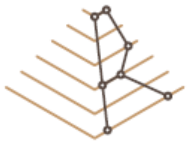

RAMSES

Internet Forensic platform for tracking the money flow of financially-motivated malware

H2020 - 700326

D4.2 Improvements over existing Economic Modelling of Malware as a Business Model

---

Lead Authors: Darren Hurley-Smith, Julio Hernandez-Castro, Edward Cartwright and Anna Stepanova (UNIKENT)

With contributions from: Alejandro Prada Nespral & Mario Recio (TREE), Luis Javier Garcia Villalba (UCM), and Michael Brengel (USAAR)

Reviewer: Michael Brengel (USAAR)

Deliverable nature:	Report (R)
Dissemination level: **(Confidentiality)**	Public (PU)
Contractual delivery date:	30/11/2017
Actual delivery date:	24/11/2017
Version:	1.0
Total number of pages:	42
Keywords:	Malware, ransomware, economics, modelling, predictive analysis

Abstract

This report is intended for public release, as agreed by the members of the RAMSES consortium. This is the second public report delivered by WP4, expanding on an initial report on the economic modelling of ransomware as a business model. Report D4.2 focuses on the shortcomings of existing economic models intended to derive the optimal profit of malware. Current models are focused on applying existing, business-specific, theories to malware but it has become clear that malware-specific business strategy must also inform this process. Attributes specific to malware, particularly ransomware, are deconstructed to inform an improved economic model. Initial costs, optimal pricing, defensive countermeasures, and the adversarial/competitive environment in which malware must attempt to make profit are discussed at length. Identifying the differences between malware as a business and traditional businesses will help us to derive appropriate improvements to the model put forth in D4.1. This, in turn, will help us progress towards a predictive model of the likely evolution of malware that has a profit motive.

As a publicly disclosed report, the information in this report is not specific to LEA provided information, nor does it contain any confidential data. All information exists in the public domain, or in research that is due to be made public by the technical partners and universities that have contributed to this work.

Executive summary

- This report is intended for public release.
- This report expands on D4.1, the direct predecessor of this report. D4.1 is also a public release document.
- This work is based on research undertaken by Julio Hernandez-Castro, Edward Cartwright, and Anna Stepanova. It has been compiled by Darren Hurley-Smith, who has added related research to help contextualise this report in line with work plan 4 objectives.
- Colleagues from USAAR, TREE, and UCM have contributed to this report. They have provided graphics, research, and clarifications on key points of discussion. Individual contributors are named on the front-page of this report.
- Michael Brengel of USAAR has provided invaluable feedback, which has helped this report move from draft to final article.
- The application of existing models (D4.1) is critically analysed in this report. Attributes unique to malware are identified, to understand what is required of a predictive economic model of malware as a business.
- Ransomware is the focus of this work, due to its overt extortion-oriented activities. Banking Trojans and other forms of malware tend to derive value through theft, or are conducted in a manner that does not directly correlate with a profit-motive.
- The key elements of this report are:
    - A specifically ransomware-related economic analysis. Where D4.1 focused on applying existing economic models, this report identifies the shortcomings of these methods. The unique attributes of ransomware, and the highly adversarial environment in which it operates, require novel approaches. The first step in identifying these processes, is to highlight where existing models do not account for context-specific phenomena.
    - The identification of ransomware-specific economic attributes allows us to formalise them. Modification of the cost-benefit calculations presented in previous work is discussed. Alternative strategies, especially near-future viable ones, are also reported on. This provides a body of work that maps the current limitations of economic models when applied to ransomware, whilst also investigating the potential future development of ransomware should operators consider the economic shortcoming of their current practices.
    - A proposal for the improvement of our economic model is produced as the main output of this report. Combining our findings regarding ransomware-specific economic attributes, and the current sub-optimal behaviours shown in current and past attacks, we suggest a set of improvements that could be used to construct a prototype predictive model of ransomware as a business.
    - As this report is intended for public disclosure, this model is not provided as part of this report – the work undertaken to define the new model and derive a likely optimal evolutionary path for ransomware will be provided in the confidential report D4.3. Both this report and D4.3 are due in M15 of this project.
    - This report can be considered a design specification for the estimated evolutionary path discussed in D4.3 and the eventual software system for the modelling of optimal ransomware (D4.4 – confidential M18).

Document Information

IST Project Number	700326	Acronym	RAMSES
Full Title	Internet Forensic platform for tracking the money flow of financially-motivated malware		
Project URL	http://www.ramses2020.eu		
EU Project Officer	Nada Milisavljevic		

Deliverable	Number	D4.2	Title	Improvements over existing Economic Modelling of Malware as a Business Model
Work Package	Number	WP4	Title	Findings on economic modelling of malware as business model

Date of Delivery	Contractual	M15	Actual	M15
Status	version 1.0		final ■	
Nature	prototype □ report ■ demonstrator □ other □			
Dissemination level	public ■ restricted □			

Authors (Partner)	University of Kent (UNIKENT)			
Responsible Author	Name	Darren Hurley-Smith	E-mail	d.hurley-smith@kent.ac.uk
	Partner	UNIKENT	Phone	+447870806745

Abstract (for dissemination)	This report is intended for public release, as agreed by the members of the RAMSES consortium. This is the second public report delivered by WP4, expanding on an initial report on the economic modelling of ransomware as a business model. Report D4.2 focuses on the shortcomings of existing economic models intended to derive the optimal profit of malware. Current models are focused on applying existing, business-specific, theories to malware but it has become clear that malware-specific business strategy must also inform this process. Attributes specific to malware, particularly ransomware, are deconstructed to inform an improved economic model. Initial costs, optimal pricing, defensive countermeasures, and the adversarial/competitive environment in which malware must attempt to make profit are discussed at length. Identifying the differences between malware as a business and traditional businesses will help us to derive appropriate improvements to the model put forth in D4.1. This, in turn, will help us progress towards a predictive model of the likely evolution of malware that has a profit motive. As a publicly disclosed report, the information in this report is not specific to LEA provided information, nor does it contain any confidential data. All information exists in the public domain, or in research that is due to be made public by the technical partners and universities that have contributed to this work.
Keywords	Malware, ransomware, economics, modelling, predictive analysis

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
15/11/2017	0.1	Darren Hurley-Smith (UNIKENT)	Initial draft of contents page updated with content from the first publication by UNIKENT.
01/11/2017	0.2	Darren Hurley-Smith (UNIKENT)	Draft updated with information provided by consortium partners. This is an iterative process with at least one more revision due.
09/11/2017	0.3	Darren Hurley-Smith (UNIKENT)	Draft updated with the last of the information provided by technical partners (UCM, SAAR, and TREE)
13/11/2017	0.4	Darren Hurley-Smith (UNIKENT)	Internal draft version, submitted to USAAR for review.
19/11/2017	0.5	Michael Brengel (USAAR)	Revision comments provided by USAAR. In-depth spelling, grammar and content corrections provided by Michael Brengel.
24/11/2017	1.0	Darren Hurley-Smith (UNIKENT)	Revised version, incorporating advice and changes suggested by Michael Brengel of USAAR. Submitted to the RAMSES CMS as an official contribution to work plan 4 (D4.2)

Table of Contents

List of figures

List of tables

Abbreviations

IoT: Internet of Things

MaaS: Malware as a Service

RaaS: Ransomware as a Service

WTA: Willingness to Accept

WTP: Willingness to Pay

Definitions

Altcoin: Any cryptocurrency that is not Bitcoin (core). This includes Ethereum, Monero and many others. These currencies may be Bitcoin-like stores of value, or may have additional properties. The most common of these additional properties is a capacity for blockchain backed contracts, known as smart contracts.

Bitcoin: The most iconic cryptocurrency currently in existence. Developed by Satoshi, Bitcoin represents the first successful cryptocurrency and has mass media exposure. This is a common currency for ransomware to request as payment, both due to its media exposure (making it likely that a victim will have heard of it) and the wealth of information and sources that make purchasing Bitcoin less arduous than other cryptocurrencies.

Blockchain: A fundamental technology that supports cryptocurrency. Effectively a distributed ledger, the blockchain provides a record of all transactions that have been agreed by a consensus among trusted nodes on a network. The most common example of Satoshi's Bitcoin blockchain, but other examples are common.

Crypto-Ransomware: A specific form of ransomware, which works by encrypting the contents of the target computer and using possession of the key as a bargaining chip. Differs from some strains of ransomware by focusing on threats of implicit data-loss instead of other forms of control. It may still rely on associated strategies, such as file deletion, but passively threatens to leave files encrypted in a manner that would make them irretrievable without the appropriate key. Generally, a cash fee is asked (though some have more exotic requirements such as pyramid schemes).

Cryptocurrency: Digital currency, backed by one of a variety of means and using blockchain technology to provide a means of undisputed exchange of currency. Usually has a focus on anonymity, community and/or niche markets that provide backing (futures, computational power etc.).

Dark Net: A network overlaying the internet. It can only be accessed with specific software and operates using non-standard communication protocols. Usually intended to be private and anonymous, these networks are attractive to criminals and play host to black markets such as Alphabay.

Fiat Currency: A government backed currency, such as the Euro. Cryptocurrency must be changed back into a fiat currency if criminals wish to take their profits into most areas of the global economy.

Malware: Malicious software. Programs that cause damage and/or disruption to a target. May involve deletion of files, spying software, ransomware or one of many other forms of attack. Some extreme examples may focus on destruction of hardware (Stuxnet).

Malware as a Service: The act of developing, maintaining and providing technical expertise in the use of malware for profit. Instead of using their software themselves, individuals and groups performing malware as a service sell their talents and outputs to others, who will then go on to use them. This may technically include distribution methods, such as botnets, which are used to send malicious software to potential victims.

Opportunity cost: The potential to make money through other means. Where the profitability of a given action falls below that of an alternative or equal or lesser complexity and cost, an opportunity cost is incurred. For example, one could report the exploit they could maliciously use, to a bug bounty program. If this would yield more profit than using it maliciously, an opportunity cost is incurred.

Ransomware: Software that focuses on seizing control of technology, software or data. The control is used as leverage in a demand for the target to exchange a sum of money for the promised (but debatable) return of the seized items. Used colloquially to reference **Crypto-ransomware**, ransomware can mean any software that focuses on seizure of assets, not just cryptographic methods.

Ransomware as a Service: See malware as a service (MaaS). This subdivision of MaaS focuses solely on the extortion-focused type of malware known as ransomware.

1 Introduction

In D4.1 an initial economic model of malware was proposed. Considering malware as a business, this model made observations regarding the current profitability of cyber-criminal activity with a focus on ransomware. Limitations, such as the technical difficulty of scaling price to target individuals for optimal ransom value, were identified. A survey was conducted to demonstrate the application of our theories to a real-world scenario. This demonstrated that current ransomware operators are asking too little, and attempting to extract value from a maximal number of victims, instead of aiming to achieve the optimal sum of all ransoms paid. The conclusion of the report states that ransomware, and the wider profit-motivated malware phenomenon, is currently in a primitive state. Law Enforcement Agencies (LEAs) and those affected by future strains of malware can expect an increase in diversity of both technical and human aspects.

The argument that malware will converge to an optimal state (highest profit) is a valid one. It is, however, not able to predict the specific actions that will lead towards the optimal state, without further information. Taking ransomware as our specific example, we are observing the very early stages of its development. Though the delivery mechanisms of malware vary in sophistication, the ransom process remains primitive. Current pricing strategies appear to be based on very general assumptions and lack nuance compared to the diverse demographics they tend to affect. The path to the optimal state is currently undefined, and as a result, it is likely that we will see significant divergence before an eventual convergence to the true optimal state. This is a consequence of individual malware operators exploring adjustments that suit their current capabilities, or improving on previous successes.

The model proposed in D4.1 is not intended to explore the minutiae of this evolutionary process. It performs well when diagnosing the shortcomings and likely changes to current ransomware, assuming the retention of current pricing strategies and delivery mechanisms. However, it does not account for attempts to reduce initial and operational costs on the part of cyber-criminals. It is not designed to address several malware-specific phenomena: target-specificity, ancillary goals (disruption/data exfiltration), and product (malware strain) lifecycles to name but a few. The D4.1 model remains a good guide of current distance from the optimal, but an unsuitable indicator of the potential next steps.

This report deconstructs our existing model, discussing its role as a macroscopic view of distance from optimal profit for a given strain of malware. Section 2 provides an overview of our previous model and the associate report, D4.1. The need for a finer-grained, malware-specific model is stated, and a deconstruction of the existing model provided to highlight areas of improvement.

In section 3, we discuss the derivation of economic attributes exhibited by malware. We address this in more specific terms than the general model described thus far. These attributes are then used to refine potential additions to the model. The outcome of this is a general discussion of a malware-specific economic model, analysing evolutionary drift and intentional change in the key costs and profits of a given strain.

