Article

A Game Theoretic Approach for Digital Forensic Tool Selection [†]

Umit Karabiyik ^{1,*} and Tugba Karabiyik ^{2,†}

¹ Department of Computer and Information Technology, Purdue University, 01 N. Grant St. West Lafayette, IN 47907, USA

² Polytechnic Institute, Purdue University, 01 N. Grant St. West Lafayette, IN 47907, USA; tugba@purdue.edu

* Correspondence: umit@purdue.edu; Tel.: +1-765-496-6877

[†] This paper is substantially extended version of the Conference on Digital Forensics, Security and Law. Association of Digital Forensics, Security and Law, San Antonio, TX, USA, 17–18 May 2017; pp. 179–195.

[‡] These authors contributed equally to this work.

Received: 4 April 2020; Accepted: 9 May 2020; Published: 12 May 2020

Abstract: Digital forensic investigations are getting harder and more time consuming everyday because of various problems including rapid advances in technology, wide variety of available devices in investigations, and large amount of data to be analyzed. In order to tackle with these issues, digital forensic tools are developed by open-source communities and software companies. These software products are released as a complete toolkit or standalone tools targeting specific tasks. In either case, digital forensic investigators use these tools based on their familiarity because of previous training experiences, available funding from their agencies/businesses, tool's ease of use, etc. Moreover, using additional tools to verify the findings is a common practice in digital forensic investigations. This is particularly common when the previously selected tools do not generate an expected output. In this paper, we propose a game theoretic approach to the tool selection problem in order to help investigators to make a decision on which digital forensic tool to use. We particularly focused on file carving tool usage when building and analyzing our model because of the available data on these tools. Our results show how important it is to investigate the dynamics of strategy changes between the tools during an investigation to increase the efficiency of the investigation using game theoretic modeling.

Keywords: game theory; game theoretic modeling; digital forensic tools; digital investigations; decision making

1. Introduction

Digital forensics, a branch of forensic science, deals with the investigation and recovery of digital information found in digital devices which are mostly found at crime/incident scenes or belonging to suspects. Digital forensics is becoming increasingly needed in this early twenty-first century because of the technological advances in today's world. Most data are now being stored in digital forms such as pictures, diaries, calendars, videos, and even our DNA information. Smart phones, tablets, computers, and wearable devices have already become a part of most people's everyday life. This inevitable change makes any device's storage a potential evidence related to a crime or an incident.

Digital evidence for a variety of crimes including child pornography, financial fraud, identity theft, cyberstalking, homicide, abduction, and rape can be collected by using digital forensic techniques and tools. Digital forensic investigators use specialized and general purpose digital forensic tools in order to collect useful information or evidence that is related to a crime or an incident.

Due to variety of digital devices and available tools, digital forensic process is generally considered to be a long and tedious process for an investigator [1]. There are many tools that an investigator must consider, both proprietary and open source. Researchers and tool developers regularly make newer tools available, particularly open source. Most of the general purpose tools have similar capabilities along with their unique properties and standards. There are also task specific open source tools which perform similar tasks on a given device. Investigators generally decide to use certain tools based on their familiarity, technical skills, previous experiences, and availability of tools that they can utilize. In addition, using additional tools to verify the findings has become part of the digital forensics best practices [2]. This is commonly the case when their preferred tools do not work as intended for a comprehensive investigation [3].

Shiva et al. [4] discuss that available solutions to the cybersecurity related problems are still far from full protection. Therefore, there is a need for holistic security solutions. The authors consider the attack and defense strategies as a game between an attacker and a defender. They strongly suggest use of game theory in cybersecurity domain for a complete quantitative decision making in order to find a solution for security problems. Hence, we believe that in order to have a full protection we should have deterministic decision making processes. When it particularly comes to the digital forensic's domain, it is crucial to help investigators on forensic tool selection problems as discussed above. The decision making process can be solved by using game theoretic modeling which John Von Neumann, Oskar Morgenstern, and John Nash invented to study interdependent decision making [5–7]. Game theory is a theory of mathematics to study strategic behavior which arises when the outcome of an individual's actions depends on actions of other individuals. Game theoretic models are analyzed by finding Nash equilibrium which is a set of strategies, one for each player, such that no player has incentive to change his or her strategy given what the other players are doing.

