A Game-theoretical Model of Ransomware

Nicholas Caporusso[1], Singhtararaksme Chea[1], Raied Abukhaled[1]

[1] Fort Hays State University, 600 Park Street, Hays, United States
n_caporusso@fhsu.edu, {s_chea2, rkabukhaled}@mail.fhsu.edu

Abstract. Ransomware is a recent form of malware that encrypts the files on a target computer until a specific amount (ransom) is paid to the attacker. As a result, in addition to aggressively spreading and disrupting victim's data and operation, differently from most cyberattacks, ransomware implements a revenue model. Specifically, it creates a hostage-like situation in which the victim is threatened with the risk of data loss and forced into a negotiation.
In this paper, we use game theory to approach this unique aspect of ransomware, and we present a model for analyzing the strategies behind decisions in dealing with human-controlled attacks. Although the game-theoretical model does not contribute to recovering encrypted files, it can be utilized to understand potential prevention measures, and it can be utilized to further investigate similar types of cybercrime.

Keywords: Game theory · Cybersecurity · Ransomware

1 Introduction

Malicious software is used by cyber-criminals to disrupt computer operations, obtain data, or gain access to network. It includes viruses, worms, trojans, rootkits, key loggers, spyware, adware, and ransomware. The latter, also known as crypto-virus, attempts to limit or deny access to data and availability of functionalities on a host by encrypting the files with a key known only to the attacker who deployed the malware [1]. Examples of ransomware include different families of viruses, such as, WannaCry, CryptoLocker, and Petya. Depending on type, ransomware can attack different devices in several forms. For instance, it can infect the hard drive partition table of personal computers and prevent the operating system to be launched [2]. Moreover, it can target mobile devices and even basic home appliances, such as, thermostats [3].

Damage caused by ransomware includes loss of data, downtime, lost productivity, and exposure of private files to further attacks. According to industry reports, the financial cost of ransomware increased from 325 million USD to 5 billion USD, between 2015 and 2017 [4]. Simultaneously, ransomware is both a virus and a business, which even non-technical attackers can enter: as they can create and customize their own version of the virus with Ransomware-as-Service (RaaS) software, which enables configuring the economic details of the threat, such as, price and payment options, in addition to providing some automation support to the negotiation process [5,

6]. As with other malware, countermeasures include frequent backups, antivirus updates, and recovery software. Correct prevention strategies minimize the risk of potential data loss. Nevertheless, depending on the type of file content, attackers might pose the threat of blackmailing victims and leaking the data, with additional consequences (e.g., reputation and litigations).

Among malicious software, ransomware is particularly interesting from a human factors standpoint: differently from other malware, which simply attack a host, ransomware encrypts data and locks down some or all functionality of the device until some ransom is paid to the attacker. Typically, the ransom translates in a payment in bitcoin. The perfect ransomware attack requires three core technologies: (1) strong and reversible encryption to lock up files, (2) anonymous communication keys and decryption tools, and (3) concealing the tracks for the ransom transaction [7].

For instance, the CTB Locker creates a Bitcoin wallet per victim so that payment can be realized via Tor gateway, which guarantees complete anonymity to both parties and enables automating key release [8]. Specifically, the strength of a ransomware attack depends on two key factors: (1) impossibility of recovering the data via backups and decryption software, and (2) sensitivity of the information contained in the data. In the worst scenario, victims might eventually enter a negotiation process in which they might decide to pay to get their data back. Moreover, the amount requested by the attacker might increase depending on the value attributed by the cybercriminal to the data [9]. Alternatively, instead of financial return, the attacker might ask to spread the infection, or to install the virus on other computers which contain more valuable information. In one of the largest ransomware attacks, the WannaCry attack infected over 230000 computers in 150 countries [10].

In addition to implementing appropriate countermeasures, initiatives for risk awareness, adequate cybersecurity training, and dissemination of information about the occurrence and dynamics of ransomware cases are crucial for preventing attacks and for addressing the threat [11]. Nevertheless, cybersecurity statistics revealed that the majority of victims, especially in the case of companies, prefer to remain anonymous to avoid any additional reputation damage. Its growth rate and economics demonstrates that ransomware is not simply a novel trend in malicious software, but rather a paradigm shift towards a more structured, lucrative model. The price of the ransom is set by the attacker, and it usually depends on the estimated value of data. Although the average ransom is 1000 US dollars, the minimum cost is one order of magnitude larger in case of businesses [12]. Global statistics show that 34% of victims end up paying ransom, whereas the rate is higher in some countries (e.g., 64% in the United States) and for businesses (i.e., 70% on average); also, victims have the option of negotiating the initial ask. Nevertheless, only 47% of those who pay actually recover their files [4].

