Dynamics of targeted ransomware negotiation

Pierce Ryan*[1,2], John Fokker[3], Sorcha Healy[4] and Andreas Amann[1]

[1]School of Mathematical Sciences, University College Cork, Ireland
[2]Artificial Intelligence Research, McAfee LLC, Ireland
[3]Trellix Laboratories, Trellix, The Netherlands
[4]Global Data Science & Analytics, Microsoft Ireland, Ireland
*pierce.ryan@ucc.ie

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ransomware

- Malware that enables a cybercriminal to extort a ransom from a victim
- Typically functions by encrypting data on an infected computer, forcing the user to pay for a key to decrypt their data
- Highly lucrative form of malware, encouraging rapid development of technology and strategy both by cybercriminals and cybersecurity providers

Classical Ransomware

- Many victims who encounter the ransomware through Internet activity
- Minimal investment in each victim
- Fixed low ransom demand, within the means of most victims
- No negotiation of ransoms
- Significantly nullified by advances in security (backups, software, etc.)

Targeted Ransomware

- Few high-value, high-security targets
- Significant investment in each target
- Increased complexity of encrypting entire networks of computers
- High ransom demands, adjusted to the value of a target's data
- Extensive negotiation process, characterised by aggressive behaviour from cybercriminals

Rules of the Game

We model the negotiations between attacker and target as a two-player game.

1. The attacker invests I_β to encrypt the target's data with value x, invests I_σ to produce an estimate of x, and make a ransom demand R.
2. The target makes a counteroffer C.
3. The attacker aggressively rejects the target's counteroffer with probability α.
4. If the attacker does not aggressively reject the counteroffer, the attacker receives C and the target receives the decryption key.
5. The target's data is successfully decrypted with probability β.

Outcome	Payoff	
	Attacker	Target
Aggressive rejection	$-I_\beta - I_\sigma$	$-x$
Decryption successful	$C - I_\beta - I_\sigma$	$-C$
Decryption failed	$C - I_\beta - I_\sigma$	$-x - C$

Modelling

The key elements of the attacker's strategy are aggression a, investment in reliability I_β, and investment in estimation I_σ. We model α as

$$\alpha = 1 - \left(\frac{C}{R}\right)^a$$

so that the attacker becomes increasingly sensitive to $C < R$ as a increases.
We introduce an economic scaling parameter I_{50} to relate the attacker's investments to their success in reliability and estimation. We model β as

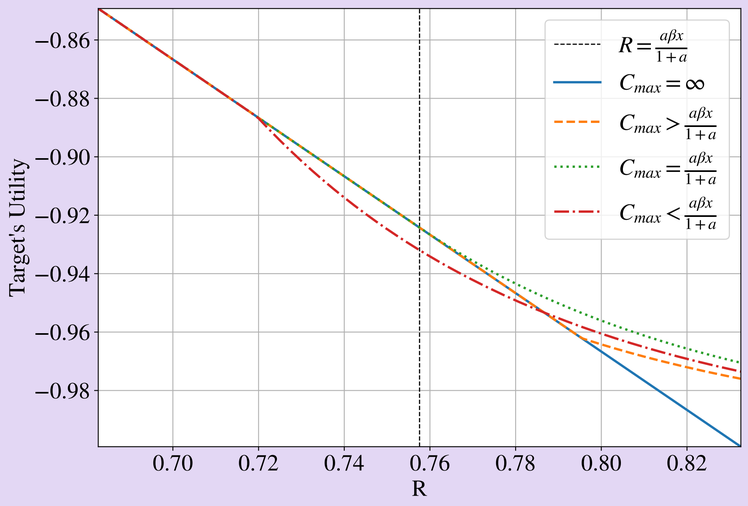
$$\beta = \frac{I_\beta}{I_\beta + I_{50}}$$

Finally we model the attacker's estimate of data value $\tilde{x}$ as a Lognormal(μ, σ) random variable with

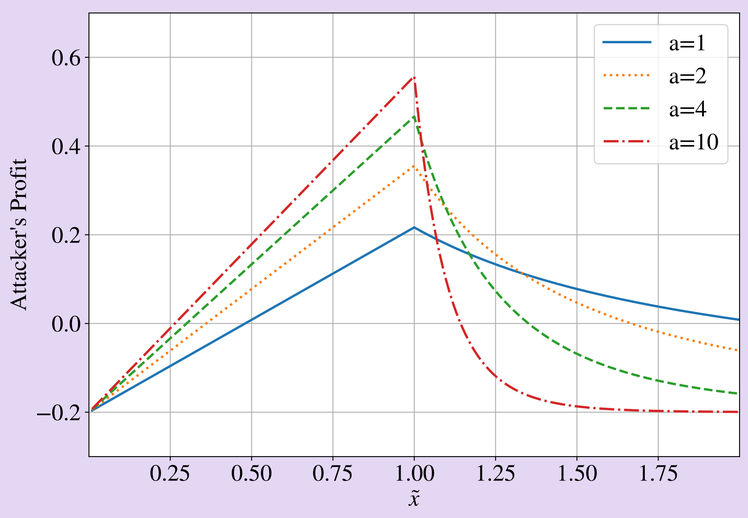
$$\mu = \ln x$$
$$\sigma = 1 - \frac{I_\sigma}{I_{50} + I_\sigma}$$

As is typical in game theory, the attacker and target are assumed to behave rationally and make decisions that optimise their profit/utility to the best of their ability.

Analysis



The target's expected utility for varying ransom demand R and different values of maximum counteroffer C_{max} where $a = 10$, $I_{50} = 0.02$ and $I_\beta = 0.1$. If the maximum counteroffer is too low, utility is reduced by provoking aggressive reactions more frequently; too high, and the target pays more than their data is worth to retrieve it.

