**2014 6th International Conference on Cyber Conflict**
P.Brangetto, M.Maybaum, J.Stinissen (Eds.)


# Low-cost active cyber defence

**Karlis Podins**
University of Latvia
Riga, Latvia
karlis.podins@gmail.com

**Iveta Skujina**
National CERT
Riga, Latvia

**Varis Teivans**
National CERT
Riga, Latvia

**Abstract:** The authors of this paper investigated relatively simple active strategies against selected popular cyber threat vectors. When cyber attacks are analysed for their severity and occurrence, many incidents are usually classified as minor, e.g. spam or phishing. We are interested in the various types of low-end cyber incidents (as opposed to high-end state-sponsored incidents and advanced persistent threats) for two reasons:

- being the least complicated incidents, we expect to find simple active response strategies;
- being the most common incidents, fighting them will most effectively make cyberspace more secure.

We present a literature review encompassing results from academia and practitioners, and describe a previously unpublished hands-on effort to actively hinder phishing incidents. Before that, we take a look at several published definitions of active cyber defence, and identify some contradictions between them.
So far we have identified active strategies for the following cyber threats:

- Nigerian letters – keep up conversation by an artificial intelligence (AI) text analyser and generator;
- spam – traffic generation for advertised domains;
- phishing – upload of fake credentials and/or special monitored sandboxed accounts;
- information collection botnets – fake data (credit card, credentials etc.) upload.

The authors analysed the proposed strategies from the security economics point of view to determine why and how these strategies might be effective. We also discuss the legal aspects of the proposed strategies.

**Keywords:** *security economics, cyber crime, active cyber defence, Nigerian letters, spam, phishing, botnet*

---

The views and opinions presented in this paper are solely those of the authors, and should not be considered as official policy of the affiliated institutions.

# 1. INTRODUCTION

The term 'active cyber defence' has been around for at least a decade (a paper by Wood *et al*. (2000) is devoted to problems around active cyber defence). The range of cyber threats can be viewed as starting from low-end (less complicated methods like spam, phishing etc.) and going to high-end (most complicated attacks e.g. Stuxnet and other state-sponsored malware). There is no clear line of division between low-end and high-end incidents; rather it is a continuous spectrum. But if we look at both ends of the spectrum there is a clear distinction between a single spam campaign and state-sponsored APT with multiple 0-day vulnerabilities. It is fairly obvious when looking at the cost estimates; an instance of spam campaign as reported by Kreibich *et al*. (2008) of 400 million emails in a three-week period, according to Goncharov (2012) would cost approx. $4000, while the Stuxnet development costs has been estimated to be around $10 million (Langner, 2010).

The authors would like to explore active cyber defence methods that are easy to implement with widely available technologies, therefore being low-cost, as the title of the paper suggests. We would like to investigate if and how such active cyber defence strategies could be applied to the occurrences of low-end cyber incidents that every netizen experiences daily.

Even though the incidents might not be technologically advanced, the sheer abundance of them causes significant losses, e.g. Rao and Reiley (2012) estimate annual spam-related costs for US companies and individuals at $20 billion, and the annual cost of advance-fee fraud to the UK economy is estimated to £150 million (Peel, 2006).

Before studying active cyber defence strategies, we look at several published definitions of active cyber defence. When comparing definitions of active cyber defence from different sources, we find some similarities and, surprisingly, some contradictions in definitions by several bodies within the US government. This paper offers both a review of published active cyber defence strategies and some novel active strategies. We have identified several active cyber defence strategies that are easy to implement and seem promising in countering some popular low-end cyber crime. As a proof of concept, we have implemented an active strategy against phishing sites and successfully tested it against two such sites.

In section 2, we list several published definitions of active cyber defence, section 3 is devoted to advance fee fraud, section 4 takes a closer look at email spam, section 5 inspects the phenomenon of phishing, section 6 deals with information stealing botnets, section 7 covers legal aspects, section 8 describes practical considerations and authors' experiments.

# 2. CONCEPT OF ACTIVE CYBER DEFENCE

Before proceeding any further, we should briefly review the concept of 'active cyber defence' as defined by other authors. A brief literature study revealed some contradictions in the definitions published so far.

The 2011 US Department of Defense (DoD) Strategy for Operations in Cyberspace (DoD, 2011) stresses the real-time property of active cyber defence:

> 'Active cyber defense is DoD's synchronized, real-time capability to discover, detect, analyze, and mitigate threats and vulnerabilities. It builds on traditional approaches to defending DoD networks and systems, supplementing best practices with new operating concepts. It operates at network speed by using sensors, software, and intelligence to detect and stop malicious activity before it can affect DoD networks and systems. As intrusions may not always be stopped at the network boundary, DoD will continue to operate and improve upon its advanced sensors to detect, discover, map, and mitigate malicious activity on DoD networks.'

