Noname manuscript No.
(will be inserted by the editor)

Ethical Awareness in Security Operations

Nils Ulltveit-Moe

the date of receipt and acceptance should be inserted later

Increased ethical awareness and improved guidelines, methodologies and tools are needed for handling private or confidential information in security operations. The current situation is that security operations in some areas may be too privacy invasive. At the same time, investigation of transnational on-line crime, is impeded both by a proliferation of attacks, lack of electronic evidence and legal hindrances across national borders. This means that new strategies are needed for attack detection that facilitates increased cooperation between organisations on attack detection, at the same time as private and confidential information must be respected. This article discusses how personally identifiable or confidential information should be managed from a legal and ethical perspective. Organisations performing monitoring of computer networks for signs of attacks can for example be expected to benefit from using quantifiable privacy and security metrics as part of a service level agreement. It is furthermore analysed whether automatic blocking of malicious traffic is better than surveillance of such activities. Last, the paper discusses how incentive compatible contractual means can be used to reduce the Moral Hazard both from a privacy and security perspective for outsourced managed security services.

Keywords Security, Privacy, Outsourcing, Managed Security Services, Ethical Awareness

1 Introduction

Globalised organised cyber-crime is a serious and rapidly increasing problem in today's society. It can ultimately threaten the IT-infrastructure of countries (Ashmore, W. C. et al., 2009). In addition, Advanced Persistent Threats (APTs), like the Stuxnet worm or Duku targeting critical infrastructures are being developed by governmental agencies as part of a cyber warfare strategy (CrySys lab, 2011). The conviction rates are low for cyber crime, because it is hard to get evidence that can be traced back to the offender.

Nils Ulltveit-Moe
University of Agder,
Jon Lilletuns vei 9
4879 Grimstad, Norway

Another reason is that it is difficult and expensive to investigate crime involving several countries with different legislation (Kshetri, 2006).

Various attack detection techniques, like Intrusion Detection Systems (IDS), spam-filters or AntiVirus are being used to detect, investigate and prevent cyber-crime both in the private and public sector.

It is legal to perform monitoring of computer networks and hosts using potentially invasive technologies like IDS in most European countries as long as the *purpose* with the monitoring is to detect cyber-attacks. There are for example explicit exceptions for measures related to detecting cyber-attacks in the EC communications directive (European Commission, 2002). This may nevertheless be problematic from a privacy or confidentiality perspective, because the *effect* of such monitoring is largely unknown.

I have had discussions with practitioners in this area, and one of the ethical dilemmas they sometimes have is that the attack detection technologies quite often trigger alarms on more than just real attacks. This is problematic from a privacy and confidentiality perspective, because this effectively is a leakage of private or confidential information that goes beyond the initial purpose the data were collected for - to detect cyber-attacks. An example scenario is when side-information from such monitoring activities by chance detects illegal or criminal activities that are not related to the core purpose of the monitoring technologies. This may put the MSS providers in ethical dilemmas on how to handle this side-information, and from a law-enforcement perspective, this may not even be considered legal evidence, because the monitoring technologies have detected activities beyond the intended purpose of identifying cyber-attacks.

A problematic area that has been identified during our research, is excessively broad IDS rules that effectively monitor usage instead of cyber-attacks [anonymised ref.]. Another similar area is IDS rules for identifying web bugs that may be a risk for privacy and data confidentiality. In the latter case, the good intention of security monitoring may be its own worst enemy, because detecting privacy leaking web plugins causes a significant privacy leakage itself. This can however, at least in principle, be handled by anonymising or pseudonymising sensitive information sent in the IDS alarms or other events, given that you know *what* to protect.

For IDS rules detecting web bugs, it is for example only interesting from a security perspective to detect the presence of such potentially malicious browser plugins. It is not interesting to view the privacy-leaking payload, addresses etc. that these plugins cause. This means that an anonymisation or pseudonymisation framework in such cases can be used to significantly reduce the leakage of private or sensitive information from this type of IDS alarms without reducing the utility from a security perspective.

In addition, the technologies used for network monitoring will typically create a significant amount of false alarms from normal benign traffic (e.g. Alharby and Imai, 2005). All these examples indicates that more private and sensitive information may leak out to semi-trusted parties like outsourced MSS providers during security operations than strictly necessary. There is furthermore a lack of technologies and techniques that can mitigate this problem in an efficient manner, without compromising attack detection efficiency.