In our prior work in [8], we defined the digital forensics tool selection for a specific task as a multi-armed bandit (MAB) problem assuming that multiple tools are available for an investigator's use on each task. We specifically chose file carving task in this study. In order to test the decision quality and determine initial probability of success of each tool, we created a dataset of 100 disk images using forensic test image generator (ForGe) [9]. We then ran widely used carving tools Scalpel [10] and Photorec [11] on these disk images by treating each tool as an individual arm used in MAB problem. Based on the results we retrieved from our analysis, we compared the average reward yielded by each algorithm based on their successes and failures against the randomization method (non-MAB method) over 1000 simulations. Our results testified that an investigator would have successfully analyzed the given disk images up to 57% more, if their decisions were based on multi-armed bandit algorithms rather than random selection. We also showed how multi-armed bandit algorithms provided different results based on the running time of an individual tool on each disk image and the number of evidence files retrieved from the disk images. In addition, we also supported our experimental results with a good level of statistical confidence by finding p -values using χ^2 test for independence.

In this study our goal is to help forensic investigators during the decision making process while they are choosing which file carving tool to use via game theory. It is important to investigate the dynamics of strategy changes among the tools (i.e., changing the software used) during an investigation to increase the efficiency. The current digital forensics literature significantly lacks game theory applications, although there are some applications of game theory in cybersecurity in general which we discuss extensively in Section 2. Particularly, to the best of our knowledge, game theory has not been applied to the decision making process in digital forensic tool selection for certain investigative tasks in the computer forensics literature. One possible reason could be the issues related to the admissibility of evidence when the decision is made by non-human entities such as mathematical models. However, we believe that it would not be an issue since investigators are not expected to know the details of tools or systems that they use during the examination as long as their findings are reproducible by others [1]. Therefore, the contributions of this study to the digital forensics literature is as follows:

1. A novel game theoretic model is developed for digital forensics tools selection problem.
2. Optimum and most efficient strategies were suggested for digital forensic investigators during their file carving processes.

This paper is organized as follows. Section 2 provides a literature review for game theory applications in cybersecurity area. Section 3 explains the experimental setup for our research to create dataset and preliminary information needed in this study. In Section 4, the game theoretic model is demonstrated in details. Our analysis and results are explained in Section 5. Finally, the last section concludes our study with some recommendations.

2. Literature Review

The current literature significantly lacks game theoretic applications in the digital forensic field. Although, we will discuss digital forensic related works in this section, we will also mainly provide a literature review of game theoretic applications in a broader cybersecurity domain.

There is limited game theory applications in the digital forensic domain in general. For instance, Stamm et al. [12] explored the interactions between a forger who uses anti-forensic technique and a forensic investigator who uses digital forensic techniques in a game-theoretic setting. They also proposed a technique to evaluate the effectiveness of anti-forensic algorithms. In order to do that, they proposed a new automated video frame deletion detection technique along with a technique to detect the use of video anti-forensics. Using these techniques with a game-theoretical framework, they found under which conditions that a video forgery will likely to be detected. By using game theory to understand the dynamics between a forger and forensic investigator, they found the Nash equilibrium which shows the forensic investigator's ability to detect forgeries. Their results show that, when anti-forensic techniques are not used, their proposed frame deletion forgery detection method achieves 95% of true positive rate with less than false positive rate of 5%. In the case when anti-forensic techniques are used, their method achieve 100% detection rate with an increased false positive rate of 10% and higher. (The results show two different tests which show independent values. In the latter, authors show that even though the false positive rate goes up, they were still able to detect the forgeries at a higher rate.) In addition, in their other work, Stamm et al. [13] also investigated the interplay between a forensic detector and a forger when the forger's use of anti-forensics can be detected. The authors were able to identify the optimal set of actions for both investigator and forger using game theoretic framework.

In the broader area of cybersecurity, another game theory application investigated software vulnerabilities in a cyber-warfare setting [14] as nations started to make decisions on whether to disclose new software vulnerabilities (zero-day vulnerabilities) or to exploit them for gain [15,16]. This research focuses on exploring the best strategies in the case of a new identified software vulnerability where any information can be revealed to the adversaries. The authors developed a game-theoretic model of cyber-warfare to calculate the optimal strategy of the players and find the Nash equilibrium of the model. As a result, they found and demonstrated the optimal conditions on how to act in variety of different scenarios.

Shiva et al. [4] proposed a holistic security approach by suggesting a new model called "game theory inspired defense architecture (GIDA)" which utilizes a game model as its brain. The authors suggest GIDA to deal with some of the challenging sides of cybersecurity issues thinking of a game between an attacker and the defender. In their research, the authors assume that players have imperfect information about the current state of the game, i.e., usually there is some error on the defender's real sensors to read the environment. In real life, there is no perfect information. Using this idea, they built a stochastic game model which aims to capture more realistic cybersecurity scenarios by computing defender's best strategy to reach the Nash equilibrium of the stochastic game with the assumption of having an imperfect defender's sensor. The authors validated their theory via experiments in simulation software. They also compared all the known imperfect information stochastic game models with the criteria of whether the model uses a zero sum game or a general-sum game.

