information systems artefact comprised of technical, informational, and social sub-artefacts.

Following Lee et al. (2015), we divide this part of our discussion into three, covering:

1. the specific IT artefacts developed,
2. the information artefacts detected and collected, and
3. the social artefact that influences, in our case, both the IT and information artefacts.

All three taken together comprise the IS artefact in which IT artefacts come together with other artefacts that are not strictly IT so that "they ultimately serve to solve a problem or achieve a goal for individuals, groups, organizations, societies, or other social units" (p. 6).

7.1. AoI and CoI are information artefacts

The three data tables collected by AoI together comprise an information artefact which serves a central purpose in the overall IS artefact. This corresponds to the definition proposed by Lee et al. (2015) wherein the artefact (a) consists of an instantiation of information and (b) it is generated by a human initiating use of a computer program in this case a FB crawler and filtering mechanism. Similarly, CoI is an information artefact. Corresponding to the definition wherein the artefact (a) consists of an instantiation of information and (b) it is generated by direct human action – in this case expert classification, combined with initiating use of a computer program in this case a classifier mechanism.

Such instantiations of information will recur repeatedly throughout the life and use of the overall system. Furthermore, the design of the overall system is such at that the information artefacts we use in the current study can be replaced, without loss of generality, by information collected from other media sources and processed by different classifiers, thus strengthening the appropriateness of the information artefact definition within the IS artefact framework. In other words, it would be misleading to consider the AoI or UoI as technology artefact, as it is not a technology making

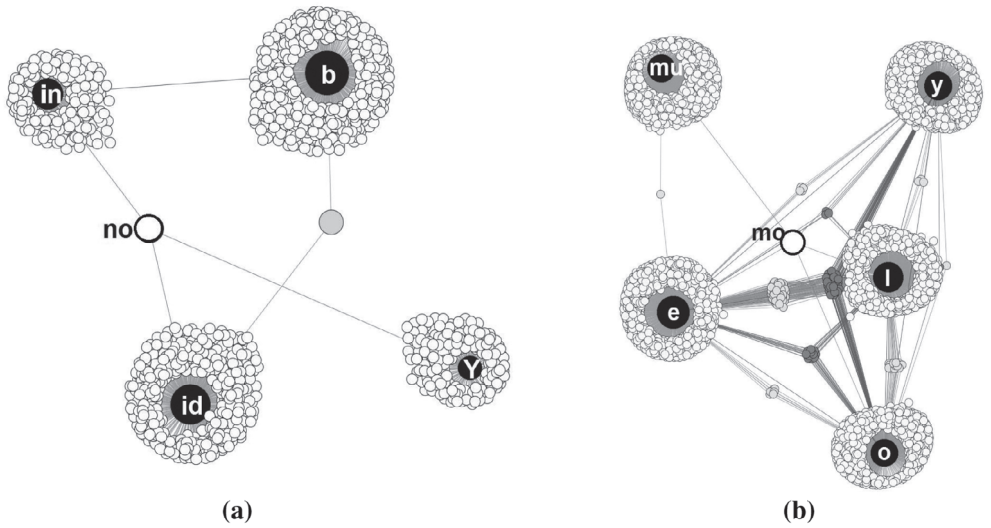

Figure 4. Networks generated by module Viz.

the decisive contribution but rather the information gathered at this stage.

7.2. UoI is a social artefact

The UoI module meets the framework's definition of a social artefact in that it reflects relationships or interactions between or among individuals, involving the social and not just the individual.

This characterisation is seen in the social-behavioral data collected and processed by this artefact which includes number of friends and FB user activity.

7.3. EgoNet and Viz are technology artefacts

In EgoNet and Viz, we have instantiations of pure IT artefacts as defined by the IS artefact framework. Both are human-created tools used to solve a problem or achieve a goal. In the case of the former the goal is to (re)build a previously unknown egocentric network, and in the case of the latter provide visualisation.