The US Defense Advanced Research Projects Agency stresses the absence of cyber offensive capabilities when describing their Active Cyber Defense program (DARPA, 2012):

> 'These new proactive capabilities would enable cyber defenders to more readily disrupt and neutralize cyber attacks as they happen. These capabilities would be solely defensive in nature; the ACD program specifically excludes research into cyber offense capabilities.'

The US Department of Defense Dictionary of Military and Associated Terms (DoD, 2010) unfortunately does not provide an explicit definition for active cyber defence, but it provides separate definitions for "active defense" and "cyberspace operations".

It defines 'active defense' as:

> 'The employment of limited offensive action and counterattacks to deny a contested area or position to the enemy.' (1)

It defines 'cyberspace operations' as

> 'The employment of cyberspace capabilities where the primary purpose is to achieve objectives in or through cyberspace.' (2)

For this paper we adopt the following definition, which can be derived by combining definitions (1) and (2):

"Employment of limited offensive **cyberspace capabilities** to deny a contested area or position to the enemy, **in or through cyberspace**."

## 3. ADVANCE FEE FRAUD

One of the most popular cyber crimes (or attempts at) that almost every netizen has experienced first-hand is advanced fee fraud. The victim is tricked into trusting a cyber criminal and sends some advance fee in prospect of receiving huge rewards. A very popular variation of this is

the so called Nigerian scam, when a message (usually email) is sent stating the victim has won a lottery or got an inheritance, or is asked for help to transfer money. It is also known as a '419 scam', referring to the article 419 of Nigerian Criminal Code dealing with fraud. Advance fee fraud is not a cyber-only phenomenon; in-depth research on advance fee fraud has been published, and for more thorough background and historic information see Smith *et al.* (1999). While advance fee fraud is not restricted to cyberspace, cyber offers the cheapest way of communicating with potential victims, so it is a crime that uses and abuses cyberspace.

Let's look at the mechanics of this scam:

1. Criminals use spam distribution channels to send out emails containing the 'hook' and asking victims to respond via email;
2. Some of the recipients respond;
3. In communication back and forth between criminals and victim, the victim is asked to pay an advance fee to enable the reception of a large monetary reward. The victim is asked to transfer the advance fee via an untraceable money transfer service, e.g. Western Union; and
4. The victim transfers advance fee, and the money is cashed out.

There are several possible passive strategies to fight advance fee fraud:

1. Stop spam distribution channels and improve spam filtering - this type of fraud requires cheap bulk distribution of scam messages, because only a few people are light-minded enough to fall for such scams. This strategy breaks the scam in stage 1, and is elaborated on further in the section dealing with email spam;
2. Stop anonymous and untraceable money transfer services. This would attack the scam scheme at stage 4 described above.

Both stopping anonymous or untraceable money transfer services and stopping spam distribution seem like difficult problems. There might be some other passive strategies the authors are not aware of, but to the best of our knowledge, passive strategies do not present an acceptable solution to the problem. That is why we move our attention to active strategies, attacking the scam as it progresses through stages 2 and 3.

We assume the following qualitative cost-estimate for operating advance-fee fraud scheme:

- sending out spam in huge quantities - cheap
- email discussion with potential victims - medium
- money transfer and cash-out - expensive

Sending spam is cheap; one could even say it is virtually free, as we will discuss in the email spam section below. Carrying out a discussion with potential victims over email and phone is more expensive, requiring manual labour and some proficiency in the target language. Cashing out is probably even more expensive because of the limited number of cash-outs a person can

do in a given amount of time. It seems reasonable to focus active strategies against the more expensive stages of the scam to maximize the damage inflicted on the scam operators.

We envision the following active strategies:

- Attack the email discussions between scammers and victim by identifying scam mail and using natural language processing algorithms to carry out conversations with the scammers. If done on a large scale, this dramatically increases the costs to scammers, as before they had 100% genuine human response that allowed for manual email conversation with victims. This ratio can easily be reduced almost to zero, forcing scammers to develop advanced mechanisms to identify genuine humans from computer generated responses, basically forcing them to solve the spam problem, which seems to be hard. This idea has been discussed before in some web forums (Halfbakery, 2004), but no references in an academic discussion could be found. A possible solution would be to have a 'scam button' in the email client or webmail similar to the spam button that would forward the email to a fully automated system for carrying on a conversation with the scammers. This does not require an AI algorithm to pass the Turing test. Withstanding a few rounds of conversations would be sufficient to substantially increase the costs for scam operators. Existing natural language processing algorithms would be sufficient for this task, an interesting research would be to use the famous ELIZA algorithm from 1966 (Weizenbaum, 1966) for such a purpose.
- Attack the cash-out. This needs active collaboration with the money transfer provider and would only work if cash-out is carried out by means of centralised money transfer provider such as Western Union, not with decentralised means like Bitcoins or similar. As a prerequisite, the money transfer service provider must be willing to cooperate with law enforcement. When requests for money transfers are received, those could be forwarded to the money transfer provider, which could generate marked transfer numbers. This number needs to be forwarded to the scammers. When cashing-out, the person walks up to the counter and presents the transfer number, the payment system can display a warning and he or she could be arrested by the police force if legislation supports it.