In this paper, we analyze the dynamics of ransomware using Game Theory (GT), to investigate the underlying human factors of attackers and victims, and to explain their decisions. GT is a mathematical model of a scenario (game) in which two or more rational decision makers (players) realize cooperative or competitive choices (strategies) to consistently pursuit their own objective (goal). Although GT has been applied

to other types of malware, there have been limited attempts to model the multifaceted dynamics of ransomware. In addition to informing victims about their options and the potential outcome of their strategies, a game-theoretical model of ransomware helps.

2 Related work

Strategies for preventing ransomware attack include using updated antivirus and firewall, and regularly backing up data so a simple restore can recover any corrupt or encrypted data, being vigilant when opening email and avoiding clicking on suspicious links and attachments, updating security software, visiting only trusted and bookmarked websites, running an antivirus scan, and enabling popup blockers [13, 14].

Nevertheless, attacks might happen because of a security breach caused by an infected device entering a network or by new versions of the virus which successfully inoculate a host using unprecedented tactics. There are several methods to recover from ransomware which do not require any negotiation with the attacker. If user receives a ransomware notice via the Internet, the most important step is to contain the attack and remove the computer from the network. As a second step, one option is to treat ransomware as a virus and attempt to scan the device and remove any threat form the computer. This usually works with least aggressive versions of malware. Conversely, a locker ransomware attack requires more sophisticated measures, and it might involve investing significant effort in research for finding the exact family and version of the virus. Among several projects, the "No More Ransom" website [15] is an initiative by the National High Tech Crime Unit of the Netherlands' police, the Europol's European Cybercrime Centre, and several organizations (including cybersecurity and antivirus companies) which aims at providing victims of ransomware with an index of tools for retrieving their encrypted data, based on the type, family, and version of ransomware.

Unfortunately, there could be situations in which none of the strategies works. Thus, in addition to waiting for a fix, negotiating with attackers is the only decision that has some possibility of getting the files back, in the short term. Although this might raise ethical concerns, the data might have a high value or be life- or mission-critical for some users (e.g., sensitive information that help treating patients, such as, medical records). Usually, this happens by paying the ransom or by reaching out to the attacker for bargaining and agree to a lower amount. The latter tactic was used by Hollywood Presbyterian Medical Center when it was hit by a ransomware attack in 2016: the initial ransom (3.7 million USD) was negotiated to 17000 (USD) [3]. In addition to the type and value of the files being kept, several factors might influence the decision of accommodating attackers' requests, such as, level of trust in the criminal established during the negotiation process (e.g., release of a portion of data), credibility of the threat (e.g., access to and use of the data), and the financial request. Indeed, the optimal solution for the victim is that the attacker honors the payment and decrypts or releases the files, or unlocks the device. However, there is very little penalty for the attacker if they do not cooperate (discussed in Section 3) and negligible

risk of being detected, regardless of the list of offenses that they might be charged with [8]. As a result, there is no guarantee about the outcome [4]. Time plays a significant role in victims' decisions: when identifying or attempting available solutions requires more time than satisfying attackers' request, entering the negotiation results as the fastest option.

Ransomware decisions are particularly interesting from an economic standpoint. The payment process is described in [8], which details how the Dark Web facilitates the attacker by reducing the risk of being discovered and, thus, by rendering affordable the cost of running a ransomware business. The authors of [16] analyzed Ransomware from an economic standpoint, focusing on price discrimination strategies for different families. Using economic models, they identified the rationale behind ransom amount and willingness to pay, in a uniform pricing condition. Also, they address the bargaining process and the attackers' motivations in lowering the price, which are primarily led by the opportunity cost of waiting rather than receiving the amount sooner. Moreover, economic theories can be utilized to identify strategies for dealing with the negotiation process, which is the last line of defense. Specifically, GT has been applied in [17] to study strategies in the distribution of malvertising. In this paper, we propose a game-theoretical model of ransomware, which particularly focuses on negotiation in the post-attack phase.

