ALLOCATION OF LIABILITY BETWEEN INTERNET SERVICE PROVIDERS AND END-USERS – A THEORETICAL PERSPECTIVE

Ronen Avraham, Joachim Meyer & Omer Pelled

ABSTRACT

Ideally, authorities should deal with cybercrime by stopping cybercriminals. Unfortunately, technological and jurisdictional constraints limit law enforcement agencies' ability to deter criminals from causing harm. Investing in security helps, but it is not enough. Cyberattacks continue to affect individuals and companies, in spite of the major efforts invested in security. Attacks often succeed because users fail to take sufficient protective actions. Service providers can warn users about possible risks (e.g., risky links or mail attachments). The question arises as to who should pay for the damage caused by a successful cyberattack. If users are required to pay, they may be more cautious or even cease to use online services with possible risks. If service providers are required to pay, they may be overly cautious and issue excessive warnings that are ignored by users. We analyze this question in a game-theoretic framework, and compute the outcomes from different allocations of the damage. These game theoretic analyses can inform system design and policy decisions regarding relevant legislation.

Working paper, please do not cite or quote without permission

January 2019

1. INTRODUCTION

In early May 2017, Google experienced a massive phishing attack. Google users received an email in their inboxes from one of their trusted contacts asking them to check out an attached Google Docs file. In reality, it was a worm. Clicking on the link made the user’s personal data vulnerable to hackers. Within hours, Google said it had "disabled" the malicious accounts and pushed updates to all its users.

Phishing attacks like this are common. What was unique about this attack is that it was unusually sophisticated – the malicious link looked very realistic and came from someone the user knew. In its efforts to kill the worm, Google disabled the accounts of innocent users, themselves victims of the attack, who clicked the link. Google essentially punished the victims of the phishing attack without giving them any warning, a hearing, or any other due process rights. There is evidence that, at least in Israel,[1] the punishment not only blocked the users from continuing to use their accounts in the future, but it was also retroactive. Google did not allow users to recover their personal data. As a result, users lost access to their emails, documents, and photos.

Many interesting questions arise. Should Google be allowed to punish its users at all for falling prey to phishing attacks? Should it be subject to some proportionality requirements of the sort that apply to governments in order to guarantee that the punishments fit the crime? Should Google be liable for failing to warn its users about the malicious software?

In this paper, we advance the ball by exploring the optimal allocation of liability between internet providers and end-users. Specifically, we are interested in exploring whether end-users should bear any responsibility, and any liability, as part of the concentrated effort to prevent cybercrime.

In Part 2 we present the problem of cybercrime and the different parties that might invest to mitigate the problem, mainly ISP providers and end-users. We show that both parties should invest in cybercrime prevention, but are not optimally incentivized to do so. We also cover current suggestion in the literature about the division of liability between the parties.

In Part 3 we develop a simple model for the interaction between Internet Service providers (ISPs) and end-user. In the model each party receives private information

[1] http://www.haaretz.co.il/captain/net/.premium-1.4157793?utm_source=mivzakimnet&utm_medium=rss&utm_campaign=mivzakimnet (in Hebrew). We could not find evidence that similar measures were taken in other countries.

about an item that might be malicious. The ISP can warn the user or refrain from doing so, and the user can open the file or delete it. We show how different liability regimes, including strict liability and comparative negligence, might affect the incentives of the parties. Specifically, we show under what conditions both parties are optimally incentivized.

In Part 4 we offer some conclusions based on the model.

2. MISALIGNED INCENTIVES OF THE PLAYERS

Cybercrime is unlawful conduct that involves one or more computers and one or more networks (usually the Internet).[2] Cybercriminals exploit some sort of security vulnerability in the computers or the networks to execute a cyberattack. Many law enforcement agencies dedicate resources to prevent these crimes and track down those responsible for successful attacks.[3] While law enforcement occasionally scores victories against attackers and criminal enterprises, there is no foreseeable conclusion to the war against cybercrime in the near future.

As time passed and Internet technology matured, the business models of cybercriminals evolved accordingly. If at the end of the 20th century a combination of curiosity and malicious intent spawned viruses and worms, the 21st century has witnessed cybercriminals finding ‘viable’ economical models to receive substantial economic gains through attacks.[4] Spam is an early example of such a model. In that case, once the ISPs stopped large portions of e-mails sent by a single computer, cybercriminals were forced to rejuvenate their business model. By creating malware that infected a multitude of computers in response, attackers were able to circumvent the restrictions imposed by the ISPs. Each of the infected computers would send only a few e-mails, but together those networks of infected computers achieved an attacker’s malicious goal. These networks – the ‘botnets’ – were also used to initiate attacks against other computers and services, thus rendering them useless for extended time periods.[5] The botnets were not

[2] Neal Kumar Katyal, *Criminal Law in Cyberspace*, 149 UNIV. OF PENN. L. REV. 1003, 1013-38 (2001).