Vidal et al. [17] discussed possible game scenarios between a team of unmanned aerial vehicles and unmanned ground vehicles who pursue a second team of evaders within an unknown environment. In this work, authors used Probabilistic Pursuit-Evasion game model and considered two pursuit policies: local-max and global-max policies. They introduced a distributed hierarchical hybrid system architecture, which emphasizes the autonomy of each agent and allows for coordination with team members. They compared the local-max and global-max policies on varying the speed, sensing capabilities of the pursuers, and intelligence of the evaders. They also made their results stronger via simulations and experimental results. Their results show that the global-max policy outperforms the local-max policy in which the dynamics of each agent are included and computer vision is used to detect the evaders. They also found that global-max policy is more robust to changes in the conditions of the games and this policy is better in catching an intelligent evader.

Moayed and Azgomi [18] focused on applying game theory to stochastic modeling. They consider large number of hackers and the diversity in their behaviors. They classified the hackers based on their motivation and skills and looking at interactions of diverse hackers and defense systems. They used continuous time Markov chains to model the system. In their work, authors introduced a game-theoretic approach for quantitative evaluation of security measures. They also proposed to use Q-learning algorithm to compute Nash equilibrium, which guarantees to converge to the optimal strategies of the players. In addition, they predicted the transition rates in stochastic models and presented an illustrative example to show how their method works on computing the security measures of an encryption system.

The work by Zhang et al. [19] focused on patrol allocation design during the opportunistic crimes. In this paper, authors were interested in interactions between attackers and opportunistic criminals, and using real data to validate their results. In this work, authors predict optimal patrolling schedule based on the previous crime data and learn from this data by using Dynamic Bayesian Network and Expectation Maximization (EM) algorithms. They also suggest an iterative learning and planning mechanism where they used real data set of criminal activity obtained from the law enforcement in California, USA. Their results show that using the suggested patrol allocation design helped the police to reduce and prevent the crimes compared to their previous scheduling mechanism.

Wu et al. [20] focused on attacker and defender's actions and consider it as a static and dynamic game during the bandwidth attacks. Authors modeled the interactions of attacker and defender for DoS and DDoS attacks. They studied both static and dynamic versions of the game and found the Nash equilibrium which shows the steady state condition of the game. They also validated the game model with simulations using NS-3 network simulation tool.

Researchers in [21] developed a software assistant agent called Assistant for Randomized Monitoring over Routes (ARMOR). Authors focused on patrolling/monitoring a problem at the Los Angeles International Airport to randomize checkpoints on the roadways entering the airport and canine patrol routes within the airport terminals. They utilized Bayesian Stackelberg game by including different types and uncertainty on adversaries and allowing defenders to weigh the different actions in terms of costs and benefits. ARMOR uses the fastest Bayesian Stackelberg game solver called DOBSS, where the dominant mixed strategies enable randomization. It also has a mixed-initiative based interface that allows police to adjust or override the automated schedule based on their constraints. In addition, it gives alert once the mixed-initiative overrides seem to degrade the desired randomization. With the similar physical security issues, Jiang et al. [22] discuss challenges in security despite of limited resources. In this paper, authors focus on deployment and allocation of limited resources to prevent or reduce crimes in several large size domains such as airports, ports, forests, transit systems etc. using game theory and machine learning. In order to understand the attacker's actions so to define best response for the defender, authors suggest use of Bayesian Stackelberg Security Game and its variations.

Shiva et al. [23] extended a previously proposed model in [24] by relaxing an assumption about the information in the stochastic game. They developed a game between an attacker and a defender

(system administrator in their case) with an assumption that players do not have perfect information about the current state of the game (i.e., Intrusion Detection System sensor is imperfect), which was not the case in previous models. The paper presents a stochastic game model with an observed Nash equilibrium of the game. Authors also validated their results via simulation experiments in MATLAB.

Finally, game theory has been also applied to steganographic security as presented in [25]. Authors explored a game theoretic model to better understand strategic embedding and detection while studying adaptive steganography. They found a unique equilibrium in mixed strategies which totally depends on the heterogeneity of the cover source.

As the current literature testifies, game theoretic models can be utilized to create new, adaptive, and efficient strategies in various security domains. This includes production of optimized strategies and mechanisms to defend cyber and physical infrastructures as well as protect public.

3. Experimental Setup for the Research

In order to test performance of digital forensic tools and also have initial knowledge about the problem studied in this research, we needed a dataset of disk images. In this section, we explain how we created our disk image dataset, collected data, and designed simulation environment in details.