2 What Is Privacy?

Privacy is a broad concept that can have different meaning in different contexts. Warren and Brandeis early on defined privacy from a legal perspective as *the right to be let alone (Warren and Brandeis, 1890)*. Other definitions focus more on privacy as an intellectual property from a utility perspective, where the data owner should be ensured *self determinism* about private data (Samuelson, 2000). This amongst others means that the data owner must be able to give and revoke consent to access private data (Cavoukian, 2009a). This has for example lead to actuarial models that aim at estimating the perceived cost of privacy leakage for insurance contracts (Gritzalis et al, 2007).

Information theorists provide a more technical definition of privacy and often define privacy as equivocation (level of ambiguity). This definition is based on the observation that some level of privacy and anonymity can be ensured by requiring that private information is hidden in a sufficiently large crowd of other information, so that the data owner cannot easily be identified. Equivocational privacy metrics are for example based on entropy metrics (Sankar et al, 2010; Clauß and Schiffner, 2006), or they directly specify a level of equivocation like k-anonymity (Samarati, 2001; Sweeney, 2002; Ciriani et al, 2007) or l-diversity, which in addition to level of equivocation considers the diversity of the data (Machanavajjhala et al, 2007).

From an information theoretic viewpoint, privacy leakage can be modelled based on the assumption that the utility of the data is inversely proportional to the level of perturbation of the data and the level of privacy can be quantified as the level of equivocation (Sankar et al, 2010).

3 How Can Privacy be Improved?

One way to illustrate the privacy against security dilemma is airport security. Most people are willing to trade convenience and privacy for some added security. It is therefore accepted in our society that all passengers undergo privacy-invasive security control checks when traveling by airplanes to increase the perceived safety. The privacy-invasive security controls aim at reducing the possibility that adversaries, like terrorists or psychologically unstable persons, bring weapons, explosives or other dangerous items on board the airplane.

There has been quite extensive research on more efficient ways to detect hidden weapons on people. One efficient technology, that recently has been deployed, is backscatter X-ray scanners (Cavoukian, 2009b). These scanners expose the person to be checked with small amounts of X-ray radiation, and uses the backscatter X-rays to produce photo-quality images that can see through clothes. This technology is used as an alternative to personal searches, since it easily can reveal hidden weapons. If a suspicious item is detected, then the security officer will perform a manual search to verify what the suspicious item is.

This technology causes a privacy concern, since it essentially shows a naked picture of the person being scanned. Privacy enhancing technologies have been suggested to deal with this problem. The techniques include using blurred pictures or stylistic images emphasising items that are not considered normal body features. Such techniques mean that the privacy of all people who are not being suspected of carrying illegal items need

not be violated, which limits the amount and degree of privacy violations as much as practically possible.

The Internet analogy of this is surveillance techniques like IDS using deep packet inspection. This means that the MSS provider effectively can see all cleartext traffic between a customer performing a service on the Internet and the service provider.

There will in some cases be a conflict between the privacy and security objectives[1], in particular for managed security services. It is frequently believed that it is not possible to achieve both perfect privacy preservation and at the same time have overview over all potential attacks. An important principle should however be that the monitoring invades privacy and confidentiality as little as possible for normal, unsuspicious traffic. However, just like in the airport example, a more thorough investigation will be required if suspicious Internet traffic is detected, to verify whether the data traffic is hostile or not. Trusted applications that are allowed to decrypt and monitor sensitive information in IDS alarms for signs of attack is a technique that can be used to limit access to private or confidential information. Much can in other words be done to improve the situation compared to current technologies, which leak more private or confidential information than strictly necessary.

4 Legal Aspects

The legal framework on privacy and data protection in Europe is mainly governed by two directives. The original European Community (EC) directive on privacy in the European Union and EEA member states is directive 95/46/EC on Data Protection (European Commission, 1995). A newer directive, directive 2002/58/EC on Privacy and Electronic Communications (European Commission, 2002), concerns the processing of personal data and the protection of privacy in electronic communications. The aim of the privacy and electronic communications directive is to protect the fundamental rights and freedoms of persons, in particular with respect to the increasing capacity for automated storage and processing of data relating to subscribers and users. This directive depends on the development of new technologies that can minimise the processing of personal data and use anonymous or pseudonymous data where possible. This means that there is a public interest in developing efficient privacy enhancing technologies that can improve the privacy handling of on-line services in general and security monitoring techniques in particular.

On the other hand, there is also a requirement in the Privacy and Electronic Communications directive that lawful interception is possible where such measures are appropriate. The EC Data Retention directive in addition requires storage of traffic information that describes the parties who participates in a communication for a period of between six months and two years (European Commission, 2006). This means that public authorities have defined laws that prohibit use of private or confidential information in general. However they also require that enough information is stored to make it possible to investigate crime and intercept communication in case of criminal investigations where a search warrant has been issued. This for example means that the providers of anonymising services like Mixes in Europe now are required to store

[1] There will also be synergies between privacy enhancing technologies and security, as will be discussed later. Aiming for such synergies is recommended by the 4. Privacy by Default principle, which states that one should aim for a win-win situation between privacy and security (Cavoukian, 2009a).

information about access to the anonymisation services due to the Data Retention directive[2].

