An Intelligent Analysis Model for Multisource Volatile Memory

Xiaolu Zhang, Liang Hu, Shinan Song, Zhenzhen Xie, Xiangyu Meng, and Kuo Zhao*
Department of Computer Science and Technology, Jilin University, Changchun130012, P. R. China
*Corresponding author, Email: xiaoluzhang1985@gmail.com

Abstract—For the rapidly development of network and distributed computing environment, it make researchers harder to do analysis examines only from one or few pieces of data source in persistent data-oriented approaches, so as the volatile memory analysis either. Therefore, mass data automatically analysis and action modeling needs to be considered for reporting entire network attack process. To model multiple volatile data sources situation can help understand and describe both thinking process of investigator and possible action step for attacker. This paper presents a Game model for multisource volatile data and applies it to main memory images analysis with the definition of space-time feature for volatile element information. Abstract modeling allows the lessons gleaned in performing intelligent analysis, evidence filing and automating presentation. Finally, a test demo based on the model is also present to illustrate the whole procedure.

Index Terms—Intelligent Analysis Model; Multisource; Volatile Memory; Game Theory

I. INTRODUCTION

Volatile memory analysis is a solution for digital forensics and it has been confirmed by researchers and experts. This paper describes an analysis model which abstract multisource data from space-time and actor difference especially for volatile memory. Both of the features are almost universal in computer media, memory, digital documents, network traces, and other kinds of digital medium, too. Model space-time feature for volatile objects can reduce redundancy from macroscopic point of view. Because capture time of two digital evidence pieces separated by a year is less important than those capture in the same day, especially for volatile data source. And for space feature, volatile memory in the same location means very different things with those distributed everywhere.

Multisource data model based on the general situation description of following element: space, time and actors. These can be divided into external factor and internal factor corresponded to space-time and actors. In particular, some special programs (malware or tools used by intruder) are also included in actors as intruders themselves. Because both of malware and intruder can cause context changes, such as an intruder can transfer or delete files, so can Virus. For better description, dynamic incomplete information game theory is applied in this situation to model action and all kind of actors is abstracted as players in game. Digital forensic is faced with a finished environment, which means investigators action would not damage the original spot. So it is different from traditional game model, investigator does not participate in game. Base on analysis above, different actor strategy is put forward.

Combine with external and internal features, the full model is provided to help understand and describe both thinking process of investigator and possible action step for attacker on a high abstract level. Due to all of above are the basic abstract for volatile multisource data, it can be employed in both analysis strategy research and intelligent forensics system implementation. In order to test the validity, we accomplished an experiment and the result indicated that this model is meaningful and useful.

Some digital forensic research emphasize time transition feature of digital evidence. Simson Garfinkel in [1] discusses the difference between two pieces of evidence by differential forensic analysis. He points out that some factor will change the piece and transform one snapshot to another. In the analysis of time, Marrington et al. [2] developed a system called CAT Detect which is based on a rule-based system to detect events in the computer activity timeline. And Schatz et al. [3] discussed an approach for correlating multiple timestamps on a computer in order to validate timestamps.

Game theory is the study of the ways in which strategic interactions among economic agents produce outcomes with respect to the preferences (or utilities) of those agents, where the outcomes in question might have been intended by none of the agents [4]. Game theory has provided many math models such as cooperative game model and non-cooperation game model. Many problems of computer science can be seen as a game and when the model can be created, it means that you find a new way to analysis this problem. Especially in a environment that has different roles in it, the game theory can make you know the things better, such as: the weight of different roles is how much? How the system can be affected by the action they made? and so on. Therefore, in the field of computer science, it is still full of benefits, for instance, it offers inspiration to researchers on the behavior and relationship analysis. And it may bring us a new way to think about the digital forensic either.