3 A game-theoretical model

Game Theory utilizes mathematical models to study situations (games) of conflict and cooperation between two or more decision makers (players) who are assumed to be rational (they make decisions consistently in pursuit of their own objective) and intelligent (they try to maximize their outcome). In the game, players choice among their available strategies to maximize expected value of their own payoff (competitive games) or the payoff of both coalitions (cooperative games). The outcome is measured on a utility scale that can be different for each player, and it does not necessarily involve a financial nature. During an instance of the game, each party will play a strategy based on the available information.

In a ransomware game, the attacker and the victim are modeled as the two coalitions: the former starts the game by successfully infecting a host. This, in turn, immediately translates in a request for ransom to the victim with the promise of unencrypting or releasing the data after payment is made. Then, victims will decide whether they want to pay the amount, while threatened with the risk of losing their data. Eventually, attackers will end the game by either releasing or by deleting the files (we use the term delete to represent the situation in which payment is not honored and files are not restored). Ransomware situations can be modeled using a finite state machine (see Figure 1). Although the steps of ransomware negotiations are simple, the process is sophisticated from a decision-making perspective, as several actual cases demonstrate. This is mainly because the attacker can both delete and release the files regardless of whether the victim pays the ransom, that is, ransomware a kidnapping or hostage situation. However, GT clarifies the available options and their outcome.

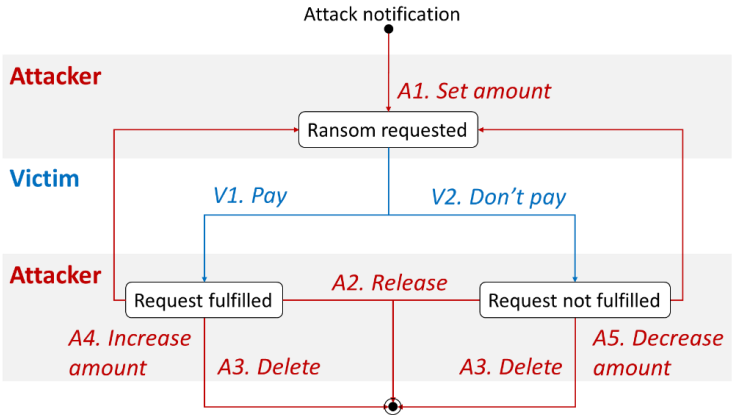

Fig. 1. A finite-state-machine representation of a Ransomware situation, showing the options available to attackers and victims. As shown in the picture, releasing the key (A2) and deleting the files (A3) take to the final state. Conversely, if the request is fulfilled by the victim, the attacker has the option of increasing the ransom amount (A4), and if the ransom is not paid, the victim can be asked for a lower amount (A5). Both events lead to a new request, which creates a loop in the game. The term delete is used to represent the case in which files are not restored.

The extensive-form representation of the game is shown in Figure 2. The game tree is the attack notification, which usually corresponds to the ransom being asked to the victim. We did not model the strategy for calculating the amount requested to the victim, primarily because either this is realized a priori, in the case of attacks targeted to organizations having valuable information, or using criteria which are independent from the actual value of the data being encrypted. Payoffs can be calculated as:

- $d - r$ for the victim and $r - c_r + T_g$ for the attacker, if the former pays the ransom and the latter releases the files
- $-d - r$ for the victim and $r - c_d - T_l$ for the attacker, if the former pays the ransom and the latter does not release the files
- d for the victim and $-c_r + T_g - t_l$ for the attacker, if the former does not the ransom and the latter releases the files
- $-d$ for the victim and $-c_d - T_l + t_g$ for the attacker, if the former does not pay the ransom and the latter does not release the files

where:

- d is the actual value of the data encrypted by ransomware, as defined by the victim;
- r is the ransom amount paid by the victim;
- c is the attacker's cost for handling data, with c_r and c_d representing the cost for releasing and for deleting the files, respectively;
- T represents the trust level in the attacker, with T_g and T_l representing the gain and the loss in trust, respectively;
- t represents the credibility of the threat posed by the attacker, with t_g and t_l representing the gain and the loss in credibility, respectively;