5 Ethical Aspects

Host and network-based attack detection techniques may as discussed earlier cause an invasion of privacy that has the potential for harm. Companies therefore need to be able to justify the practice from an ethical perspective. The situation description in the previous section shows that *security* interests can come in conflict with other important human rights like *privacy, freedom of expression* and *the right to be presumed innocent until proven guilty.*

Teleological principles have an account of the good which is fully independent from the right, and a fully dependent theory of the right, as that which maximises the good (Ronzoni, 2009). The best known example of a Teleological principle is perhaps Utilitarianism introduced by Jeremy Bentham and John Stuart Mill (Mill, 1863; Bentham, 1843), which deems the moral action as the one that aims at maximising a given good.

It is for example possible to define the best moral choice of monitoring techniques as the choice that provides the highest employee utility in form of minimising losses due to non-work related Internet usage, security incidents and liabilities for example from downloading pirated software or music (Charters, 2002). However, such a narrow definition of the best moral choice is problematic. The objective for the moral choices should rather be to reduce the harm for the society at large based on accepted standards like the Human Rights (United Nations, 1948), than to focus on narrow definitions of the moral choice.

The general concern with monitoring without regard to privacy or other human rights, is that the on-line community would end up being like an electronic Panopticon where the inspector could see anything and the inspected would be aware of being monitored, potentially at any time, but not *when* they were monitored (Bentham, 2003). Another way to describe it is as an Orwellian society that observed and controlled all on-line information (Orwell, 1949). Even though security may be improved, the overall utility would be lower, since other human rights like privacy and free speech would be reduced.

This is not only of theoretical interest. Mandatory surveillance and censorship of on-line behaviour is for example frequently performed in totalitarian regimes. One example is China, where Internet Service Providers are required to perform the monitoring of the citizens (Walton, 2001). There is also a pressure towards widening the scope of on-line surveillance also in democratic regimes both in Europe and elsewhere in the world, for example to detect certain types of crime (Brown and Korff, 2009).

Awareness of such monitoring causes self-censorship, which means that people are afraid of telling the truth because of the risks of punishment or retaliation from parties responsible for the monitoring. Another risk is that the monitoring organisation may not act morally right and abuse acquired knowledge from private or confidential information. Corrupt insiders in the monitoring organisation may for example sell private or confidential information, extort the information owner or use the information

[2] JAP implementation of data retention http://anon.inf.tu-dresden.de/dataretention_en.html.

for own advantage (anonymised ref.). This could in the worst case lead to a legalistic society where everything was dictated by law, and the inner freedom of ethical choice was reduced to little or nothing. This means that security monitoring as a means in general may be problematic from an ethical perspective, unless ethical guidelines and policies are being used to control how the security monitoring is being done.

This paper will not go into a detailed moral discussion on whether security monitoring can be considered acceptable or not. It is sufficient to notice that using security monitoring practices, like IDS, in general are legally accepted as means for detecting cyber-attacks. Given this reality, then the objective should be to limit the harm of such monitoring from a privacy and human rights perspective.

Intrusion detection systems are not without their problems. The monitoring is challenging because of the proliferation of new attack vectors and the use of obfuscation techniques, which make detection difficult (Polychronakis et al, 2009). The reason behind the proliferation of malicious software (malware) is partially due to malware creation kits (Ollmann, 2008). These kits can create Trojans or entire phishing web sites that are intended to lure users to install malware on their computers. To avoid detection, malware uses techniques like cryptographic obfuscation and self-mutating code (M. Sharif and Lee, 2008). It is therefore easy for adversaries to create new attacks that typically go undetected by anti-virus and IDS systems, so called zero-day exploits, which open up a window of opportunity for the adversary to attack the system. Zero-day vulnerabilities and exploits are considered valuable by cyber criminals, and are frequently traded on underground black markets (anonymised ref). A problematic aspect with these markets from an ethical perspective, is that not only cyber criminals, but also grey market security companies and governmental backed agencies participate in the trade of zero day exploits (Greenberg, 2012). This means that cyber criminals, governmental services and grey market actors have an economic incentive to keep information about such vulnerabilities secret, instead of disclosing them which in general would provide better overall utility to the society.

