WISP 2018 Proceedings

Pre-ICIS Workshop on Information Security and Privacy (SIGSEC)

Winter 12-13-2018

Cyber insurance for correlated risks from phishing attacks: A decision-theoretic approach

Arunabha Mukhopadhyay
Indian Institute of Management, Lucknow, arunabha@iiml.ac.in

Baidyanath Biswas
International Management Institute Kolkata, b.biswas@imi-k.edu.in

Gaurav Gupta
Indian Institute of Management Calcutta, gauravg13@email.iimcal.ac.in

Recommended Citation

Mukhopadhyay, Arunabha; Biswas, Baidyanath; and Gupta, Gaurav, "Cyber insurance for correlated risks from phishing attacks: A decision-theoretic approach" (2018). *WISP 2018 Proceedings*. 32.
https://aisel.aisnet.org/wisp2018/32

Cyber insurance for correlated risks from phishing attacks: A decision-theoretic approach

Arunabha Mukhopadhyay (arunabha@iiml.ac.in)
Indian Institute of Management Lucknow, India

Baidyanath Biswas (b.biswas@imi-k.edu.in)
International Management Institute Kolkata, India

Gaurav Gupta (gauravg13@email.iimcal.ac.in)
Indian Institute of Management Calcutta, India

ABSTRACT

Phishing attacks contribute to a variety of cyber incidents such as data breaches, and ransomware attacks. These attackers regularly discuss cyber sensitive topics and keywords, share exploits, and ransomware kits through messages in online forums that act as communities of practice. The research on correlated cyber risk from phishing attacks is in its infancy. In this research-in-progress paper, we propose a framework for the assessment of phishing risks in an organization and subsequent mitigation through balanced investments in IT security and complimentary cyber insurance. First, our framework employs binary classifiers to determine an expert phisher, who can launch phishing attacks and the misdetection of phishing URLs in an organization. Second, our framework identifies the optimal cyber insurance premium to indemnify the correlated loss from undetected phishing attacks. In this manner, the results of this study will assist CTOs to plan for balanced cybersecurity investments, and guide cyber insurers to design differentiated insurance products under various risk attitudes of organizations.

Keywords: Information security; dark forums; cyber risk assessment; cyber risk mitigation; cyber insurance; Copula; utility models.

INTRODUCTION

Phishing involves social engineering of user data over the Internet to acquire personal and business information from innocent users. Attackers exploit the user's susceptibility to deception (Goel et al., 2017) and trick them to divulge critical information unintentionally (Leukfeldt et al., 2016; Wright et al., 2014). Such information includes login credentials for social networks, banking applications, credit cards, and healthcare through emails, corrupt URLs and multimedia messages.

Recently, phishing attacks have contributed to a variety of secondary cyber disasters: data breaches, ransomware attacks, business email compromises, tech support frauds, and tax refund scams (FBI IC3, 2017). In 2017, the FBI received more than 25,344 complaints about phishing attacks where the victims suffered a combined loss of $720 million. In 2015, the US-CERT announced a phishing alert for attacks through Kill Disk malware that was delivered via spear-phishing emails to power, oil and gas companies in Ukraine (ICS-CERT, 2016). Although malicious agents execute these attacks through a set of technical steps, security researchers consider phishing as more of an economic problem than a technological one (Akerlof and Shiller, 2015). Thus, phishing attacks are a vital concern among organizations that necessitates immediate attention and proactive intervention from CTOs (APWG, 2017).