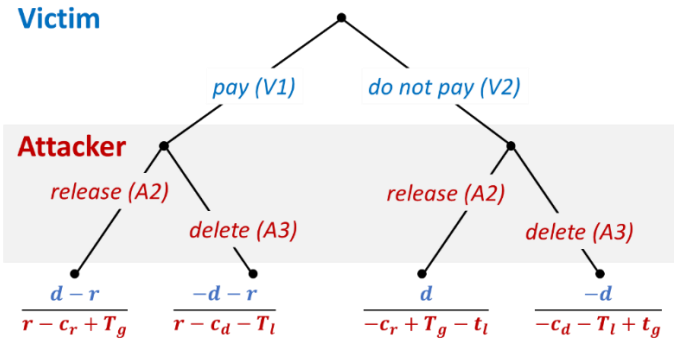

Fig. 2. Extensive-form representation of the basic model of the game. Strategies for renegotiation of the amount are not included. The victim's and attacker's outcomes are represented above and below the line, respectively.

3.1 Characteristics of the game

Indeed, each Ransomware game is competitive, as the attacker and the victim focus on their individual outcome and they primarily consider how their decisions will affect the distribution of payoff within their own coalition. Specifically, as discussed later, attackers have the option of enacting a non-cooperative strategy to enforce the credibility of their threat (i.e., deleting the files in case the ransom is not paid) to influence the victims, in situation of repeated games and semi-imperfect information. In this regard, attackers can also opt for a cooperative strategy (i.e., releasing the files if the ransom is paid) to gain trust.

Ransomware creates an inherently asymmetric situation, as victims are forced by the other party into a game in which they engage in response to an initial attack. Moreover, each coalition has a strategy which involves completely different choices which have no overlap with the decision space of the other player. The payoffs for the two coalitions are dissimilar, with the attacker being in a more powerful position, because they can only gain (as discussed below); contrarily, each strategy of the victim is aimed at limiting their loss.

Although being similar to a kidnapping situation, the consequences associated with this type of cybercrime involve significantly lower risks. Regarding payoffs, the potential gains for attackers (ransom amount) exceed losses and costs for running the game. Also, a gain by the victim (release of the encryption key) does not necessarily correspond with a loss by the attacker, because the costs for releasing and deleting files are both negligible. Especially if calculated over the total spread of cyberattack, this gives a measure of the profitability of the ransomware business model. Conversely, the victim is in a zero-sum situation only if they decide not to pay the ransom; otherwise, the net payoffs for the game involve a loss.

A Ransomware negotiation typically involves a sequence of moves which are known to the players: the attacker initiates the game by encrypting data on the target device (even when the malware infects the host similarly to a computer virus. Then, the victim plays their turn, in which they can decide to pay or negotiate the amount, or ignore the request. The latter decision does not actually end the game, because attackers still have their turn, in which they can decide whether they will disclose the encryption key. The strategy of releasing the data is dominated, to a certain extent (see below), by the option of deleting the files. Although this attacker behavior is very unlikely, because it undermines the credibility of the threat, it was reported in several cases, in which files were returned even if no ransom was paid. As a result, the game can be described by its extensive form and analyzed using backward induction (see Figure 2).

Independently from individuals' computer literacy and cybersecurity awareness, which can depend on several factors (e.g., background, education level), ransomware is a typical example of a perfect information scenario: rules, potential moves, and possible outcomes of the game are known a priori to each player. Furthermore, each coalition usually takes into consideration the strategy chosen by the opponent in the previous turn, with the only exceptions of situations in which the attacker has a "no matter what" approach. Specifically, this is the case of virus-like attacks that focus on disrupting the victim (discussed later), in which no decision making is involved (files are deleted anyway). Alternatively, the attacker might implement a strategy in which the outcome goes beyond the single instance of the game and, thus, they might decide to use the current negotiation for gaining trust (release the key no matter what) or for enforcing the credibility of their threat (delete the files no matter what).

Moreover, players know their own utility function, that is, victims are aware of the actual value of the data and attackers know the cost for releasing key or deleting the files. However, each party is not able to precisely evaluate the utility function of the other coalition, though the attacker can estimate it by exploring victim's data. As a result, ransomware is an interestingly rare case of game with incomplete but perfect information. Nevertheless, the type and amount of information available to each player increases when either coalition renegotiates the amount: this creates loops that tend to reveal the utility function of the bargaining party. For instance, if a victim accepts the initial attacker's request, they might suggest that the data being encrypted has a value. Additionally, any interaction with the attacker might reveal the nature and source of the attack, and might change the level of trust or perceived threat.