[3] R. Anderson, C. Barton, R. Böhme, R. Clayton, M.J.G. van Eeten, M. Levi, et al., *Measuring the cost of cybercrime*, 11th Annual Workshop on the Economics of Information Security, Berlin, Germany, at 2 (2012).

[4] Jonathan Zittrain, THE FUTURE OF THE INTERNET–AND HOW TO STOP IT (Yale Univ. Press 2008), at 43-51.

[5] Jonathan Zittrain, THE FUTURE OF THE INTERNET–AND HOW TO STOP IT (Yale Univ. Press, 2008), at 45-47.

Coordinated attacks from many computers that are part of a ‘botnet’ are called DDOS [distributed denial of service] attacks. Such attacks can be used to extort businesses into paying for not being attacked. (*See* T. Luis De Guzman, *Unleashing a Cure for the Botnet Zombie Plague: Cybertorts, Counterstrikes, and* the most effective way for the "botnet herder"[6] to extract payment from the infected computers' owners, and with time attackers discovered more efficient ways to monetize cyberattacks.

One recent breakthrough to the hacker's economic model has been through the growing use of virtual-currency (e.g., BitCoin).[7] By using virtual-currencies, attackers are not bound to the traditional monetary intermediaries and can accept payments from around the world without being easily traced.[8] With the new ability to receive money straight from end users, the popularity of some attacks has surged. For example, attackers have applied "ransomware"[9] to extort payment directly from an infected computer's owners.[10]

Cybercriminals exploit the lack of meaningful enforcement against them and invent new ways to attack. The enormous potential earnings generated through cybercrime motivate them to continue their malicious deeds and to impose high crime costs on society.[11] Since current law does not deter the cybercriminals, we need to examine other ways to reduce the damages from cyberattacks.

A. The ISPs

Various policy proposals have been advanced to address the rising costs of cyberattacks. Some proposals advocate reducing the legal hurdles that law enforcement agents face

Privileges, 59 CATH. U. L. REV. 527, 528-31 (2009); Adam J. Sulkoswki, *Cyber-Extortion: Duties and Liabilities Related to the Elephant in the Server Room*, U. ILL. J. L. TECH. & POL'Y 19, 21-25 (2007)).

[6] The person who controls the botnet.

[7] Matthew Kien-Meng Ly, *Coining Bitcoin's Legal-Bits: Examining the Regulatory Framework for Bitcoin and Virtual Currencies*, 27 HARVARD J. L. & TECH. 587, 589-91 (2013).

[8] Derek A. Dion, *I'll Gladly Trade You Two Bits on Tuesday for a Byte Today: Bitcoin, Regulating Fraud in the E-Conomy of Hacker-Cash, U. Ill. J. L. Tech. & Pol'y* 165, 166 (2013).

[9] An attack that encrypts the infected computers' files and would release them only using a key that the attacker will supply for a ransom; see A.J. Gazet, Comput Virol (2010) 6:77. doi:10.1007/s11416-008-0092-2 pp 78-79/

[10] A. Kharraz, W. Robertson, D. Balzarotti, L. Bilge, E. Kirda, (2015) *Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks*, In: Almgren M., Gulisano V., Maggi F. (eds) Detection of Intrusions and Malware, and Vulnerability Assessment. DIMVA 2015. Lecture Notes in Computer Science, vol 9148. Springer, Cham. pp 13-18; Scaife, N., Carter, H., Traynor, P., & Butler, K. R. (June 2016). Cryptolock (and drop it): Stopping ransomware attacks on User data. In Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on (pp. 303-312). IEEE. pp 303.https://regmedia.co.uk/2016/10/27/scaife-icdcs16.pdf

[11] Even though that the social costs of cybercrime are high, it is extremely challenging to quantify them. Paul Hyman, *Cybercrime: It's Serious, But Exactly How Serious?* 56 COMMUNICATIONS OF THE ACM 18-20 (2013).

when trying to police an international network.[12] Others suggest tackling cyberattacks from different perspectives. Since the Internet is basically a network of computers, activities usually involve many intermediaries, ranging from various ISPs to software manufacturers. If those intermediaries have abilities to reduce the potential harm and likelihood of cyberattacks, there are strong policy reasons to assign some form of liability on them to incentivize optimal reduction of costs from cyberattacks.[13]