It is well known that game theory is a decision analysis theory based after action. Because of the worthy to understand and model attack-defense, it is widely
applied in internet security research. In this area, ordinarily, there are two kind of relationship: first is the relationship between offenders and security officers, and the second is the relationship between offenders and target computers. And both of them can we explain by the suited model of game theory. In recently years, many researchers take effort in employing the game theory to the field of information security especially the intrusion detection models. For instance, Bencsath et al. [5] applied game theory and client puzzles for the purpose to devise a defense against denial of service (DoS) attacks. In the same year, a framework to model intrusion detection that based on game theory has been constructed by Kodialam et al. [6], and they developed some sampling schemes that are optimal in the game theoretic setting. After that, Animesh Pacha.et.al [7]have employed a game theory to modeling intrusion detection in mobile and hoc net works, and in the research they establish a model for two player( attacker and individual nodes) non-cooperative game .They believe that the game-theoretic modeling is more realistic than other modeling techniques. Most of their work, the typical sense is to run quantitative cost analysis of active defense and uncertain reasoning for intrusion detection. Kaiquan Shi.et.al created a attack and defense model based on repeat game theory for wireless sensor network and it received a good result[9].Intrusion detection plays a very important role in network security system and the solution that it used can be available to digital forensic either because some work are same[10]. However, there is few researchers use this way to resolve problems in digital forensics so that we expect that our work could be a new attempt. Intelligent analysis method also is employed in activity recognition; Justin et al. [8] apply dynamic Bayesian networks in video stream recognition.

II. SPACE-TIME MODEL FOR MULTISOURCE VOLATILE DATA

External factor or the context of memory dump is described as those nature properties, such as its capturing time, location, size and format. For one piece of data, its space and time attributes are a constant which may not get more attention than its size or format. But when considering several memory dumps, space-time difference become more important than before. So the external factor can be represented by space-time model. To keep it simple, let us compare two snapshots firstly. There are different with each other has three meaning: first, they come from different locations but capture at the same time or time range; second, they are at the different time but in the same host; third, both of the space and time are different. Obviously as in the second case, the relevance between the two images is more like a kind of changing. But in the first case, it can be defined as relationship. The last case is including more uncertainly and closer to the practice.

A. References

Moreover, space or time is no longer an atom meaning, and the concept of group will be introduced. In order to explain clearly, all the forensic snapshots are regarded as main memory images. All the basic definition is shown as follows:

- Each image \( M \) has one, and only one external factor \( E \), it is a quadruple: \( E = (\theta, S_g, T, \tau) \);
- \( S \) is the location information of \( M \), it should be a N-dimensional string or an encoding big integer;
- \( S_g \) is the group location information of \( M \), it should be a N-1-dimensional string or an encoding big integer; if \( S = (s_1, s_2, ..., s_n, \theta) \), then \( S_g = (s_1, s_2, ..., s_n, \theta) \);
- \( T \) is the time information of \( M \), it is an accurate timestamp; \( \tau \) is an timestamp, too;
- The time group \( T_g \) is an time range, it is a set of \( t : T_g = \{ t | T - \tau < t < T + \tau \} \).

B. Definition of Space-time Diagram

Although an external factor is a four-dimensional, it can be simplify in two-dimensional as a space-time diagram which including space information \( S \) and time information \( T_g \). Based on these definitions we are able to broadly classify multiple images as a matrix visualized in Figure. 1. The horizontal axis shows location as \( S_g \) or \( S \) and the vertical axis shows time as \( T \) or \( T_g \) discussed in previous section. Then looking at two main memory images \( M_1 \) and \( M_2 \) for example, \( M_1 \) is capture at the timestamp of Sat-May-05:04:06:48-2012 which is the value of \( T_1 \), and has location information shown as IP address of 192.168.0.1 which is the value of \( S_1 \). The same situation is for \( M_2 \). Considering the network delays, machine time difference and other factors, \( T_g \) of \( M_1 \) has a useful meaning that if \( T_1 - \tau < T_2 < T_1 + \tau \), they can be considered capturing in the same time although they are not equal to each other. This feature has more value for those time sensitive objects such as processes or connections.
III. MULTISOURCE ANALYZING MODEL

As all the relevance is created by internal factors, in the analysis environment of volatile memory, internal factors can be distinguished as intruder and malware. For a generalized understanding, an actor can be a human or a tool. A malware is a program so that it must running under a series of rules, but an intruder may just follow his mood. Although in most cases an intruder may be the chief of the tools, defining them separately would provide a clear analysis procedure. Attack procedure is implemented by actor, of which possible attack paths constitute an attack graph including a set of attack action sequence [10]. When investigator begins to analysis digital evidence pieces, attack graph was hided or partly hided in them, and one of the most significant tasks is to reduction attack process.

A. Dynamic Incomplete Information Games

Game theory is a decision analysis theory based after action. Because of the worthy to understand and model attack-defense, it is widely applied in internet security research. Typical sense is to run quantitative cost analysis of active defense and uncertain reasoning for intrusion detection.