Each ransomware negotiation involves a limited number of turns. Usually, the game ends after 3 steps, summarized as infection (attacker), ransom payment (victim), and data release (attacker). Moreover, assuming that players are rational and focus on the current instance of the game only, as dominant strategies are equivalent to not taking any action (i.e., not paying the ransom and deleting the files), victims can end the game in second turn. Conversely, bargaining creates loops that increase duration.

3.2 Real-life decision making

When ransomware utilized the standard mail services and bank accounts, we could consider the cost $-c_p$ in all attackers' payoffs, to represent the risk of being discovered and the effort of handling a payment. Also, $c_r > c_d$ because storing and releasing the data would involve more traceable operations. This is different than traditional kidnapping situations, in which failing to honor the promise coincides with physically eliminating the detained individual, which involves more severe felonies and substantial charges. In modern ransomware, $r > c_r > c_d$ still holds, even though we could perceive that $c_r \cong c_d \cong 0$, considering the inexpensiveness of computational power and the advantages offered by cryptocurrencies and the Dark Web. This is because releasing the data involved the effort of writing the software function for processing the request, whereas $c_d \cong 0$. This would lead to having a situation in which the attacker would delete the files no matter what, because $r - c_d > r - c_r$ for $V1$ and $-c_d > -c_r$ for $V2$.

In addition, victims evaluate the threat t posed by the attacker, and their trust level T, in their decision. Although ransomware games occur in single instances and there is very limited information about the attacker, a rational and intelligent player would perceive and attribute $0 \cong T < t$ in a competitive game. Although there is no opportunity of achieving certain estimate of t and T until the end of the game, a victim with basic cybersecurity awareness would understand that $c_r > c_d$, primarily because c_d involves no action. Therefore, an attacker must invest some effort in convincing the other player to accommodate their requests. This is achieved by setting the ransom amount accordingly. A hostage or kidnapping situation requires planning and it is an individual action. As a result, the payment requested r is proportional to the value of the detained entities d and to the risk c_p. Instead, in a ransomware situation, any value greater than c_p (which is especially low) is worth, especially considering the scale of ransomware attacks. However, a very low amount would impact the credibility of the threat, whereas a ransom close to the value of the data would affect the willingness to pay of the victim. This is compatible with the findings of [16], which found that the average ransom is below 1000 USD, and with data from reports about ransomware attacks. Applying backwards induction to the extensive-form model of the game leads to more users paying the ransom, given the affordable amount, and hoping that the data will be returned, which is very unlikely to happen, because $c_r > c_d$. This situation fits actual data more realistically. Nevertheless, price discrimination strategies might influence the decision of the victim, but they do not change the dynamics of the game unless other factors are taken into consideration.

Specifically, although attackers play one instance of the game with each victim, ransomware can be considered as having multiple instances over a population of different players, which share and get access to a little piece of the information "leaked" from each game, thanks to statistics, reports, users sharing their stories, and awareness and training initiatives. In this regard, attackers' decisions in a single game influence the perceived trust and threat of victims in the next iterations, though $0 \cong T < t$ holds. If the ransom is paid, attackers experience a trust loss T_l or gain T_g depending on whether they honor or disregard their promise, respectively. This information is

shared indirectly between players, and available to individuals in the population in a nonuniform fashion. As a result, if $T_g \cong T_l > c_r > c_d$, for any trust gain or loss greater than the cost of releasing the files, holds for an attacker, they will release the files. Vice versa, they will not honor their promise. There is no threat gain or loss in attackers' payoffs in addition to the component included in pricing strategies. Conversely, if the victim pays the ransom, the attacker will release the files only in case $T_g > t_l + c_r$, assuming that $T_l \cong t_g \cong 0$. This model results in a better representation of real-life cybercrime scenarios, in which attackers honor their promise even if deleting the files would seem the most convenient choice. Also, this model fits several cases in which the key was released even if no ransom was paid: this is because $0 \cong T < t$, and therefore, attackers need to gain trust. Also, as ransomware can work as a simple virus, cybercriminals must counteract the impact of situations in which the key will never be released, "no matter what", which diminishes the level of trust.