Without any form of legal incentives, there is a dearth of reasons for the intermediaries to invest in Internet safety. Market failures are an issue that often go overlooked. One major impediment to producing safe software and services is the amount of time needed to do so. Network externalities incentivize companies to be the first to market their services and software, dis-incentivizing companies from taking the time to ensure their software is as safe as it could be. The first to enter the market will gain a large market share and may even subsequently enjoy a monopoly in its field.[14] Even when no network externalities are involved with a product or service, the market cannot provide secure solutions to consumers because consumers have no means to understand if the service or the product that they receive is secure. Reverse engineering of the code is usually prohibited, thus creating informational asymmetries that cause a "market for lemons."[15] Even if ISPs and software manufacturers could demonstrate that their services are secure, transaction costs and other switching costs can undermine competition. Changing service providers or software is not trivial, as it requires one to acquire technical knowledge about the implications of the new service on the users' Systems. As one might infer, the process of transferring information from one ISP to another might be extraordinarily time consuming and can unfortunately act as a deterrent to user mobility with regard to provider choice.[16]

[12] For cooperation in law enforcement, *see* Ross J. Anderson et al., *Security Economics and the Internal Market* 79-81 (2008) (study commissioned by ENISA). For a suggestion of a wide jurisdiction over cybercrime, *see* Kelly A. Gable, *Cyber-Apocalypse Now: Securing the Internet Against Cyberterrorism and Using Universal Jurisdiction as a Deterrent*, 43 VAND. J. TRANSNAT'L L. 105 (2010).

[13] *See also* Jonathan Zittrain, *A History of Online Gatekeeping*, 19 HARVARD J. OF L. & TECH. 253, 256 (2005-2006); Lawrence Lessig, *The Constitution of Code: Limitations on Choice-based Critiques of Cyberspace Regulation*, 5 COMM. LAW CONSPECTUS 104 (1997); Ronald J. Mann & Seth R. Belzley, *The Promise of Internet Intermediary Liability*, 47 WM. & MARY L. REV. 239, 267-69 (2005).

[14] Ross Anderson &Tyler Moore, *The Economics of Information Security*, 314 SCIENCE 610, 611 (2006); Douglas A. Barnes, *Deworming the Internet*, 83 TEX L. REV. 279, 290 (2004).

[15] Ross Anderson &Tyler Moore, *The Economics of Information Security*, 314 SCIENCE 610, 612 (2006); *see also* George A. Akerlof, *The Market for Lemons: Quality Uncertainty and the Market Mechanism*, THE QUARTERLY JOURNAL OF ECONOMICS 488-500 (1970).

[16] Aaron S. Edlin & Robert G. Harris, *The Role of Switching Costs in Antitrust Analysis: A Comparison of Microsoft and Google*, 15 YALE J. L. & TECH. 176-184 (2012).

Given the market cannot provide an adequate solution to the problem, and that bad actors are out of the law's reach and intermediaries are unable to contractually agree with end-users on the optimal level of security in the service provided, some scholars have suggested imposing tort liability on the ISPs, while other have called for assigning liability to software manufacturers.

Rustad and Koenig suggested creating a new tort for software vendors for the negligent enablement of cybercrime. Their proposed tort would allocate responsibility to both software vendors and end-users. The vendors would be liable for negligent products that contain preventable security flaws, while end-users would be liable for taking inadequate security measures or failing to implement vendor security updates.[17]

Lichtman and Posner proposed imposing strict liability on ISPs. The two scholars justify assigning stringent liability to ISPs because these actors are in the best position to take precautions that would reduce the risk and harm associated with cyberattacks.[18] As for the end users' incentives in maintaining security, Lichtman and Posner asserted that a tailored threshold of liability that would leave some incentives on subscribers to take additional care is needed. This liability could be cast as comparative negligence or strict liability with the defense of contributory negligence when the target of the attack failed to exercise due care.[19]

The idea of imposing some form of liability on intermediaries has been the subject of substantial research in the last decade. Moore recommended that ISPs take on a role in fighting malware in the network.[20] Researchers for a project commissioned by the European Network and Information Security Agency ("ENISA") recommended that the European Union introduce a statutory scale of damages against ISPs that do not respond promptly to requests for removal of user computers compromised with malware. If the ISP disconnects a user that wants an infected computer reconnected to the network, the user would assume full liability for any resulting damages.[21]

[17] Michael L. Rustad & Thomas H. Koenig, *The Tort of Negligent Enablement of Cybercrime*, 20 BERKELEY TECH. L. J. 1553, 1561 (2005).

[18] Douglas Gary Lichtman & Eric Posner, *Holding Internet Service Providers A*ccountable (John M. Olin Program in Law and Economics, Working Paper No. 217, 2004); Lichtman & Posner at 18; *see also* Michael L. Rustad & Thomas H. Koenig, *Rebooting Cybertort Law*, 80:2 WASH. L. REV. 315, 386 (2005); Jonathan Zittrain, *Internet Points of Control*, 44:2 BOSTON COLLEGE L. REV. 653-688 (2003); Perset, K. (2010), "The Economic and Social Role of Internet Intermediaries", OECD Digital Economy Papers, No. 171, OECD Publishing, 4-5; Hadi Asghari, Michel van Eeaten & Johannes M. Bauer, *Economics of Cybersecurity, in* HANDBOOK ON THE ECONOMICS OF THE INTERNET 262, 271-272; Ross J. Anderson et al., *Security Economics and the Internal Market* 49-50 (2008) (study commissioned by ENISA).