Attacking cash-out would work only for a limited number of cash-out schemes. Recently a popular scareware in Latvia asked victims to pay by purchasing PaySafeCard prepaid vouchers and sending the codes printed on the vouchers to the scammers (CERTLV, 2013). Scammers could use the code received to purchase easily resalable goods such as iTunes gift cards or electronics in an online shop. In such cases there is no physical cash-out vulnerable to attack.

# 4. EMAIL SPAM

Abuse of electronic messaging systems to send unsolicited bulk messages is a daily and annoying occurrence for any netizen. Stopping spam is a hard task, and industrial-grade spamming has

been a problem for almost 20 years (Cranor, 1998) and 69% of email traffic in 2012 was spam (Symantec, 2013). Although the spam ratio in global email has slightly decreased over the last few years, we cannot consider spam fighting a success story. We expect that for the foreseeable future the costs of sending spam will remain negligible, a report by Trend Micro estimates the black market price for spam distribution at $10 for a million spam emails (Goncharov, 2012).

Although spam filtering is constantly improving, some spam is always likely to get to recipients' mailboxes. The regular email user is not aware of the amount of spam filtered out by the ISPs and email providers, and the costs associated with developing and running the spam filtering software.

We would like to focus on unsolicited commercial email, i.e. email with commercial content that is sent to a recipient who has not requested it (Hoffmann, 1997) which, for example, may advertise an online shop selling fake pharmaceuticals. When taking a look at the economics of unsolicited commercial email operation, we could identify several stages:

1. Bulk email sent out;
2. Spam delivered to inbox;
3. A few users fall for the advertised product or service and purchase through the advertised website.

To maximize sales in stage 3, spammers must send out as much as possible in stage 1 (not taking into account advanced dynamic spam filtering). Taking into account the small percentage of users that fall for the advertised goods or services (less than 0.00001% according to Kanich (2008)), the amount of emails sent to become a reasonably profitable operation must be huge.

To influence stage 1, the number of infected hosts on the internet needs to be reduced – that seems to be a hard task. A lot of effort is spent in spam filtering to influence step 2 and have less spam delivered to the email inboxes. It is also possible to attack the spamming operation at step 3 with at least the following methods:

- Blacklist advertised websites;
- Community DDoS the advertised websites - if instead of deleting the spam emails at ISP/email provider/user level, requests could be generated to the advertised websites; this would create some costs that would grow proportionately to the amount of spam sent out. Currently, when email is categorized as spam, spammers do not receive any penalty, this could change just by modifying the functionality behind the spam button in your email environment, be it email client or web mail. This approach does not involve automatic detection of spam – there are lots of research towards this goal, and it seems a difficult problem. We rely on users as the final spam filter, and merely suggest that the 'spam' button in the email client would have more advanced functionality than simply deleting the email.

Spammers could probably evade blacklisting by frequently changing domain names for their web shops or even using disposable domain names uniquely created for each spam message.

Similar ideas have been implemented in production by a commercial company named Blue Security (PCWorld, 2005) but abandoned for various reasons. Such actions were quickly labelled as internet vigilantism. Taking into account that the current lack of law enforcement on the internet resembles that of the Wild West, such a term might be used without negative connotations.

# 5. PHISHING

Phishing is aiming at getting users to disclose sensitive information such as passwords, financial account information, or social security numbers (Ramzan, 2010). Let us look at phishing via email as a prominent example. A user might be asked to disclose some sensitive information in reply to an email, and enter it in a malicious website, or some other way. An interesting phenomenon are malicious websites which are not advertised in phishing emails but use typo squatting, i.e. use domain names which are very similar to the legitimate site, and attract users who make a typing mistake when entering the address of the legitimate website.

When phishers gather credentials and other sensitive information these can be, among others, sold on the black market, or used to access some services to steal valuables (money, stocks, in-game items etc.). The phishers must provide a way for phished users to submit the sensitive information, and this interface looks like an attractive target to attack.

We see two distinctive scenarios for such an attack:

- Flood phishing interface with fake data;
- Submit credentials corresponding to monitored/sandboxed accounts.

