

# A knowledge-assisted visual malware analysis system: Design, validation and reflection of KAMAS

## *Supplemental Material*

Markus Wagner <sup>a,b,\*</sup>, Alexander Rind <sup>a,b</sup>, Niklas Thür <sup>a</sup> and Wolfgang Aigner <sup>a,b</sup>

<sup>a</sup> St. Poelten University of Applied Sciences, Austria

<sup>b</sup> Vienna University of Technology, Austria

### **Abstract**

IT-security experts engage in behavior-based malware analysis in order to learn about previously unknown samples of malicious software (malware) or malware families. For this, they need to find and categorize suspicious patterns from large collections of execution traces.

Currently available systems do not meet the analysts' needs described as: visual access suitable for complex data structures, visual representations appropriate for IT-security experts, provide workflow-specific interaction techniques, and the ability to externalize knowledge in the form of rules to ease analysis and for sharing with colleagues.

To close this gap, we designed and developed KAMAS, a knowledge-assisted visualization system for behavior-based malware analysis. KAMAS supports malware analysts with visual analytics and knowledge externalization methods for the analysis process.

The paper at hand is a design study that describes the design, implementation, and evaluation of the prototype. We report on the validation of KAMAS by expert reviews, a user study with domain experts, and focus group meetings with analysts from industry. Additionally, we reflect the gained insights of the design study and discuss the advantages and disadvantages of the applied visualization methods.

It is very interesting that the arc-diagram was one of the preferred visualization techniques during the design phase but it did not provide the expected benefits for pattern finding. In contrast, the seemingly simple looking connection line was described as supportive finding the link between these tables.

---

### **Keywords**

malicious software, malware analysis, behavior-based, prototype, visualization, visual analytics, interactive, knowledge generation, design study

---

## Description

As supplemental material, we added five different analysis clusters to the package, containing between 10 and 17 analyzed malware samples. These samples are from different malware families (e.g., the cluster C000-0031 contains Graybird, IRCBot and Koodface samples) and the analysis cluster contains between 61 and 794 generated rules. All the used samples were collected by our collaborators from the IT-security department in 2014. Overall they collected a sample set with 800 different samples from different malware families (worms, trojans and bots) for their analysis tests.

## Cluster C000-0031

This cluster contains 16 malware samples and 794 rules:

- 3x Graybird (Trojan)
- 1x IRCBot (Bot)
- 12x Koobface (Worm)

The screenshot displays the KAMAS software interface, which is used for analyzing malware rules. The interface is divided into several panels:

- Knowledge Base:** A tree view on the left showing categories like Malicious Activity, Preparation, Recon, Execution, Exploitation, and TestCases.
- Rule Exploration:** A central table with columns for Occurrence, Calls in Rule, and Length. It shows a list of rules with their respective occurrence counts and lengths. A blue arrow points from a rule in this table to the Call Exploration panel.
- Call Exploration:** A table on the right showing the details of a specific call, including its occurrence and ID. It lists various system calls and their parameters.
- Test Menu:** A panel at the bottom left with buttons for 'Reset Interface' and 'Start Logging'.
- Rule Filter:** A panel at the bottom center with a histogram and a list of filter options, including '# Samples: 16', 'multiples only', 'equal only', 'not known rule (150)', 'partial known rule (0)', and 'full known rule (0)'. It also includes 'Occurrence' and 'Length' filters with corresponding histograms.
- Call Filter:** A panel at the bottom right with a 'Call Filter' input field and a 'Call (Regex):' field.

## Cluster C000-0073

This cluster contains 17 malware samples and 195 rules:

- 8x Bagle (Worm)
- 2x Bifrost (Trojan)
- 1x Brontok (Virus)
- 1x IRCBot (Bot)
- 5x Prorat (Trojan)

The screenshot displays the KAMAS malware analysis interface, divided into several panels:

- Knowledge Base:** A tree view on the left showing categories like Malicious Activity, Preparation, Recon, Execution, Exploitation, and TestCases.
- Rule Exploration:** A central table with columns for Occurrence, Calls in Rule, and Id. It shows a list of rules and their associated calls. A blue arrow points from a rule entry to the Call Exploration panel.
- Call Exploration:** A table on the right showing a list of system calls with their Occurrence and Call details. A blue arrow points from a call entry back to the Rule Exploration panel.
- Test Menu:** A panel at the bottom left with buttons for 'Reset Interface' and 'Start Logging'.
- Rule Filter:** A panel at the bottom center showing a bar chart and filters for '# Samples: 17', 'multiples only', 'equal only', 'not known rule (113)', 'partial known rule (0)', and 'all known rule (7)'. It also displays 'Occurrence: 17' and 'Length: 1'.
- Call Filter:** A panel at the bottom right showing a bar chart and filters for 'Occurrence: 1' and 'Length: 68'. It also displays 'Call (Regex):' and 'Id: 100133, 100089'.

# Cluster C000-00134

This cluster contains 10 malware samples and 61 rules:

- 10x Hybris (Worm)

The screenshot displays the KAMAS interface with the following components:

- Knowledge Base:** A tree view on the left showing categories like Malicious Activity (1), Preparation, Recon, Execution, Exploitation, and TestCases (1). A specific rule path is highlighted: `RtDnsPathNameToPathName_U -> HOpenFile`.
- Rule Exploration:** A central table with columns for Occurrence, Calls in Rule, and Length. It lists 61 rules, with a red bar highlighting a specific row. A blue callout box points to a rule in this table.
- Call:** A detailed view of a selected rule, showing its ID and the specific call signature: `RtUnicodeStringToAnsiString`.
- Call Exploration:** A table on the right showing various system calls and their IDs, such as `DllMain` (101561) and `RtlUnicodeStringToAnsiString` (107995).
- Rule Filter:** A bar chart at the bottom left showing the distribution of rules across 10 samples. It includes checkboxes for "multiples only", "equal only", "not known rule (43)", "partial known rule (0)", and "not known rule (1)".
- Call Filter:** A section at the bottom right with input fields for Occurrence (set to 1) and Call (set to 5). It also includes a "Call (Regex):" field and a "Case Sensitive" checkbox.



# Cluster C000-00199

This cluster contains 15 malware samples and 748 rules:

- 13x Fizzer (Worm)
- 2x IRCBot (Bot)

The screenshot displays the KAMAS interface with the following components:

- Knowledge Base:** A tree view on the left showing categories like Malicious Activity (5), Preparation, Recon, Execution, Exploitation, and TestCases (5).
- Rule Exploration:** A central table with columns for Occurrence, Calls in Rule, and Length. A blue callout box highlights a specific rule entry.
- Call:** A table on the right showing details for the selected rule, including LocalAlloc, NtAccessCheck, and LocalFree.
- Call Exploration:** A table on the far right listing various system calls with their occurrence counts and IDs.
- Test Menu:** Buttons for 'Reset Interface' and 'Start Logging' at the bottom left.
- Rule Filter:** A section at the bottom center with checkboxes for '# Samples: 15', 'multiple only', 'equal only', and 'Full known rule (1)'. It includes two bar charts for Occurrence (0 to 1897) and Length (1 to 10).
- Call Filter:** A section at the bottom right with a filter value of 435 and a 'Case Sensitive' checkbox.