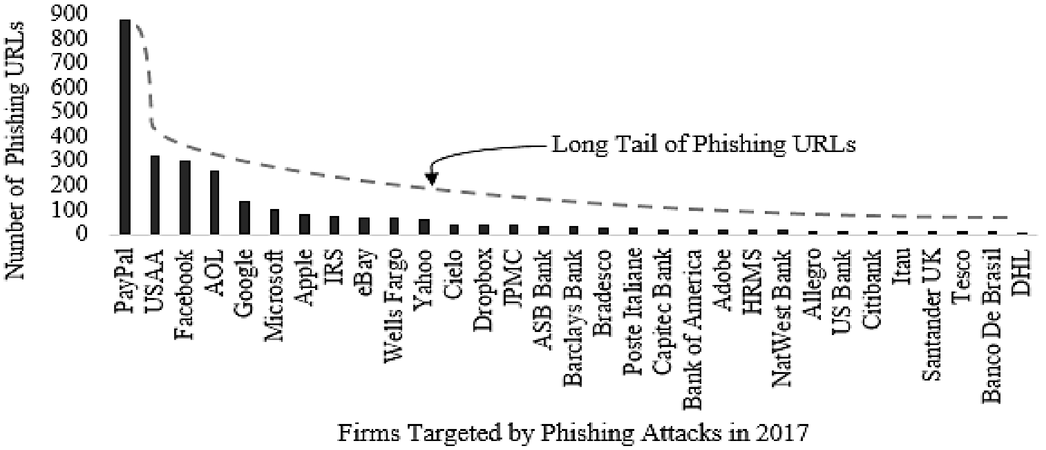

Figure 1. — Top firms facing phishing attacks in 2017

Figure 1 shows the number of phishing URLs used by malicious attacks to target users in 2017, and verified by PhishTank. We noted that (i) hackers targeted naïve customers by fake PayPal domains and counterfeit tech support messages, (ii) a high number of phishing emails with technology firm domains such as Facebook, Google, AOL, Microsoft, Apple and Adobe. (iii) the number of phishing URLs follow a long tail where the number of highly impacted firms are smaller in number, whereas a large number of firms with minimal impact exist (Anderson, 2008).

PROPOSED FRAMEWORK TO ASSESS CORRELATED PHISHING RISK AND MITIGATION THROUGH CYBER INSURANCE

Theoretical Foundation

According to the Opportunity Theory of Crime (Cohen and Felson, 1979), a phisher needs to target a gullible victim using Deception (Akerlof and Shiller, 2015) whose IT system does not have any effective spam filters, or antivirus to block phishing URLs (Hannon, 2002). This is in line with Social Conflict Theory (Pitcher et al., 1978) where expert attackers and naïve users act as instigating factors while anti phish filters and perimeter security installations act as inhibiting factors for phishing attacks (Biswas and Mukhopadhyay, 2016). The Rational Choice Theory (Ehrlich, 1996) suggests that a phisher will weigh the cyber risks versus returns before executing an attack. Hence, an accurate estimation of the likelihoods of phisher expertise and efficient detection of suspicious URLs is crucial to minimize phishing attacks. Based on the accuracy of these predictions, CTOs can resort to IT security investments and complimentary cyber insurance under different risk-taking abilities of a firm (Kahane et al., 1988). Figure 2 shows our proposed conceptual framework PRAMCI consisting of phishing-risk assessment, followed by risk mitigation.

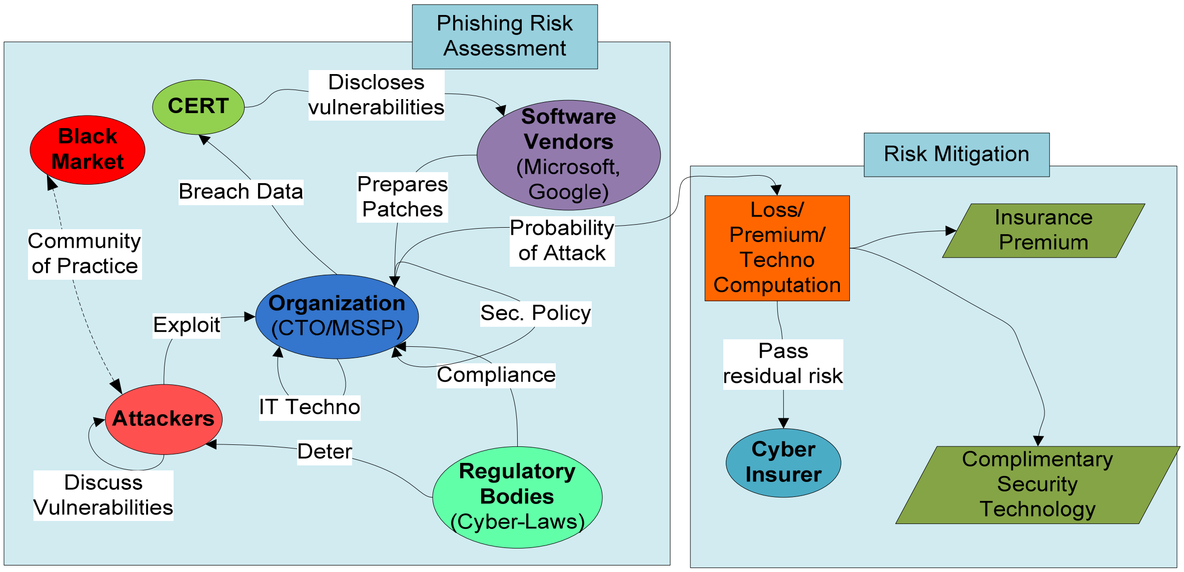

Figure 1. – Correlated phishing risk assessment and subsequent mitigation using cyber insurance

Cyber risk assessment for correlated phishing attacks

Cyber risk assessment consists of (i) estimation of expert attackers by mining Darknet messages; (ii) estimation of the likelihood of phishing attacks; (iii) misdetection of attacks; and expected loss under correlated misdetection and phishing attacks.