## *Flood the interface with fake data*

We assume that under typical modus operandi, the 'phished' data is of very high quality. Only genuine users enter their login credentials, social security numbers, credit card details etc., so the only source of bad data is users – either typos or memory error (e.g. not remembering the correct password). Since phishing interfaces are publicly accessible, it is easy to attack them by submitting lots of fake data. The phishers are now faced with a large volume of low-quality data; this data needs to be checked, which might involve some costs depending on the type of data collected: for example, a username/password might be automatically checked for free just by logging in the legitimate service (it could be checked during the phishing phase performing some sort of man-in-the-middle scenario), while checking passport numbers might not be free.

If checking phished data involves some cost, it is possible to submit fake data to reach a threshold when phishing stops being profitable. Irrespective of checking cost, generation of fake data could be brought to such high level that either network, CPU or storage resources could be overwhelmed, and the phishing site would become inaccessible. Such idea is also proposed by Shah *et al.* (2009). As a likely response to the proposed strategy, phishing operators could introduce captcha mechanisms, just like the way operators of legitimate sites are fighting bots nowadays.

### *Submit credentials corresponding to monitored/sandboxed accounts*

Credentials corresponding to monitored/sandboxed accounts could be submitted to the phishing interface. Once criminals use those accounts, they can be automatically tracked and some further information might be extracted, depending on the case. For example. when dealing with banking credential theft, monitored bank accounts could be used to find out ways to transfer the stolen funds. The monitored/sandboxed account would have the same look and feel as a genuine internet bank account to trick phishers in trying to transfer the funds to their associates. A popular way to transfer money from a compromised account is to use a money mule, a person who wittingly or unwittingly forwards the received money to the phisher's account in a way that is not transparent to law enforcement. We assume money mules are an expensive resource for cyber crime to acquire (possible mules need to be attracted, recruited, some cover story for the company employing the mules needs to be maintained etc.), and effectively disclosing the identities of money mules and reporting to law enforcement would be a major setback for phishers. The monitored/sandboxed accounts should be designed to withstand scrutiny by the attackers, the account should have a legitimate-looking transaction history, test transactions should be at least simulated (e.g. attackers could send or receive small amounts of money in the account as a test, before proceeding to cash-out).

Serial numbers, IMEI, MAC and other unique numbers for goods purchased could also be recorded by sellers or forwarded to banks. Once banks detect fraudulent payments, lists of unique numbers identifying stolen goods could be produced. Such lists could be used by law enforcement when inspecting grey markets. The internet could also be searched to locate the devices to better understand the geography of cyber crime.

## 6. INFORMATION STEALING BOTNETS

When facing botnets that search infected machines for information such as login credentials, credit card numbers etc., again a fake information submission strategy could be used. Usually drones gather the information and upload it to a dropzone either automatically or when instructed via a command & control (C&C) channel. Depending on the specifics of the C&C protocol the botnet uses, bot herders could find it impossible to distinguish between genuine information and fake information uploaded using the same bot id (this assumes the bot C&C/upload to the dropzone protocol is reverse-engineered).

If uploads are digitally signed, private keys need to be extracted from infected machines, making it much more complicated. So far the authors have not found sources stating that botnets use public key cryptography to sign uploaded data, but it is safe to assume that bot herders would implement it if such methods gained popularity.

If encountered with an unknown C&C protocol or digitally signed uploads, one voluntarily gets infected in a controlled environment. It is possible to run the malware in a sandbox/virtualised environment, supply the bot with fake data and use the original bot code to upload the fake data to the dropzone. If a large pool of diverse IP addresses is available, enough fake bots could join

a botnet and spoil the gathered data. Bot herders would need to check the data to distinguish genuine data from fake, inflicting costs on their operations.

Botnets in general have been a popular area to apply active strategies. The Conficker botnet was such a challenge that security professionals organised a Conficker Working Group to coordinate the takedown attempts (CWG, 2010), but it can be disputed whether it was a low-cost endeavour, taking into account the number of expert man-hours spent on this task. Some estimate of the economic dimension of this is the $250,000 bounty (not awarded so far) that Microsoft announced for information on persons behind the Conficker botnet (Microsoft, 2009). Some successful botnet takedown operations have had a major impact on global amount of spam, like those of Waledac and Rustock takedowns: this topic is elaborated by, among others, Dittrich (2012) and Czosseck et al. (2011).

# 7. LEGAL CONSIDERATIONS

This paper identifies several main issues from a legal perspective worth considering. The authors analyse the Latvian law and regulations that could be applied in proposed active response strategies. The objective is to offer a framework in which the discussions on proposed preventive measures could be evolved further.