Section 4 discusses a range of improvements to the model, based on the detailed identified in previous sections. Analysing the impact of these optimisations on other attributes of malware is a primary goal of this economic model of malware as a business. Section 5 concludes the work, discussing how this report will inform the development of a predictive economic model of malware as a business. Future work is discussed in specific terms; WP4 is due to finish March 2018 and the predictive model is a key deliverable of that body of work.

2 Economic Modelling of Malware as a Business

D4.1 presented the application of economic theory to malware. The goal of this work was to identify attributes that improved or diminished the profitability of a given malware strain. In this it succeeds, a general economic model of malware as a business is presented. A game theoretic approach allows the exploration of adversarial scenarios, wherein the operator (attacker), victim and Law Enforcement Agencies (LEAs) can be accounted for. This model, however, is limited. It is a general overview of the path to optimal profitability, but can only offer domain-wide optimisations at this time. To predict more specific evolutions in malware, both technically and in the human aspects employed to encourage payment, a finer-grained model is required.

This section provides an overview of our findings from D4.1. The key findings are summarised and further discussion regarding the current limitations of the model is provided. This will serve as a primer for the subsequent sections, which will address these requirements.

2.1 Profit-motivated malware

It is the assumption of this report that profit is the primary motive for a malware operator. There are many additional motivators for malware operators. Disruption, political statements, and curiosity are but three of these motives. However, the focus of this report is on the motive of profit. Malware has evolved from a hobby and research activity, to a profit driven enterprise. Initial malware strains include Bob Thomas' Creeper system (1997), an experimental self-replicating system. This program was essentially harmless, taunting users with a message but having no other effect. 1975 saw the first Trojan in the wild, ANIMAL. John Walker developed this program carefully, to avoid causing damage to infected machines. In 1989, Joseph Popp was responsible for the creation and dissemination of the AIDS Trojan, an early example of a cryptovirus with an accompanying ransom demand ($189 at time of request in 1989). This DOS scrambler was distributed via floppy disk. The 1990's saw the internet develop. This resulted in a high degree of connectivity between all sectors of modern society, potentially lucrative targets have become more accessible to canny malware operators. The concept of public key cryptography for malicious purposes was introduced by Young and Yung in 1996, and since May 2005, examples of extortionate ransomware have become prominent. Cryzip and Archiveus are examples of ransomware of this time, and the use of advanced RSA schemes with increasing key sizes was observed. After a brief lull, this type of ransomware returned in 2012 with Reveton, followed by the well-known CryptoLocker ransomware in 2013.

The increased connectivity afforded by the internet has allowed all manner of services and conveniences to be extended to the population. Online shopping, bill payment and banking are but a few examples. However, this also means that confidential information, financial records and even access to personal or institutional wealth are increasing digital. This place a great deal of information into the sights of malware operators, who may seek to steal, lock or delete such data. When the data has personal, institutional, emotional, and/or financial value, malign individuals may equate that with the opportunity to extract a price for the data. This may be through the sale of stolen data, or the return of such data to the original owner – the constant in the situation is the perpetration of cyber-crime for profit.

Ransomware is a good example of profit-motivated malware. Infected computers have their file-systems encrypted with strong cryptographic algorithms, with the ransomware operator holding the private key required to decrypt the files. In this manner, the operator exerts control over a victim, taking something that they assumedly require, and charging a ransom fee for its return. The assumptions in this transaction are two-fold when considering cases in which payment is forthcoming: the operator assumes that the victim wants what has been locked, and the victim must assume that the files will be returned. This creates a transactional dynamic between the operator and their victims.

It is this transaction-oriented phenomenon that encourages the use of economic theory to begin to explain how individuals using malware as part of a for-profit activity might improve their product. Technological sophistication is only one aspect of such enhancement, the human driven aspects of malware are many and increasingly sophisticated social-engineering and target prioritisation methods may return significant gains for

a marginal increase in costs. Threats, negotiation, bargaining, and extensive *customer* support are well documented elements of many ransomware attacks. The incorporation of these features may enhance or diminish the profitability of the ransomware strain, but the interplay between social and technical elements is poorly documented at this time.

Technically sophisticated malware may fall short of its optimal profit by being too lenient when bargaining with victims. A good pricing strategy is useless without malware of sufficient sophistication or ingenuity to bypass the defence mechanisms of at least some victims. This co-dependence and subsequent profitability modifying after-effects is non-trivial to model, but by viewing the malware attack as part of a larger business, one can begin to quantify the effects of key choices throughout an attack's execution.

2.2 Costs

Cost is a constant in all businesses. Hardware, software, employees, training, and offices all have inherent costs, often substantial. The capabilities that they offer the business are a key consideration – it is not worth spending money on items that will not provide measurable return unless they have some other organisational benefit. When discussing malware, these costs may be incurred through some non-traditional routes, but they are business-related costs nonetheless.

Initial costs for ransomware have become more complex to analyse in recent times, whilst facilitating a more streamlined purchasing process for malware operators through phenomena like Ransomware as a Service (RaaS). Several years ago, one could consider the development and dissemination costs as the primary initial costs. Development may be considered a function of the opportunity cost incurred by the developer spending time on the creation/adaptation of ransomware instead of putting their skills to other use. This may be mitigated by co-opting code written for associated (malicious or legitimate) development tasks, but will still incur some degree of cost for which the individual will want to see a return. Dissemination costs can vary and are unavoidable. The use of botnets has been shown to dramatically increase the spread of malware (Hernandez-Castro, Cartwright and Stepanova, 2017). Use of an existing botnet owned by the operators may be considered a developmental or opportunity cost. Purchase of a botnet, email spam bots, or more sophisticated social media propagation vectors represents a simple initial cost. In many cases, existing botnets will sell their services, allowing those without access to such assets to make use of them without incurring their own set up costs.

Ransomware as a Service (RaaS) is the cause of the increase in complexity when calculating initial costs. Malicious developers have been documented offering their services through Darknet markets and other anonymity-preserving channels. Complete software, dissemination services, and bespoke development are all available for purchase. This has led to a new stratum forming in the malware ecosystem – the service industry has arisen among cyber-criminals. Ransomware purchases are often supported with supplementary materials and guidance, further reinforcing the existence of a service sector in dark markets.

RaaS is not the only service cost that a malware operator may incur. Customer service, as unintuitive as it may seem, is a key part of many ransomware attacks. The goal of a for-profit attack is to get victims to pay for the cessation of the attack or return of locked files. If the technical skills of a customer are insufficient, they may struggle with the concept of having to pay using cryptocurrency, Bitcoin being the preferred payment medium of most ransomware operators. In such cases, it is helpful to have associates who can guide customers who are willing to pay, but find it difficult to do so, through the process. Unlike initial costs, this is an ongoing expense throughout the operation. So, one must consider operational costs alongside initial expenditures, when assessing the cost of a ransomware attack.

A final category of costs involves so-called adversarial costs. These costs are incurred by opposition – by the victims or by LEAs. Victims can protect themselves against attack, by keeping back-ups of their files or other forms of security best practice. LEAs inform the public of appropriate prevention techniques. They follow up on attacks, conversing with victims and encouraging behaviours that reduce the effectiveness of the criminal activity in question. These efforts all work in opposition to the interests of the cyber-criminal's profit motive. Successful defence denies the opportunity to extract a ransom from a victim. Detection and apprehension are

terminal costs for the individuals – the organisation may not be undermined completely, but it is likely that the individuals involved will face costly fines and likely prison sentences. These represent an extreme cost – the inability to make more profit from the current effort and the opportunity cost of being incarcerated are so significant that they are likely to destroy the profitability of a malware attack from the perspective of affected individuals. One must consider, however, that individual apprehension may have a variable cost to larger criminal organisations depending on their importance and knowledge.

In conclusion, the costs associated with malware depend on several key factors:

- Initial costs
  - Development or purchase of malware
  - Development or purchase of dissemination method(s)
- Operational costs
  - Opportunity costs – it may be more profitable (on occasion) to comply with companies and LEAs. Bug bounties are an example of legitimate submission of exploits to authorities for reward. There's an opportunity cost incurred if the less profitable action of those possible is selected.
  - Providing service to assist victims who are willing to pay but find doing so technically challenging
- Adversarial costs
  - Potential victims have appropriate security countermeasures, denying the opportunity to extract ransom
  - Victims refuse to pay
  - LEAs and technical experts derive information that negates the malware
  - LEAs apprehend individuals involved in the attack

These factors should inform the pricing of the attack, as they represent an estimate of the break-even point for the malware strain in question. In the case of the more extreme adversarial costs, prevention of these eventualities is a more likely focus than incorporation of their costs into the pricing model. As criminal proceeds, where accessible, will be seized it is in the criminal interest to instead expend resources on minimising this eventuality. The expectation is that malware operators will seek to maximise their profits, although our findings to date suggests that many are more interested in maximal instances of payment, rather than achieving the optimal sum of all payments.

2.3 Pricing

D4.1 discusses the pricing model formulated as part of this work in detail. A brief overview of the model is discussed here for the sake of readability and context.

As with any transactional process, two key actors must be considered when discussing the price of a ransom: the attacker and the victim. The attacker seeks to maximise their profit, while the victim has their own valuation of the assets or services being held to ransom. Both parties have far more complex ideals additional to this simple cost-benefit analysis.

The attacker must consider:

- How much they believe the victim will pay (willingness to pay - WTP) based on logical (market value, where relevant) and emotive factors. This varies greatly and unpredictably.
- Maximising ransom value vs. maximising number of victims who pay

These two points are further affected by whether the attack is targeted, and whether the attacker has any intelligence that could inform them of the likely value of the captive assets.

The victim is most likely to consider:

- How much they believe the captive assets are worth
- Personal ethics regarding the payment of criminals
- The level of trust that they have that the files/services will be restored

Victim's don't need to concern themselves with whether they have been targeted – the fact that they have been infected and are being held to ransom is their sole concern. Previous experience, whether personal or communicated, can have an enormous impact on their likelihood of paying. Bad experiences, where files were deleted regardless of payment, will weigh heavily against the likelihood of paying. It is therefore in the interests of ransomware operators to cultivate a good reputation if they intend to carry out a prolonged attack (or future attacks).

Considering the victim's motivations, we can express these as a simpler attribute: *willingness to pay* (WTP). This is unique to each victim, but can be abstracted as v_i - the willingness to pay of each victim *i*..

$$\Pi = \left(\sum_{i=1}^{N}(p_i - c)1_i\right) - F$$

N is the number of victims. p_i represents the ransom demanded of the victim. 1_i is an indicator variable that takes value 1 if $p_i \leq v_i$ and 0 otherwise. F is the fixed cost of operating malware, and c is the cost of dealing with any ransom money (laundering fees, for example). Costs are considered post-attack – variable c depends on the laundering methodology and other choices made during an attack, but is assumed to be known for this discussion. The equation deals specifically with fixed pricing – price discrimination requires individuals to be classified into as many subsets as there are strata within the pricing scheme.

Fixed, or uniform, pricing is the easiest method of setting a ransom. It is also the most common, with almost all major ransomware strains making use of fixed ransoms. These ransoms may be modified by negotiation or by a regular interval at which the ransom increases if it has not yet been paid. Pepall, Richards and Norman (2008) state that it is most optimal to set fixed prices where demand is elastic. This leads to the somewhat counter-intuitive conclusion that the optimal price is one that less than half of victims are willing to pay. This is because the comparatively fewer individuals willing to pay the higher price are willing to pay substantially more than the other half of the population. This makes up for the opportunity cost of victims that will not pay by increasing the total profit across the set of victims.

Price discrimination takes this a step further, seeking to classify victims into bands of WTP. By finding the price than an individual is willing to pay, one may dramatically increase profit by always asking a price that the victim will pay if there are no other mitigating factors (ethical considerations or bad prior experiences). This is known as first-degree price discrimination. However, it is technically demanding to derive such information without targeting the attack. Gathering intelligence on individuals as they are infected is possible, but would require sophisticated software to process and value the contents of the target machine. It is easier to either target a specific institution that one has prior intelligence of, or create a coarse-grained pricing model based on assumed factors that affect WTP (geographic location, hardware specifications, etc.). This more realistic model is known as third-degree price discrimination.

Some ransomware strains have been observed employing price discrimination techniques. Shade uses a Remote Access Trojan (RAT) to spy on victims and estimate the ransom that can be paid (Atanasova, 2016). This is an inefficient way of discriminating WTP, as it is technically demanding and requires a lot of time to gather the associated intelligence. Furthermore, detection of the RAT may alert the victim prior to the ransomware portion of the attack, sinking the costs associated with the RAT portion of the attack. It is likely that more malware operators will attempt this kind of profit maximisation strategy in the future, but the reconnaissance phase will have to become stealthier and faster to mitigate the risk of discovery. At present, it is easier to make assumptions about one's victims or choose them, than it is to derive value alongside an attack.