Taken together, and given the interactions and interdependencies between the two information artefacts, one social artefact, and two IT artefacts, we have a prototypical example of an IS artefact as per the Lee et al. (2015) framework.

Other than defining three of many possible artefact types, Lee et al. (2015) provide little guidance on determining artefact categorisation and we note that such categorisation may not be obvious. For example, one might argue that a given artefact such as our AoI is a technology artefact rather than an information artefact when observing that it is a software technology enabling automated identification of the relevant articles. However, in making this assessment we look to the *main contribution of the artefact to the design* which, in the case of AoI is not in the *technological way* in which the information was achieved, but in the essence of *having collected the information* itself.

Based on our experience we have found that the concept of "artifact contribution" can be essential in determining whether an artefact should be viewed as social, information, or technology. We emphasise that the distinction is not always clear cut and requires careful consideration. Table 4 summarises.

The Lee et al approach to IS artefacts is not without controversy. Iivari (2017) points out, that Lee et al. "simply interpret IT artifacts as purely technical ones" which he considers a potential shortcoming. We have found, however, that this narrow definition when used alongside the complementary social and information artefacts, enables a rich and precise descriptive and analytical discourse. Among other critiques, Iivari (2017) questions how Lee et al. (2015) might associate their work with extant approaches to design science as this is not explicitly addressed in their work. We believe that our work has provided an initial answer to this question, and we have found and demonstrated that the Lee at al. framework can coexist quite nicely with traditional design science. Finally, Iivari (2017) questions how design science research would make direct contributions to the non-IT artefacts in the Lee et al. framework. We have found that the characterisation of certain artefacts as social or information eases their placement and analysis within an overall IS design project. Rather than opening an unmanageable distance between the artefact and DSR, it forces us to think in terms of the actual contribution of each artefact to the IS rather than limiting our assessment to technological contribution.

7.4. Meeting the objectives and assessing contribution

Our system design goals specified three sets of objectives: modularity; automation, and accuracy; visualisation of leakers' social networks.

Table 4. Determining artefact type.

Artefact	Type	Determining contribution
Articles of Interest & Comments of Interest	Information	Q. Does the fact that a technology is used to gather the information not make these technology artefacts? A. No, as the technological implementation is not what contributes to the system design.,There are multiple, perhaps equally valid, technological approaches. The contribution of the artefact lies in the information, not the technology. Different information sources plugged into this artefact might require alternative technologies
Users of Interest	Social	Q. The user data is provided by FB which is a technology implementing a social network, so perhaps this is a technology artefact? A. No, as the technological implementation is not what contributes to the system design. There are multiple possible social networks that might be analysed as part of the information system. The contribution of this artefact to the overall information system lies in the social structures provided, not in the technology that supports it. Different social networks plugged into this artefact might require alternative technologies, and provide different social insights
EgoNet	Technology	Q. This artefact is meant to provide information about network structures, so perhaps this is an information artefact? A. No, as the contribution of this artefact to the system was to enable the creation of the required network structures where no such capability previously existed
Viz	Technology	Q. This artefact presents social structures in a visual manner so perhaps it is a social artefact not a technology artefact? A. No, as the primary contribution of this artefact is to determine a visually effective way to present network data. This contribution is technological in can be repurposed for use in different network domains

The *modularity* goal is achieved by the choice of architecture. Most important in this respect is the ability to replace *Module 1*: AoI (Articles of Interest) and *Module 2*: CoI (Comments of Interest) with alternatives that can process other forms of social media. Separating out *Module 3*: UoI (Users of Interest) further extends the desired flexibility for different OSN structures.

Automation and per-module *accuracy* goals have been partially achieved at the *proof-of-concept* level as demonstrated by the experiments. *Accuracy* is measured via the AUC measure for the statistical module, yet cannot be measured for the Viz module, as we later discuss in the limitation section. Further experimentation will be required in these areas.