Active response strategies can be initiated and performed by public authorities, industry, or private individuals. All of them are subject to national law and should act in the framework of national regulations. Public authorities will act only if the attack meets certain criteria set forth in law and only in accordance to procedures defined by the law. These procedures in some cases make the process slower than it is needed to prevent or even investigate the cyber threats. Public authorities can act and apply either the public and administrative or criminal law if they are notified by the victim.

## *Legal capacity of public authorities*

The scope of administrative law in Latvia covers administrative violations which are acknowledged as unlawful actions or inactions which must be committed with intent or negligence and must endanger state or public order, property, the rights and freedoms of citizens, or management procedures specified in the law.[1] However the law does not issue regulation and does not provide liability regarding violations of computer systems or computer data. Although there is liability for unfair commercial practice, unsolicited distribution of an advertisement or commercial information, which in certain situations could be applied to spam, the limits of jurisdiction restrict the regulation to territory of Latvia and international agreements. Such cyber threats in most cases will be multijurisdictional in their nature which would hinder cooperation and effect of the law in the field of administrative violations. E.g., a citizen of State A sends spam letters. State A does not qualify spamming as violation. Recipients in six other states receive the spam letters and all of these states have different regulations regarding spam. State B regulates unsolicited e-mails, but to qualify it as violation, a certain threshold of damage must be met. The victims from State B, that have suffered damage, can make a claim to the public authorities, but in order to meet the threshold of damage the authorities need more

1 The Administrative Violations Code 1984 (Latvia), s 1

claims. As states do not have obligations to cooperate in these cases the state can not gather enough cases to act and even if it does, there is still State A that does not qualify spamming as violation and does not have internationally binding obligation to act.

Another part of public law, criminal law, does have the regulation aimed at the protection of society against cybercrime, but the criteria which must be met in order to qualify an act as the criminal offence demand the establishment of all the constituent elements of an offence set out in the law.[2] In most cases the harm done to a single individual may be comparatively small and will not qualify as criminal offence, although the nature and harm of the threat to the interests of a person or to society is substantial and, if gathered, can be subject to criminal law. Cyber threats are aimed at victims without considering the factor of territory of the state, so the limits of jurisdiction apply to harm as well. This substantially hampers the ability to identify enough victims in order to apply the criminal law.

In some cases public authorities may use our proposed measures in the investigative process to identify and to stop the source from going after other victims and causing greater damage. These actions could be used as preventive measures in order to face potentially harmful behaviour.

## *What can private individuals and legal persons do*

The lack of regulation of cyber threats which by their nature are less harmful then criminal offences and cannot be dealt within the scope of administrative law, provide a favourable setting for the victims of those acts to seek out different defence techniques which they can employ themselves.

We should note that an active response could not only cause positive results and lessen the crime, but can also cause undesirable effects. The need to defend the network may occur when something worse comes back as revenge against the actions taken. Collateral damage may occur in the process of active response (e.g. if the spam letters are sent by competitor in the name of a company which actually is not an initiator of these unsolicited e-mails) or provoked attacks (e.g. if we use active response strategies to stop unsolicited messages we can expect that systems of our ISP can suffer from DDoS attack as well). Orin S. Kerr (2005) draws attention to several such undesirable outcomes.

There is no doubt that person can and even must reduce all possible risks in order to avoid the threats and attacks. But when the threat or the attack occurs and the harm is done, the victim has a choice either to notify state authorities or act on his own. In both cases law and regulations apply.

A former official at the National Security Agency and the Department of Homeland Security, Stewart Baker (2012), suggested that within the national legal system applicable law can have ambiguous wording and as a result the victim can argue his rights not only to protect himself, but actively engage in defence against the threat '… to conduct at least limited surveillance of a machine that is, after all, directly involved in a violation of the victim's rights'. Fred C. Stevenson Research Professor of Law Orin S. Kerr points out several undesirable effects this conclusion can cause 'As long as someone believes that they were a victim of a computer

2 The Criminal Law 1998 (Latvia), s 1 (1)

intrusion and has a good-faith belief that they can help figure out who did this or minimize the loss of the intrusion by hacking back, the hacking back is authorized' (Kerr, 2012a). This discussion points out the grey area of regulation where the victim takes risk to violate the rights of attacker. The same arguments would apply to proposed active defence methods if the use of false data would cause violation of attacker's or third party rights.

Proposed measures involve use of false information. Latvian law and regulations establishes certain conducts where provision of false information is deemed to be illegal. Firstly, if person has an obligation to provide certain information to public authorities, then provision of false information is a breach of the law. As the attacker does not represent the public authority and there is no regulation under which there is obligation to provide data to attacker, this regulation cannot be applied to the proposed situation.