Pfleeger and Caputo (2012) state that individuals tend to be risk loving over losses, and risk averse over gains. This suggests that ransomware operators are best served by threatening the victim, instead of compromising with them. This flies in the face of intuition, which would instead suggest that compromise will lead to a higher likelihood of payment. But with risk appearing more appealing in a threatening situation than in a gainful one, the victim is more likely to pay the requested ransom under pressure, than when treated well. This combines with the previous statement regarding fixed pricing strategies – if the price to demand is attractive to less than 50% of victims, than it is in the attacker's interests to ensure that price is paid.

Any compromise undermines all other ransom attempts, as it is likely that any reduction will be communicated eventually, leading to a general expectation of a lower price (which will directly reduce WTP). The most successful attackers exhibit *irrational aggression* when bargaining is attempted, destroying some or all files to ensure that not only does the victim in question understand the situation, but that communication of that event reinforces the *payment begets reward, refusal begets loss* narrative that the attacker wishes to cultivate to extract their target price. Many ransomware operators choose to negotiate instead of exercising such strong-arm tactics (F-Secure, 2017), but this is a logical fallacy borne of their intention to receive as many payments as possible, instead of aiming for optimal total profit.

2.4 Game theory and kidnapping as a transaction model

Selten (1988) explores the application of game theory to kidnapping. His findings are applicable to this scenario – instead of an individual or group being held to ransom, we consider the computer system infected with malware to be the focal element of the transaction process. Lapan and Sandler (1988) provide an extended model, in which the victim's potential defensive measures are considered. In this manner, the model becomes a two-sided measure of cost-effectiveness. The victim measures the cost of acquiring and maintaining a defence vs. the cost of it failing to prevent an attack, while the attacker weighs up the costs of launching an attack against potential payment.

Table 2 shows the matrix of payoffs for the attacker and victim in a Lapan and Sandler style game of kidnapping. *E* is the cost of establishing and maintaining defensive measures. *F* is the cost of failing an attack, which may vary depending on if the defensive measures allow the source or nature of the attack to be derived. It may be assumed that the more detections there are of an attack, the less time the attacker must exploit their propagation vector before it is discovered and mitigated. *A* is the cost to the victim of their files being locked. *C* is the ransom value. *L* is the cost to the attacker of any effort expended on securing payment that is then not forthcoming (e.g. failed negotiation). *W* is the cost of the files being destroyed, as perceived by the victim.

This table doesn't account for the possibility of capture. Selten's model does account for this and can be found in report D4.1. For the sake of brevity, the above table provides the salient points for this discussion: there is a mutual cost-reduction objective that both attackers and victims are in pursuit of. Only the attacker has a direct profit motive during the attack, the defender must attempt to preserve their unrelated (but threatened) profit motive. It is therefore in the interests of the attacker to make it seem that *C* is the lowest cost available to any victim that has been infected despite defensive countermeasures. It is in the interests of the victim to encourage as many individuals as possible to invest in defensive measures, to starve attackers of potential income. This would raise the initial costs and time-to-launch of an attack, as weaknesses must be found in the defences in place. Educational campaigns informing victims of the risks of paying ransoms would enhance this effect.

Table 1 - The payoff of different outcomes in a game of kidnapping with the possibility of attack deterrence

OUTCOME	PAYOFFS	
	Attacker	**Victim**
NO ATTACK	0	*-E*
FAILED ATTACK	*-F*	*-A-E*
RELEASE OF FILES FOR RANSOM C	*C*	*-C-E*
RANSOM NOT PAID	*-L*	-W-E

The inclusion of game theory at this high-level of abstraction is very useful when seeking the optimal solution for the involved parties. However, it is not able to inform malware-specific improvements. Increasing one's ransom and encouraging payment to increase one's total profit towards the optimal makes sense, but the best method of achieving that goal remains unclear. For LEA's, knowing the steps towards this optimal is important as it will allow them to get ahead of criminally funded development and sink the costs associated with such development. If malware is being operated as a business, we must look to ways of forcing them out of business.

2.5 Limitations of the current economic model

The current model can achieve the following:

- Identify the optimal profit for a given malware strain
- Examine ransom strategy and suggest profit maximisation techniques
- Evaluate the cost of defence vs. cost of ransom for victims, reinforcing the fact that the constant cost of defence is exceeded by the initial and subsequent costs of paying a ransom
- Expose negative externalities harmful to the attackers – such as encouraging good data-security practices and running LEA-led public awareness programmes

It is a high-level economic model, which accounts for the transaction process, initial costs and ongoing expenses of a ransomware attack. However, it is not equipped to differentiate between infection strategies, scope, or changes in the market. RaaS and the proliferation of anonymity-focussed altcoins as an alternative to traditional cryptocurrency laundering are very specific events with a significant impact on the profitability of malware. The model must be extended to accommodate variables that consider the reduction of cost or increased efficiency of cost-bearing precursors to an attack.

Victim selection is coarsely addressed in the current model. Institutionally targeted malware is dealt with in the same way as randomly proliferating strains, as the model accounts for WTP in terms of individuals within a set of victims. Though this approach stands, and is the only reasonable way of estimating likely profitability, it can be improved. Extrapolating the costs associated with targeted attacks, and incorporating the ability to price discriminate based on known or soon-to-be-known intelligence, will allow us to make judgements about which improvements are most likely to yield the greatest returns.

Evaluating the cost and scope of a given infection vector is another way of improving the model. By identifying the attributes of botnet, email spam and other propagation methods, the most cost-effective methodologies may be identified. The current model only assumes flat costs, which can be improved with a better understanding of the cost of acquisition and scope of infection a given method is capable of.

The current model also accounts only for individual attacks. If we consider ransomware strains as products, we can extend the model to view the attackers as business with variable-term investment in the malware marketplace. This will allow us to make more detailed observations regarding the evolution of malware outside of one-shot enterprises seeking to maximise temporal profit over longer-term gains. The model also ignores the possibility that the attacker may seek to add value to their attack through data exfiltration and resale (among other profit-bearing criminal endeavours).

Section 3 will analyse the economic aspects of malware. Attention will be given to the technical requirements of development and maintenance, as well as the other initial costs associated with launching a malware-oriented business venture. We will also discuss the proliferation, pricing and value-maximisation strategies available to malware operators.

3 Deconstructing the Economic Attributes of Malware

Section 2 has provided an overview of the current economic model of malware as a business. This is important to our end goal, predictively modelling future malware development. We have identified the limitations of the current model, and provided an overview of the existing state-of-the-art presented in D4.1. Using this information, we now turn out attention to the gaps that our model does not yet address.

This section will provide a detailed analysis of malware, identifying attributes that contribute or detract from its profitability. As in D4.1, the discussion in this section will focus on ransomware, but is applicable to the wider malware context. The product of this section will be a comprehensive break-down of factors that must be accounted for in our improved economic model.

3.1 Initial costs and technical challenges

There are two sides one must consider when discussing the costs and challenges of malware (particularly ransomware). The attacker and defender have opposing objectives, but also share the burden of cost. The desire to infect machines to extract profit is on the attacker's side, which incurs the cost of devising and deploying malware that will yield profit. This places a sympathetic pressure on the defender, who must defend against the possibility of attack, not just existing attacks.

The attacker

The attacker must ensure the completion of multiple objectives, to ensure the success of their attack. They must secure a means of spreading their malware, infecting as many target (or random) machines as possible. Where attackers desire to infect many machines, zero-day vulnerabilities are required. Identifying these vulnerabilities may require extensive research or appropriate technical skills, which represent a significant opportunity cost to the individual tasked with exploring avenues of infection. Zero-day threats can, and often will, simply be bought. Many companies and government institutions will pay for reports leading to actionable, full-featured exploits. This represents a legitimate means of monetising an exploit – but once one submits the exploit it should be considered useless from that point. This means that any such payment is a one-off – the exploit loses further value once patched (depending on the efficacy of the patch). It is also possible that information may be changed upon, or derived, but an organised attacker is likely to expend resources on seeking out such issue.

A separate issue is the need for the attacker to ensure the correct functionality of their malware. Ransomware differs from most other malware, in that its objective is to be cryptographically sound. This requires in-depth cryptographic understanding, but releases the attacker from considerations such as persistence and even reduces the importance of stealth during the encryption phase of an attack. Attackers can choose between symmetric and asymmetric cryptography. Symmetric cryptography uses a single key for both encryption and decryption, and is faster than asymmetric cryptography. Asymmetric cryptography uses a dedicated key for encryption, and another for decryption. Both approaches pose significant technical challenges for the attacker.

There is a preference among attackers, for the use of symmetric cryptography with distinct keys for each encrypted file. However, this is an inefficient and unwieldy approach, as one would have to manage many thousands of keys per infected machines. This is not viable when one is already seeking to infect (at least) thousands of potential victims. There is the additional issue of key derivation – if the defender can derive key information they might be able to decrypt their files. As a result, an additional layer of encryption is desired, to protect against this eventuality. Hybrid encryption is the answer to this issue, allowing the attacker to use both symmetric and asymmetric cryptography. The process that follows is:

- Generate a public and private key after each file is encrypted with a per file symmetric key
- The symmetric key is then encrypted with the public key and appended to the file
- The private key is encrypted with a separate public key that is hardcoded in the ransomware, which is then sent to a server controlled by the attacker

This approach limits the communication between the attacker and victim (only the final encrypted key is sent), and protects against trivial discovery of the symmetric keys associated with each file. Though apparently simple, cryptographic systems are hard to implement correctly. Many ransomware strains have severe issues with then cryptographic implementation, leading to key derivation or complete loss of the encrypted data. Some primitive attacks have no intention of encrypting properly, and simply overwrite existing data with random data.

The defenders (victims and LEAs)

Many of the identifying characteristics of malware are well-known, and so easy to identify and mitigate if systems remain up-to-date and security best practice is observed. However, ransomware does not abide by most malware requirements as its strategic objectives are far simpler, despite their cryptographic complexities. Defenders, including victims and LEAs, do not currently have a robust definition of the characteristics that define ransomware attacks (Rajput, 2017). Recently, academic and industrial interest has turned to this issue, but we are not yet equipped with a complete picture of ransomware detection. This has, however, led to the development of definitions for several ransomware families and the ability to protect certain file types.

Slow detections and false negatives can be catastrophic when defending against ransomware. Failing to detect ransomware will allow it to encrypt the target file-system. Slow detection will allow it to partially encrypt the file system, which can still represent a significant cost if the encrypted files are important. As a result, the cost of poor detection capabilities is far higher when considering ransomware as opposed to other forms of malware. In the case of traditional malware, damage might be undone by reinstalling the machine, whereas ransomware will ensure that any data that is not backed up is irretrievable

Human identification of malware is incredibly important, due to the lack of robust autonomous detection methods. A Bitdefender study revealed that less than half of users can't accurately identify ransomware as a type of malware that prevents or limits access to computer data. Two-thirds can identify that it can harm computers. Figure 1 outlines the number of individuals affected by ransomware, as identified in the Bitdefender

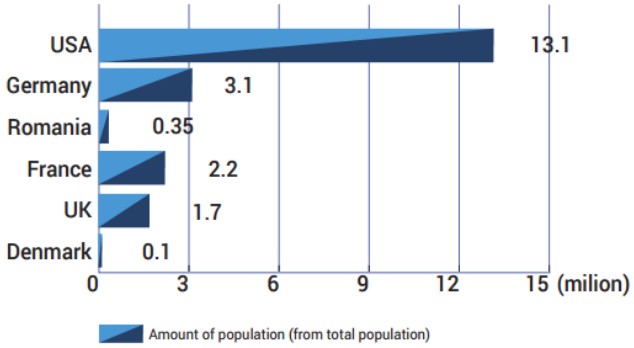

Figure 1 – Number of individuals affected by ransomware (Bitdefender)

report.

When one considers the lack of knowledge of ransomware possessed by the public, it is not difficult to extrapolate that this is where much of the cost to the defenders (as a whole) is incurred. Prevention is difficult when one doesn't have any knowledge of what is being prevented. Furthermore, appropriate response to ransomware attacks is difficult when one is being instructed by the attacker to pay the ransom to unlock files, and one lacks a frame of reference on which to judge their trustworthiness. The same report states that 32% of those unaffected by ransomware believe that it is improbable that they will be affected.

These findings show that the public is ill-informed regarding ransomware. They are unlikely to understand the appropriate response to ransom demands, and even have a high chance of paying a ransom. Figure 2 shows the stated likelihood of payment and the actual rate of payment of ransoms.