In addition, backdoors and control channels to large bot-nets of compromised computers are increasingly using encrypted communication. It is not feasible to intercept the malicious communication in these cases, since the attacker is the only person who knows the decryption keys. Furthermore, attacks are hard or impossible to trace due to no or limited retention of traffic data on a worldwide basis. This means that more comprehensive strategies are required for efficient monitoring of malicious activities in the future.

One example of a more invasive strategy is to combine security monitoring with decoy systems, so called Honeynets (Barford et al, 2010). Honeynets present themselves as a network of vulnerable hosts and is able to run malicious programs in a controlled environment to see if it performs security violating changes of the system. Distributed Honeynets have been proposed as one approach to quickly identify new attack signatures that can be distributed to IDS systems. However Honeynets are problematic from an ethical perspective, since they effectively are based on deceiving the adversary. Despite this, security vendors use Honeynets to a large extent to get an overview over the threat landscape and identify new attacks on the Internet (Fossi et al, 2008).

There is in other words an arms race between adversaries (malware producers, organised crime, governmental agencies and hackers) and the computer security industry. Traditionally hackers running bot-nets are opportunistic and will pick the targets that are easy to attack using any attack vectors that give a reasonable success rate (anonymised ref). The attackers can target software vulnerabilities using exploits,

or social vulnerabilities using Trojans, or both. The aim of the hackers is to a large extent monetary gain (Fossi et al, 2008). They are harvesting information from the hacked computers that can be used for financial fraud, identity theft, password logging or extortion.

The advent of ATPs changes this picture, since governmental agencies may use a large amount of resources for attacking a target critical infrastructure as part of a war or cyber intelligence strategy. It is much harder to protect oneself against such threats, since the attackers may have political reasons and sufficient funding for choosing a given target almost at any cost, and will therefore not necessarily go for an easier or less protected target, as cyber criminals frequently do. This means that better and more resilient methods are needed for protection of private or confidential information, as well as to protect availability and integrity of information infrastructures against cyber-attacks, than what currently is available. For example using a defence-in-depth strategy which assumes that more than one barrier need to break to access private or confidential information.

A rhetorical question is whether it is worse that privacy is being violated by criminals for criminal purposes, or by governmental agencies for war or intelligence reasons, than by a presumably reputable security monitoring organisation running a trustworthy and legitimate business? Many people will accept quite privacy invasive surveillance if they perceive that the benefits outweigh the disadvantages, for example to avoid that their information gets stolen by hackers invading their computers (Tsai et al, 2012). The increasing problem with cyber-crime, where it is relatively easy for adversaries to create new attack variants, causes a demand for better and more invasive security monitoring systems that can detect criminal activity as early as possible. This again causes a pressure on privacy interests to allow more invasive monitoring in the interests of computer security. However the monitoring organisations can and should take privacy considerations during their operation, something one cannot expect that criminals do.

An important principle, is that the monitoring should invade privacy and confidentiality as little as possible for normal, unsuspicious traffic. However, just like in the airport example, a more thorough investigation will be required if suspicious Internet traffic is detected, to verify whether the data traffic is hostile or not. Much can be done to improve the situation compared to the current security monitoring technologies used, which leak more private or confidential information to the security analysts monitoring the traffic than strictly necessary. So although information sharing and less anonymity may help to prevent crime, it can have serious implications for personal privacy and other human rights.

It is therefore in many cases necessary to limit either privacy or security somewhat to come to a compromise that ensures the greatest benefits to society at large. The situation description shows us that on-line crime is a significant threat to our society. It can even disrupt the on-line infrastructure of countries. Law enforcement is in many cases without effective means to prosecute globalised on-line crime due to lack of evidence, resources and jurisdictional issues between countries (Kshetri, 2006). Security monitoring is therefore in most countries considered legal and morally acceptable, as long as the purpose is to detect cyber-attacks.

Privacy-enhancing technologies and privacy metrics should however still be used to allow security monitoring being performed as precisely as possible, in order to minimise the privacy and confidentiality impact. A challenge is that security monitoring needs to be implemented within a commercial organisation that mainly aims to maximise the profit for its owners. This means that a customer of a MSS provider will have a limited

budget available for security investments. It has for example been suggested that only a fraction of the expected loss due to security breaches (max 37%) should be spent on security investments for a risk neutral firm (Gordon and Loeb, 2002). This also means that privacy and security interests need to compete on the funding to implement the best possible security and privacy handling for the managed security service with the given funding.

There are in other words practical limits for how much money and effort that a monitoring company can put into both security and privacy to improve the service. This is at the moment a major hurdle, since technologies for privacy-enhanced security monitoring not yet are readily available. There are however research efforts underways both in EU and the US to mitigate this deficiency[3].