The *visualisation* of leakers' social networks goal has been achieved as illustrated by the experiments and accompanying graphs.

Beyond meeting the objects set at the outset of the system design process, we briefly address DSR knowledge contribution as discussed in Gregor and Hevner (2013). Their framework assesses contribution on the axes of *x:problem maturity* and *y:solution maturity*. Scaling regions of *high* and *low* for each axis gives the four quadrants of:

1. Routine design (high, high) applying known solutions to known problems, resulting in no major knowledge contribution;
2. Exaptation (low, high) extending known solutions to new problems, resulting in research and knowledge contributions;
3. Improvement (high, low) developing new solutions to known problems, resulting in research and knowledge contributions; and
4. Invention (low, low) inventing new solutions for new problems, resulting in research and knowledge contributions.

Identifying leakers of material non-public information is not a new problem, as we see from the analogy to insider trading and organisational information leakage described in our introduction. However, the shift of this problem from inside the organisation to the broad context of social media has introduced significant new complexities to the problem, changing important problem characteristics particularly with regards to scale and scope. Therefore, on the problem maturity axis DUIL would score a mid-range value rather than a clear *high* expected of an improvement and clear *low* expected of an invention. On the solution maturity axis the DSR contribution of DUIL is clearly in the *low* range owing to the novelty of the approach and previously un-attempted combination of artefacts in a complex information system artefact. Therefore, based on the criteria set out by Gregor and Hevner (2013) our DSR contribution lies in the upper right of the *improvement* quadrant extending slightly into the *invention* quadrant. We present this graphically in Figure 5 modelled after the Gregor and Hevner framework.

In testing their framework Gregor and Hevner (2013) presented a table of 13 design science articles classified by knowledge contribution type. They document the classification of a single contribution per article, into a distinct quadrant of the grid. We suggest that when following the multi-artefact approach to IS artefacts, research value can be revealed by mapping the contribution of each component artefact when presenting an information system. We therefore, have enhanced the original DSR contribution framework diagram to show

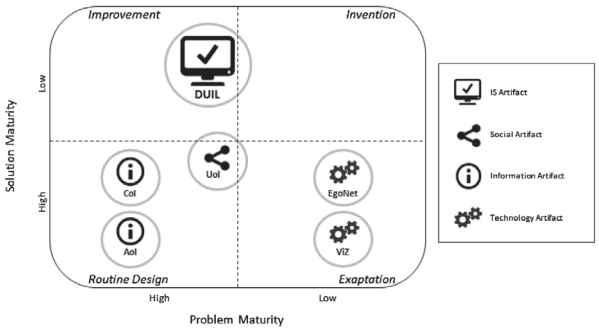

Figure 5. DSR knowledge contribution framework with multiple artefacts (after Gregor & Hevner, 2013).

the positioning of the sub-artefacts that comprise the DUIL information system artefact. Each artefact type is represented by a different symbol.

The two information artefacts, AoI and CoI appear in the *routine design* quadrant. These artefacts use a combination of human intervention and text processing that are well-known in the art and applied to many text classification problems. Therefore, they rank high on both axes of problem maturity and solution maturity. This means that taken on their own these artefacts provide no research and knowledge contribution.

The two technology artefacts, EgoNet and Viz, appear in the *exaptation* quadrant. Here, we have examples of two technologies that have been used successfully in other domains, being re-purposed to solve a new problem. Therefore, they rank high on the solution maturity axis and low on the problem maturity axis.

The social artefact, UoI, appears straddling the *routine design* and *exaptation* quadrants, with a slight advance upward into the *improvement* and *invention* quadrants. This indicates a solution design that has elements of routine use (tracing the known social networks of commenters, for example), elements of exaptation (finding new uses of the social structures for leak containment), while being applied to a newly complex instantiation of an existing problem.

Following the DSR knowledge contribution guidelines and enhancing the graphical presentation for multiple artefacts, helps to effectively express the knowledge contributions of DUIL.