3.1. Forensic Disk Image Creation

When we built a game theoretic model for the file carving problem, we needed to have initial probability of success of each file carving tool (Scalpel and Photorec) based on their performances in a series of investigations. Therefore, we created 100 disk images using ForGe, a forensic test image generator. Each disk image contains a set of files (including forensically interesting files) in allocated space and unconventional areas such as slack space, unallocated space (for deleted files), and free space.

As the first step, we setup a “case” in which we generate disk images of 250MB in size with sector size of 512 bytes and cluster size of 8 sectors (4KB). Our case is set up to create an NTFS file system on the disk image as the type of the file system has minimal effect on the file carving process. In the next step, we created the “trivial strategy” which represents the files normally found in the file system’s allocated space. We filled the allocated space with randomly chosen files gathered from Digital Corpora [26] using similar methods discussed in [27].

At the final step of disk creation, we hid horse pictures (representing files containing illegal contents) in the above-mentioned hidden areas on disks. We had various sizes of horse pictures as we needed files less than 4KB in size to be stored on the slack space. As a result, we created 100 disk images with NTFS file system which contain various file types stored on allocated, slack, unallocated, and free spaces (Reports on the details of disk image creation process and hidden files with their locations are available for interested researchers and tool developers upon request).

3.2. Tool Execution and Data Collection

After we created all disk images to test carving tools, we wrote a Python script to run Scalpel and Photorec on each disk image and recorded the processing time of each tool on each disk as well as the carved files. Both tools are configured to carve only picture files such as jpeg, png, and gif. Based on the collected results with respect to execution time and number of carved horse pictures, we calculated the initial probability of success value of each tool for the following three cases.

- **Time sensitive case:** In this case, the time is critically important for an investigator. Therefore, more weight is given to the faster tool when assigning reward values.
- **File sensitive case:** In this case, the total number of successfully carved illegal files is critically important. Therefore, the tool that carves all the illegal files is considered successful and thus given more weight.
- **Equally sensitive case:** In this case, time and file sensitivity are considered equally important and equal weights are given to both tools.

These three cases are chosen mainly because, depending on the digital forensics case, investigators may be interested in finishing the investigation faster than finding all the possible evidence especially when at least one evidence is enough, and vice versa. An example of this can be when a digital forensic investigator searches for illegal pictures during the execution of a warrant. Time is critical if there are many computers to look at. In such a situation, the only thing the investigator wants is to determine if there is an illegal picture on the computer or not to motivate its seizure. Once the device is seized, it will be transferred to the lab and the file sensitivity will become important to prove the existence of incriminating evidence in seized devices.

In order to calculate the initial probability of success of each tool we need to assign reward values (1 for success and 0 for failure) to each tool depending on their results. In the first case, we assigned reward value to the faster tool as 1. When one tool succeeds, it is also possible for other tool to be successful. Therefore, we compared the running time of the slower tool (t_s) to faster tool (t_f), and assigned slower tool's reward value (r_s) as:

$$r_s = \begin{cases} 0, & \text{if } (t_f)^2 < t_s \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

After assigning all the reward values to both tools, we calculated the initial probability of success of each tool for all three cases mentioned above (see Table 1 for initial probability of success values). Note that, we will refer to the probability of success of each tool as their reward value in the rest of this paper.

Table 1. Initial reward values of Scalpel and Photorec for all three cases.

Sensitivity	Scalpel	Photorec
Equal Sensitive	0.16	0.5
Time Sensitive	0.23	1
File Sensitive	0.85	0.5

As Table 1 shows, Photorec was given reward value of 1 in time sensitive case, because it outperformed Scalpel in most of the test cases with respect to execution time. In rare test cases, Scalpel outperformed Photorec, however Photorec's execution time was never small enough to satisfy the first condition in Equation (1). Thus, in such test cases, both tools were considered successful with respect to their execution times, hence both were given a reward value of 1.

In the file sensitive case, Scalpel was more successful than Photorec with respect to the number of successfully carved files. The only condition that Scalpel failed was when files were located in the slack space and fragmented. In such cases, some of the fragments were carved however we considered tools being successful if the pictures contained large portion of horse bodies. When horse bodies were not recognized, we considered tools being failed. On the other hand, Photorec was failed carving the files from all locations including allocated, slack, and unallocated spaces in different test cases. We do not further discuss the reasons of failures or successes as tool testing is out of the scope of this work.