Estimation of expert attackers by mining Darknet messages

Dark forums and hacker communities provide an easy and simple mechanism for malignant users to share and discuss technical knowledge and distribute malicious source codes and files (Benjamin and Chen, 2012). We employ binomial logistic regression model to classify into two hacker roles – “expert” and “novice” (Biswas et al., 2018). Logistic regression can handle such a classification problem. We use the maximum likelihood estimator (MLE) and it ensures that the probability of hacker follows $0 < p_{ex} < 1$ in (1), (2).

$$P(Y = \text{Expert}|\ X_1, X_2, \dots, X_{11}\ ) = p_{ex} = \frac{1}{1+e^{-W_1}} \quad (1)$$

$$P(Y = \text{Novice}|\ X_1, X_2, \dots, X_{11}\ ) = 1 - p_{ex} \quad (2)$$

where X_1, X_2, ... X_{11} are the predictor variables extracted from hacker messages (Samtani, 2016).

Estimation of the likelihood of phishing attacks

Each of the N_t organizations out of Y_t had installed security technologies and complied with industrywide standards in t th year (Sec_t) (Mukhopadhyay et al., 2017). Additionally, they conducted IT security audits, up-to-date IT security policies (Dhillon and Backhouse, 2000), reported intrusion(s) to enforcement agencies and legal counsels (Reg_{t-1}) (Fischer, 2013). Yet Y_t out of N_t CTOs reported to the CSI-FBI about a cyber-attack in year t, while $(N_t - Y_t)$ CTOs were not attacked. So, Y_t has two possible states (i.e., attack: $Y_t = 1$ or no attack: $Y_t = 0$) with a probability of p_{ph} and $(1 - p_{ph})$ respectively, as shown in Eq. (3). The Logit link function approximates the probability p_{ph} of phishing with an exponential form that maps the likelihood of attack p_{ph} to [0, 1]. We consider a lag of one year (i) from the time of security deployment Sec_{t-1} and impacted users, Y_t, (ii) between the legal, regulatory factors Reg_{t-1} , and Y_t. Y_t has two states $(Y_t = 1\ ; Y_t = 0)$, and fits a binomial distribution.

$$P(Y_t = 1) = {}^{N_t}C_{Y_t} p_{ph}^{Y_t} (1 - p_{ph})^{(N_t - Y_t)} \quad (3)$$

$$M1: P(Y_t = 1|\ X = t, Sec_{t-1}, Reg_{t-1}) = p_{ph} = \frac{1}{1+e^{-V_1}} \quad (4)$$

where $V_1 = \beta_0 + \beta_1 t + \beta_2 Sec_{t-1} + \beta_3 Reg_{t-1}$ and p_{ph} = likelihood of phishing attacks.

Estimation of the likelihood of misdetection of phishing attacks

The CTO in a bid to prevent the phishing attacks wishes to examine the URLs in emails (Valecha et al., 2018) based on features such as address-bar, abnormality, HTML/JavaScript, and web-site statistics. Each of the features is encoded as [-1, 0, +1] for [phishing, suspicious, and legitimate] URLs. The CTO intends to classify URLs as phishing (-1) and genuine (+1). Therefore, the detection of phishing and genuine URLs is a two-class classification problem (5).

$$p_d\,(Y = 1|Z_1 = \alpha_1, Z_2 = \alpha_2, \dots, Z_{30} = \alpha_{30}) = p(Y)\prod p(Z_i|Y)/p(Z_i) \tag{5}$$

where $Z_1, Z_2, Z_3, Z_4, \dots Z_{30}$ = URL features (Mohammad et al., 2014).