The legal provisions of most criminal offences that involve the use of false information are specific to the obligation under the law to provide information. Regulation of computer systems related criminal offences on the other hand defines computer fraud as an action taken knowingly by entering false data into a computer system for the acquisition of the property of another person or the rights to such property, or the acquisition of other material benefits, in order to influence the operation of the resources.[3] Active response by which false information is provided to source of threat falls under the scope of this section. However the active response strategy does not acquire any property or the rights to such property, or other material benefits. The motive should be taken into account as well – the active response eliminates the threat. As a result active response cannot be deemed a criminal offence according to this section unless the missing elements occur.

The second well-represented view which confers on the victims a right of active defence refers to affirmative defence. It is important to note, that in this legal concept the self-defence does not deny the fact of offence - the criminal act is done, but it asserts a defence of the offender that would negate the legal effect of the offence. In the authors' view the reason why the self-defence is the last and least preferred course of reasoning by legal practitioners is the difference in the consequences of criminal procedures; in the self-defence case the offender has to admit that he committed the offence and after that has burden of proof of circumstances which exclude his criminal liability, but if the self-defence argument is not used then the offender just denies all allegations and the fact of the offence, so the burden of proof solely lies on the law enforcement.

Eugene Volokh, Gary T. Schwartz Professor of Law at UCLA, points out some common reasons why digital self-defence should be viewed without negative connotations. He writes that generally speaking the use of force is allowed '… the law has never treated defense of property as improper "vigilantism"', he continues '… the right to defend yourself and your property (subject to certain limits). By using this right, you aren't taking the law into your own hands. You're using the law that has always been in your hands' (Volokh, 2012). The opposing view is represented by Orin S. Kerr (2012b) and draws attention to the lack of precedent regarding 'cyber self-defence' as the law, at least in the US, does not have clear wording or case law that interpret the rights to defend property to be applicable to cyber defence.

3 Criminal law 1998 (Latvia), s 177.1 (1)

The criminal law of Latvia establishes several circumstances which exclude criminal liability. Self-defence is one of those admissible conditions. The criminal law provides:

> Necessary self-defence is an act which is committed in defence of the interests of the State or the public, or the rights of oneself or another person, as well as in defence of a person against assault, or threats of assault, in such a manner that harm is caused to the assailant. Criminal liability for this act applies if the limits of necessary self-defence have been exceeded.[4]

Private individuals and legal persons may use necessary means to defend their interests and rights, and in some instances the interests and rights of others, but they must act so under specific circumstances allowed by law. Necessary self-defence can be used as a defence providing that:

- The threat itself is illegal;
- The threat is actual and has already occurred;
- Actions to protect the property can be taken only to protect lawful rights and interests;
- Actions taken are proportional to the nature and the danger of threat (using reasonable force); and
- Only the source of threat suffers damage.[5]

It is important to note, that a threat must be on going and this requirement rules out any claim to use active response to prevent a potential threat. When the threat has been averted, there are no further grounds for self-defence. This shows the limits of the active response. The active response can be taken only for the time period while the threat is occurring.

Thus, it is useful to know the extent to which active response is reasonable. The Supreme Court of the Republic of Latvia explains 'defense disproportionate to the nature and the danger of the threat must be recognized as evident, if objectively there was no need for use of such means and methods to avert the threat'.[6] The question is if such cyber threats as Nigerian letters, spam, phishing, and information collection botnets is ignored by their potential victims, then is there still a need for active self-defence?

The last criterion establishes the amount of damage that can be incurred. If interpreted literally, the criterion implies that necessary self-defence is not a passive defence limited to deletion of undesirable content, but rather actions taken to cause damage to the attacker. So if a potential victim ignores the cyber threat the actions of the victim cannot be considered as necessary self-defence for several reasons: the victim's actions are legal, the actions taken are passive and the attacker has not suffered any damage. If the victim uses active defence measures, then it is important not to exceed the limits of necessary self-defence.

The proposed active defence mechanisms as a well-weighed instrument can be used by public authorities, CERTs or technicians within the scope of law. The authors have not found any court decisions on this issue, but by evaluating legal regulations it is possible to conclude that in

4 Criminal law 1998 (Latvia), s 29 (1)

5 Uldis Krastiņš, Valentīna Liholaja, Aivars Niedre, Kriminallikuma komentāri 1.grāmata Vispārīgā daļa, (1999 Riga AFS) 125

6 Case-law in application of necessary self-defence [1995] Plenary Session of the Supreme Court of Republic of Latvia 3, 1996 Latvijas Republikas Augstākās tiesas plēnuma lēmumu krājums 1990-1995

certain situations the activities may become illegal and there is a risk that the limits of necessary self-defence could be exceeded.