[19] Lichtman & Posner at 27.

[20] Thomas Moore, *The Economics of Cybersecurity: Principles and Policy Options*, 3 INT. J. CRIT. INFRASTRUCT. PROT. 103, 110-13 (2010).

[21] Ross J. Anderson et al., *Security Economics and the Internal Market* 4 (2008) (study commissioned by ENISA).

B. End users

End-users are consumers that receive services from ISPs. As discussed above, the role of intermediaries to shield users and other parties from malicious attacks is intertwined with the need of end-users to also take some precautions. Unfortunately, this task is easier said than done. The users' characteristics vary from the most sophisticated security engineer to the less tech-savvy grandparent browsing the Internet at home. Since end-users are human beings, they can suffer from vulnerabilities inherent to all individuals that undermine efforts at enhancing cybersecurity. Even if the end-users were perfectly rational, they will still lack incentives to take sufficient security measures.

Negative externalities pose a significant problem. Not all actions taken by users affect only themselves. This reduces the users' incentives to take efficient defensive measures. Lack of internalization of the risks by the user can even promote dangerous activities like illegal file-sharing, whereby the user will get free software or media, while putting her computer at risk of being infected by malware.[22]

Positive externalities also pose a hurdle towards efficient System security. When users take security measures, not only the users themselves benefit from the precautions. Indeed, when sufficient number of users have security measures installed, other not-secured users can enjoy the added security of the network, similar to 'herd immunity' in vaccination.[23] But without adequate incentives, end-users will likely not install security to the benefit of the entire network.[24]

Not all security flaws stem from end-users' rational decisions. Some attacks take advantage of the human irrationality. The human factor of end-users is key to the success of certain attacks. Hackers are particularly keen to take advantage of those vulnerabilities because the human factor is both easier to exploit and harder (if not impossible) to "patch." The use of such vulnerabilities is a practice referred to as "social engineering." Social engineering involves the use of pretexting[25] and phishing[26] to bypass a given System's security measures. Users often believe what is written in a phishing message, similar to the psychological factors that have enable con-artists and scammers.

[22] Huw Fryer, Roksana Moore & Tim Chown, *On the Viability of Using Liability to Incentivise Internet Security*, Twelfth Workshop on the Economics of Information Security 13-14 (WEIS 2013).

[23] Neal Katyal, *Community Self-help*, 1 J. L. ECON. & POL'Y 33, 64 (2005); Neal Kumar Katyal, *Criminal Law in Cyberspace*, 149:4 UNIV. OF PENN. L. REV. 1003, 1081 (2001).

[24] Derek E Bambauer, *Ghost in the Network*, 162 UNIV. OF PENN. L. REV. 1011, 1065 (2014).

[25] Pretexting is a form of scam in which the attacker provides the victim with some pretext in order to convince the victim to give the attacker the data wanted. According to Anderson, pretexting targets mostly corporations while phishing attacks target the consumers. (SECURITY ENGINEERING at 21).

[26] Ross J. Anderson, SECURITY ENGINEERING: A GUIDE TO BUILDING DEPENDABLE DISTRIBUTED SYSTEMS (John Wiley & Sons 2008) at 17-22.

As explained above, dealing with end-users is no easy task. Above all else, end-users often adversely influence network security. Not all bad results from the lack of security of the network, however, should be attributed to the end-users, as they are not in the best position to mitigate these risks.

Calls for some sort of end-user liability are meant to internalize the negative externalities that users produce. If liability is imposed on intermediaries, moral hazard behavior by end-users will ensue. Without user liability, assigning liability on intermediaries will increase the risks that users will take, as they will be reimbursed for some of the damages incurred. Therefore, the call by Professors Posner and Lichtman, as well as that of Rustad and Koenig, to allocate liability between users and intermediaries is only logical.[27] Others think that it is essential to impose some liability on users to incentivize them, even if intermediaries will not compensate end-users.[28] Some even say that without some sort of user incentives it is challenging to impose any form of liability on intermediaries.[29]

3. A MODEL OF ISP AND END-USER INTERACTION

The last Part of the Paper showed that the cost of cybercrime should be split between the ISP and the end-user. The goal of this Part is to examine the optimal regime for dividing liability between the parties.