3.3. Simulation Results

The simulation in this study was implemented in Java programming language and source code of our program is available upon request. We performed 1000 simulations for all five tool selection strategies (simple randomization and four multi-armed bandit algorithms: ϵ -greedy, ϵ -first, Softmax, and UCB1) on given 100 disk images for both carving tools. By the law of large numbers, running 1000 simulations was sufficient to confirm that success rates of Scalpel and Photorec were converging to certain numbers (see Table 1).

In Table 2, we present the average number of disk images successfully analyzed (shown in column **Avg Total Reward**) over 100 disk images for 1000 simulations. According to the our results, it is clear

that if bandit algorithms would be followed during the decision making process, average number of successfully analyzed disk images would increase up to 42% and 57% in equal sensitive and time sensitive cases respectively. Similarly, success increase for file sensitive case would be up to 18%.

In the next sections, we describe our game theoretic model and its analysis in details, and discuss how our results support the findings described in Table 1.

Table 2. Number of disk images successfully analyzed.

Sensitivity	Algorithm	Avg Total Reward
Equal Sensitive	ϵ -greedy	44.693
	ϵ -first	41.378
	Softmax	46.818
	UCB1	43.236
	Randomization	32.96
Time Sensitive	ϵ -greedy	92.03
	ϵ -first	86.408
	Softmax	96.507
	UCB1	93.514
	Randomization	61.619
File Sensitive	ϵ -greedy	79.598
	ϵ -first	72.148
	Softmax	74.909
	UCB1	78.299
	Randomization	67.701

4. The Model

Suppose we have two investigators who are making decisions to select which file carving tool to use during an investigation. We assume to have Scalpel and Photorec as two strategies for an investigator during the file carving process in an investigation. Using each carving tool has its own benefit/value (V). There are also costs with respect to processing time (C_t), and the cost with respect to being unsuccessful to carve some files (C_f) in the game-theoretic model. From our prior findings about the initial reward values of file carving tools from Table 1, we know that Scalpel is more successful regarding the number of carved files but significantly slower with respect to processing time compared to Photorec and vice versa. Therefore, there is a tradeoff for an investigator to make a strategic decision. Our goal is to explore the best strategy to maximize the investigator's effectiveness during the file carving process using game-theoretic modeling.

Suppose we structure the investigator's tool selection problem as a two player game with Scalpel and Photorec being the players' distinct strategies. We also assume that the players (investigators) are willing to cooperate over file carving process to maximize performance and outcome of their investigation. We determine strategic interactions between two players who are using Scalpel and Photorec (denoted as S and P , respectively) as their strategies. Moreover, we assume that the investigator who uses either S or P strategies will get a benefit of carving with a different level of success as well as cost of processing time and unsuccessful file carving. In this 2×2 game, we use the parameters shown in Table 3:

Table 3. Definition of the parameters.

V :	Value (benefit) of file carving process
E_1 :	Effectiveness of file carving using Scalpel
E_2 :	Effectiveness of file carving using Photorec
C_t :	Cost of processing time
C_f :	Cost of unsuccessful file carving

In order to explore the interactions between two investigators (players) using S and P strategies, we create a payoff matrix. In the payoff matrix A depicted in Table 4, the element in row i and column j (a_{ij}) shows the payoffs to the individuals who uses strategies i and j against to each other. If a player who uses Scalpel cooperates with a player who uses Photorec, Scalpel player gets the benefit of VE_1 however pays the cost of processing time C_t while Photorec player gets the benefit of VE_2 and pays the cost of unsuccessful file carving C_f . Therefore, the expected payoff of a player who plays S becomes $VE_1 - C_t$ and similarly the expected payoff of a player who plays P becomes $VE_2 - C_f$. Table 4 shows the payoff matrix A for Scalpel S and Photorec P strategies. Here in matrix A , each entry has two elements separated by a comma. The first element (left hand side of the comma) represents the payoff of the row player (commonly referred as first player) the second element (right hand side of the comma) represents the payoff to column player (commonly referred as second player). For example, if we look at the a_{12} element of the payoff matrix A shown in Table 4, $VE_1 - C_t$ is the payoff to row player who uses Scalpel as a strategy and $VE_2 - C_f$ is the payoff to column player who uses Photorec as a strategy.

Table 4. Payoff matrix A for Scalpel-Photorec (S - P) game.