Expected loss under correlated misdetection and phishing attacks

Once the organization can distinguish between phishing and genuine URLs, it needs to address the residual cyber risks from the undetected phishing URLs.

$$P\,(\text{Attack}) = f\{P(\text{Expert})\ , P\,(\text{Phishing}), P(\text{Detection Failure})\} \tag{6}$$

$$p_a = f(p_{ex}) * f(p_{ph}|\ p_{ex}) * f(p_d\ |\ p_{ph}, p_{ex}) \tag{7}$$

So, the expected loss $E(L)$ according to Sklar's Theorem (Sklar, 1959) is given by (8).

$$E(L) = F_{p_a,L}(p_a, L) = F_1(p_a) * F_2(L) * C[F_1(p_a)F_2(L), \rho(p_a, L)] \tag{8}$$

where $F_{p_a,L}$ represents the joint c.d.f. of the expected loss; $F_1(p_a)$ and $F_2(L)$ are univariate and continuous marginal c.d.f-s; Copula $C: [0,1]^2 \rightarrow [0,1]$ describes the link in a two-dimensional unit-square space and C is unique and differentiable; $\rho(p_a, L)$ denotes the correlation between p_a and L. We choose Archimedean Copula for the c.d.f of the joint distribution of E(L) with the marginal distributions for p_a and L being exponential (Wolpert, 2000) as shown in (9) and (10).

$$F_{p_a,L}(p_a, L) = C[F_1(p_a)F_2(L)] \quad \text{for } (p_a, L) \in (0,1) \times (0, \infty) \tag{9}$$

$$f\,(p_a, L) = f_1(p_a)\,f_2\,(L)\,c[F_1(p_a)F_2(L), \rho(p_a, L)\,] \tag{10}$$

Mitigation through complimentary cyber insurance and security technology

We propose a cyber-insurance model built on utility theory (Strecker et al., 2011) to recommend to the CTOs on possible cyber risk management – transfer outsourced, or in-house (Gordon et al., 2003; Dhillon et al., 2017). The PRAMCI framework shows that expert hackers could launch phishing attacks, and CTOs can block those attacks using our proposed phishing filter. If the filter fails, then the firm faces a possibility to suffer losses. If the CTO has paid a premium (I), and received L as full indemnification, assuming the cyber-insurance coverage pays him the entire loss suffered. However, if there is no cyber-insurance, the CTO suffers a loss L. Even if no phishing attack occurs, an insured organization would still pay I. Therefore, the CTO compares between (i) procuring third-party cyber-insurance or (ii) manage the cyber risk on his own. Eqn. (11) – (13) present the decision-making problem for the organization, as illustrated in Figure 3.

Objective: To find $(I, \Delta CT)$ for risk-neutral, risk-averse, and constant-risk types

$$\text{s.t. } E\left[U_{insured}\right] \geq E\left[U_{not-insured}\right] \text{ and } B = (I + CT) \tag{11}$$

$$E\left[U_{insured}\right] = p_{ex}p_{ph}p_d U(R - CT - I) + p_{ex}p_{ph}(1 - p_d)U(R - CT - I) + p_{ex}(1 - p_{ph})p_d U(R - CT - I) + p_{ex}(1 - p_{ph})(1 - p_d)U(R - CT - I) \tag{12}$$

$$E\left[U_{not-insured}\right] = p_{ex}p_{ph}p_d U(R - CT) + p_{ex}p_{ph}(1 - p_d)U(R - CT - L) + p_{ex}(1 - p_{ph})p_d U(R - CT) + p_{ex}(1 - p_{ph})(1 - p_d)U(R - CT) \tag{13}$$

where utility $U(x)$ can take a linear, quadratic or exponential functional form for risk-neutral, risk-averse, and constant-risk types respectively, and ΔCT = Per unit cost of security.

CONCLUSION AND FUTURE STUDY

Through our preliminary study, we explored and built decision-theoretic models to compare the feasibility of proposing cyber insurance in comparison to complimentary security investments. In future, we aim to compute the probabilities in (1)-(7) and estimate the parameters of the bivariate Archimedean copula of E(L) with empirical data. Our initial analysis revealed that E(L) showed a longer tail than exponential (Mukhopadhyay et al., 2017), with almost four times variance.