A. Game parameters

We are interested in the reactions of ISPs and end-users to a known risk of malicious software "hiding" in what might appear as a benign file or message. Since the creator of the malicious software is not of interest, we treat the risk as exogenous.

We model the interaction using two players – a System and a User. The System provides items to the User (files, e-mail, messages, etc.), where items might be infected. Both the System and the User receive a private signal about each item. The System can issue an

[27] Lichtman & Posne; Rustad & Koenig.
See also Robert W. Hahn & Anne Layne-Farrar, *The Law and Economics of Software Securit*y, 30 HARVARD J. L. & PUB. POL'Y 283, 343 (2006).

[28] Susan W. Brenner, *Toward a Criminal Law for Cyberspace: Product Liability and Other Issues*, 5 PITTSBURGH J. OF TECH. L. & POL'Y 1, 50-51 (2004).
T. Luis De Guzman, *Unleashing a Cure for the Botnet Zombie Plague: Cybertorts, Counterstrikes, and Privileges*, 59 CATH. U. L. REV. 527, 554-56 (2009).

[29] Fryer, Huw, Roksana Moore, and Tim Chown. "On the Viability of Using Liability to Incentivise Internet Security." Twelfth Workshop on the Economics of Information Security (WEIS 2013), June 2013, pp. 17 & 22.

alert to the User when an item is suspected as malicious. The User then can choose whether to access the item or not. After the User decides whether to access the item, payoffs are realized.

We assume that all items of the same type have the same payoffs, a negative payoff for malicious items, and positive payoff for benign items. If an item is not accessed, both parties receive a payoff of zero, whether the item is malicious or not.

B denotes the sum of the benefits for both parties from accessing a benign item. We assume that the shares of the System and the User are fixed. Let B_s denote the payoff for the System and B_u the payoffs of the User.[30]

C denotes the overall costs of accessing a malicious item. We assume that the parties bear together the entire costs. That can either mean that the initial harm is inflicted on the User and the System shares the costs by paying damages, or that both User and System are required to compensate third parties for their harm. Either way, we examine ways in which the law can divide the overall costs caused by the malicious item between the parties.

The legislature decides on an initial allocation of the costs, under which the System share is $R = [0,1]$ of the costs and the User's share is $(1 - R)$. Under a regime of strict liability, the initial allocation of costs would not change ex-post.

To compare strict liability with two fault-based regimes we first examine the case where the court can change the initial allocation of the costs if the System issued a warning and the User accessed the item. In that case, the User would be at fault, and the System's share of the cost would be reduced by a factor of $r = [0,1]$. That is, the System would pay rRC and the User would pay $(1 - rR)C$. We can think of r as a measure of comparative fault, or a reduction in the System's liability since the User assumed the risk of the item being malicious when she chose to open it despite the warning. We then turn to examine an alternative fault-based regime, which is based on the System's alarm-sending policy, and not on whether it decided to send an alarm in a particular case.

Both the User and the System are aware that files might be infected with malware, with probability P_M. That is, items are drawn randomly from two categories – malicious and benign. We assume that the System and the User receive a private signal about each item that allows them to update their beliefs about the probability of the item being malicious.

The game proceeds in three stages.

First, we assume that the System can pass the item to User accompanied with an alarm, or refrain from sending an alarm with the item. Since the system's choice is binary, the

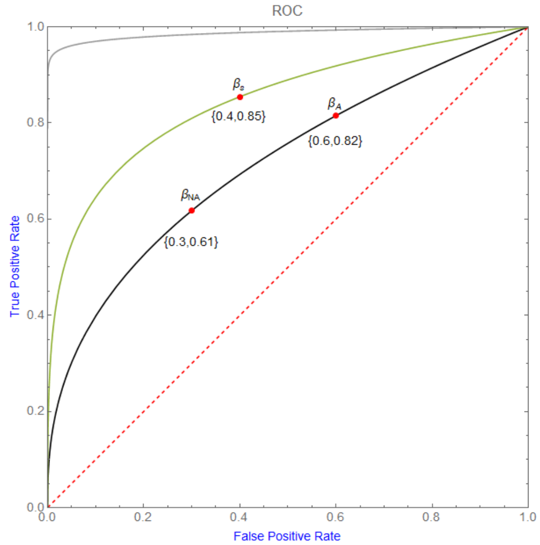
[30] Where $B = B_s + B_u$

information it passes to the User depends on the sensitivity (true positive rate) and specificity (true negative rate) of the alarm. System adopts a threshold β_s, such that it would send an alarm only if the odds that the signal, s, originated from a malicious item crosses the threshold, or formally when $\beta_m < \frac{P(s|M)}{P(s|B)}$.