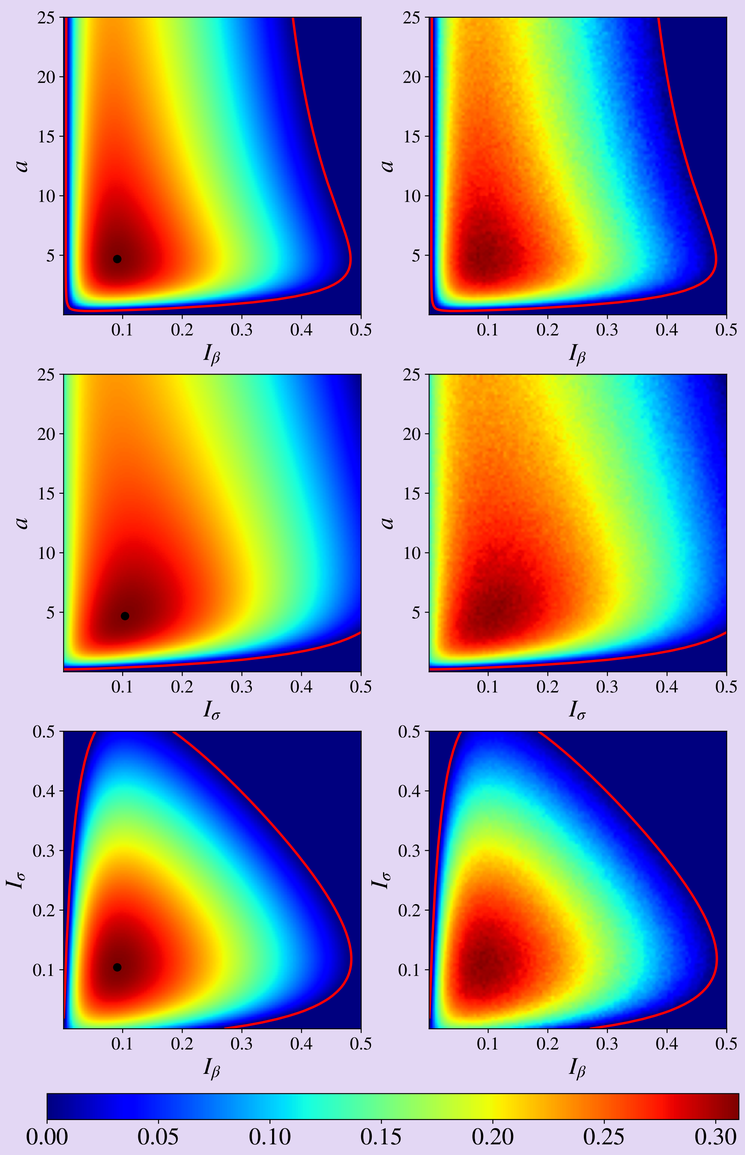


The attacker's expected profit as a function of estimated data value $\tilde{x}$ for varying aggression a where $x = 1$, $I_{50} = 0.02$, and $I_\beta = I_\sigma = 0.1$. Underestimating the value of the target's files decreases expected profit by setting a low ransom demand, while overestimation sets the attacker up to be angered by the counteroffer.

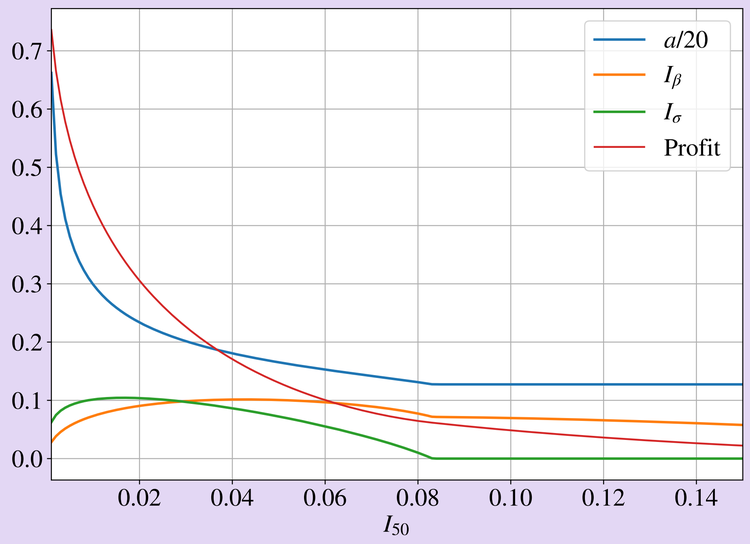
Discussion

Cybercrime develops extremely rapidly, allowing us to observe strategy evolve in real time. Recent evolutions in ransomware include mixed strategy exfiltration and extortion, refusal to engage with external negotiators, and development of complex criminal ecosystems. This creates a wealth of interesting phenomena to analyse using game theory and other modelling techniques, which we hope to study further in the future.

Results



Attacker's profit for varying strategy parameters (a, I_β, I_σ) where $I_{50} = 0.02$, calculated numerically (left), and through simulating the game (right). In each plot, the hidden parameter is set to its optimal value. The red curve marks where the the mean profit is equal to 0.



The attacker's optimal strategy (a, I_β, I_σ) and expected profit under that strategy for varying I_{50} and fixed $x = 1$. As the cost of operating targeted ransomware increases, the relative importance of investing in reliability and accuracy changes.

This work was supported by the Irish Research Council and McAfee LLC through the Irish Research Council Employment-Based Postgraduate Programme under Grant EBPPG/2018/269.