3.3 Amount renegotiation

Attackers and victims have a third option, which is bargaining, which creates a subgame that has been extensively discussed in the literature, and will not be detailed in this paper. However, from a cybercrime perspective, initiating a conversation has several benefits. Several ransomware situations merely involve operations disruption or they might result in the spread of an outdated version of the virus which is not followed-up by the initiator. Thus, contacting the attacker helps identify whether the communication channel is still active, which is crucial for a payment decision. Moreover, attackers know that $0 \cong T$ and they will release part of the data to improve their trust. Also, time is playing against them, because the victim might be attempting alternative situation and because of the cost opportunity of receiving r sooner. As a result, they will probably lower their request to increase victim's willingness to pay. Moreover, bargaining might reduce the risk of being asked for more after accommodating the first request.

4 Conclusion

A ransomware game is discrete, and it begins after a successful attack, and it proceeds sequentially, with the victim playing the first move and deciding their strategy, which can be either cooperative (i.e., negotiate or pay the ransom) or competitive (i.e., avoid paying). As the game is non-zero-sum and asymmetric, the attacker and the victim have different objectives, strategy sets, and payoffs. Furthermore, the incomplete-information nature of ransomware can be mitigated with outcome from previous games, which might help victims taking decisions.

In this paper, we focused on post-attack game dynamics that occur between two human players. Specifically, we use GT to address situations in which alternative strategies failed, and negotiation is the last resort. We did not describe the sub-game of amount renegotiation, which will be addressed in a follow-up work. Indeed, ransomware can be utilized as a traditional virus, to simply disrupt victims' operations.

This case is similar to a “no matter what” situation, because the victim is there is only player, and files will not be released. In addition to the structure of the game, which would lead to attackers not honoring their promise, available information, education, and human factors play a significant role in real-life decision-making: by incorporating them in the evaluation of payoffs, we can achieve a more robust model.

References

1. O'Gorman, G., & McDonald, G. (2012). Ransomware: A growing menace. Symantec Corporation.
2. Palisse, A., Le Bouder, H., Lanet, J. L., Le Guernic, C., & Legay, A. (2016, September). Ransomware and the legacy crypto API. In International Conference on Risks and Security of Internet and Systems (pp. 11-28). Springer, Cham.
3. Richardson, R., & North, M. (2017). Ransomware: Evolution, mitigation and prevention. International Management Review, 13(1), 10.
4. Hammill, A. (2017). The rise and wrath of ransomware and what it means for society (Doc-toral dissertation, Utica College).
5. Nieuwenhuizen, D. (2017). A behavioural-based approach to ransomware detection. White-paper. MWR Labs Whitepaper.
6. Tuttle, H. (2016). Ransomware attacks pose growing threat. Risk Management, 63(4), 4.
7. Hampton, N., & Baig, Z. A. (2015). Ransomware: Emergence of the cyber-extortion menace.
8. Upadhyaya, R., & Jain, A. (2016, April). Cyber ethics and cyber crime: A deep dwelved study into legality, ransomware, underground web and bitcoin wallet. In Computing, Communication and Automation (ICCCA), 2016 International Conference on (pp. 143-148). IEEE.
9. Kharraz, A., Robertson, W., Balzarotti, D., Bilge, L., & Kirda, E. (2015, July). Cutting the gordian knot: A look under the hood of ransomware attacks. In International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (pp. 3-24). Springer, Cham.
10. Floridi, L. (2017). The Unsustainable Fragility of the Digital, and What to Do About It. Philosophy & Technology, 30(3), 259-261.
11. Luo, X., & Liao, Q. (2007). Awareness education as the key to ransomware prevention. Information Systems Security, 16(4), 195-202.
12. Formby, D., Durbha, S., & Beyah, R. (2017). Out of control: Ransomware for industrial control systems.
13. Pathak, D. P., & Nanded, Y. M. (2016). A dangerous trend of cybercrime: ransomware growing challenge. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume, 5.
14. Fanning, K. (2015). Minimizing the cost of malware. Journal of Corporate Accounting & Finance, 26(3), 7-14.
15. “No more ransomware” project. Online: https://www.nomoreransom.org
16. Hernandez-Castro, J., Cartwright, E., & Stepanova, A. (2017). Economic Analysis of Ransomware.
17. Huang, C. T., Sakib, M. N., Kamhoua, C., Kwiat, K., & Njilla, L. (2017, May). A game theoretic approach for inspecting web-based malvertising. In Communications (ICC), 2017 IEEE International Conference on (pp. 1-6). IEEE.