Game Theoretic Analysis of Ransomware: Identifying and Mitigating Motivators to Pay

Information Security Group
Darren Hurley-Smith

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Outline

- Economic Modelling of Ransomware
- Willingness to Pay
- Blockchain-based Extortion
- Price Discrimination
- Disincentivising Payment: A matter of cost

Introduction

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Research Projects:
  - REVOKE: Key Revocation to Mitigate Extortion in Ethereum Proof-of-Stake Validators. Ethereum Foundation, 2022-2023, Dan O'Keefe, Darren Hurley-Smith, Alpesh Bhudia. https://blog.ethereum.org/2022/07/29/academic-grants-grantee-announce
  - RAMSES H2020 (2016-2020): Identifying and Tracking the money-flow of Financially Motivated Cybercrime. https://ramses2020.eu

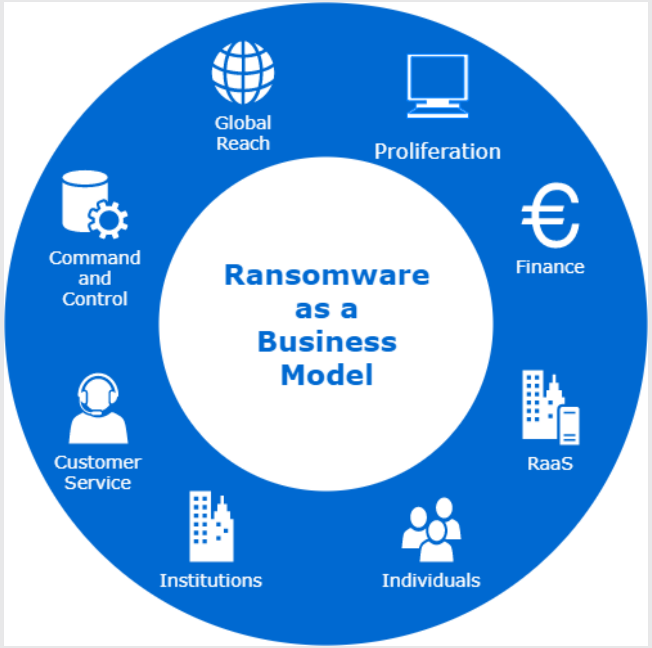
Introduction

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Recent publications:
  - Bhudia, A., Cartwright, A., Cartwright, E., Hernandez-Castro, J. and Hurley-Smith, D., 2022, September. Identifying Incentives for Extortion in Proof of Stake Consensus Protocols. In The International Conference on Deep Learning, Big Data and Blockchain (DBB 2022) (pp. 109-118). Cham: Springer International Publishing.
  - A. Bhudia, A. Cartwright, E. Cartwright, J. Hernandez-Castro and D. Hurley-Smith, "Extortion of a Staking Pool in a Proof-of-Stake Consensus Mechanism," 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS), Barcelona, Spain, 2022, pp. 1-6, doi: 10.1109/COINS54846.2022.9854946.

Ransomware: An Economic Perspective

- Ransomware groups are increasingly organised
  - Brand recognition and rebranding
- Technology and service ecosystems have developed
  - RaaS, BaaS, initial access brokers,
- Symbiotic relationships have evolved
  - Negotiation and insurance complicate measures of 'Willingness to Pay'
  - Despite this demands, and payments, continue to rise in value
  - Average ransom value ~$247,000 in 2021 (up 45% from 2020) [1]
  - Highest demand $240,000,000 ($30,000,000 in 2020) [1]



[1] Group-IB Ransomware Uncovered 2021-2022

Game Theoretic Modelling of Ransomware

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Consider a Game of Ransomware
  - An extortionist wants to extract the maximum ransom to restore continuity of service
  - Their victim wants to restore their operations, but may not wish to pay
  - We disregard concurrent attacks (data theft) where they do not advantage the extortion attempt
  - Specific classes of victim may benefit from Law Enforcement or Negotiators (opposition)

Willingness to Pay

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Extortionists are highly motivated to identify **WtP**
  - Lower cost of attack, less complicated negotiations
- **Initial Access Brokers (IABs)** provide access and intelligence
  - Insiders highly valued as a result
- Since 2018, Ransomware has increasingly been synonymous with **data theft**
  - **Cyber Insurance** and **chatlogs** can signal **WtP** in specific contexts