Game theory describes such a case: there are n-player participates in game and everyone can make rational decision. Each player has a sequence of strategies to make choice. Game result is determined by player’s payoff [11]. However there is not always so pure game in real world. Considering the type of player and uncertain information, game model can be extended in complete information and incomplete information. For details when the player only knows possible distribution of the other players’ strategies and he must apply subjective judgment to make a decision, this is called game of incomplete information. Most of network security analysis on game theory is based on it, no matter intrusion detection or digital forensic, to judge attacker’s action is exactly match this description.

Pierpaolo Battigalli presents a dynamic games model with incomplete information as follow [12]:

A game of incomplete information with observable actions is a structure:

$$\Gamma = (N, \{\Theta_i | i \in N\}, \{A_i | i \in N\}, \{H_i | i \in N\}, \{U_i | i \in N\}, P)$$

given by the following elements: $N$ is a non-empty, finite set of players. For each $i \in N$, $\Theta_i \subseteq R^q$ is a non-empty set of possible types for player $i$ and $A_i \subseteq R^r$ is a non-empty set of possible action profiles for player $i$ ($R^q$ is the k-dimensional Euclidean space). $\{H_i | i \in N\}$ is historical action profiles for player $i$. $\{U_i | i \in N\}$ is payoff for player $i$. $P$ is probability distribution defined in $\Theta$.

B. Different Actor Strategy

Obviously, applying incomplete information games model to multisource digital evidence analysis scenario and combining with the discussion at the start of this section, we present a simplified model called different actor strategy as follow:

Build a possible player set $N = (n_1, n_2, ..., n_p)$ and a possible type set $\Theta = (\Theta_1, \Theta_2, ..., \Theta_q)$. Set function $Member(\Theta_i) \Rightarrow n_i, n_{i+1}, ..., n_j, i < j < s, n_s \in N$ is an array of player of which has a type of $\Theta_i$.

Transform player set and type set into a $q \times p$ adjacency matrix, of which two axes are type set and player set, if $Member(\Theta_i) \Rightarrow n_i, n_{i+1}, ..., n_j$, point $(i,i),(i,j),...(i,s)$ has value of payoff $U_i$ in matrix. In particular, if every player is limited to have one type, the adjacency matrix can be transformed into an adjacency list.

Action profiles search procedure:

1. Build a directed search graph $G$, the nodes of $G$ is chosen from $N$ at beginning.
2. Begin until all player nodes are marked “finish”.
3. Choose a non-finish player node $n_i$ (next will discuss how to choose one), which is 0 in-degree.
4. Extend subsequence nodes of $n_i$, if $A_i$ is not empty choose those $A_i$ as sub-nodes of $n_i$, add them in $H_i$, and wipe $A_i$, then link those sub-nodes in graph $G$ and mark $n_i$ is “finish”. If $A_i$ is empty, update $U_i$ and go to step 3.
5. End.

Different player choose strategy is mainly based on payoff for every player. But it is not enough as payoff is dynamic changing through action search procedure, and this can be formal description as an aging function: $U_i = Aging(U_i, P, \Delta')$. What must be highlight for different player type is that impact factor can be a vector $\Delta = \langle \delta_1, \delta_2, ..., \delta_s \rangle$ which may own quite difference for different player type. So the aging function and impact factor is the key of strategy. One helpful suggestion is for those perform known regular behaviors should have higher pay off because they has more reliability, and this can be called experience vector $\Delta E$; for those perform random behaviors or including unknown knowledge should rely on investigator to make decisions, and this can be called investigator vector $\Delta I$.

For most situations, investigator vector is as significant as experience vector, because volatile digital evidence cannot ensure the integrity of evidence. When relying on experience is unable to recover key path of forensic, subjective assume of investigator will extend new element in possible action set $A_i$ and let the loop in action profiles search procedure continue. The whole procedure is presented as a flowchart in Figure. 2

IV. HYBRID MODEL FOR VOLATILE MULTISOURCE DATA ANALYSIS

Now we can model the scenario for volatile multisource data analysis. There are key points in this issue: one is multisource, which apply space-time model to deal with identity definition, environmental description,
classification of snapshots and macroscopically deciding for different analysis strategy chosen; the other is volatile data, which is quite different from persistent data that may not provide investigator integrity data to run automatic analysis based on experience database. So the subjective judgment of investigator is imported to accomplish the whole forensic procedure. Another reason of bringing uncertainty in is attacker’s behavior, which is hard to judge in advance. Although network intrusion must follow some specific steps, it is still quite different from operational procedure of malware.