7.5. Limitations

Though DUIL reaches its design and security goals, there are some unavoidable limitations and shortcomings. The main limitation of DUIL is reliance on human intervention for initialisation, thus can be considered a semi-automated information system. Specifically, four actions are done manually:

1. In module AoI, system user defines the context, formulated as a list of regular expressions. Alternatively, page administrator can flag articles that need to be monitored for potential leakers.
2. Initialisation of CoI involves experts annotation of a suf ficient training set. Expert annotation is customary in classifying UGC.
3. In modules CoI and UoI, system users can, but are not obligated to, control the list of data features and classifiers, according to their domain knowledge and experience.
4. Modules EgoNet and Viz provide an overview of the potential leak and involved leakers, yet it is up to the user to carry on further analysis of the threat.

The second limitation of the system relates to the expert annotation, which is needed since no *ground truth* is available. In practice a comment can be misclassified by reviewers as *leak enabling*, while in fact it is not.

The third limitation stems from the use of statistical models in a Swiss-Cheese fashion. Modules AoI (if pre-post labelling was chosen), CoI, and UoI may introduce some statistical errors (false positives and false negatives) into the system. These error are carried on through the modular design, resulting in potentially increased error rate.

7.6. Future research

Given the great popularity of OSN, they have become one of the most common means to share and discuss information, including news articles. We find that comments and commenters' OSN leak information, originally withheld in news articles. To underline the importance of this issue, in this paper we present DUIL, a system we designed and implemented to analyse news comments in order to detect information leakers and assess the risk associated with such leaks. We ran real data experiment on military news articles, the results underlines the effectiveness of our approach in finding leakers among UoI and (re)building their social network. Moreover, our tool was able to de-censor the information in some of the articles.

This study, aside from the primary contribution of knowledge created by our specific IS artefact, has secondary theoretical contributions. It provides a crisp illustration of how the IS artefact framework espoused by Lee et al. (2015) can effectively aid in the description and analysis of a complex system, and has extended the use of Gregor and Hevner's (2013) DSR knowledge contribution framework.

Detection of leaked information, identification of a leaker, and quantification of the impact of a leak are three distinct yet interrelated challenges faced by organisations in the age of OSN. DUIL was designed and implemented with the holistic view of addressing all three of these challenges. Through DUIL and the systems that follow it, organisations will be able to assess

and address their exposure to the risks of information leakage. With the inherent structure of OSNs turning more users into bearers of material non-public information, addressing these challenges will continue to grow in importance.

Funding

Mauro Conti is supported by a Marie Curie Fellowship funded by the European Commission under the agreement No. PCIG11-GA-2012-321980. This work is also partially supported by the TENACE PRIN Project 20103P34XC funded by the Italian MIUR, and by the Project "Tackling Mobile Malware with Innovative Machine Learning Techniques" funded by the University of Padua. This research was partially funded by Israel Ministry of Science and Technology research grant 3-9770 Data Leakage in Social Networks: Detection and Prevention.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Giuseppe Cascavilla http://orcid.org/0000-0002-0724-3772
Mauro Conti http://orcid.org/0000-0002-3612-1934
David G. Schwartz http://orcid.org/0000-0002-2125-2069
Inbal Yahav http://orcid.org/0000-0002-1513-017X

References

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 183–194). Palo Alto, CA: ACM.

Bermingham, A., & Smeaton, A. (2011). On using twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)* (pp. 2–10). Chiang Mai: Asian Federation of Natural Language Processing.

Bishop, M., & Gates, C. (2008). Defining the insider threat. In S. Frederick (Ed.), *Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research: Developing Strategies to Meet the Cyber Security and Information Intelligence Challenges Ahead* (Vol. 15, pp. 1–3). CSIIRW'08. ACM.