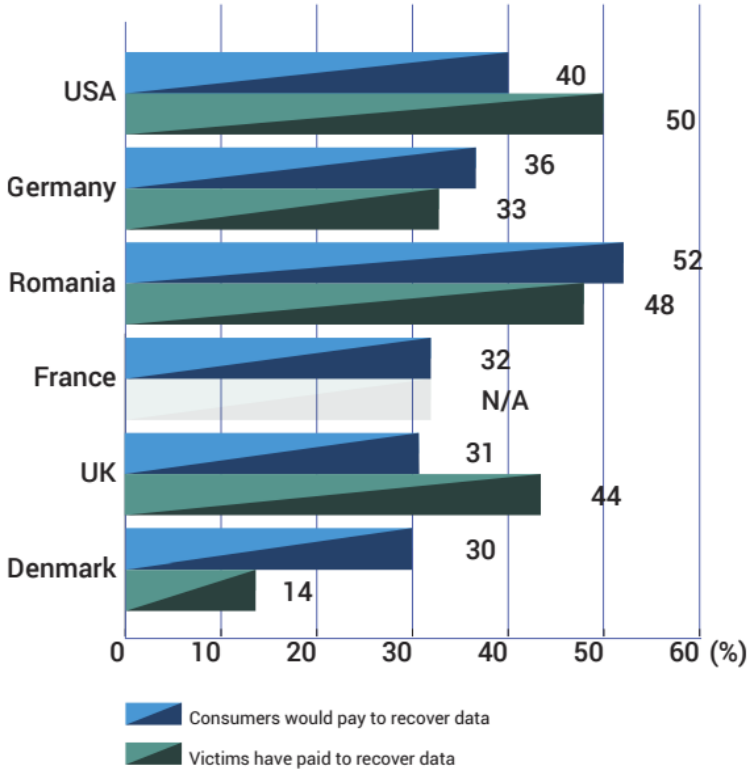

Figure 2 – Likelihood of ransom payment and actual rate of payment (Bitdefender)

Educational campaigns, using media preferred by malware operators, are likely to reduce these rather high figures. With 50% of infected US citizens paying ransoms and 52% of Romanians stating that they are likely to pay, individuals feel that this is a preferable option to the alternatives. Comparing the cost of ransom payment to that of offsite backups, or using social media campaigns to inform potential victims of how they may be infected, is a worthwhile economic consideration. The cost of such campaigns would be measured per failed attack or refusal to pay a ransom, and could be considered a means of magnifying the effects of negative externalities brought about by non-payment or implied bad faith (file deletion).

Associated costs and benefits

The costs associated with the challenges listed above can be represented simply as the variables *F* and *E* as shown in section 2. However, this does not allow us to make modifications to the costs based on the latest information, when we don't possess the full picture. To improve the model, one must account for objects in the sets *F* and *E*. We modify these to be indexed sets, which are indexed by the victims (*i*). :

Where *F* is the set of all costs associated with the attacker failing to infect a target (maximal cost):

- One must account for the possibility of failure by key derivation, where F_i is incurred because of subsequent decryption, not a failure to infect the target
    - This may not prevent payment if a technically inept target doesn't know keys are retrievable, but should heavily penalise the theoretical ransomware
- Where a sub-standard cryptographic implementation destroys all files on the machine of *i*, the cost of

the attack if the ransom is not paid (L_i) is substituted in place of F_i to reflect the destruction of files regardless of ransom payment
- The cost of ransomware may be trivially derived where RaaS involvement is identified – even if the strain has been co-opted or stolen, the current asking price for a given strain provides a guide-price

Where E is the set of all defensive costs and A is the cost associated defending against a specific attack:

- The cost E should be expanded to incorporate hardware countermeasures, such as offline storage
- E_i will vary depending on the nature of the target machine
    - Well-funded institutions should have a respectively higher E_i
    - Individuals will likely have a very low (or non-existent) E_i
    - Identifying the benefits of collective defence through negative externalities (successful defence and unwillingness to pay) may be possible through analysis of individual ransomware defence costs
- An additional variable should be considered alongside E to represent the role of LEA/Academic/Industry initiatives
    - This should account for awareness initiatives and preventative measures
- A_i will vary with each type of ransomware attack, and be modified heavily in exceptional circumstances:
    - Where the attacker deletes all files by accident, A_i becomes equivalent to W_i, as if all files had been deleted after not paying the ransom
    - Where the attacker deletes all files by accident, and the victim pays, W_i is supplemented by C_i
    - A variable approximating the impact of potential negative externalities is required to accurately model future effects
- A new component of E should be formalised, to reflect the cost of national and social media campaigns
    - The cost of such campaigns is universal, but should be modified upwards by failed infection and/or refusal to pay a ransom
    - This cost may have to be abstracted or a large body of ransomware strains, reflecting an organisational scale and defence against multiple ransomware strains

Increased detail in the initial costs calculation will allow more accurate predictions to be made. Differing adversarial scenarios, with defenders of varying competence and cost, can be used to identify potential strategies to enhance opposition of current strains. One may also play a form a wargame with these figures, creating hypothetical ransomware that is effective against current countermeasures. It may be possible to identify how effective communication and awareness strategies are when defence and recovery are unlikely.

Arguably, the most important ransomware-specific revelation that this research has uncovered, is the possibility for file destruction despite all intentions otherwise. The competence of an attacker can incur cost not only on their part, but it may have an extreme negative impact on the defender. This can also lead to negative externalities, wherein defenders assume that the deletion of files was due to malice, not incompetence. This can reduce the WTP for individuals affected by (or knowledgeable of) such attacks, as the trustworthiness of the ransomware operator may be called into question.

3.2 Propagation and control

Ransomware, like all malware, must infect machines to carry out its function. Unlike malware, ransomware is often non-specific in its target-selection. Malware will frequently aim to compromise a specific system or company, but ransomware targets the file-systems of infected computers. As such systems are ubiquitous, it is in the interests of ransomware operators to infect as many machines as possible. This interest may not be shared by all such criminals, some will have ancillary motivations (political/social/personal). As such, it is possible for ransomware operators to forsake optimal profit to achieve ancillary goals, and this will likely express itself at the propagation phase.

Viewing ransomware as a product instead of the entirety of a business helps to put such actions into context. Competition is not merely played out between attacker and defender in a ransomware infection. Ransomware operators exist in a finite marketplace, and compete for victims. The most widespread ransomware will likely outperform more targeted ransomware unless there are specific benefits to reducing the scope of an attack.

There are several scenarios in which a reduced scope may be of benefit:

- The exploit is system-specific, so limits the attacker to infecting that system
- The attacker possesses substantial intelligence regarding a given institution or demographic, helping them to utilise price-discrimination
- The attacker has non-profit motives for limiting their attack, such as disruption, political statements or personal motives

The first point is the most likely, but the number of systems of a given type is substantial. Even if an exploit is only able to affect a given type of machine, peripheral or other point of ingress, it is likely that many tens of thousands of those devices are in circulation. It is the province of malware, not ransomware, to attack specific integrated systems such as microcontrollers – as such this possibility may be ignored.

The second point ignores the possibility of running a two-tiered attack. The assumption here is that the attacker will restrict their activity to the entity about which they have substantial information. Some of the cost associated with gathering intelligence may be mitigated through reasonable assumptions. It may be reasonably assumed that schools and universities possess a large body of identifying information and operationally critical records. The same may be assumed of utilities companies, with the addition of bank details for customers. The impact of a ransomware attack can be assumed in such cases – operations requiring the locked data will cease and there may be considerable damage to the institution's reputation as employee/customer/student data is assumed to be compromised. It is cost effective for attackers to make educated assumptions, and target entities that can be safely assumed to hold data of value.

	Industry sector	% attacked with ransomware
1	Education	23
2	IT/Telecoms	22
3	Entertainment/Media	21
4	Financial Services	21
5	Construction	19
6	Government/public sector/defence	18
7	Manufacturing	18
8	Transport	17
9	Healthcare	16
10	Retail/wholesale/leisure	16

Figure 3 – Sectors affected by ransomware (Kaspersky Security Bulletin 2016)

Figure 3 shows the industries affected by ransomware. Academic institutions are a common target. A BitSight Insights study placed educational institutions as the #1 target, 1 in 10 having experienced some form of attack. Budgetary constraints and smaller IT teams have been cited as the contributing factors making these especially attractive targets. Any institution that possesses a large amount of personal and/or financial information, with an obligation to protect that data, is vulnerable (Mansfield-Devine, 2016).

This explains why targeted attacks are so attractive to current ransomware developers, but does not explain the lack of efforts to conduct targeted campaigns as part of a larger scatter-shot attack. It is theoretically possible to launch a fixed-price ransomware campaign against a general population, whilst dedicating increased effort to running a price discrimination effort against your actual target. There is no reason to limit oneself unless forced by circumstance – any such action will have a negative impact on achieving optimal profit.

The proposed model assumes a profit-driven motive. As a result, we will assume that other motivations are of minimal impact, but advise users of the predictive model to look out for anomalous cases. Where attackers appear to ignore all common optimisations that are found in contemporary malware, it may be worth investigating for ulterior motives.

Propagation channels

The scope of propagation can be limited by choice, or by defensive actions. How malware is propagated can be varied, and defensive measures are usually highly specific. As previously mentioned, we currently lack a comprehensive method of identifying new ransomware strains. This can make it very hard to identifying new strains until they have begun to infect machines, by which time damage is already done. Ransomware shares malware's many potential infection vectors.

Traffic redirection is achieved by covertly changing the destination address of a web query, or luring a victim to a malicious site through social engineering. The malicious site hosts malware, which is forwarded to the victim's machine using exploits in their browser, or by socially engineering them to download and execute a program. Once installed, the payload begins to exploit the target system's vulnerabilities to carry out its objective.

Email attachments work in much the same way, but instead rely predominantly on social engineering. These are common attacks against institutions, where intelligence about the format and content of expected mails may be derived easily. Individual targets are harder to reach through this method, but lotteries, scam businesses and other easy money schemes are attempted and do work on occasion.

Botnets are themselves supported by a form of malware that is usually first introduced to a machine by one of the other methods listed above. The machines co-opted into a botnet can be used to launch DDOS attacks, or other forms of automated mass-communication based attacks. A relevant example would be the use of a botnet to propagate malicious links via social media, or coordinate a mass email scam.

RaaS doesn't stop at the purchase of malware. Though it is possible to buy code then propagate it oneself, there's little reason to do so. For an additional fee on top of expenses, RaaS distributors are willing to coordinate any number of different propagation methods on the behalf of their customers. This is like cloud services offered by legitimate companies – the ransomware operator pays for access to software and hardware relevant to their needs and receives expert guidance in how best to utilise it.

Social media as a propagation and control channel

Social media is a potential propagation method. The sheer size of modern social networks, and their reliance on reactive moderation, means that malicious links may be spread far and wide before eventually being shut down. It is also possible for more sophisticated attackers to incorporate hidden contents to images, hiding malicious links within images or redirecting them to malicious sites when the image is clocked (a common feature to enlarge an image).

This propagation channel stands apart from those previously mentioned. Hiney at al. (2015) suggest that social media combines an ideal environment for social engineering and a platform for other forms of propagation. Malicious links on Twitter or Facebook have been an increasing source of infections. In 2016, more than 100,000 Facebook users were tricked into downloading an allegedly pornographic file, which contained a Trojan known as Magnet. This attack used the assumed trust between friends to convince people to click on links provided by friends who had already been infected. Multimedia content can be modified to appeal to certain demographics. This isn't an exact process – there's no guarantee that a given individual will interact with the file – but it does increase one's chances of infecting those who would be interested in the offered content. Masquerading as a product or service is a common tactic, when attempting to entice people to click on links or download files. Phishing emails, social media posts, and hacked accounts are all used to enable campaigns relying on this form of propagation.

Social media can serve a purpose beyond the initial propagation of malware. Command and control (C&C) can be established over Twitter, Facebook, Imgur and other forms of media sharing. A team from Cisco Talos (2017) published a report explaining the ROKRAT malware. This malware exploited Twitter, Yandex, and Mediafire as C&C communication channels. Twitter accounts were used to receive commands, while Yandex and Mediafire were used to store malicious files and stolen data. Locky ransomware was propagated through a Facebook spam campaign. Notably, this ransomware was embedded into the SVG image format. Ransomware embedded in images, specifically Locky, has also been identified on LinkedIn (Brewer, 2016).

Social media is an effective meeting of all the tools and circumstances required to launch effective attacks against a diverse audience. It is common for users of all technical levels to have social media accounts, and many offer insights to attackers regarding their likely income and job. Organisations also have a large presence, with professional accounts being a common aspect of many individual's social media activity. This makes target selection possible, but it is just as simple to rely on the inherent connectedness of social media to propagate infection through careless clicks and assumed trust.