The monitoring organisation may also see benefits in better privacy handling from an economic perspective, for example if reduced handling of private or confidential information has side effects like reduced operating costs from handling fewer false alarms. In addition, improved privacy handling should reduce the chance of liabilities from privacy leakages, and it will improve the trustworthiness for customers where privacy and confidentiality is paramount. One example of such customers is health institutions who, due to very strict privacy requirements, will not allow sensitive data from IDS alarms to leave the corporate network.

Better privacy handling of IDS alarms is also in-line with the 4. foundational Privacy by Default principle (Cavoukian, 2009a), since integrating privacy enhancing technologies can create a win-win situation by supporting both the privacy and security objectives.

A similar Utilitarian way to solve the moral dilemma between privacy and utility, has recently been proposed based on information theory (Sankar et al, 2010):

> *"For a data source with private and public data and desired utility level, maximum privacy for the private data is achieved by minimising the information disclosure rate sufficiently to satisfy the desired utility for the public data."*

This way to solve the ethical dilemma implies that private or confidential information is disseminated strictly on *a need to know* basis. An advantage of this approach, is that it may be possible to quantitatively analyse the optimal solution and compare how close a real solution is to the optimal one, given that some objective criteria or *metrics* for the information disclosure rate are identified.

A disadvantage with this model may be that it does not consider the semantics and therefore also not the *value* of revealed private data. Some data are typically considered more sensitive and therefore also more valuable than other. Econometric or actuarial models typically aim at modeling the cost of revealing data (Gritzalis et al, 2007; Yannacopoulos et al, 2008). The practical challenge with these economic models, is that it may be difficult to get representative cost distributions, since they are based on peoples' subjective value of private data.

I did some preliminary experiments as part of my research where security analysts attempted to classify the privacy leakage of IDS alarms. They found it very difficult to do this. In many cases they found it hard to understand, or even found it purely hypothetic, that the sampled IDS alarms even would contain any significant information that was sensitive from a privacy or confidentiality perspective. The information they

[3] For example EU project (anonymised), NIST framework for reducing cyber-risks to critical infrastructures: https://www.federalregister.gov/articles/2013/02/26/2013-04413/developing-a-framework-to-improve-critical-infrastructure-cybersecurity.

sampled, was after all open (i.e. not encrypted/protected) in their opinion. This could mean that security analysts are less privacy conscious than others, for example that they are blinded by operating routinely on sensitive information. It could also mean that there actually is less really private or confidential information in the IDS alarms than one should think.

The perceived value of private information is after all highly subjective (Gritzalis et al, 2007), so it is not given that the valuation by security analysts, which are the only people that have security clearance to access the IDS alarms, would give a representative picture of the privacy leakage. In practice, the only stakeholders that can give the correct valuation of the privacy impact from IDS alarms, is the users themselves. And it is in most cases not trivial and also not desirable to connect the users to the underlying data from a privacy perspective.

One possible way to get around this problem, to get realistic measurements of the privacy impact may be the following: Assume that the Privacy Officer compiles a top ten list of the most privacy concerning IDS alarms, for example from a given web service. The Privacy Officer then needs to ask a representative random sample of users in an anonymous poll, presented during use of the service, what they think their privacy is worth in monetary value, given that a security company may see how they used a given set of web pages. The results from this poll could then be used to estimate the impact factor as a random variable for each given IDS alarm.

It may however in practice not be feasible to do this, because it would be difficult to get permission to do such an experiment in an outsourced scenario where you would have to consider the business concerns of both the MSS provider and the service provider being monitored. It is hard enough to get consent from the MSS provider to do research on IDS data, and may be even harder to get consent from the service provider that used intrusion detection services from the MSS provider, due to concerns that such a detailed poll would affect the reputation of the service being monitored. This means that it will be challenging at best, maybe not even possible, to get a representative cost distribution for the IDS rules that seem to leak most information, not to mention getting a representative cost distribution for the entire IDS ruleset consisting of several thousand rules.

It is therefore better to focus on measuring information leakage in IDS alarms based on objective criteria that correlate with the disclosure rate of sensitive information, for example based on Shannon entropy Shannon (1948), so that optimising the security monitoring from a privacy perspective effectively will reduce the information disclosure rate.