Bowman, E. K. (2016). Content-based multimedia analytics: Rethinking the speed and accuracy of information retrieval for threat detection. In *STO Meeting. NATO Science and Technology Organization*(Vol. 18, pp. 1–10). Virginia: NATO. (Visited on 05/22/2017)

Burattin, A., Cascavilla, G., & Conti, M. (2015). SocialSpy: Browsing (supposedly) hidden information in online social networks. In *Lecture Notes in Computer Science* (pp. 83–99). Trento: Springer International Publishing.

Casanovas, P. (2017). Cyber warfare and organised crime. A regulatory model and meta-model for open source intelligence (OSINT). *Ethics and Policies for Cyber Operations* (pp. 139–167). Cham: Springer.

Cascavilla, G., Conti, M., Schwartz, D. G., & Yahav, I. (2015, August). Revealing censored information through comments and commenters in online social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on* (pp. 675–680). IEEE.

Claudio, C.-R. (2009). Modelling deterrence in cyberia. *NATO Science for Peace and Security Series - E: Human and Societal Dynamics 59. Modelling Cyber Security: Approaches, Methodology, Strategies* (pp. 125–131).

Constine, J. (2010). *Facebook announces friendship pages that show friends mutual content*. Retrieved November 11, 2016, from http://www.insidefacebook.com/2010/10/28/friendship-pages-mutualcontent

Facebook Inc. (2015). *Facebook reports third quarter 2015 results*. Retrieved November 11, 2016, from http://investor.fb.com/releasedetail.cfm?ReleaseID=940609s

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–356.

He, R. C. (2013). *Facebook developers page - Introducing new like and share buttons*. Retrieved November 24, 2016, from https://developers.facebook.com/blog/post/2013/11/06/introducingnew-like-and-share-buttons

Hevner, A., & Chatterjee, S. (2010). *Design research in information systems: Theory and practice* (1st ed). Springer US.

Hevner, A. R., March, S. T., & Park, J. (2004). Design science in information systems research. *MisQuarterly, 28*(1), 75–105.

Huddart, S. J., & Ke, B. (2007). Information asymmetry and cross-sectional variation in insider trading*. *Contemporary Accounting Research, 24*(1), 195–232.

Huhtämki, J., Russell, M.G., & Sill, K. (2016). Processing data for visual network analysis. In B.Elliot & C. Sacha*Visual Analytics for Management: Translational Science and Applications in Practice* (Chap. 5); Oxford: Routledge.

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems, 24*(1), 107–115.

Iivari, J. (2017). Information system artefact or information system application: That is the question. *Information Systems Journal, 276*), 753–774. https://doi.org/10.1111/isj.12121

Im, G. P., & Baskerville, R. L. (2005). A longitudinal study of information system threat categories: the enduring problem of human error. *ACM SIGMIS Database, 36*(4), 68–79.

Jasper, S. E. (2017). US cyber threat intelligence sharing frameworks. *International Journal of Intelligence and Counter Intelligence, 30*(1), 53–65.

Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., & Fowler, J. H. (2013). Inferring tie strength from online directed behavior. *PLOS ONE, 8*(1), 1–6.

Kandias, M., Stavrou, V., Bozovic, N., & Gritzalis, D. (2013). Proactive insider threat detection through social media: The YouTube case. In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society. WPES'13* (pp. 261–266). Berlin: ACM.

Karsch, M. (1984). The insider trading sanctions act: incorporating a market information definition. *Journal of Comparative Business and Capital Market Law, 6*(3), 283–305.

Lee, A. S., Thomas, M., & Baskerville, R. L. (2015). Going back to basics in design science: from the information technology artifact to the information systems artifact. *Information Systems Journal, 25*(1), 5–21.

Lyytinen, K., & King, J. L. (2004). Nothing at the center?: Academic legitimacy in the information systems field. *Journal of the Association for Information Systems, 5*(6), 220–246.

Madejski, M. M., Johnson, M., & Bellovin, S. M. (2012). A study of privacy settings errors in an online social network. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops* (pp. 340–345). IEEE.