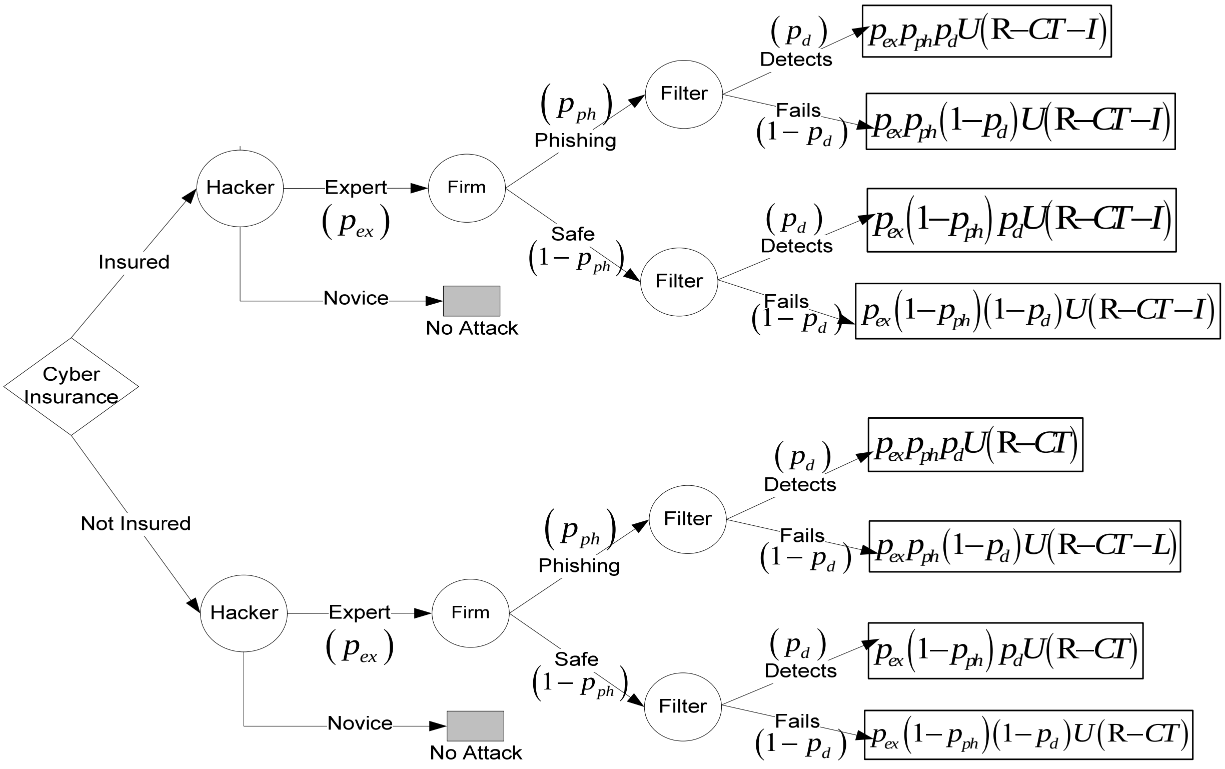
Cyber Insurance
Insured
Not Insured
Hacker
Expert
(p_{ex})
Novice
No Attack
Firm
Phishing
(p_{ph})
Safe
$(1-p_{ph})$
Filter
Detects
(p_d)
Fails
$(1-p_d)$
$p_{ex}p_{ph}p_dU(\mathrm{R}-CT-I)$
$p_{ex}p_{ph}(1-p_d)U(\mathrm{R}-CT-I)$
$p_{ex}(1-p_{ph})p_dU(\mathrm{R}-CT-I)$
$p_{ex}(1-p_{ph})(1-p_d)U(\mathrm{R}-CT-I)$
$p_{ex}p_{ph}p_dU(\mathrm{R}-CT)$
$p_{ex}p_{ph}(1-p_d)U(\mathrm{R}-CT-L)$
$p_{ex}(1-p_{ph})p_dU(\mathrm{R}-CT)$
$p_{ex}(1-p_{ph})(1-p_d)U(\mathrm{R}-CT)$

Figure 3. – Decision tree for the insured and uninsured states of an organization

REFERENCES

Akerlof, G. A., & Shiller, R. J. (2015). Phishing for phools: The economics of manipulation and deception. *Princeton University Press*.

APWG(2017). "AntiPhishing Working Group (APWG) Attack Trends Reports (2004-2017)," (https://www.antiphishing.org/resources/apwg-reports/, access on 01 August, 2018).

Benjamin, V., & Chen, H. (2012). Securing cyberspace: Identifying key actors in hacker communities. *Proceedings of the International Conference on Intelligence and Security Informatics (ISI),* IEEE, pp. 24-29.