5.1 The Effect of Outsourcing Security Monitoring

Outsourcing security monitoring to Managed Security Services (MSS) providers has gained popularity for two main reasons. First, the cost of providing 24x7 monitoring is only a fraction of what such monitoring would cost in-house (Ding et al, 2005). Second, MSS providers have in general got more experience in handling security incidents and more updated monitoring technology, by specialising in this area, than the average customer. A large client base also contributes to service quality improvements, because an MSS provider, monitoring a large set of networks, easier can correlate attacks and identify new attack patterns. They can also share information about attacks and attack mitigation strategies between its customers, which is one of the factors that have been

shown to reduce the risk of attacks from adversaries (Schechter and Smith, 2003). One concern firms have when considering to outsource security services, is that the MSS provider may shirk (avoid doing its duties) secretly to increase profits. In economics this behaviour is commonly referred to as the Moral Hazard problem. The optimal way to avoid such behavior on a contractual basis is to use a performance-based contract, however the degree of performance dependence may decrease if the reputation effect becomes significant (Ding et al, 2005).

It should be noted that the Moral Hazard problem not only is applicable to the security area of security monitoring. It is also applicable to the privacy and confidentiality of the monitored data. Both in the sense of handling more private and confidential information than strictly necessary and in the sense of potentially leaking or abusing private or confidential information. This does in the end mean that a principal (here the customer of security services) should require that both the *security* and *privacy* performance for outsourced MSS should be part of a performance-based contract with the MSS provider. This means that the MSS provider should be *accountable* both for the privacy and security part of the operation which means that suitable performance metrics and activity logging procedures are needed for both privacy and security, so that the performance in these areas can be reported and audited if necessary. This is in line with the 6. foundational Privacy by Design principle, that the privacy-enhanced design must ensure transparency (Cavoukian, 2009a).

5.2 Attack Prevention or Surveillance, Which Is Better?

Intrusion Prevention Systems (IPS) is a network monitoring technology that extends IDS with the possibility to automatically enforce a computer security policy. A question is then: when is it acceptable from an ethical/moral perspective to automatically enforce a computer security policy, and are there any cases where it can be considered better to automatically enforce the policy than to use traditional monitoring techniques like IDS? A related question is whether blocking of undesirable content is more acceptable than surveillance covering use of undesirable content?

In general, IPS, firewalls and IDS may all leave electronic evidence in the form of system logs or alerts sent to a central security operations centre. It is possible to define rules that enforce a security policy without leaving electronic traces, however this is not common to do. The reason is that system logs are useful to detect and improve rules that perform poorly or incorrectly. It can also be useful to verify correct system operation.

Logging of what is being monitored may also be important for accountability, to audit what is being monitored either by the network owner or by third party quality certification organisations. It should however be noted that such logs also may contain private or confidential information. They should therefore be cryptographically protected both against unauthorised modifications by the MSS provider as well as against external attacks, and should use privacy enhancing technologies, for example anonymisation or pseudonymisation, to avoid showing private or confidential information in cleartext to unauthorised personnel. Since IPS rules typically perform automated actions, then there should normally not be a need to view detailed information from such events in cleartext. IPS alerts should therefore be suitable candidates for anonymisation/pseudonymisation.

A problem with automatic enforcement using IPS, is however that monitoring rules typically are neither perfect, meaning that false alarms may occur, nor complete, meaning that the rule is able to catch all attacks (Flegel, 2007). This means that using an IPS causes a risk that some legitimate traffic also will be denied. On the other hand, one should not be complacent because of having an IPS implemented, since the rule definitions typically are not complete, and may not detect all attack scenarios. A common way for IPSs to enforce preventive actions, is to block traffic from the attacker either permanently or for a given time interval. This can be problematic both from an ethical and business perspective since it may cause legal traffic to be blocked out. There is also a risk for targeted Denial of Service attacks against the IPS or firewall if the adversary uses forged attack traffic to disrepudiate a given user or to block the entire service. This shows that automatic filtering of attack traffic based on blocking traffic that matches given rules can be problematic from both an ethical, business and security perspective, although it clearly is more cost effective than manual 24x7 monitoring of IDS alerts. It is also more efficient since it actually may prevent an ongoing attack, given that the IDS rule is sufficiently precise, to detect and deter the attack vector without harming innocent third parties.

A somewhat related area, is permanent blocking of traffic from certain hosts assumed under control by adversaries, or even censorship of web sites providing content that in a given legislation is deemed illegal. Is automatic enforcement of security policies, for example via rules that deny access to certain on-line resources in this case more acceptable than security monitoring? Is it for example worse to block inappropriate web sites or web sites that may be risky from a security perspective, than if humans investigate such events? This is a discussion on censorship against surveillance - which is better or worse. Content filtering is cheaper and may be a better choice from a purely economical perspective, however one may risk liabilities from legitimate users and customers.