Traditional way described intrusion process is focus on regarding attacker and malware as a unified whole. In this paper, the two is separate from each other. An important reason is if malware do something it will probably leave its trace such as a client or a copy, but the action of an attacker is largely hidden, which can only deduce or guess by investigator himself.

Considering how an investigator executes his reasoning, something potential exist may draw his attention to, one mentioned in this post is those changing with time. Combine time and space as a consideration gain analyses comprehensively, and separate attacker and his tools add decision flexibility. All of above bring a better understanding of whole scenario due to abstract model.

A. Examples

Consider in a fine grit view, we can provide a more common example for the model above in Figure 3. There are four main memory images M1, M2, M1’ and M2’. M1 and M1’ has the same location information (see section 3.1), which means they are captured in the same host in this scenario, so do M2 and M2’. Circle means processes in an image and diamonds represent files information in an image. Dotted line is the certain action such as obviously existed or getting from experience vector (see section 4.2). All the shape and line in bold and shadow is get from investigator vector (see section 4.2).

Along the transverse can get follow things. In M1, we can get information that P1 and P2 has process creation relationship, which can get from process list directly, so do P3 and P4 in M2. Between M1 and M2, P2 and P3 has communication relationship, which may be defined under the experiment vector of server process and client process. In M2, F3 is a file exploited by P4, which may be a creation. Then we go upper layer to see the next time frame. In M1’, P1 visits F1. In M2, there are only P3 and F3’, but P3 and F3’ do not have any relationship just depend on experiment vector.

Along the longitudinal can get other information. Compared M1 and M1’ can see that P2 has been destroyed and F2 is a new file created right now. Compared M2 and M2’ can get similar information that P4 has been destroyed and F3 has been changed to F3’ (they have the same location information, so make the

---

**Figure 2.** Space-time Diagram for volatile memory images

**Figure 3.** Example of hybrid model for processes and files in main memory
assumptions). By now, experiment vector cannot carry forward the procedure of forensic one more step. If follow the flow chart of different actor strategy, it is the point of $A_i = \emptyset$ (see section 4.2). Then we can begin to use investigator vector.

The investigator may make a possible judgment as follow: $F_2$ is a copy of $F_1$ executing by $P_1$; $P_2$ has communication with $P_3$; $P_4$ accesses $F_3$ and changes it. And add the history action in experiment vector for $P_1$ and $P_2$ has creation relationship, so do $P_3$ and $P_4$. The investigator may find a “path” from $F_2$ to $F_3$, then depend on this, he may make more heroic assumptions that $F_3'$ include some data of $F_2$ or $F_1$.

All the description above is quite nature for a human. The abstract model help computer or forensic analysis system can run such a procedure independently. In return, computer can return an intuitive, friendly results for humans those do not need special training.

B. Use Cases and Test Demo

The model can be implemented as an automating system, which needs a database to store experiment vector (see section 4.2). The input of this system is several piece of evidence classified by space-time model. Results of classify determine the alternative of experiment vector and investigator vector. Then along with two-axes of space-time matrix run different actor strategy separately.

Investigator is a key point in the strategy which have two implementation way. One is leave interactive port for those used the system with the purpose of receiving users’ subjective decision. The other is coding intelligent algorithm to imitate a human’ thinking process and make a decision.

An automatic demo is applied based on rules and assumptions in Table 1, in which a twenty nodes module is supposed. Based on our space-time model, one node acquired in different time is regarded as different node. And the relationship defined in these two nodes is time relationship. The elements have been defined before and their detailed descriptions are in table, either. This is a fully automatically demo for the reason of no investigator take part in whole process. A simple strategy used by the program is that higher pay off one should always be chosen.

The information of nodes and calculate results of pay off is shown in Table 2. Node ID is from 0 to 19. If two nodes has network communication or created relationships, they are defined on the action of space, which is marked as S. If one node is changed as time going on, the former one and later one are defined on the action of time, which is marked as T. The action set is present randomly. Redundancy definition is not included, which means one relationship is only defined once. Action set in this demo is defined by two attribute: type and destination node. Because both space and time relationship is a kind of relative binary relation, and one node itself can be regard as source node.

From the test results, pay off calculate by the program represents critically of each node, which means higher pay off indicates more changes or interactive. When program choose next node to investigate, refer to the pay off at that time, program will give priority selection to the unfinished highest pay off one, just as described by algorithm in Fig. 2. The experiment result provides that all the suspicious nodes are priority processing.