Second, the User receives the item with or without the alarm, and a private signal - u. We assume that the User knows β_s, i.e., the sensitivity and specificity of the System's alarm, so the User's updated belief about the odds that the item is malicious, based on the System's decision is correct. The User then sets two thresholds for accessing the item or discarding it, one threshold, β_{NA}, determines which items the User would access if the System has not issued an alarm, and the second threshold, β_A, determines if the User would access an item given that the System has issued an alarm.

We can think of the System's and the User's thresholds as a classification mechanism, and that in choosing the thresholds, the System and the User should optimize the rates of true positive (identifying malicious items) and false positive (misidentifying benign items). Figure 1 illustrates the thresholds on an ROC curve –

Figure 1: Receiver Operating Charecteristics



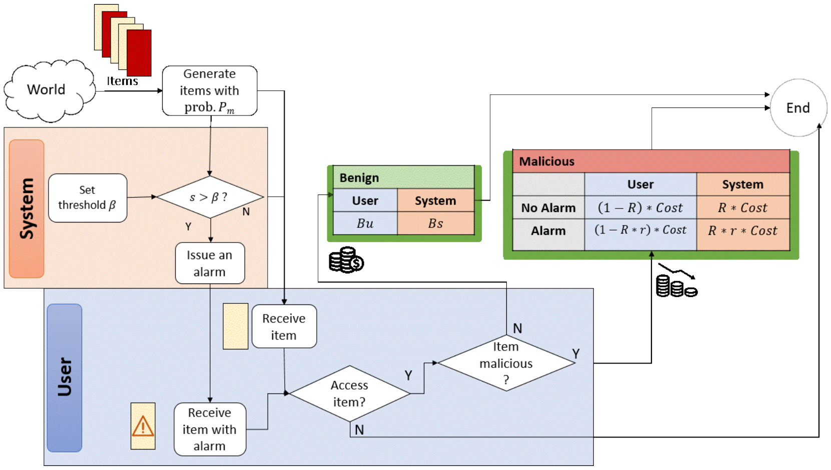
The X-axis presents the false positive (FP) rate. In this context, the FP rate is the number rate the System and the User would think an item is malicious when it is benign. The Y-axis presents the true positive (TP) rate, which is the rate the System and the User correctly classify a malicious item. The dashed diagonal line represents random classification. The top curve (gray) illustrates a nearly perfect signal, so the rate of TP is close to 1, while the rate of FP is low. The green curve illustrates the system's classifications abilities. The dot, marked as β_s, shows the rate of TP and TN used by the System. The System's signal, s, allows it to place β_s anywhere on the curve. The black curve represents the User's classification capabilities, based on the User's signal, u. The

User sets two thresholds, one if she receives the item with an alarm, and the other in case the item was received without an alarm.

After the User decides to access the item or discard it, payoffs are realized. If the item is benign the User and the System share the benefit according to their predetermined shares, and if the item is malicious the parties share the costs according to the legal regime.

For convenience, the following flowchart presents the stages of the game:

Figure 2: Flowchart of the System-User interaction



B. Strict liability

In our model, players face sequential decisions – the System sets a threshold, and only then does the User set her thresholds, based on the decision of the System.

The System's threshold determines the probability that it would send an alarm if the item is malicious – $P(TP)$ (System's true positive), the probability that it would not send an alarm if the item is malicious – $P(FN)$ (System's false negative), the probability that it would send an alarm if the item is benign – $P(FP)$ (System's false positive) and the probability that it would not send an alarm if the item is benign – $P(TN)$ (System's true negative).

The User then decides to access the item based on the information she has. Since the User's decision is based on the noisy signal she receives, we can describe the User's choice in terms of the probability that the User would open a malicious item if an alarm was sent – P(O|A,M) and if it was not – P(O|NA,M), as well as the probability that the User would open a benign item if an alarm was sent – P(O|A,B), and if it was not – P(O|NA,B).

The social welfare function is described by the following function –

$$(1)\ SW = B * (1 - P_m) * \big(P(FP) * P(O|A, B) + P(TN) * P(O|NA, B)\big) \\ -C * P_m * \big(P(TP) * P(O|A, M) + P(FN) * P(O|NA, M)\big)$$

The model presents a case of a bilateral accident, since both the System and the User can reduce the frequency of accidently opening a malicious item (and of accidently leaving a benign item closed) by taking precautions. In most models of bilateral accident, a strict division of the costs between the parties, that does not change based on the parties' actions, cannot produce efficient incentives (Kornhauser and Revesz 1989). Under this model, however, it is possible to induce both parties to act optimally using strict liability.

First, notice that under our definition of strict liability, the legislature allocates the costs between the parties (formally, the legislature sets the level of R). The following matrix describes the payoffs to the System and to the User**:**

Table 1: The payoffs to the System and the User

		System	
		Malicious	**Benign**
User	**Discards**	0 / 0	0 / 0
	Opens	$(1 - R)C$ / RC	B_u / B_s

From Table 1 and Exp. (1) we can formulate the parties' payoff function.