Content filtering can be considered better from a privacy perspective provided that IPS alarms are properly anonymised. However it is not necessarily better from an anti-censorship/free speech perspective. Knowing that systems in general log what is being filtered, then it can be discussed whether content filtering is a good argument from a privacy perspective, although it certainly is possible to create IPS rules that either anonymise or encrypt sensitive information or do not log any information at all. Also, a censored environment may give a deceptive perception of reality, something that is morally questionable.

Content filtering using IPS or firewall technologies is in other words useful and can be morally acceptable if used against attack scenarios, provided that the MSS provider aims at minimising the impact of the rules both from a privacy and freedom of speech perspective, as well as ensuring that the rules do not harm innocent third parties. However a potential risk is that content filtering is vulnerable to denial of service attacks.

6 Conclusions

There is a need for good privacy metrics to provide organisations with an incentive for improving the operation from a privacy and confidentiality perspective. This would allow a continuous improvement process where privacy and security could be improved according to needs and also budgetary limitations. Metrics should be used to limit

leakage of private or confidential information to the bare minimum necessary from an operational perspective. Said from another perspective, the security analysis done by IDS/IPS should aim at operating strictly according to the need-to-know principle.

This will also give increased productivity through fewer false alarms and may give the monitoring company a better reputation for handling of private and confidential information, something that is required in certain business cases, for example for health institutions or for critical information infrastructures.

Content filtering using IPS or firewall technologies can be considered useful and also morally acceptable against on-line attack scenarios, provided that the rules aim at minimising the impact both from a privacy and freedom of speech perspective. However a potential risk with automatic content filtering based on known attacks is being vulnerable to denial of service attacks, something that may harm legal use of the system.

A risk to consider for outsourced MSS, is Moral Hazard (or shirking). The network owner cannot assume that security monitoring outsourced to a MSS provider will be performed in a way that aims at reducing privacy invasiveness as much as possible, and at the same time with as good security as possible. Contracts to perform outsourced managed security services, as well as related metrics and indicators, should therefore be *incentive compatible*, which means that payment should be related to auditable performance metrics both related to security and privacy to give the MSS provider an incentive to improve both privacy and security. This also means that transparency is required on how the MSS operation is being performed. However access to information required to provide auditability and nonrepudiability should be protected against unauthorised access, since such information also may contain private or confidential information.

Certification and auditing is one way to improve adherence to the ethical guidelines. Businesses that provably perform unethical security monitoring would risk to loose their quality certification, which would be detrimental for the reputation of a company in the security business. This can either be enforced via governmental regulation or voluntary regulation, as part of a quality accreditation.

Future work involves designing and implementing an architecture, methodology and tools for monitoring critical infrastructure for signs of cyber-attacks based on these ethical guidelines.

Acknowledgements

This work has been partially supported by the project "PRECYSE - Protection, prevention and reaction to cyber-attacks to critical infrastructures", funded by the European Commission under the FP7 frame programme with contract number FP7-SEC-2012-1-285181 (www.precyse.eu).

References

Alharby A, Imai H (2005) IDS false alarm reduction using continuous and discontinuous patterns. Lecture Notes in Computer Science 3531:192–205

Ashmore, W C et al (2009) Impact of alleged russian cyber attacks. Baltic Security & Defence Review 11

Barford P, Chen Y, Goyal A, Li Z, Paxson V, Yegneswaran V (2010) Employing honeynets for network situational awareness. In: Cyber Situational Awareness, pp 71–102

Bentham J (1843) The Works of Jeremy Bentham. Edinburgh: W. Tait; London, Simpkin, Marshall

Bentham J (2003) Panopticon. or, the inspection-house, &c. Criminological perspectives: essential readings p 25

Brown I, Korff D (2009) Terrorism and the proportionality of internet surveillance. European Journal of Criminology 6(2):119–134, DOI 10.1177/1477370808100541, URL http://euc.sagepub.com/content/6/2/119

Cavoukian A (2009a) Privacy by design. http://privacybydesign.ca/about/principles/

Cavoukian A (2009b) Whole Body Imaging in Airport Scanners: Activate Privacy Filters to Achieve Security and Privacy. http://www.ipc.on.ca/images/Resources/wholebodyimaging.pdf

Charters D (2002) Electronic monitoring and privacy issues in Business-Marketing: the ethics of the DoubleClick experience. Journal of Business Ethics 35(4):243–254, DOI 10.1023/A:1013824909970

Ciriani V, di Vimercati SC, Foresti S, Samarati P (2007) k-Anonymity. In: Secure Data Management in Decentralized Systems, pp 323–353

Clauß S, Schiffner S (2006) Structuring anonymity metrics. In: Proceedings of the second ACM workshop on Digital identity management, ACM, Alexandria, Virginia, USA, pp 55–62