Propagation and control costs

Traffic redirection and email scams are simple to incorporate into the economic model. The cost per infection is equivalent to the running cost of hardware, the initial cost of purchase in the case of RaaS, and/or the cost of acquiring the botnet(s) required to support one's propagation strategy. This is divided by the total number of infections to obtain the cost per infection. The cost per machine is likely to be very low, in the range of cents per infection. This makes the cost of infecting additional targets after the first extremely small, and encourages ransomware operators to infect as many targets as they can.

Black and Opacki (2016) identify extensive diversification of the dark markets associated with malware, noting increased divergence in recent years. Caballero et al. (2011) highlight the rise of pay-per-install (PPI) services. These act as an intermediary between attackers and their victims, and are an extension of the RaaS economy. The low-cost per infection is, like email and redirection, low. This allows malware operators to iterate on malware quickly, directing most of their efforts to focus on defeating countermeasures developed in the wake of successful attacks. This is a frequent occurrence, highlighting their concern about anti-virus signatures. This also goes some way to explaining why it has been so difficult to build a comprehensive library of ransomware stairs and associated signatures.

There has also been an increase in the use of steganography and image manipulation to distribute malware. In addition to the above propagation techniques, malware such as Stegoloader possesses significant advantages over previous iterations. Stegoloader scans the infected system to determine if it is an analysis machine (or honeypot) before downloading a PNG image, from which it extracts its main module. This is achieved using least-significant-bit steganography – a means of hiding information within other innocuous data (images or videos, usually). The ability to bypass detection by honeypots and other cyber-security analysis methods is a strong addition to any malware that seeks to cause a significant amount of damage (and commensurate profit).

Associated costs and benefits

Considering these points, the key economic factors associated with propagation and control can be summarised as:

- Per-infection costs are likely to be very low, but increase as the infection-prone demographic decreases in size
    - Information campaigns may be an effective means of combating infections that rely on assumed trust and careless activity
    - The economic model should account for the cost of education and awareness campaigns in defensive costs
    - Hybrid attacks, where targets are selected from an otherwise random propagation strategy, should be accounted for

- Social media is a one-stop shop for malware propagation
    - It will be useful to identify the efficacy of attacks that propagate via social media vs. those that don't
    - If a cost difference (effectively number of infections) can be identified, this should be incorporated into a classification system within the model
- Command and control can be conducted over free (social media) channels and in plain sight
    - Steganography can be used to maintain command and control in plain sight, by hiding commands inside images
    - The cost of malware development is reduced by keeping the product as simple as possible: assessing the impact of simple, hidden control mechanisms on the cost of malware development would enrich the model
    - The cost of command and control over alternative channels should be identified as a cost and as a variable when determining the likelihood of LEAs uncovering infrastructural elements of the malware – this will impact the likelihood of getting caught or shutdown (a significant terminal cost)
- Stealthy infection and avoidance of cyber-surveillance prior to the execution of an attack is highly valuable
    - This may be achieved by using precursor malware that scans systems to determine if they are used for analysis
    - Steganography and file manipulation are the most common means of delivering the payload of the malware, should the scout malware report that a system is viable for infection
    - The additional cost of development needs to be compared to the likely increase in infection rates prior to cyber-security specialists releasing a solution to the attack
    - This approach will also reduce the likelihood of being detected during initial propagation of the malware – which could destroy profitability if specialists derive a countermeasure before the malware spread significantly

The key point of this, for the economic model, is that infection costs decrease with the number of infections. There's a current trend for malware operators to run targeted attacks against businesses. The advice given in D4.1 was that the additional cost of infection is negligible, so should be ignored. This should be reconsidered, as it is likely that the propagation strategy chosen by the malware operator will have a significant impact on the associated costs. An improved model would be able to identify where selecting a smaller target demographic, for reasons of increased intelligence or higher ransom demands, is effective or not. This may inform LEAs about future trends, should the current spate of targeted attacks turn out to be unsustainable.

3.3 Intelligence gathering

Intelligence gathering can be active or passive. Active intelligence gathering requires research and/or infection of target systems, it is also dependent on the focus of the attack being on targets that one has intel for. Passive intelligence gathering can be as simple as educated assumptions (a utilities company is likely to have critical information such as customer bank details), and as sophisticated as a long-term study of the behaviour of previous targets. Intelligence gathering can be its own form of attack, such as the RAT used by Shade to gather information about infected machines prior to follow up attacks (Atanasova, 2016).

The main costs incurred by an attacker seeking to gather intelligence are encountered during acquisition and consolidation. The active acquisition of intel incurs many of the costs associated with mainstream malware attacks (development/acquisition and propagation). Analysis of a target's file system, hardware, or known traits (income or economic class) are all examples of active intelligence gathering. Such research may be performed in advance, but is most likely to become a real-time, machine learning assisted process.

Passive acquisition is less costly, but both incur the cost of processing. Processing costs reflect the time spent developing sophisticated data processing systems, or the time spent by individuals trawling through collected data for useful information about their targets. An example of this process would be the incorporation of data gathered about previous attacks into the development of future ones. It is likely that more organised groups

will begin to iteratively improve on their malware strains, using previous performance. Such groups will possess an estimation of the number of infections achieved, and specific numbers relating to ransoms paid at a given value.

The benefit of intel about one's target(s), is the ability to price discriminate more accurately, dramatically increasing profit towards the optimal. Such activity can be viewed as an additional cost incurred to increase the likelihood of payment and optimise the lifetime profitability of the ransomware strain for a target. The cost of gathering highly specific intelligence makes it unlikely that multiple targets will be considered, but is not outside the realm of possibility.

Associated costs and benefits

The key economic factors associated with intelligence gathering can be summarised as:

- Active intelligence gathering should be considered either as an additional cost variable, or as a component of *F*
    - If incorporated into *F*, this should reflect the specific target selection of the ransomware being modelled
    - When considering the evolution of ransomware, a separate variable is preferred, as this will allow analysts to account for the longevity of intelligence – longer-lived intel will reduce future costs that target the same demographic, so long it remains viable
- The impact of poorly sourced or misleading intelligence should be considered
    - It is in the interests of victims to provide false intelligence if there is a high possibility of targeted attacks
    - False intel will waste resources on the initial gathering effort, and may lead the operator to sub-optimally price discriminate
    - Poorly sourced intelligence may have a similar impact, but with a much lower possibility of LEAs/target demographics capitalising on the wastage

3.4 Adding and/or recouping value beyond the ransom

Extracting a ransom may be the primary motivation of ransomware operators, but it is far from the only monetisation strategy at their disposal. Malware in general has a diverse array of objectives, even when limiting oneself to the profit-bearing options. As a result, malware may take on the trappings of ransomware to hide its true goals, and vice versa. This can make deriving the goal of a given malware operator difficult, but merely adds further variables to the evaluation of the economic viability of the malware in question.

Theft and sale of data

The most obvious source of supplementary income, when considering ransomware, is the file system being held to ransom. Ransomware has effectively unimpeded access to unencrypted files on a victim's computer, and the operator will have the private keys required to decrypt any files that they wish to extract after the ransomware attack has begun. It is possible to convince victims to remain online even if common sense would suggest that they turn off their computer. This can be achieved by implementing a heartbeat function in the malware, which will delete the encrypted files if it isn't in regular communication with the operator. Should this convince a victim to remain online, then the operator may continue to interact with the target machine. Such steps do represent an additional cost, in terms of additional sophistication in the malware itself, and in the workload that operators will have to undertake or delegate.

The main issue with this approach is the labour-intensive sifting process required to find files that may have black market value. Financial records, personal details, credit card numbers; these have value but are not trivially derived from a raw dump of data from potentially thousands of machines. Such an attack would have to target specific individuals if the operator seeks to accrue profit in line with the true value of the files stolen. Otherwise, the operator may instead sell the data dump. This ignores the individual value of potentially high-

profit files, but offload labour onto the purchaser who pays an appropriately lower price per file. This directly correlates the potential profit of a data-theft to the chosen granularity of the attack (specific or non-specific files stolen). Targeted attacks have a further correlation with the level of intelligence the operator possesses about the chosen files/institutions. One must note that even within targeted attacks there is a high degree of granularity. Coarse grained attacks targeting medical or financial data have been shown to be highly effective and attractive to contemporary malware operators. More specific, institution targeted attacks are becoming more common, showing an increasing overlap with corporate espionage.

The sale of data may be a potential end goal, but getting there is not easy. Malware operators must exfiltrate the data without exposing themselves. Depending on the quantity of data this may be a challenge, but recent attacks against the NHS and other medical institutions demonstrate the capability of cyber-criminals in this regard. This means one can effectively discount the issues inherent in retrieving and storing data. The issue instead lies in the stealthy exfiltration of data. Figure 4 shows the basic flow of steganography, a means of hiding data within other data, such as images or videos. Steganography, among other forms of image and video manipulation, has been used to stealthily exfiltrate data for criminal resale (Sharma et al. 2016).

Steganography has been observed in malware attacks targeting companies, banks and medical institutions. Some of the more recent attacks present themselves as ransomware, but also extract data that is then forwarded via image or video to the operators. Image steganography is now falling behind video in the scope and depth of use by cyber-criminals. There exists a large body of steganalysis tools for image, but steganography over video and other data (VOIP, TCP/IP, etc) is poorly documented.

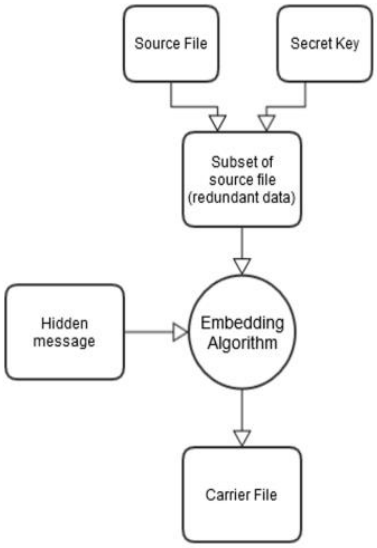

Figure 4 – The general process of steganography (Sloan and Hernandez-Castro 2015)

The use of steganography has increased, due to its ability to facilitate stealthy data extraction. Uploaded video and image content does not tend to flag as a security issue for most institutions. The transfer of such data is normal for the functioning of most companies, and so does not appear to be malicious. However, company financial records and personal medical details have been exfiltrated using video steganography. As a result, this is a viable method of increasing the profit of a cyber-criminal endeavour, by combining a ransomware attack with the opportunity to increase the income from people who pay ransoms, and recoup some value from those who do not.

The benefits of data exfiltration, and subsequent sale are many. Darknet markets are rife with such activity, proving that there is a demand for stolen data. This means that there is value in data exfiltration, but there are risks associated with it. The flow of stolen data to malware operators must, where possible, be masked to prevent LEAs from locating them and putting an end to their activity. It is also in the interests of the malware operator to stealthily exfiltrate data, to prevent premature shutdown of target machines. Coercion tactics may

also assist in keeping exfiltration channels open. This places additional cost burdens on the operator: technical sophistication (heartbeat/kill switch and steganography), marketing and handling of data and the potential for data exfiltration to reveal information about their operation. It is in the interests of victims to minimise the theft of data as costs can accrue rapidly. Loss of confidence, customer lawsuits and loss of competitive advantage can all result from the sale of confidential data.

One must also consider the use of data as an additional bargaining chip in ransom negotiations. With an increase in data theft, and the difficulty of determining whether data has been stolen prior to its disclosure, attackers could claim to possess information held by a company. They could claim that payment of the ransom would lead to the destruction of the stolen data, thus preserving the company's reputation. However, it is not possible to prove, beyond any doubt, that the stolen data has been deleted. The assumption that it has been is too dangerous, meaning that it is unlikely a victim will consider this an affordable risk. As a result, simple ransom of encrypted files is likely to remain an effective strategy, moving forwards.

Sale of intelligence

The sale of stolen data is one thing, but one must consider the sheer quantity of incidental data a malware operator will end up with concerning the efficacy of their own methods. RaaS as a growing phenomenon suggests that there will be an increased demand for data that leads to more effective strains. As groups specialise in producing, instead of deploying, malware, they will likely seek any advantage they can over their competitors.

Academic coverage of malware itself presents a source of information that may be used for iterative improvements. However, such data is of limited utility to malware operators – responsible disclosure prevents such publicly released information being useful to criminals for long. However, confidential data collected during attacks, supplemented by an elevated level of understanding of the wider malware market, could inform operators of potential improvements and better profit-bearing strategies. It is likely that there will be an increase in demand for statistics derived from successful malware attacks – specifically statistics regarding the ransom demands and payment/refusal ratios of ransomware strains. It is not just intelligence regarding victims that is valuable – the information that will improve the next ransomware strain exists in the data associated with the current and previous strains.