	S	P
S	$VE_1 - C_t, VE_1 - C_t$	$VE_1 - C_t, VE_2 - C_f$
P	$VE_2 - C_f, VE_1 - C_t$	$VE_2 - C_f, VE_2 - C_f$

5. Game-Theoretic Analysis and Results

In game theory, Nash equilibrium, named after the Nobel prize winner mathematician John Forbes Nash Jr., is used to analyze and find the proposed solution which assumes that each player knows the other players' strategies and has no incentive to change its own [28]. Simply, if no player can gain benefit by changing their own strategies while the other players keep theirs unchanged, then the current set of strategies are the Nash equilibrium of the game. From the Table 4, (S, S) is a Nash equilibrium if the expected payoff of playing S against S is greater than that of playing P against S . Thus, if the Equation (2) below holds,

$$VE_1 - C_t > VE_2 - C_f \quad (2)$$

(S, S) is a Nash equilibrium. If we divide Equation (2) by V , we get the Equation (3) below.

$$E_1 - \frac{C_t}{V} > E_2 - \frac{C_f}{V} \quad (3)$$

Similarly, if the expected payoff of playing P against P is greater than that of playing S against P , then (P, P) is a Nash equilibrium if the Equation (4) holds.

$$VE_2 - C_f > VE_1 - C_t \quad (4)$$

If we divide Equation (4) by V , we get the Equation (5) below.

$$E_2 - \frac{C_f}{V} > E_1 - \frac{C_t}{V} \quad (5)$$

Let us denote $\Delta_S = E_1 - \frac{C_t}{V}$ and $\Delta_P = E_2 - \frac{C_f}{V}$ which represent the differences between effectiveness of file carving and cost-value ratios using Scalpel and Photorec, respectively. Figures 1 and 2 show the equilibrium diagrams for the one-shot S – P game.

It can be said that playing Scalpel is the best strategy whenever the difference of the effectiveness of file carving and cost-value ratio (we will call this as net gain) using Scalpel is bigger than net gain using Photorec as shown in Figure 1.

Figure 1. Equilibrium diagram for the S – P game with parameters Δ_S and Δ_P .

Figure 2. Equilibrium diagram for the S – P game with parameters Δ_S and Δ_P .

On the other hand, playing Photorec is the best strategy whenever the net gain of using Photorec is bigger than the net gain of using Scalpel as shown in Figure 2.

Until here, we have focused on playing (choosing) only S or P which refers to the pure strategies of the game. However, we can elaborate on the analysis by focusing on finding the mixed strategy Nash equilibrium of the game. A mixed strategy Nash equilibrium involves at least one player playing a randomized strategy and no player being able to increase their expected payoff by playing an alternate strategy. If a player randomizes over strategy S or strategy P , then both of these strategies

must produce the same expected payoff. Otherwise, the player would prefer one of them and would not play the other.

Here we investigate whether there is a mixed strategy Nash equilibrium (MSNE) for the one-shot S - P game. Let σ represents the probability that a player plays a particular pure strategy S or P . Namely, we use σ_S and σ_P as probabilities player 1 plays S and P , respectively which can be written as

$$\sigma_S + \sigma_P = 1 \quad (6)$$

Using this notation, we can write player 2's expected utility of playing S as a pure strategy and the function of player 1's mixed strategy σ_S

$$EU_S = \sigma_S(VE_1 - C_t) + (1 - \sigma_S)(VE_1 - C_t) = (VE_1 - C_t) \quad (7)$$

and player 2's expected utility of playing P as a pure strategy is

$$EU_P = \sigma_S(VE_2 - C_f) + (1 - \sigma_S)(VE_2 - C_f) = (VE_2 - C_f) \quad (8)$$

We are looking for a mixed strategy from player 1 that leaves player 2 indifferent between their pure strategies. We want to find a σ_S and σ_P such that

$$EU_S = EU_P \quad (9)$$

$$VE_1 - C_t = 1 - VE_2 - C_f \quad (10)$$

If we divide Equation (10) by V we get the Equation (11).

$$E_1 - \frac{C_t}{V} > E_2 - \frac{C_f}{V} \quad (11)$$

As discussed in Section 5, Equation (11) can be rewritten as

$$\Delta_S = \Delta_P \quad (12)$$

which means there is a MSNE if $E_1 = E_2$ and $C_t = C_f$ (see Figure 3). This result shows that if the tools are indifferent in terms of the effectiveness of file carving using either Scalpel or Photorec as well as costs of processing time and unsuccessful file carving then investigator can choose any tool he/she finds convenient or easy to use. Even if we calculate the mixed strategy for player 2 that leaves player 1 indifferent between his/her two pure strategies S and P we will end up having the same equations as Equations (7)–(12) defined earlier. Therefore, we have exactly the same conclusions here.

Figure 3. Mixed strategy Nash equilibrium for the one-shot S - P game.