A Simple Game of Ransom

1. The criminal decides if they will infect the victim's machine
2. Criminal sets ransom demand D > 0
3. Victim receives demand and may propose counter-offer C
4. The criminal may irrationally destroy files, resulting in a payoff of –Y < 0 for the criminal, and –W < 0 for the victim
   i. Y represents the cost of time spent by criminal
   ii. W represents the victim's valuation of their files
5. Criminal may release files for C. If C < M (a minimum acceptable offer held secretly by the criminal), the files will be destroyed
6. The criminal may be caught with probability *q*. It is less costly to be caught having not destroyed files.
   i. -X is a reduction of cost –Z for the criminal for potential cooperation with authorities or perceived 'good' behaviour

Outcome	Payoffs	
	Criminal	Victim
Criminal doesn't infect computer	0	0
Release of files for C	C	-C
Files destroyed	-Y	-W
Criminal caught after release of files	-X	0
Criminal caught Files destroyed	-Z	-W

Table 1: Payoffs to different outcomes Simple games of kidnapping [2]

[2] Hernandez-Castro, J., Cartwright, A. and Cartwright, E., 2020. An economic analysis of ransomware and its welfare consequences. *Royal Society open science*, *7*(3), p.190023.

Opposed Game of Ransom

ROYAL HOLLOWAY UNIVERSITY OF LONDON

1. Victim chooses how much to spend E on defensive measures
2. Criminal chooses whether to attack
   i. This incurs additional cost A on the victim, representing active countermeasures
3. The attack fails with probability $\theta(E)$
   i. θ is a continuous monotonically increasing function of E
   ii. With probability 1- θ(E) the attack succeeds
   iii. A failed attack costs the criminal –F (effort/resources expended)
   iv. A failed attack costs the victim –A-E (combined cost of defense)
4. If successful, criminal demands C as ransom
   i. Victim can choose whether or not they pay
   ii. If they pay, they regain their files. Criminal gets C and victim pays costs –C and -E
   iii. If they don't pay, their files are destroyed, and they incur costs –W (victim's valuation of files) and -E

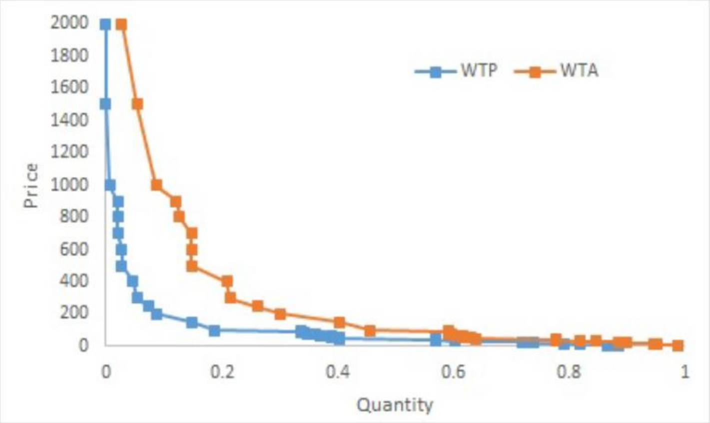
Outcome	Payoffs	
	Criminal	Victim
No attack	0	-E
Failed attack	-F	-A-E
Release of files for ransom C	C	-C-E
Ransom not paid	-L	W-E

Table 2: Payoffs to different outcomes Kidnapping with possible deterrence [2]

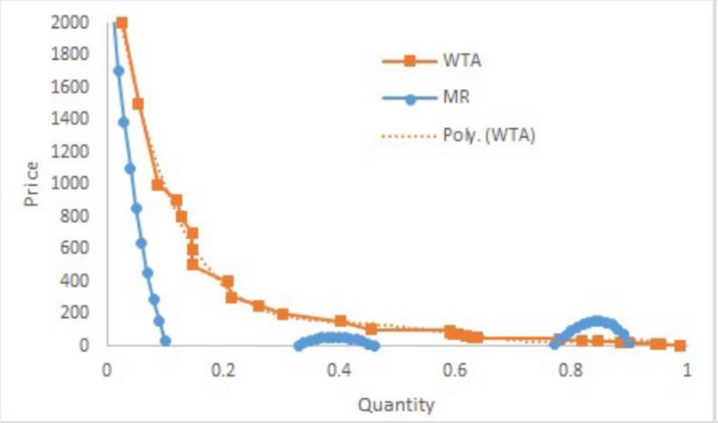
[2] Hernandez-Castro, J., Cartwright, A. and Cartwright, E., 2020. An economic analysis of ransomware and its welfare consequences. *Royal Society open science*, *7*(3), p.190023.