The System's payoff function is –

$$(2)\ SPF = (1 - P_m) * \big(P(FP) * P(O|A, B) + P(TN) * P(O|NA, B)\big) * B_s - C * R * \\ * P_m * \big(P(TP) * P(O|A, M) + P(FN) * P(O|NA, M)\big)$$

The System's payoff is equal to the probability that a benign item was sent to the User, times the probability that the item would be opened, times the benefit it derives from that event, minus the probability a malicious item was sent to the User, times the probability that the item would be opened, times the System's share of the costs.

The User's payoff function is-

$$(3)\ UPF = B_u * (1 - P_m) * \big(P(FP) * P(O|A, B) + P(TN) * P(O|NA, B)\big) - C * \\ * (1 - R) * P_m * \big(P(TP) * P(O|A, M) + P(FN) * P(O|NA, M)\big)$$

Exp. (3) is identical to Exp. (2) except for the size of the benefit (which might be different for the User) and the User's share of the costs.

Notice that if $R = \frac{B_s}{B_s+B_u}$ then $(1-R) = \frac{B_u}{B_s+B_u}$ and we can rewrite Exp. (2) and Exp. (3) as follows –

(4) $SPF = \frac{B_s}{B_s+B_u} * \left[B * (1-P_m) * \left(P(FP) * P(O|A,B) + P(TN) * P(O|NA,B)\right) - C * P_m * \left(P(TP) * P(O|A,M) + P(FN) * P(O|NA,M)\right)\right]$

(5) $UPF = \frac{B_u}{B_s+B_u} * \left[B * (1-P_m) * \left(P(FP) * P(O|A,B) + P(TN) * P(O|NA,B)\right) - C * P_m * \left(P(TP) * P(O|A,M) + P(FN) * P(O|NA,M)\right)\right]$

Notice that the function in the square brackets in both Exp. (4) and Exp. (5) is identical to the social welfare function $(\text{Exp. }(1))$. Since the value outside the square brackets is constant, each party would maximize its own payoff by acting optimally. Thus, by dividing the cost between the parties according to their respective shares in the social benefit, the law can induce both parties to act efficiently.

C. Negligence

In Section 3.B. we found that to create optimal incentives to both the System and the User, the parties have to internalize a fixed share of the costs, in proportion to their share in the benefits.

The question arises then, can we create optimal incentives to both parties under a fault-based regime, which allocates costs based on the parties' decision.

We assume that the court cannot observe the private signals the parties have received, so the fault-based regime cannot be based on the parties' reaction to the specific signals they received. The court, however, can see if the System has sent an alarm to the User or not. If the User is supposed to be more careful when the item is accompanied with an alarm, it might make sense for the law to allocate a larger share of the costs when she has received an alarm, and a smaller share if she did not.

The r parameter represents the reduction in the System's share of the costs when it has sent an alarm. Table 2 presents the payoffs of the parties under a negligence regime.

Table 2: The payoffs to the System and the User

		System			
		Alarm		No Alarm	
		Malicious	Benign	Malicious	Benign
User	Discards	0 / 0	0 / 0	0 / 0	0 / 0
	Opens	$(1-Rr)C$ / RrC	B_u / B_s	$(1-R)C$ / RC	B_u / B_s

It is easy to see from Table 2 that it is impossible to incentivize the User to act efficiently under negligence regime that is based on whether she received an alarm. First, notice that the game is sequential. Hence, the User faces one of two nodes in the game – either she received an alarm, and bears $(1 - Rr)$ of the costs, or she did not receive an alarm, and she bears $(1 - R)$ of the costs. However, from Exp. (5) we know that the User acts optimally if she has to pay for $\frac{B_u}{B_u + B_s}$ of the costs, whether the System sent an alarm or not. Therefore, the User's share in the costs cannot depend on the System's decision to send an alarm. For example, if $(1 - R) < \frac{B_u}{B_u + B_s}$, then the User would access too many items when the System did not send an alarm. Alternatively, if $(1 - Rr) > \frac{B_u}{B_u + B_s}$, the User would access too few items when the System sent an alarm. Whenever $r < 1$, it must be the case that either $(1 - R) < \frac{B_u}{B_u + B_s}$ or $(1 - Rr) > \frac{B_u}{B_u + B_s}$, proving that this type of fault-based regime can never create optimal incentives for the User.

It is possible to create a fault-based regime that is based on the System's alarm sending test, i.e., the choice of true-positive and true-negative rates. Under such a regime, the law would allocate a large share of the costs to the System $\left(R > \frac{B_s}{B_s + B_u}\right)$ when it does not behave efficiently, and an optimal share of the costs otherwise $\left(R = \frac{B_s}{B_s + B_u}\right)$.