6. Conclusions

In the field of digital forensics, practitioners/examiners/investigators often have several tool options to use during their investigations for a given task. In order to conduct their investigations, examiners prefer to utilize one tool over the others depending on various reasons such as execution time required to use the tool and success of the tool for particular tasks. It is also common case that digital forensic tools do not have exactly the same performance on different cases involving different tasks or devices. Needless to say that examiners also do not have extra time to test the tool performances. Due to all of these issues, examiners usually prefer to use their tools depending on their proficiency and previous experiences unless the tools fail to perform as expected.

It is clearly shown that examiners significantly deal with a decision making problem with respect to tool selection. In this paper, we aim to discuss and show that game theoretic modeling can significantly ease this decision making process for investigators. This is an important contribution as Beebe [29] discusses that it is always welcomed by the digital forensics community. We strongly believe that our proposed work can also be modified and integrated into open source tools (e.g., Autopsy [30] and Digital Forensics Framework [31]) which perform similar tasks in order to automatically suggest tool selection.

In this paper, we proposed a novel game theoretic model for the digital forensics tool selection problem during the file carving process. This game theoretic model suggests optimum and the most efficient strategies for digital forensics investigators. Our results show that if the net gain of using Scalpel is greater than net gain of using Photorec, then the investigators should use Scalpel for file carving process, and vice versa. Moreover, our results also suggest that if the tools are indifferent in terms of the effectiveness of file carving using either Scalpel or Photorec as well as costs of processing time and unsuccessful file carving then an investigator can choose any tool he/she finds convenient or easy to use.

An obvious limitation of this work is the comparison of only two carving tools in the model. However, this was intentional and due to the sake of simplicity in game theoretic modeling and its analysis. Nevertheless, as a future work, our goal is to increase the number of tools for comparison.

Although we chose data carving tools for this study, similar game theoretic model can be developed for other digital forensic tool functions such as deleted file recovery from a filesystem or parsing unsupported application data on a mobile device. In addition, we believe this work will open new application areas regarding adaptation of game theoretic modeling to various digital forensics related areas. For instance, it could be applied to SCADA/ICS and Internet of Things (IoT) forensics to use available tools with automated selection.

Author Contributions: Conceptualization, U.K. and T.K.; methodology, U.K. and T.K.; validation, U.K. and T.K.; formal analysis, T.K.; investigation, U.K.; writing—original draft preparation, U.K. and T.K.; visualization, U.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MAB	Multi-armed Bandit Algorithm
GIDA	Game theory inspired defense algorithm
ARMOR	Assistant for 160 Randomized Monitoring over Routes
MSNF	Mixed strategy Nash equilibrium
IoT	Internet of Things