# 8. PRACTICAL CONSIDERATIONS AND PROOF OF THE CONCEPT

The authors have chosen the strategy of feeding phishing pages with fake information to implement as a proof-of-concept to test the proposed strategy in real life and demonstrate feasibility of such an approach.

The amount of fake information should be substantial in comparison with genuine phished traffic. The volume should exceed the genuine data by several times. If the aim is to overwhelm the resources of phishing site, there is no upper limit for fake data. In the case when a phishing site is suspected to be hosted on a compromised server also providing legitimate services, the fake data stream needs to be limited in order to prevent the server from crashing. A further in-depth study should be done to determine the trends of malware/phishing site hosting: whether the bad guys still use hacked servers for hosting their services, or move to use more reliable dedicated Virtual Private Servers, possibly in bulletproof hosting companies.

We have implemented the strategy as a set of python scripts and tested it on two phishing websites, both of which closed down in a reasonably short timeframe after the feed started. This is by no way a scientific proof of universal effectiveness of the proposed strategy though.

There are several points that need to be considered in order to make the generated data difficult to filter from the genuine phished data:

## *Content*

The generated content needs to look authentic and legitimate, and that depends on the target of the phishing operation. Usernames and passwords should follow the same pattern as those of the target – usually not fully random but contain the names and surnames of the targeted country (if applicable). Passwords could be taken from popular password lists or generated from dictionaries of the target language. Credential data leaked to the public in some previous incidents could be reused in modified or unmodified form, because it does not add harm, or some algorithm to generate credible usernames and passwords could be devised. We gathered usernames, emails and passwords from public websites hosting leaked credential data. When generating ‘fake’ data, usernames and passwords were picked randomly from leaked lists of different incidents so the original leaked username and password pair was not reused. Domain names for email addresses might be changed to adjust to the targets of the phishing campaign (e.g. phishing campaign with Italian targets would not expect too many mail.ru email accounts). Re-use of leaked credential data offers a way to generate fake data that is difficult to filter out by the phishing site operators, contrary to randomly generated strings.

## *Metadata*

Metadata such as user-agent strings in http/https connection headers, time, time zone, and

counters in various protocol headers/footers should not be static or enable effective filtering in any other way. Our implementation randomly picks user-agent fields from pre-defined list.

### *Infrastructure and other considerations*

Feeding the phishing sites should be done from a large pool of source IP addresses from various autonomous systems and randomized in time. Use of IP addresses should make sense; if a Latvian bank website is being phished for, most source IPs should be in the Latvian IP space but some should come from abroad, otherwise attackers can filter away Latvian source IPs to obtain genuine data, although much less in volume. In our implementation we used proxy services to randomize the source IPs. Submission of fake data should be carried out for long periods of time rather than have peaks of activity; it should be randomized in time rather than predictable. If fake data is sent in peaks or scheduled at regular intervals (e.g. first minute of every hour), phishers will have an easy time filtering out the generated submissions.

## 9. CONCLUSIONS

The authors have reviewed a range of active strategies that deal with some popular sorts of low-end cyber crime. All the proposed strategies can be promptly implemented with the available technical know-how and infrastructure, without need of any R&D investments. We have succeeded in providing several low-cost active cyber defence strategies, contrary to the popular belief that active cyber defence is limited to huge budget projects. As proof of concept, we implemented one of the described active strategies and most likely made the internet just a little bit better place by closing two phishing sites. We can assume that spammers, scammers and other evildoers will adapt their modus operandi once active cyber defence measures will start to exert noticeable pressure. This means the security community must constantly innovate to keep in touch in such arms-race like circumstances.

## ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their scrutiny and valuable comments, enabling us to substantially improve the paper.

## REFERENCES

Baker, Stewart , 2012, ‘RATs and Poison Part II: The Legal Case for Counterhacking’ (The Hackback Debate, November 2, 2012) http://www.steptoecyberblog.com/2012/11/02/the-hackback-debate (accessed 14 February 2014)

CERT.LV, 2013, CERT.LV brīdina par ‘Policijas’ izspiedējvīrusa izplatību., https://cert.lv/resource/show/251 (accessed 12 February 2014)

Cranor, Lorrie Faith, and Brian A. LaMacchia, 1998, ‘Spam!.’ *Communications of the ACM* 41.8 (1998): 74-83.