Self-reported Willingness to Pay

ROYAL HOLLOWAY UNIVERSITY OF LONDON



Demand curve elicited using
Willingness to Accept and Willingness to Pay
(Hernandez-Castro, Cartwright & Stepanova 2017)



Demand curve elicited using
Willingness to Accept and Marginal Revenue
(Hernandez-Castro, Cartwright & Stepanova 2017)

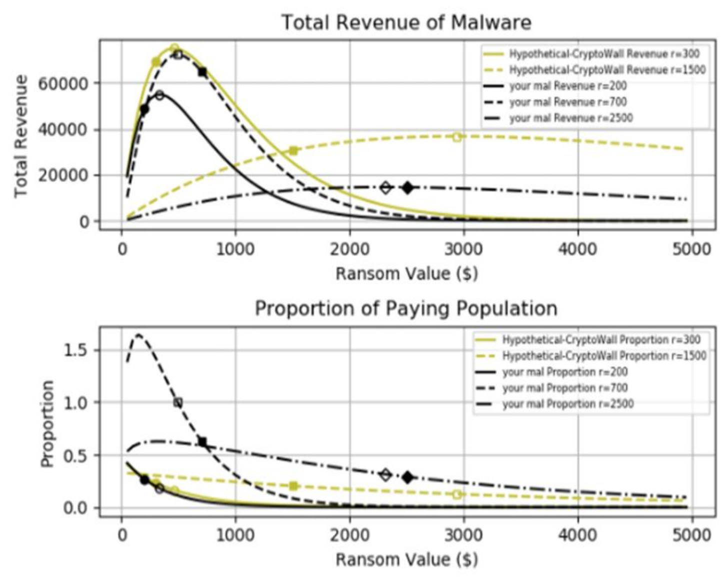
Software modelling of Ransom Games

- An output of RAMSES:

  [3] https://github.com/DarrenHurleySmith/RAMSES_OEMSR

  [4] https://ramses2020.eu/wp-content/uploads/sites/3/2019/09/D4.4-Optimal-model-system.pdf

- Focuses organizations on WtP
  - WtP derived from survey data (defaults in [4])
  - Reconfigurable by sector/organization
- Development restarted:
  - Concurrent attacks
  - Insider threat and intelligence modelling
  - Novel Ransomware targeting Blockchain



Price Discrimination is Key

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Year on year, ransom payments rise – indicating poor understanding of WtP
  - 2020: CWT Global pays $4.5 million (Colonial Oil was $4.4m in 2021)
  - 2022: Insurance giant CNA pays $40m to restore files
- Predictable WtP allows for optimal initial demands
  - Rising demands indicate that extortionists are still identifying optimal ransom values for CNI and large Enterprise (unpredictable WtP)
  - Emphasis on IABs and bribing insiders indicates that WtP is a consideration
- However, some sectors leak WtP values by their very nature…