Biswas, B., & Mukhopadhyay, A. (2016). "Phish Detection and Loss Computation Hybrid Model - A Machine Learning Approach," *ISACA Journal*, (8:1), pp. 22-29.

Biswas, B., Mukhopadhyay, A., & Gupta, G. (2018). "Leadership in Action: How Top Hackers Behave - A Big-Data Approach with Text-Mining and Sentiment Analysis," *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 1752-1761.

Cohen, L. E., & Felson, M. (1979). "Social change and crime rate trends: A routine activity approach," *American Sociological Review*, pp. 588-608.

Dhillon, G., & Backhouse, J. (2000). "Technical opinion: Information system security management in the new millennium," *Communications of the ACM*, (43:7), pp. 125-128.

Dhillon, G., Syed, R., & de Sá-Soares, F. (2017). "Information security concerns in IT outsourcing: Identifying (in) congruence between clients and vendors," *Information & Management*, (54:4), pp. 452-464.

Ehrlich, I. (1996). "Crime, punishment and the market for offenses," *Journal of Economic Perspectives*, (10:1), pp. 43–67.

FBI IC3 (2017). “FBI Annual Internet Crime Report 2017 - Internet Crime Complaint Center,” (https://pdf.ic3.gov/2017_IC3Report.pdf, access on 01 August, 2018).

Fischer, E. A. (2013). “Federal Laws Relating to Cybersecurity: Overview and Discussion of Proposed Revisions,” *Congressional Research Service, Library Of Congress*, Washington DC.

Goel, S., Williams, K., & Dincelli, E. (2017). “Got phished? Internet security and human vulnerability,” *Journal of the Association for Information Systems*, (18:1). pp. 22-44.

Gordon, L.A., Loeb, M.P., & Sohail, T. (2003). “A Framework for Using Insurance for Cyber-Risk Management,” *Communications of the ACM*, (46:3), pp. 81-85.

Hannon, L. (2002). “Criminal opportunity theory and the relationship between poverty and property crime,” *Sociological Spectrum*, (22:3), pp. 363–381.

ICS-CERT (2016): “Cyber-Attack Against Ukrainian Critical Infrastructure,” (https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01, accessed on 01 August, 2018).

Kahane, Y., Neumann, S., & Tapiero, C. S. (1988). “Computer backup pools, disaster recovery, and default risk,” *Communications of the ACM*, (31:1), pp. 78-83.

Leukfeldt, E. R., Kleemans, E. R., & Stol, W. P. (2016). “Cybercriminal networks, social ties and online forums: social ties versus digital ties within phishing and malware networks,” *British Journal of Criminology*, (57:3), pp. 704-722.

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). “Predicting phishing websites based on self-structuring neural network,” *Neural Computing and Applications*, (25:2), pp. 443-458.

Mukhopadhyay, A., Chatterjee, S., Bagchi, K. K., Kirs, P. J., & Shukla, G. K. (2017). “Cyber Risk Assessment and Mitigation (CRAM) Framework Using Logit and Probit Models for Cyber Insurance,” *Information Systems Frontiers*, pp. 1-22.

Pitcher, B., Hamblin, R. & Miller, J. (1978). “The Diffusion of Collective Violence,” *American Sociological Review*, (43), pp. 23-35.

Samtani. S. (2016). “Hacker Web Forum Collection: Hackhound Forum Dataset,” (http://www.azsecure-data.org/, access on 01 August, 2018).

Sklar, M. (1959). Fonctions de repartition an dimensions et. leurs marges. Publ. Inst. Statist. Univ. Paris, 8, pp. 229-231.

Strecker, S., Heise, D., & Frank, U. (2011). “RiskM: a multi-perspective modeling method for IT risk assessment,” *Information Systems Frontiers*, 13, pp. 595–611.

Wolpert, R.L. (2000). “Exponential Families,” (http://www2.stat.duke.edu/courses/Spring11/sta114/lec/expofam.pdf, access on 01 May, 2018).

Valecha, R., Chakraborty, R., Rao, H. R., & Upadhyaya, S. (2018). A Prediction Model of Privacy Control for Online Social Networking Users. In *International Conference on Design Science Research in Information Systems and Technology* (pp. 300-315). Springer, Cham

Wright, R.T., Jensen, M.L., Thatcher, J.B., Dinger, M., & Marett, K. (2014). “Research Note — Influence Techniques in Phishing Attacks: An Examination of Vulnerability and Resistance,” *Information Systems Research*, (25:2), pp. 385-400.