CrySys lab (2011) Duqu: A stuxnet-like malware found in the wild, technical report. URL http://www.crysys.hu/publications/files/bencsathPBF11duqu.pdf

Ding W, Yurcik W, Yin X (2005) Outsourcing internet security: Economic analysis of incentives for managed security service providers. In: Internet and Network Economics, LNCS, vol 3828, Springer, pp 947–958

European Commission (1995) Directive 95/46/ec. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

European Commission (2002) Directive 2002/58/ec of the european parliament and of the council of 12 july 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (directive on privacy and electronic communications). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:NOT

European Commission (2006) Directive 2006/24/ec directive 2006/24/ec of the european parliament and of the council of 15 march 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending directive 2002/58/ec. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:01:EN:HTML

Flegel U (2007) Privacy-Respecting Intrusion Detection, 1st edn. Springer

Fossi M, Johnson E, Mack T, Turner D, Blackbird J, Low MK, Adams T, et al (2008) Symantec Global Internet Security Threat Report. Trends for 2008, vol XIV. Symantec

Gordon LA, Loeb MP (2002) The economics of information security investment. ACM Trans Inf Syst Secur 5(4):438–457, DOI 10.1145/581271.581274

Greenberg A (2012) Shopping for zero-days: A price list for hacker's secret software. URL http://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/

Gritzalis S, Yannacopoulos A, Lambrinoudakis C, Hatzopoulos P, Katsikas S (2007) A probabilistic model for optimal insurance contracts against security risks and privacy violation in IT outsourcing environments. International Journal of Information Security 6(4):197–211, DOI 10.1007/s10207-006-0010-x

Kshetri N (2006) The simple economics of cybercrimes. IEEE Security and Privacy 4(1):33–39, DOI 10.1109/MSP.2006.27

M Sharif JG A Lanzi, Lee W (2008) Impeding malware analysis using conditional code obfuscation. NDSS'08

Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M (2007) l-diversity: Privacy beyond k-anonymity. Cornell University p 52

Mill JS (1863) Utilitarianism. London: Parker, Son, and Bourn

Ollmann G (2008) The evolution of commercial malware development kits and colour-by-numbers custom malware. Computer Fraud & Security 2008(9):4–7, DOI 10.1016/S1361-3723(08)70135-0

Orwell G (1949) Nineteen eighty-four. Secker and Warburg

Polychronakis M, Anagnostakis KG, Markatos EP (2009) An empirical study of real-world polymorphic code injection attacks. In: Proceedings of the 2nd USENIX conference on Large-scale exploits and emergent threats: botnets, spyware, worms, and more, USENIX Association, Berkeley, CA, USA, LEET'09, pp 9–9

Ronzoni M (2009) Teleology, deontology, and the priority of the right: On some unappreciated distinctions. Ethical Theory and Moral Practice URL `http://dx.doi.org/10.1007/s10677-009-9209-z`

Samarati P (2001) Protecting respondents' identities in microdata release. IEEE Transactions on Knowledge and Data Engineering 13:1010–1027

Samuelson P (2000) Privacy as intellectual property? Stanford Law Review 52(5):1125–1173, URL `http://www.jstor.org/stable/1229511`

Sankar L, Rajagopalan S, Poor H (2010) Utility and privacy of data sources: Can shannon help conceal and reveal information? pp 1 –7, DOI 10.1109/ITA.2010.5454092

Schechter SE, Smith MD (2003) How much security is enough to stop a thief? the economics of outsider theft via computer systems and networks. Financial Cryptography 2742:122–137

Shannon CE (1948) A mathematical theory of communication. Bell System Technical Journal 27:379–423, 623–656

Sweeney L (2002) k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10:557–570

Tsai J, Kelley PG, Cranor LF, Sadeh N (2012) Location-sharing technologies: Privacy risks and controls. SSRN eLibrary URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1997782`

United Nations (1948) The universal declaration of human rights. `http://www.un.org/en/documents/udhr/`

Walton G (2001) China's golden shield - corporations and the development of surveillance technology in the people's republic of China. `http://www.ichrdd.ca/site/_PDF/publications/globalization/CGS_ENG.PDF`

Warren S, Brandeis LD (1890) The right to privacy. Harvard Law Review 4(5)

Yannacopoulos AN, Lambrinoudakis C, Gritzalis S, Xanthopoulos SZ, Katsikas SN (2008) Modeling privacy insurance contracts and their utilization in risk management for ICT firms. In: Proceedings of the 13th European Symposium on Research in Computer Security: Computer Security, Springer-Verlag, Málaga, Spain, pp 207–222