In other words, the negligence mechanism that might work is one that depends on the System's threshold β_S. If the System's optimal threshold is β_S^*, there can be an equilibrium under which the System chooses $\beta_S = \beta_S^*$. Consider, for example, the following legal regime – $R = 1$ if $\beta > \beta^*$, otherwise $R = \frac{B_s}{B_s + B_u}$. Under this regime the System has a strong incentive to choose the optimal threshold, and the User's share of the cost incentivizes the User to open items only when it is socially optimal.

To determine the optimal threshold for the System, we should first determine the socially optimal threshold for the User, given the threshold of the System.

The User should access items whenever the expected benefit from accessing it outweighs the expected costs. That means that the User should access the item when the odds of it being malicious are lower than $\frac{B}{C}$.

If the System has sent an alarm, the posterior odds that the item is malicious equals the prior odds, times the odds that the System would send an alarm if the item is malicious over the probability that it would send an alarm if the item is benign $\frac{P_m * P(TP)}{(1-P_m) * P(FN)}$. Similarly, the odds of the item being malicious if the System did not send an alarm is given by $\frac{P_m * P(FN)}{(1-p_m) * P(TN)}$.

Thus, the User's socially optimal threshold given alarm is $\beta_A = \frac{P_m * P(TP) * C}{(1-P_m) * P(FP) * B}$. The User's socially optimal threshold given no alarm is $\beta_{NA} = \frac{P_m * P(FN) * C}{(1-p_m) * P(TN) * B}$.

As we have seen, The System's threshold would determine the true positive and false positive rates. Notice that the rate of true positive equals $\frac{P(TP)}{P(TP)+P(FN)}$, and similarly the rate of false positive equals $\frac{P(FP)}{P(FP)+P(TN)}$. Thus, knowing the User's and the System's ROC curves (Figure 1) allows the court to determine the optimal thresholds for both the User and the System.

4. CONCLUSIONS

Traditionally, tort liability has adopted one of two regimes – strict liability, which places the entire costs of accidents on one party, and comparative negligence, which examined how each party has acted and places liability according to the blameworthiness of each partys' actions.

Legal economists have long argued that in cases of bilateral accidents, sharing the costs of accidents between the parties cannot create optimal incentives, unless the allocation is based on fault.[31]

As the model shows, in situations of sequential precautions, where the actions of one actor influences the effectiveness of the other's precautions, current legal regimes would not create optimal incentives. Placing strict liability on one of the parties would result in one party taking excessive care and in the other taking inadequate care. A negligence regime that is based on each party's actions (sending an alarm or accessing the item) is no better – when one party can observe the other's actions the negligence regime acts as an insurance, resulting in moral hazard for the second player.

In this paper we find two liability regimes that have not been sufficiently examined in the literature. The first regime is proportional liability based on the parties benefit. Dividing costs according to benefits allows us to construct a strict liability regime that

[31] John J. Donohue III, *The Law and Economics of Tort Law: The Profound Revolution*, 102 HARV. L. REV. 1047, 1073 (1989) (states that strict liability only works when there is a clear injurer and a clear victim. For example, strict liability has a clear outcome in an accident involving a car and a pedestrian but not when the accident involves two cars.); Lewis A. Kornhauser & Richard L. Revesz, *Sharing Damages Among Multiple Tortfeasors*, 98 YALE L. J. 831, 856-860 (1989) (when sharing the costs of accident strict liability, there will be under-deterrence even if joint tortfeasors are held jointly and severally liable.)

creates optimal incentives for both parties, while alleviating the need to examine the parties' investment in care, or the standard of care. This insight can be implemented in a variety of situations covered by tort law, as well as to other areas in private law.[32]

A second alternative that was discussed in the model is the adoption of rule-based negligence instead of action-based negligence. There are situations in which injurers choose only the mean care level, and the actual care level is stochastically distributed around the mean.

In that case, injurers might set an efficient level of precaution, and still act negligently from time to time. In a unilateral accident model, the standard of care can be set in a way that induces efficient care levels. Under a bilateral accident model, precautions are sequential, and the victim knows the care level of the injurer prior to investing in care. In that case, injurers' liability might cause victims to underinvest in care. In these types of cases, courts should place liability according to the rate of accidents and not according to the behavior in a particular case.[33]

[32] *See* Omer Pelled, *The Proportional Internalization Principle in Torts, Contracts and Unjust Enrichment* (Unpublished manuscript, 2019)

[33] Similar suggestions have been raised with regard to motor vehicle accidents and medical malpractice cases. *See* Robert Cooter and Ariel Porat, *Lapses of Attention in Medical Malpractice and Road Accidents*, 15 THEORETICAL INQ. L. 329 (2014)