Conficker Working Group, 2010, ‘Conficker Working Group: Lessons Learned’, http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf, (accessed 4 September 2014)

Christian Czosseck, Gabriel Klein, Felix Leder, 2011, On the Arms Race Around Botnets – Setting Up and Taking Down Botnets, in Proceedings of the 3rd International Conference on Cyber Conflict, Tallinn, Estonia 7-10 June 2011

DARPA, 2012, Active Cyber Defense (ACD), http://www.darpa.mil/Our_Work/I2O/Programs/Active_Cyber_Defense_%28ACD%29.aspx, (accessed 13 December 2013)

Dittrich, David, 2012, 'So you want to take over a botnet.' *Proceedings of the 5th USENIX conference on Large-Scale Exploits and Emergent Threats*. USENIX Association, 2012.

US Department of Defense, 2010, US Department of Defense Dictionary of Military and Associated Terms, http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf, (accessed 12 February 2014)

US Department of Defense, 2011, US Department of Defense, Strategy for Operations in Cyberspace *www.defense.gov/news/d20110714cyber.pdf (accessed 12 February 2014)*

Goncharov, Max, 2012, 'Russian underground 101.' *Trend Micro Incorporated Research Paper* (2012).

Halfbakery, 2004, Advance fee fraud (419) reply-bot, http://www.halfbakery.com/idea/Advance_20fee_20fraud_20(419)_20reply-bot (accessed 13 December 2013)

Paul Hoffmann, 1997, Unsolicited Commercial Email: Definitions and Problems, Internet Mail Consortium report, http://www.imc.org/imcr-002.html (accessed 13 December 2013)

Kanich, Chris, et al., 2008, 'Spamalytics: An empirical analysis of spam marketing conversion.' *Proceedings of the 15th ACM conference on Computer and communications security*. ACM, 2008.

Kerr, Orin S., 2005, 'Virtual Crime, Virtual Deterrence: A Skeptical View of Self-Help, Architecture, and Civil Liability'. *Journal of Law, Economics & Policy*, Vol. 1, January http://papers.ssrn.com/sol3/papers.cfm?abstract_id=605964 (accessed 24 February 2014).

Kerr, Orin S., 2012a, 'The Legal Case Against Hack-Back: A Response to Stewart Baker' (The Hackback Debate, November 2, 2012) http://www.steptoecyberblog.com/2012/11/02/the-hackback-debate (accessed 14 February 2014).

Kerr, Orin S. 2012b 'A Response to Eugene Volokh' (The Hackback Debate, November 2, 2012) http://www.steptoecyberblog.com/2012/11/02/the-hackback-debate (accessed 14 February 2014).

Kreibich, Christian, *et al*., 2008 'On the Spam Campaign Trail.' LEET 8 (2008): 1-9.

Langner, Ralph, 2010, 'The short path from cyber missiles to dirty digital bombs', http://www.langner.com/en/2010/12/26/the-short-path-from-cyber-missiles-to-dirty-digital-bombs/ (accessed on 12 February 2014).

Microsoft, 2009, Microsoft Collaborates With Industry to Disrupt Conficker Worm, http://www.microsoft.com/en-us/news/press/2009/feb09/02-12confickerpr.aspx, (accessed 12 February 2014).

PCWorld, 2005, Spam Slayer: Bringing Spammers to Their Knees, PCWorld, http://www.pcworld.com/article/121841/article.html, (accessed 13 December 2013).

Peel, Michael, 2006, *Nigeria-related financial crime and its links with Britain*. London: Chatham House

Ramzan, Zulfikar, 2010, 'Phishing Attacks and Countermeasures.' *Handbook of Information and Communication Security*. Springer Berlin Heidelberg, 433-448.

Rao, Justin M., and David H. Reiley, 2012, 'The economics of spam.' *The Journal of Economic Perspectives*, 87-110.

Shah, Ripan, et al., 2009, 'A proactive approach to preventing phishing attacks using Pshark.' *Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on*. IEEE

Smith, Russell G., Michael N. Holmes, and Philip Kaufmann, 1999, '*Nigerian advance fee fraud.*', in Trends & Issues in Crime and Criminal Justice, July 1999.

Symantec, 2013, Symantec, Internet Security Threat Report 2013, http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v18_2012_21291018.en-us.pdf, (accessed 13 December 2013).

Volokh,, Eugene, 2012, 'The Rhetoric of Opposition to Self-Help' (The Hackback Debate, November 2, 2012), http://www.steptoecyberblog.com/2012/11/02/the-hackback-debate (accessed 14 February 2014).

Weizenbaum, Joseph, 1966, 'ELIZA—a computer program for the study of natural language communication between man and machine.' *Communications of the ACM* 9.1 (1966): 36-45.

Wood, Bradley J., Saydjari, O. Sami and Stavridou Victoria, 2000, 'A proactive holistic approach to strategic cyber defense.' *SRI International. http://www. cyberdefenseagency. com/publications/Cyberwar_Strategy_and_Tacti cs. pdf* (accessed 12 February 2014).