These facts lead one to the conclusion that even if such data is not being freely traded at present, it will become increasingly desirable as criminals specialise in the production of malware, instead of its use. This gravitation towards specialisation has the dual purpose of maximising one's ability to profit from one element of the malware lifecycle, whilst insulating oneself from the risks in adjacent markets.

Political activity

Although political motive is discounted as a primary concern in this report, it is worth discussing the impact of politically motivated attackers on profit-driven malware. It is not in the interests of a profit-motivated operator to suggest that they have political motives. Such claims bring the affiliations of the victim in to play, complicating the dynamic between operator and victims. Ransom payment is most likely if the victim's prejudices against the attacker extend only to the attack itself – allowing further factors to apply themselves is counter-productive. Positive alignment with political statements is unlikely to increase the possibility of payment, while opposition may diminish the victim's likelihood of paying.

On the other hand, malware is used by political activists, states and other actors with political or social messages. They use their attacks to bring attention to their cause, attack the machinery that supports their opponents or otherwise further their agenda. RanRan is a curious example of this type of ransomware, encrypting the victim's machine before demanding a political statement in exchange for a decryption key. This was a relatively unsophisticated attack, using symmetric cryptography and a re-used key. This allowed cyber-security experts to undo most of the damage caused by this attack. However, it does demonstrate that profit is far from the only motivation of malware operators. RanRan is a particularly overt example of politically motivated malware.

Sometimes politically motivated malware operators may not wish to disclose their political affiliation, and instead perform attacks that appear to be another form of attack. Ransomware attacks with low profitability, but a prominent level of visibility and disruption are perfect for this. It has been suggested that Wannacry was politically motivated. If the attention of the attacker is to diminish public opinion of affected institutions and sap confidence, there is no need to reveal a political motive. Outright destruction of data, or failure to support a robust ransom strategy, seem counter-productive what one assumes a profit motive. In such cases, it is likely that the attack is either conducted at a novice level, or has ulterior, likely political, motives.

Such attacks are most likely to harm the profitability of malware, if the operators are bad actors. If they do not honour ransoms, victims affected by profit-motivated malware are less likely to pay in the future. The use of malware as a cover for other activities will reduce profitability when operators do not act in accordance with profit-motivated best practice.

Associated costs and benefits

There are several immediately accessible paths to increasing value above and beyond ransom payments. Ransomware operators possess control over the host file system, allowing them to exfiltrate data so long as there's a viable connection to infected machines. They can constantly evaluate their own performance and collect data during an attack. As RaaS develops into a standalone enterprise, self-assessment and investigation of peers (whether passively or by active spying) will become vital to the iterative improvement of ransomware. With the development of effective malware being the sole focus of such groups, there will be a greater expenditure of resources to this end.

- The theft and sale of data is already the focus of some forms of malware
    - Currently ransomware and data exfiltration occur separately much of the time
    - The cost F of an attack is increased by adding further requirements to the malware
    - The cost of discovery must similarly increase
    - The potential gain is no longer restricted to C_i if data is successfully harvested
    - A new variable B_i is required to account for all files extracted where i is a file that has been successfully exfiltrated and has worth on the black market
- Improvement will not remain unguided, especially not with the increased specialisation of RaaS
    - The collection of data regarding malware efficacy can be done by operators and observers
    - One may publicly observe malware through the lens of media, institutional and academic reports
    - Cyber-criminals may share or steal such information
    - A variable representing the cost of intelligence gathering can be incorporated into the development costs of subsequent malware
    - There doesn't need to be a new variable to track the increase revenue generated by improvements – this must be tracked generationally, accounting for differences between collaborative/intelligence-fed development and isolated efforts
- The use of malware for political/social disruption will most likely work against the profit motive
    - Malware operators may have ulterior motives that make them bad actors relative to malware-for-profit, which can reduce confidence.
    - Modelling such malware as for-profit is still of use, as this can help determine the impact on global profitability after an observed 'bad faith' malware attack regardless of whether it served non-profit motives

Most of these factors can be incorporated into the existing model, with only one new variable required. This new variable is needed to provide a potential value for stolen files. Whilst the WTP of all victims determines the elasticity of demand, it is the Darknet market's WTP that determines the value of stolen files. Where files are used to incentivise payment, one may assume that WTP will increase with ancillary costs (loss of custom, confidence, employment). This is an advanced threat and should be considered an element of C, not of the file's resale value. One may consider a combination of these tactics, where the attacker sells the files regardless

of payment. This should be considered an act of bad faith. Negative impact will occur whether the attacker promised to not disclose/sell files, but will be higher if they did.

3.5 Extraction of value

Assuming some ransoms are paid, the average ransomware operator will now possess some form of cryptocurrency in several wallets. Prior to the invention of cryptocurrencies, wire-transfer and online-credit systems were used to facilitate the payment of ransoms. Contemporary attacks demand payment in cryptocurrency. Converting this back to fiat currency, or goods and services desirable to the operator and their colleagues, is a critical phase of an attack. Staff must be paid, future attacks financed and the desire for personal wealth is at the centre of any profit-motivated attack.

Bitcoin is the perennial favourite of malware operators (Meiklejohn et al, 2013). This cryptocurrency is well known and well regarded by both its users and many financial institutions. Despite reservations about the inherent instability of this digital currency, the value has increased year on year and continues to do so. Some ransomware attacks have taken advantage of this, holding Bitcoin in affiliated wallets prior to cashing out during a spike in value. Wannacry exhibited this behaviour even though it was a poor example of a profit-driven attack (and was likely not one).

The association of Bitcoin with malware is now so commonplace that many banks and well-funded companies hold Bitcoin as part of their cyber-security provision (Parker, 2016). Recent studies have also shown that many employees (up to 69% of those affected by ransomware and subsequently surveyed) will pay out of their own pockets rather than disclose the attack to their employers, if possible (Help Net Security, 2017). This is possible due to the speed of acquisition and accessibility of Bitcoin as a medium of exchange. These observations indicate a concerning willingness to pay, especially among businesses.

Report D6.2 provides detail on altcoins. This is suggested reading if one wishes to know more about these specific cryptocurrencies. This report focuses purely on economic factors, such as cryptocurrency friction, laundering, and exchange.

The costs of using Bitcoin

All cryptocurrencies have costs associated with their use, the most prominent being:

- Exchange fees when converting from fiat to cryptocurrency and vice versa
- Transaction fees

There are additional fees incurred during trading, but these can be summarised as exchange fees and brokerage. Both types of cost may be ignored for the purposes of our economic model – they reflect a separate phase of value increase which goes beyond the extraction of value from malware itself. If one wishes to model such activity, tracking the wallets associated with a malware attack may provide a means of tracking such activity. Cryptocurrency markets are generally unstable, so one must base any assumptions on the known or predicted proceeds of cyber-crime. This instability can be caused by a variety of factors, but in September of 2017, China's ban of initial coin offerings (ICOs) had a significant impact on Ethereum and related currencies. Events like this cause significant shifts in value, but many cryptocurrencies vary by double digit percentages on a weekly (if not daily) basis.

Instead, focus should be placed on the transaction fees as the primary source of cost during the initial stages of value extraction. Commonly, cyber-criminals will want to evade the scrutiny of LEAs and other interested parties. Bitcoin's erroneously propagated reputation for anonymity has been shattered this year, with several high-profile publications detailing how individuals may be identified by their transactions. A high degree of operational security is required on the part of malware operators, should they wish to extract their earnings whilst avoiding the attention of LEAs (Ruffing et al, 2014).

This leads back to transaction fees as the primary source of cost. Bitcoin tumblers (also known as mixers) are a way of obfuscating the flow of money by creating increasingly complex transactions from an origin wallet to an arbitrary number of additional wallets. This process can be repeated for any number of transactions, though it is theoretically capped by two variables: the time the criminal is willing to wait for the process to finish, and the amount of currency they are willing to lose to transaction fees and tumbler charges. de Balthasar and Hernandez-Castro (2017) have recently published work as a part of this project, analysing Bitcoin laundry services. Figure 5 shows the workflow of the Darklaunder service. This is one of the smaller Bitcoin laundries, but demonstrates the basic attributes of this service.

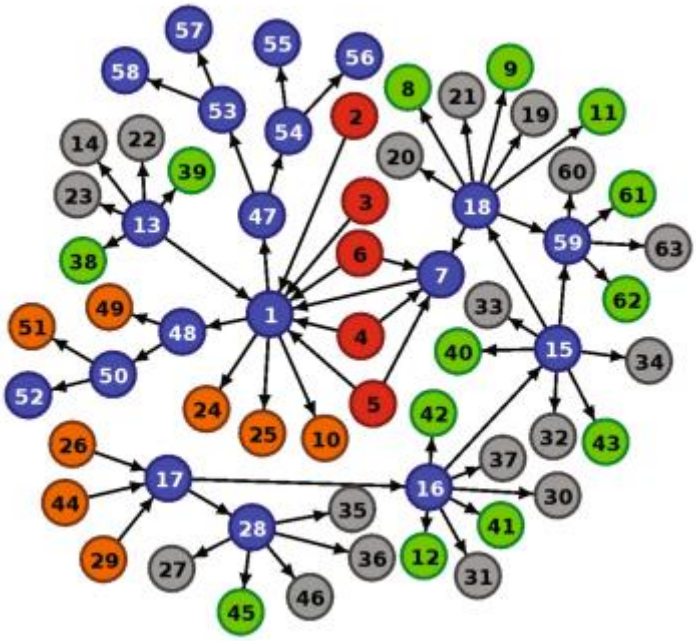

Figure 5 – Darklaunder workflow (de Balthasar and Hernandez-Castro, 2017)

This service is but one of many, but exhibits the traits shared by most laundry services. Bitcoin is sent to one central (1), and one change (7) address. This change address will receive credit at the end of the withdraw chain and send it to the central address. The central address may send Bitcoins to local Bitcoins directly, or form a peeling chain, as shown in Figure 6.

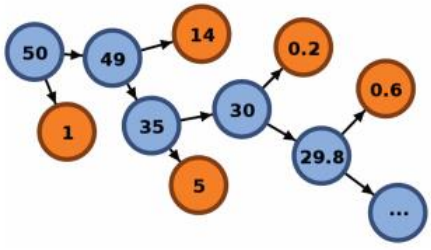

Figure 6 – How a peeling chain works (de Balthasar and Hernandez-Castro, 2017)

A peeling chain is the basic structure of a Bitcoin laundering service. The blue line shows the withdrawal chain. Bitcoin flows through this chain, occasionally being siphoned to orange nodes. Orange nodes are usually linked with other withdrawal chains, effectively mixing the currency from one chain with another. This makes it hard to trace where the Bitcoin is going, and what Bitcoin is associated with the original transaction. However, it is possible to follow this process, given a sufficiently developed understanding of the mixing methodology and the computational resources required to trace the transactions. It is also extremely costly.

The current average transaction fee demanded by the Bitcoin network is approximately $6.8 dollars (bitinfocharts.com, 24/11/2017 – value subject to change). Not all transaction fees are this high, whilst some may be higher, due to the highest-value-mined-first nature of mining. But this average has been growing as the network becomes increasingly congested. This is also a fee that is applied regardless of transaction value (but affected by the size of the transaction in bytes). One can begin to imagine the costs associated with Bitcoin laundering, when this figure is applied to a transaction graph like that shown in Figure 5.

These services are run by individuals who take a cut of whatever they launder, adding an additional cost to the already high transaction fees that Bitcoin demands (Moser et al., 2013). This means that using a tumbler may represent a significant (likely single-digit) percentage loss of profit. Some may consider this the price of safety, and wouldn't be far wrong, but there are alternatives. There is also a risk of theft – bad actors may run a laundering front that steals Bitcoins, or a malicious party may perform an attack against a laundering service that redirects credit to their own wallets. Reliance on third parties in the cryptocurrency space is a huge risk – especially for cyber-criminal endeavours.

With increasingly well-equipped LEAs, vulnerabilities in known major and minor laundry services, and prohibitive costs of operation; it is becoming clear that the time of Bitcoin mixing is short (Neilson et al., 2017). It may persist for some time, especially among privacy advocates and less informed cyber-criminals, but it is being superseded by low-cost, low-complexity altcoins that provide the same effect much more economically.

Altcoins and their benefits/downsides

For the purposes of this report, a Bitcoin mixer is assumed to stay single-currency throughout its mixing. Altcoins, however, offer another way of obfuscating one's activity. Report D6.2 highlighted two very important additions to the altcoin family: Monero and ZCash.