<table>
<thead>
<tr>
<th>Node id(order)</th>
<th>Type</th>
<th>Dest node</th>
<th>Space info</th>
<th>Time info</th>
<th>Test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>N0</td>
<td>S</td>
<td>1</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>U</td>
</tr>
<tr>
<td>N1</td>
<td>N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N2</td>
<td>S</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>N3</td>
<td>N</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>N4</td>
<td>T</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>N5</td>
<td>S</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>N6</td>
<td>T</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>N7</td>
<td>N</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>N8</td>
<td>N</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>N9</td>
<td>N</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>N10</td>
<td>N</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>N11</td>
<td>N</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>N12</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>N13</td>
<td>N</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>N14</td>
<td>N</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>N15</td>
<td>T</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>N16</td>
<td>S</td>
<td>1</td>
<td>1</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>N17</td>
<td>N</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>N18</td>
<td>N</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>N19</td>
<td>N</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Action, pay off and processing order for test nodes set

For a clearly visualization result, all the information can be organized into a hybrid model like Fig. 3. The
figure can be implementation automatically by the program with the employ of Graphviz [14], see Fig. 4. Action type, pay off and processing sequence are marked on the line, and the ellipse represents node. Those nodes in the same square are acquired in the same space and at the same time, which means they has same space time characteristic. Combined with Fig. 4 and Table 2 we can get a better understand for the implementation of program and algorithm. From the line and order number, it is clear that the algorithm employs searching process following relationship chain, which is nearly similar with a real investigator.

V. CONCLUSION AND RESEARCH ISSUES

In a distributed system, data may have different source such as a net organization of DDoS. In this paper, we use the space-time feature of multirsource with a two-dimensional matrix to model the location and timestamp difference for piece of volatile evidence. Except for providing an intuitive description and identification of data, the space-time model also points out how to choose analysis policy along location information defined in abscissa axis or time through in vertical axis. To model the internal factor, game theory is applied to definite actor or player’s behavior process. Different from other application in network security by game theory to represent attack process, we believe that separate attacker from malware is more realistic than previous modeling techniques. As part of our future work, we intend to extend our model to cloud computing environment analysis since the cloud forensic is a hot issue these years and it still faced a lot of problems. From our exited work we hope that we can make a better solution [15]. And this work may be meaningful for acquiring digital evidence and the validity need our future work to test [16].

ACKNOWLEDGMENT

This work is supported by the National Grand Fundamental Research 973 Program of China under Grant No. 2009CB320706, the National High Technology Research and Development Program of China under Grant No. 2011AA010101, the National Natural Science Foundation of China under Grant No. 61073009 and 60873235, the Youth Foundation of Jilin Province of China under Grant No. 201101035, and the Fundamental Research Funds for the Central Universities of China under Grant NO.200903179.

REFERENCES


Xiaolu Zhang was born in Changchun, China. He is a research student of Jilin University of China. Currently, he is studying for Ph.D. in computer architecture at Jilin University of China. His interests are in digital forensics and computer networks & information security.

Liang Hu was born in Changchun, China, on Feb. 1968. He received his Ph.D. degree in computer software and theory from
Jilin University of China in 1999. He is currently a professor in the Department of Computer Science and Technology, Jilin University, China. As an author or co-author, he has published more than 180 research papers, including 80 journal papers, 8 books, and has 1 National invention patent of China as well as 18 computer software copyright registration certificates of China. His research interest covers network security and distributed computing, including related theories, models, and algorithms of PKI/IBE, IDS/IPS, and grid computing.

**Zhenzhen Xie** A master’s candidate of the college of College of Computer Science and Technology, Jilin University, China. Her current research areas are digital forensic technology for cloud computing environments and behavior modeling for digital forensics.

**Shinan Song** A master candidate of the College of Computer Science and Technology, Jilin University, China. Her current research areas are digital forensics technology for windows memory environments and behavior modeling for digital forensics.

**Xiangyu Meng** A master candidate of the College of Computer Science and Technology, Jilin University, China. His current research areas are digital forensics technology for windows memory environments and behavior modeling for digital forensics.

**Kuo Zhao** received the B.E degree in Computer Software in 2001 form Jilin University, followed by M.S degree in Computer Architecture in 2004 and Ph.D. in Computer Software and Theory from the same university in 2008. He is currently Associate Professor in the College of Computer Science and Technology, Jilin University. His research interests are in operating systems, computer networks and information security.