References

1. Nelson, B.; Phillips, A.; Steuart, C. *Guide to Computer Forensics and Investigations*, 5th ed.; Course Technology Press: Boston, MA, USA, 2015.
2. Sammons, J. *Digital Forensics: Threatscape and Best Practices*; Syngress: Rockland, MA, USA, 2015.
3. Seigfried-Spellar, K.C.; Leshney, S.C. The intersection between social media, crime, and digital forensics: #WhoDunIt? In *Digital Forensics*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 59–67.
4. Shiva, S.; Roy, S.; Dasgupta, D. Game theory for cyber security. In Proceedings of the 6th ACM Annual Workshop on Cyber Security and Information Intelligence Research, Oak Ridge, TN, USA, 21–23 April 2010; p. 34.
5. Nash, J. Non-Cooperative Games. *Ann. Math.* **1951**, *54*, 286–295. [CrossRef]
6. Nash, J. The Bargaining Problem. *Econometrica* **1950**, *18*, 155–162. [CrossRef]
7. Von Neumann, J.; Morgenstern, O.; Kuhn, H.W.; Rubinstein, A. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*; Princeton University Press: Princeton, NJ, USA, 1944.
8. Karabiyik, U.; Karabiyik, T. Digital Forensics Tool Selection with Multi-Armed Bandit Problem. In Proceedings of the Conference on Digital Forensics, Security and Law. Association of Digital Forensics, Security and Law, San Antonio, TX, USA, 17–18 May 2017; pp. 179–195.
9. Visti, H. ForGe Forensic test image generator. Available online: <https://github.com/hannuvisti/forgel> (accessed on 3 May 2020).
10. Richard, G.G.; Marziale, L. Scalpel. Available online: <https://github.com/sleuthkit/scalpel> (accessed on 3 May 2020).
11. Grenier, C. PhotoRec. Available online: <http://www.cgsecurity.org/wiki/PhotoRec> (accessed on 3 May 2020).
12. Stamm, M.C.; Lin, W.S.; Liu, K.R. Forensics vs. anti-forensics: A decision and game theoretic framework. In Proceedings of the IEEE 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 1749–1752.
13. Stamm, M.C.; Lin, W.S.; Liu, K.R. Temporal forensics and anti-forensics for motion compensated video. *IEEE Trans. Inf. For. Secur.* **2012**, *7*, 1315–1329. [CrossRef]
14. Bao, T.; Shoshitaishvili, Y.; Wang, R.; Kruegel, C.; Vigna, G.; Brumley, D. How Shall We Play a Game?: A Game-theoretical Model for Cyber-warfare Games. In Proceedings of the IEEE 2017 30th Computer Security Foundations Symposium (CSF), 2017; Santa Barbara, CA, USA, 21–25 August 2017; pp. 7–21.
15. Fung, B. The NSA hacks other countries by buying millions of dollars' worth of computer vulnerabilities. Available online: <https://www.washingtonpost.com/news/the-switch/wp/2013/08/31/the-nsa-hacks-other-countries-by-buying-millions-of-dollars-worth-of-computer-vulnerabilities/> (accessed on 1 April 2020).
16. Zetter, K. US used zero-day exploits before it had policies for them. Available online: <https://www.wired.com/2015/03/us-used-zero-day-exploits-policies/> (accessed on 1 April 2020).
17. Vidal, R.; Shakernia, O.; Kim, H.J.; Shim, D.H.; Sastry, S. Probabilistic pursuit-evasion games: Theory, implementation, and experimental evaluation. *IEEE Trans. Robot. Autom.* **2002**, *18*, 662–669. [CrossRef]
18. Moayedi, B.Z.; Azgomi, M.A. A game theoretic framework for evaluation of the impacts of hackers diversity on security measures. *Reliab. Eng. Syst. Saf.* **2012**, *99*, 45–54. [CrossRef]
19. Zhang, C.; Sinha, A.; Tambe, M. Keeping pace with criminals: Designing patrol allocation against adaptive opportunistic criminals. In Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems; Istanbul, Turkey, 4–8 May 2015; pp. 1351–1359.
20. Wu, Q.; Shiva, S.; Roy, S.; Ellis, C.; Datta, V. On modeling and simulation of game theory-based defense mechanisms against DoS and DDoS attacks. In Proceedings of the 2010 Spring Simulation Multiconference, Orlando, FL, USA, 11–15 April 2010; p. 159.
21. Pita, J.; Jain, M.; Marecki, J.; Ordóñez, F.; Portway, C.; Tambe, M.; Western, C.; Paruchuri, P.; Kraus, S. Deployed ARMOR protection: the application of a game theoretic model for security at the Los Angeles International Airport. Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial Track; Estoril, Portugal, 12–16 May 2008; pp. 125–132.
22. Jiang, A.X.; Jain, M.; Tambe, M. Computational game theory for security and sustainability. *J. Inf. Proc.* **2014**, *22*, 176–185. [CrossRef]

23. Shiva, S.; Roy, S.; Bedi, H.; Dasgupta, D.; Wu, Q. A stochastic game model with imperfect information in cyber security. In Proceedings of the 5th International Conference on i-Warfare and Security, Dayton, OH, USA, 8–9 April 2010; pp. 308–318.
24. Lye, K.W.; Wing, J.M. Game strategies in network security. *Int. J. Inf. Sec.* **2005**, *4*, 71–86. [[CrossRef](#)]
25. Schöttle, P.; Böhme, R. Game theory and adaptive steganography. *IEEE Trans. Inf. For. Sec.* **2015**, *11*, 760–773. [[CrossRef](#)]
26. Garfinkel, S.; Farrell, P.; Roussev, V.; Dinolt, G. Bringing science to digital forensics with standardized forensic corpora. *Digit. Invest.* **2009**, *6*, S2–S11. [[CrossRef](#)]
27. Karabiyik, U.; Aggarwal, S. Advanced Automated Disk Investigation Toolkit. In *IFIP International Conference on Digital Forensics*; Springer: Berlin, Germany, 2016; pp. 379–396.
28. Nash, J.F. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 48–49. [[CrossRef](#)] [[PubMed](#)]
29. Beebe, N. Digital Forensic Research: The Good, the Bad and the Unaddressed. In *Advances in Digital Forensics V: IFIP Advances in Information and Communication Technology*; Peterson, G., Sheno, S., Eds.; Springer: Boston, MA, USA, 2009; Volume 306, pp. 17–36, doi :10.1007/978-3-642-04155-6_2.
30. Carrier, B.D. The Sleuthkit. Available online: <http://www.sleuthkit.org/autopsy> (accessed on 3 May 2020).
31. ArxSys. Digital Forensics Framework. Available online: <http://www.digital-forensic.org> (accessed on 3 May 2020).

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).