These currencies focus on anonymity as their primary goal. Their networks are far smaller than Bitcoin's, but benefit from advances in technology that allow them to demand lower transactions fees. This is aided by their lower rate of use – miners accept lower payments because they have less competing transactions to process. This can change during times of high activity, such as August 2017 where Monero experienced several days of delays to transactions due to high throughput. However, the fees for Monero transaction (at time of writing) have a media of $1.62 and a range of $0.38 to $12.45 (www.monero.how/Monero-transaction-fees, 2017). This is substantially lower than Bitcoin. ZCash demands even lower transaction fees, with an average of $0.02 (Bitinfocharts.com/zcash/, 2017).

This is a substantial advantage over Bitcoin. However, these currencies are harder to acquire – they do not have direct fiat-to-cryptocurrency options in nearly as much abundance as Bitcoin. For this reason, malware operators tend to use Bitcoin as their requested payment. They are not under any obligations to keep their ill-gotten-gains in that currency, though.

Services such as Shapeshift.io allow small-value exchange between cryptocurrencies without identification of any kind, unlike most trading platforms. They do not offer a fiat-conversion service, but do allow individuals to convert Bitcoin to Monero and ZCash for a nominal fee in addition to the transaction fee. This service has also implemented a well-documented API, that online stores (legitimate or otherwise) can use to convert Bitcoin to altcoins automatically. There is no documented evidence of such a method being employed by a malware operator yet, but there is evidence of the Wannacry wallets being converted to Monero after a series of Bitcoin transactions.

The number of transactions required to erase records of one's activity in Monero or ZCash is far smaller than that of a traditional Bitcoin mixing service. Theoretically, a single Monero-to-Monero transaction will offer almost complete anonymity, as all transactions are bundled with transactions of equal value made in the same epoch. Stealth addresses obfuscate the sending address, and the grouping of transactions with identical value transactions makes it very hard to derive which inputs match their respective outputs. The result is a fungible

currency – each Monero is its own entity and cannot be associated with previous transactions beyond the immediately previous. ZCash offers a similar benefit, but through use of zero-knowledge proofs which allow transactions to be resolved without either party having to reveal any identifying information throughout the process.

The effectiveness of these currencies is improved by conducting many transactions between unaffiliated wallets. It is also dependent on good operational security – any LEA can trace wallets that have identifying information associated with them if they gain access to hardware or individuals associated with criminal activity. The cost of multiple ZCash and Monero transactions is low, and operational security is a concern even when using Bitcoin (which has a far higher rate of information leakage considering its transparent blockchain).

It is our expectation that ransomware will continue to demand Bitcoin. Operators will, however, increasingly convert to Monero and ZCash to hide their value extraction practices from LEAs. This will be hard to counter – the cost to malware operators is low while the cost of increased digital forensics and traditional police work to deal with increased anonymity is high. The cooperation of exchanges and Shapeshift has been given in the past, but to limited effect. The decentralised nature of cryptocurrency, the feature that initially attracted cyber-criminals to it, will continue to develop and provide services that neither exchanges nor advocates wish to support but have no recourse to prevent without substantial evidence of wrongdoing.

Costs associated with cryptocurrency and less technically adept users

The technical and operational costs of laundering one's ill-gotten Bitcoin can be considered the penultimate cost of the malware monetisation process. The only thing left to do is pay the exchange rates and broker's fees required to turn one's cryptocurrency into fiat currency. However, there is a cost that permeates this entire process, and it may not be the only option available. An increasing number of services accept cryptocurrencies – Darknet markets prefer taking this form of payment wherever possible. As a result, the question isn't just 'How do attackers extract fiat value?' but 'What are the alternatives that can also result in material gain?'. There is a very real possibility that ransomware could be used to finance subsequent criminal activity. Currently, the extraction of fiat value from cryptocurrency is common to profit-driven attacks, but this may not be the case in the future – especially if legitimate services show an increasing acceptance of cryptocurrency.

Technical skill cannot be assumed in any technological endeavour. This is no more applicable than it is with a random selection of individuals from a variety of national, educational, and professional backgrounds. Individual exposure to what ransomware, malware, and cryptocurrency is diverse. Some may be relatively up to date with one or more of these topics, while others have zero knowledge until they become victims of a ransomware attack. This is the problem of the ransomware operator, should they wish to achieve optimal profits.

Technically inept victims may not know how to acquire Bitcoin, their trust will be exceedingly low and even written instructions may be met with confusion or scorn. It is therefore in the interests of ransomware operators to provide a form of customer service. This was discussed in D4.1, where F-Secure articles from 2014, 2015, and 2016 showed that many of the major ransomware strains had invested in some form of webchat or telephone support. This additional cost has not yet been quantified, but it is likely to increase the likelihood of payment form uncertain victims.

The technical complexities of cryptocurrency are also a key factor when determining why malware operators seem to stick with Bitcoin despite better, anonymous currencies exist. Bitcoin is known in popular culture; most people know of it even if they don't understand it. It is very well documented, and has been designed over the last 6 years to be increasingly usable to the public. As a result, it is the default and most realistic choice when determining what to ask for when holding someone's files to ransom. An interesting mutual reinforcement effect can be observed between companies holding Bitcoin as part of their cyber-security strategy and ransomware operators – cyber-criminals are more likely to request Bitcoin that is accessible to the victim instead of a currency that may take more effort to acquire.

When considering the economics of malware as a business model, ease of acquisition, payment and resolution are key. The operator wants the transaction to go smoothly, and as quickly as possible, to increase their likelihood of being paid and to accrue as much profit as possible before the lifespan of their malware expires (due to its propagation/infection vector being closed off). But extraction of value allows more exotic strategies to flourish, and increasing diversity among cryptocurrencies, alongside opaque blockchain implementations, offers a wealth of opportunity for cyber-criminals to cut costs and maximise profits in the closing stages of an attack.

Associated costs and benefits

The choice of cryptocurrency and laundering method affects the entire transactional process for any malware. As a result, these choices and costs must be reflected in an accurate economic model of malware as a business. A summary of these attributes can be found below:

- Any cryptocurrency will incur transaction fees, but this varies greatly between currencies. This may represent a pressure to migrate between currencies, rather than a set cost.
    - Bitcoin has high fees, but high exposure and popularity
    - The Lightening Network Bitcoin Fork will have essentially zero fees, which will have an impact on the operational costs associated with laundering services
    - If public perception, trust, and familiarity with a cryptocurrency is sufficiently high; the cost of transaction fees may be justified by lower costs associated with customer support
    - It is almost certain that a ransomware strain will make demands in Bitcoin – Ethereum and Monero have been requested, but have thus far not been as successful as Bitcoin-based ransoms
- Bitcoin and other *transparent blockchain* solutions require laundering to prevent personal details exposed at the point of conversion to fiat being related back to the criminally tainted cryptocurrency
    - Transaction fees will be incurred at every step of a laundering attempt, making even DIY laundering a costly proposition
    - Monero and ZCash (among others) offer a low-cost, low-complexity alternative, but are not well-known, limiting their exposure among potential victims
    - *C* should not be considered the sum of all profit, but the sum of all ransoms paid
        - A final variable, *P*, should be considered, resulting from *C* minus all transaction, laundering and broker's fees (*T*)
        - *T* could also incorporate the cost of asset/individual seizure by LEAs or exchanges, should this prove useful to the model's user
- The technical skill of a victim has a direct impact on the likelihood of a ransom being paid
    - WTP should be supplemented with an Ability to Pay (ATP) modifier based on the ability of a victim to pay and the cyber-criminal's support network to assist in that process
    - Expenditure on such support infrastructure can be assumed to be an element of *F*
- The technical challenges an LEA is likely to endure result directly from the value extraction strategy of the ransomware operator
    - The likelihood of funds being intercepted, based on expenditure of prevention, should be considered
    - This is an inversion of the operator-victim dynamic – the operator becomes the potential victim and they invest time and resources into countermeasures against LEA intervention in their value extraction process

This is a complex element of the malware monetisation process. It demands special attention in the proposed model. As this is the point at which criminals are most likely to expose themselves to detection by the authorities, it also represents some critical cost-reduction/risk-mitigation strategies.

3.6 Organisational considerations

The final element that we consider in this section, is the transition from a malware-as-business model to a malware-as-product model. Malware strains may be relatively long-lived, but tend towards short operational lifespans for individual attacks. This makes them far more akin to product launches than full businesses.

As this is currently a poorly documented phenomenon, this section will be kept brief, but will explain our intention to move towards a *business model of malware*.

- Identifying malware operators is difficult, so the abstraction of malware as a product must make assumptions based on best-evidence
    - Where identification of individual malware operators is not possible, the model must be able to abstract both one-shot and sustained business models
    - WTP is calculated per attack based on current factors associated with the victim, but the model must account for the effect of previous good/bad actors on the perception of all malware operators
    - Distinct strategic differences may be considered as classes of business
- As malware operators are in competition, game-theoretic and competition models apply as they would in traditional economic scenarios
    - Competition will initially drive diversification – successful malware will build on success, while newcomers will look for new strategies to maximise their exposure, mitigate risk and increase profitability
    - Convergence to optimal profit may be assumed, but there's no indication that this is a mature marketplace – there is likely to be much diversification at this point
    - Diversification may be technical, target-oriented, and/or strategic
    - It is likely that we have not yet observed the optimal strategy for ransomware at this stage of development – low returns in recent attacks suggest ulterior motives or primitive monetisation strategies
- Diversification has already been observed in this area, with complementary services arising and profiting
    - Diversification is not restricted to malware, it can also apply to the supporting market
        - RaaS is an example of a specialist business within the wider marketplace
    - Likely sectors that will result in their own specialist industries may include:
        - Intelligence gathering
        - Propagation/preliminary infection with supporting malware
        - Malware development
        - Attack management (orchestration/customer service)
        - Laundering and money-handling
- Criminal gangs are likely to take an interest in malware and may become a dominant force
    - This will drive diversification as such organisations are highly compartmentalised
    - Compartmentalisation mitigates risk, by ensuring that one aspect of the malware production and deployment process doesn't compromise the next (at least not completely)
    - Criminal organisations have the benefit of central leadership, conducted in a decentralised manner
        - This makes the organisation hard to dismantle
        - This preserves knowledge even when some elements of the organisation may be compromised
    - Higher levels of organisation and more intense competition with other organisations will drive a higher pace of improvement in all malware-related sectors

The higher level of organisation within criminal gangs will likely focus malware operators on achieving cost effective, high profit results in their respective fields. This will also weed out misguided, idealistic, or hidebound approaches – tolerance for sub-par performance is likely low and attacks will iterate on previous ones quickly. Greater awareness of the cyber-crime marketplace will also force malware operators into closer

proximity. It is currently possible for operators to act in ignorance of one-another, but criminal gangs are likely to want to directly affect the competition, or at least carve out as much space as possible in the market for themselves.

This bring section 3 to a close. Having identified key economic attributes of malware as a part of a larger business, we now turn our attention to the specific improvements planned for the model proposed in report D4.1 (summarised in section 2).

4 Proposed Improvements to the Malware-specific Economic Model

Section 3 outlined attributes that have yet to be incorporated into our economic model. These improvements are the focal point of this report. However, their incorporation, and the extensions that must be made in both the process and output of the model, requires specific attention. Where the initial model provided a general idea of the optimal state, it did not present any detailed information regarding the next logical steps one may take to improve malware profitability. Such a model would require additional inputs, processes and outputs to deliver a strain-specific analysis of the likely modifications, their potential costs, and benefits.

This section summarises proposed enhancements to the existing economic model of malware as a business. These enhancements are aimed towards D4.3, which will provide a confidential report on the improved model and its predictions. The model proposed here will be iterated on, culminating in a software system (D4.4) that employs the principles presented here to allow LEAs to map out the likely evolution of malware strains when considering profitability as the key motivator.

4.1 Extending the existing economic model

Work on an improved economic model of ransomware is ongoing, but this subsection details some suggested improvements in brief. The current model derives the optimal price of an existing ransomware strain, indicating the likely next step in development with a focus on the demanded value of a ransom. It is a game theoretic model, derived from work undertaken by Selten, and Lapan & Sandler (1988). As we have reported in this document, there are many factors to consider in this optimal price calculation.

The optimal ransom is one that will deliver the greatest total return over the lifetime of a given ransomware strain. An extension of the previous model should account for the minutiae that go into identifying the optimal ransom (be it fixed or price discriminatory). It should also be scalable, applicable on individual strains and able to incorporate the performance and externalities of a given strain into a wider model of the ransomware marketplace. The goal, therefore, is to model the business practices supporting malware as a for-profit product – malware can be viewed as what the business offers, not the business.

A key assumption that must be implemented, is that of the increasing rewards of good behaviour. Ransomware operators gain nothing by not returning files. The more bad behaviour is perceived by victims, the lower the ransom demand must be to be considered an acceptable risk.