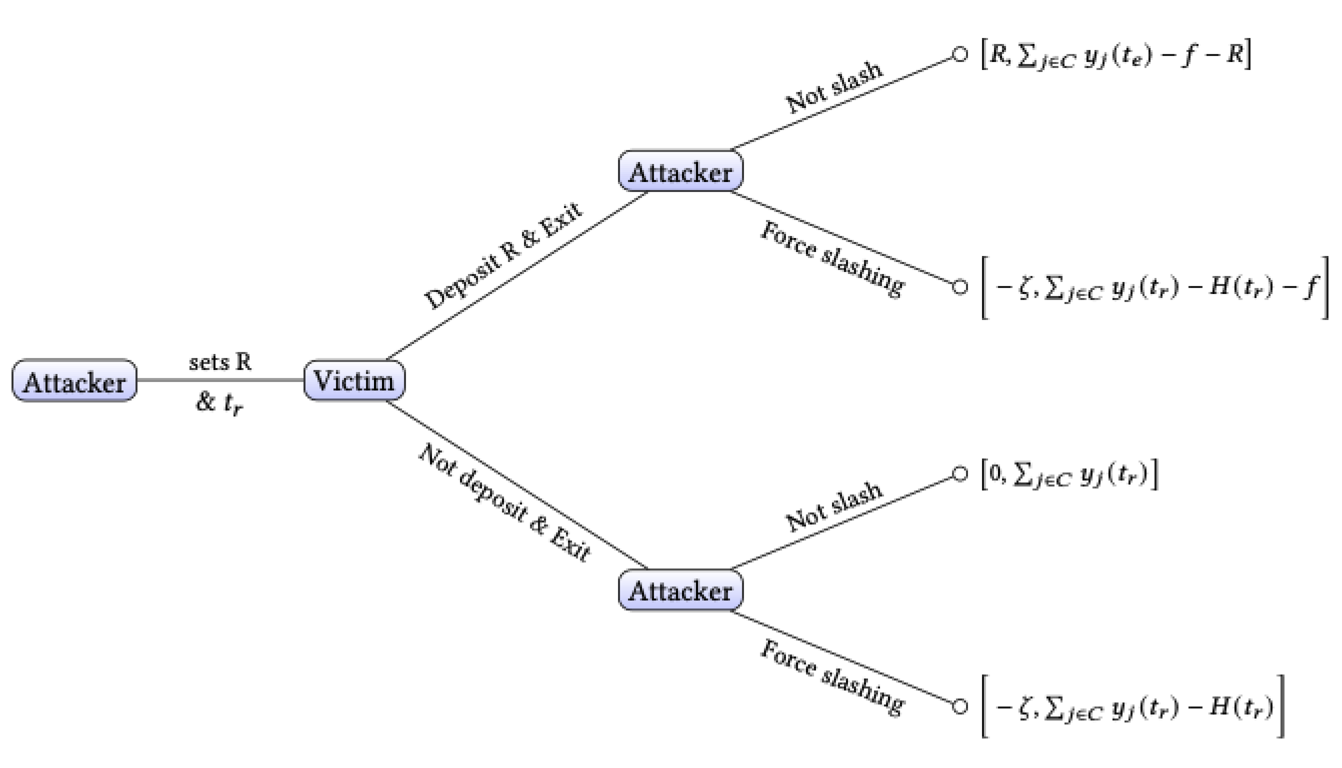
Ethereum 2.0 – A Target for Extortion

- We identified that Proof of Stake (PoS) cryptocurrencies leak high-quality WtP data
- All transactions logged on the blockchain
- All validator balances and actions viewable online
- Enumeration more difficult, but internet-wide scans effective
  - Ports 13000 TCP and 12000 TCP are associated with PoS ETH2.0 nodes
  - Many validators are located in data centres (already a favourite target)

Ethereum 2.0 – A Target for Extortion

- Attacks focus on Proof-of-Stake Validators
  - Attacker holds the signing key to ransom
  - Well-known, pre-existing, exploits to obtain keys (e.g., CVE-2023-28834)
  - Various strategies possible once key obtained (e.g., Pay and Exit, Pay or Slash)
  - Validator funds can't be withdrawn but…
  - Being slashed for any reason delays exit by 36-days

A Pay and Exit Game



A Pay and Exit Game

- Pay and Exit Strategy
  - Validator wishes to leave the network to prevent further key misuse
  - High opportunity cost: exit and re-enrolment takes time
  - Risks associated with re-enrolling: potential loss of 32ETH
- Importantly - this forces the attacker to reacquire and exploit the validator
- REVOKE: proposes a key rotation mechanism instead of exit

What about Willingness to Pay?

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- WtP, in pure economic terms, is leaked by ETH2.0 Validators:
  - All validators hold ~32ETH
  - **Slashing is trivially computed**
  - Penalties increase with concurrent slashings in a window
  - Extortionists are aware of current slashings (they cause) in the last 36 days
  - Opportunity cost incurred for 36 days prior to exit.
- <u>This doesn't include moral or psychological disinclination to pay</u>

Cost of Refusal to Pay for ETH 2.0

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Initial Penalty
  - $\frac{1}{32}\ ETH$
- Correlation Penalty
  - $C = \min(B, \frac{3SB}{T})$ ETH
  - B = effective balance, T = total increments
  - S = sum of increments over 36 days (18 before this validator is slashed, and 18 after)
  - Zero if 3SB < T due to implementation
  - Attacker intelligence about S is limited (cannot predict next 18 days)

- Inactivity Penalties
  - Up to $8192 \frac{14_26}{64} 32b = 0.0827$ETH

- Total penalty between 1.0827 and 32ETH
  - Attacker knows $C = \min(B, \frac{3XB}{T})$
  - X is number of validators slashed by the attacker in the last 18 days
  - Avg. annual ROI for ETH Validator is 6% or 1.92 ETH
  - **Losses are unlikely to be compensated**

Mitigations

- Perimeter Defence & Resilience
  - REVOKE provides key rotation, but this only provides so much long-term resilience
  - Community tools/groups are largely sector specific
- Obfuscation of WtP
  - Doesn’t prevent probing and scaling attacks (the current status quo in CNI)
- Regulation* (e.g., Criminalising ransom payment)
  - SMEs disproportionately affected by inability to expedite return of services
  - Negotiation is an effective intel-gathering tool
  - Not paying ransom in some scenarios may be more damaging (CNI)

Questions?

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Thank you!

Contact:

darren.hurley-smith@rhul.ac.uk

Twitter: @DSmith_PhD