Magazine, C. R. (2012). *Facebook & your privacy*. Retrieved November 11, 2016, from http://www.consumerreports.org/cro/magazine/2012/06/facebook-your-privacy

Markus, L. M., Majchrzak, A., & Gasser, I. (2002). A design theory for systems that support emergent knowledge processes. *MIS Quarterly, 26*(3), 179–212.

Nunamaker, J. F., Twyman, N. W., Giboney, J. S., & Briggs, R. O. (2017). Creating high-value real-world impact through systematic programs of research. *MIS Quarterly, 41*, 2.

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the IT in IT research a call to theorizing the IT artifact. *Information Systems Research, 12*(2), 121–134.

Orlikowski, W. J., & Iacono, C. S. (2006). The artifact redux: Further reflections on the IT in IT research. In J. L. King, & K. Lyytinen (Eds.), *Information Systems: The State of the Field* (Chap. 12, pp. 287–292). Hoboken, NJ: John Wiley & Sons.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45–77.

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 327*(1241), 475–484.

Schwartz, D. G. (2014). Research commentary The disciplines of information: Lessons from the history of the discipline of medicine. *Information Systems Research, 25*(2), 205–221.

Schwartz, D. G., Yahav, I., & Silverman, G. (2017). News censorship in online social networks: A study of circumvention in the commentsphere. *Journal of the Association for Information Science and Technology, 68*(3), 569–582.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly, 35*(1), 37–56.

Spagnoletti, P., Resca, A., & Sæbø, Ø. (2015). Design for social media engagement: Insights from elderly care assistance. *The Journal of Strategic Information Systems, 24*(2), 128–145.

Strudler, A., & Orts, E. (1999). Moral principle in the law of insider trading. *Texas Law Review*, *78*, 375–438.

Twyman, N. W., Lowry, P. B., Burgoon, J. K., & Nunamaker, J. F. (2014). Autonomous scientifically controlled screening systems for detecting information purposely concealed by individuals. *Journal of Management Information Systems, 31*(3), 106–137.

Wakefield, R., & Wakefield, K. (2016). Social media network behavior: A study of user passion and affect. *The Journal of Strategic Information Systems, 25*(2), 140–156.

Warkentin, M., & Willison, R. (2009). Behavioral and policy issues in information systems security: The insider threat. *European Journal of Information Systems, 18*(2), 101–105.

Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T., & Nunamaker, J. F. (2004). A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems, 20*(4), 139–165.

Appendix 1. Implementation of EgoNet Strategy 4

1. EgoNet receives as input the list of UoI.
2. For each user in the list, EgoNet obtains the set of publicly available albums, and public pictures within these albums (line 1 of Algorithm 1).
3. For each picture the tool collects the identities of users who commented or liked the picture (line 4; the users who left a comment or pressed the Like button are defined, respectively, in the algorithm with U_i^c and U_i^l.)
4. Using MCP (Constine, 2010), EgoNet
   - checks whether the UoI and a given user are friends *"Facebook friends since **[date]**"* (line 7);
   - if yes, retrieves the list of common friends (line 8).
1. Lastly, EgoNet, returns the list of the *"Friends Found"* (line 9).

Algorithm 1 implements the steps taken for each user in the UoI set.

Algorithm 1: Algorithm of Strategy 4

Data: UoI *uoi*
Result: Set of friends of *uoi*

```
1 I ← set of public images of uoi
2 CandidateFriends ← ∅
3 foreach i ∈ I do
      /* Add candidate friends set all users that liked or commented the image */
4     CandidateFriends ← CandidateFriends ∪ U_i^l ∪ U_i^c
5 FriendsFound ← ∅
6 foreach c ∈ CandidateFriends do
      /* Check friendship with Mutual Content Page */
7     if AreFriends(c, uoi) then
8         FriendsFound ← FriendsFound ∪ {c}
9 return FriendsFound
```

(Facebook Inc, 2015; Magazine, 2012)