Importantly, the improved model must account for the adversarial nature of the malware marketplace. LEAs and competing malware operators are likely to act against a given operator. This exceeds the level of adversarial activity one might encounter in standard models of competition – with far higher stakes. LEAs will act to decrease the likelihood of ransom payment, but their end-goal will always be to shutdown illegal activity where possible. As a result, malware operators must adopt both attacking and defensive postures as part of a complete and robust profit-bearing operation.

The extended model that will be developed as part of the ongoing work for WP4, will add to the current model in the following ways:

- Profit optimisation will be considered at a product and business level
    - Individual malware strains will be viewed as products
    - Product profitability can be used as the entire output of a one-shot venture, or part of an ongoing business strategy
    - Businesses of varying size and complexity can be modelling by increasing their repertoire of hypothetical (or real) malware strains and supporting activities
- Malware-specific economic attributes, as discussed in section 3, will replace more general assumptions
    - Initial costs will be modelled to account for the inherently adversarial nature of this market
    - Criminal-specific risks, such as capture and closure, will be modelled
    - As malware operates outside of legal markets and fiat currencies, the risks and costs associated

with cryptocurrency must be considered
    - This must also account for the possibility of exchanging via multiple altcoins as part of an LEA avoidance strategy
- The possibility for increased compartmentalisation in the marketplace will be accounted for

4.2 Properties of an optimal ransomware from an economic perspective

This sub-section summarises the findings of section 3. The properties that optimal ransomware should exhibit, are summarised as follows:

- Criminals should actively search for the optimal ransom demand
    - The current demand is too low
    - Criminals could seek optimal profits by way of increasingly sophisticated analysis techniques:
        - Rudimentary analysis of the data that they generate through their own attacks (how many pay the ransom out of the total infected population)
        - Actively varying the ransom amount between successive attacks to compare differing price points and their effect on total profit
        - Perform market analysis by conducting the previous two points in cooperation with multiple sources. This is a potential data market
- Criminals should price discriminate
    - Intelligence gathering is key to this approach, increasingly detailed intelligence will allow increasingly sophisticated price discrimination
    - Sophistication correlates with granularity of discrimination:
        - A blunt pricing mechanism could be based on the amount of data encrypted and improved through trial and error
        - Machine learning could be incorporated to update the optimal pricing algorithm, attempting to establish WTP based on hardware cost and encrypted file types/contents (Abrams, 2016)
        - Identifying information about the victim or institution may be used to perform additional analysis to determine likely WTP and ability to pay
- The distinction between targeted and ransom attacks should diminish
    - As sophistication of analysis techniques improves and the body of experience increases in size, operators will be able to make real-time valuations of their targets
    - This will allow random attacks to harness the benefits of intelligence-fed targeted ransoms
    - There is no reason why this will not be witnessed in a transitional phase – where easily derived information informs a primitive learning algorithm and a flat ransom is offered in cases where the algorithm is unable to accurately predict a higher ransom
- Ransomware and profit-driven malware could merge with other forms of cyber-crime
    - Price discrimination isn't the only possibility when one has intelligence about the target's files – resale and theft of data are examples of added value
    - Protection racketeering is another possibility – with operators contacting companies to elicit protection monies prior to an attack, rather than consequently
- The cultivation of a good reputation is in the interests of individual malware operators
    - Improved customer service is expected
    - More robust and reliable encryption should be expected
    - Improving the reliability of ransomware, to prevent accidental destruction of files, is beneficial to operators as it helps maintain trust in the long run by ensuring files can be released back to paying victims
    - Criminals are more likely to honour payments and return files
    - This may lead to some form of self-regulation within the malware marketplace
        - Good reputation is everything, so bad actors must be removed
        - Increased documentation and communication about good ransomware practice may result
- User experience and interface design can be used to present the victim with information more likely

to elicit a payment
    - Increasing focus on aesthetics and appropriate communication strategies should be expected
    - Current ransomware screens are very basic, but conform to the basic tenets of good UI design (present clear, concise information where possible)
- More specialisation will be encountered in future malware attacks
    - RaaS is evidence of this
    - Further specialisation is likely:
        - Design
        - Propagation
        - Programming
        - Communication
        - Laundering/monetisation
        - Long-term strategy
- Malware operators will be best served viewing their malware as a product with a finite lifespan
    - This means it is likely that malware operators will view their work as part of a larger business
    - This will support specialisation as discussed above
    - It is possible that multiple malware strains could be part of a larger organisation benefitting from multiple income streams
- Criminals will switch from Bitcoin to other forms of cryptocurrency
    - Current studies suggest that criminals will ask for ransoms in Bitcoin, but exchange to anonymity-preserving currencies
    - The limitation of Bitcoin is its transparency
    - Bitcoin is high-cost for transactions
    - Once public knowledge of alternative currencies rises, it is likely that switches to other currencies will be successful
        - At present it is unlikely that Monero would be a likely candidate for the ransom collection phase due to lack of public knowledge
        - Ethereum is better known, but doesn't yet show any signs of widespread adoption in ransomware – it also lacks anonymization features, making it worse than Bitcoin in terms of popularity, with little incentive to adopt as it doesn't boast any particularly useful features for ransomware operators

Our current assumption is that criminals do not yet have this mentality. This is borne out by current studies, which suggest that ransomware is still in its infancy as an effective money-making crime. The sophistication and pace of ransomware deployment will likely increase to put additional strain on LEA and cyber-security institutions. This may be intentional, or a consequence of groups rushing to exploit new gaps in the cyber-crime market as they become well-known. As the organisational structures conducting such attacks become more sophisticated, such incidental occurrences are likely to become more intense and intentional.

5 Conclusion

This report has provided a detailed breakdown of areas in which the current economic model of ransomware, proposed in D4.1, can be improved. Specific improvements will be detailed in D4.3 (confidential), and incorporated into D4.4 (a predictive software implementation of the model).

The limitations of the current model have been presented. The current model is useful as an indicator of the current state of malware as a for-profit enterprise. In this, is has excelled, helping the consortium to identify that ransomware (in particular) is at a very primitive stage in terms of negotiation strategy and extraction of value. Current ransomware operators display some basic understanding of WTP and the effect that ransom price has on likelihood of payment, but have overcompensated or made erroneous assumptions based on this data. Many ransoms are set too low, with a focus on increasing number of ransoms paid, not the per ransom payment. Furthermore, price discrimination is performed to the exclusion of random attacks – operators will select targets and focus on them, rather than using partial knowledge to allow them to price discriminate when dealing with targets about which they possess intelligence, whilst applying best-guess flat rate ransoms to unknown targets.

Though the existing model provides evidence of these shortcomings in contemporary ransomware, it doesn't model the minutiae of cyber-criminal activity. The highly adversarial environment in which cyber-criminals operate makes for a unique risk-model. This will have a significant impact on any economic model of malware as a business. Furthermore, malware strains are short-lived. This makes them better suited to being viewed as a product or service, instead of as a business unto themselves. The current model does not facilitate this.

An improved model will account for the unique forces that act on malware-based profit-motivated organisations. A capacity for higher levels of organisation, sophisticated intelligence gathering, and LEA/cyber-security avoidance strategies will be required. The improved model must support the possibility of competing entities, as well as the dual attack/defence postures that must be adopted by ransomware operators when dealing with victims and LEAs respectively. The competing demands of maximising exposure and propagation of malware, whilst avoiding early detection and premature prevention of the attack must be represented. The risks and costs associated with extract real money value from the cryptocurrency-based ransom process must also be modelled, and this report provides guidelines as to how these variables may be incorporated into a next-generation economic model of malware as a business/product.

This document stands as an analysis of the requirements of the improved model, and tentative design document. The full details of the improved model are yet to be formalised, but will draw upon the work presented in this report.

5.1 Future work

D4.3 and D4.4 will draw on this document for their requirements. Using D4.1 and this document, both deliverables will incorporate our findings into a more sophisticated model. The initial model will be capable of per-strain analysis and prediction. This will allow LEAs and academics to identify the shortcomings in existing malware strains, and determine the likely developments that will provide the greatest profit for the least additional cost. Cost cutting measures may also be identified, where high-cost practices are being used in the place of existing, low-cost alternatives. This model will be iterated upon, with the D4.3 model being developed into a software implementation (D4.4) that will provide more detailed predictions. It is our intention to provide a model that is adaptable to changing conditions, supporting the likely increase in specialisation and organisation in the groups engaged in this type of malware.

References

http://www.ramses2020.eu

Atanasova, S. (2016). The Shade Ransomware with New RAT Features to Determine Worthwhile Victims Virus Guides' Computer Security News of August 12, 2016. http://virusguides.com/shade-ransomware-new-rat-features-determine-worthwhile-victims/

Abrams, L. (2016). Fantom Ransomware derives Ransom Amount and Address from Filename Bleeping Computer News of September 21, 2016. https://www.bleepingcomputer.com/news/security/fantom-ransomware-derives-ransom-amount-and-addressfrom-filename/ ransomware-hit-you

Black, P., and Opacki, J., "Anti-analysis trends in banking malware," *2016 11th Int. Conf. Malicious Unwanted Software, MALWARE 2016*, pp. 129–135, 2017.

Brewer, R., 2016. Ransomware attacks: detection, prevention and cure. *Network Security*, *2016*(9), pp.5-9.

Caballero, J., Grier, C., Kreibich, C, Paxson, V., and Berkeley, U. C., "Measuring Pay-per-Install : The Commoditization of Malware Distribution," *USENIX Secur. Symp.*, pp. 13–13, 2011.

de Balthasar, T. and Hernandez-Castro, J., 2017, November. An Analysis of Bitcoin Laundry Services. In *Nordic Conference on Secure IT Systems* (pp. 297-312). Springer, Cham.

Hernandez-Castro, J., Cartwright, E. and Stepanova, A., 2017. Economic Analysis of Ransomware.

Hernandez-Castro, J., Cartwright, E. and Stepanova, A., 2017. Ransomware and Game Theoretic Insights on Kidnapping.

Hiney, J., Dakve, T., Szczypiorski, K., and Gaj, K., "Using Facebook for Image Steganography," *2015 10th Int. Conf. Availability, Reliab. Secur.*, pp. 442–447, 2015.

Lapan, H. E., & Sandler, T. (1993). Terrorism and signalling. European Journal of Political Economy, 9(3), 383-397.

Lapan, H. E., & Sandler, T. (1988). To bargain or not to bargain: That is the question. The American Economic Review, 78(2), 16-21.

Mansfield-Devine, S., 2016. Ransomware: taking businesses hostage. Network Security, 2016(10), pp.8-17.

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S.: A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference, pp. 127–140. ACM (2013)

Moser, M., Bohme, R., Breuker, D.: An inquiry into money laundering tools in the Bitcoin ecosystem. In: eCrime Researchers Summit (eCRS), 2013, pp. 1–14. IEEE (2013)

Muthoo, A. (1999). Bargaining theory with applications. Cambridge University Press.

Neilson, D., Hara, S. and Mitchell, I., 2017, January. Bitcoin forensics: a tutorial. In International Conference on Global Security, Safety, and Sustainability (pp. 12-26). Springer, Cham.

Parker, L. (2016). Large UK businesses are holding bitcoin to pay ransoms. Bravenewcoin.com News, 9 June 2016. http://bravenewcoin.com/news/large-uk-businesses-holding-bitcoin-to-pay-ransoms/

Pfleeger, S. L., & Caputo, D. D. (2012). Leveraging behavioral science to mitigate cyber security risk. Computers & security, 31(4), 597-611.

Rajput, T.S., 2017. Evolving Threat Agents: Ransomware and their Variants. *International Journal of Computer Applications*, *164*(7).

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. Econometrica, 50: 97-109. Selten, R. (1977). A simple game model of kidnapping. Lecture Notes in Economics and Mathematical Systems 141: pp 139-155.

Ruffing, T., Moreno-Sanchez, P. and Kate, A., 2014, September. CoinShuffle: Practical decentralized coin mixing for Bitcoin. In European Symposium on Research in Computer Security (pp. 345-364). Springer, Cham.

Sandler, T. (2003). Terrorism & game theory. Simulation & Gaming, 34(3), 319-337.

Sandler, T., & Arce, D. G. (2007). Terrorism: a game-theoretic approach. Handbook of Defence economics, 2, 775-813.

Selten, R. (1988). A simple game model of kidnapping. In Models of strategic rationality (pp. 77-93). Springer Netherlands.

Sharma, V., Jha, S., and Bharti, R. K., "Image Forgery and it ' s Detection Technique : A Review," *Irjet*, vol. 3, pp. 756–762, 2016.

Varian, H. R. (1989). Price discrimination. Handbook of industrial organization Volume 1, R. Schmalensee and R. Willig (Eds.) Elsevier: North Holland, pp. 597-654.

Young, A., & Yung, M. (1996). Cryptovirology: Extortion-based security threats and countermeasures. Security and Privacy Proceedings IEEE Symposium. IEEE.