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

## *Supplemental Material*

Markus Wagner [a,b,*], Alexander Rind [a,b], Niklas Thür [a] and Wolfgang Aigner [a,b]

[a] St. Poelten University of Applied Sciences, Austria
[b] 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

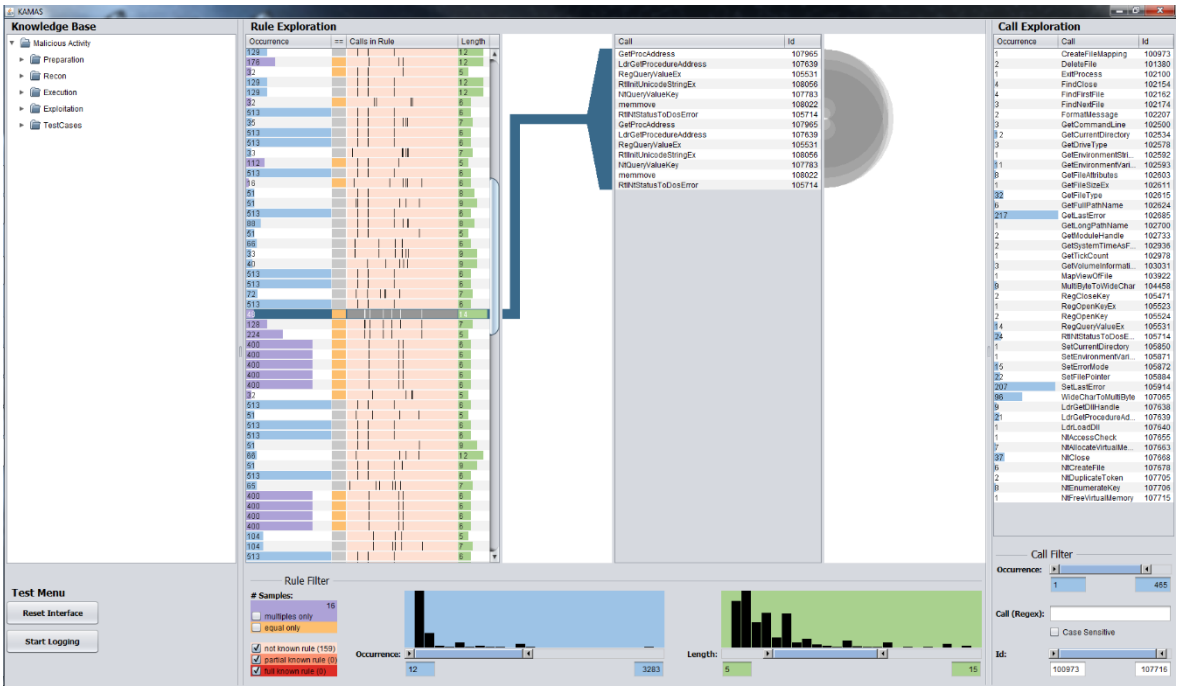
## Description

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:

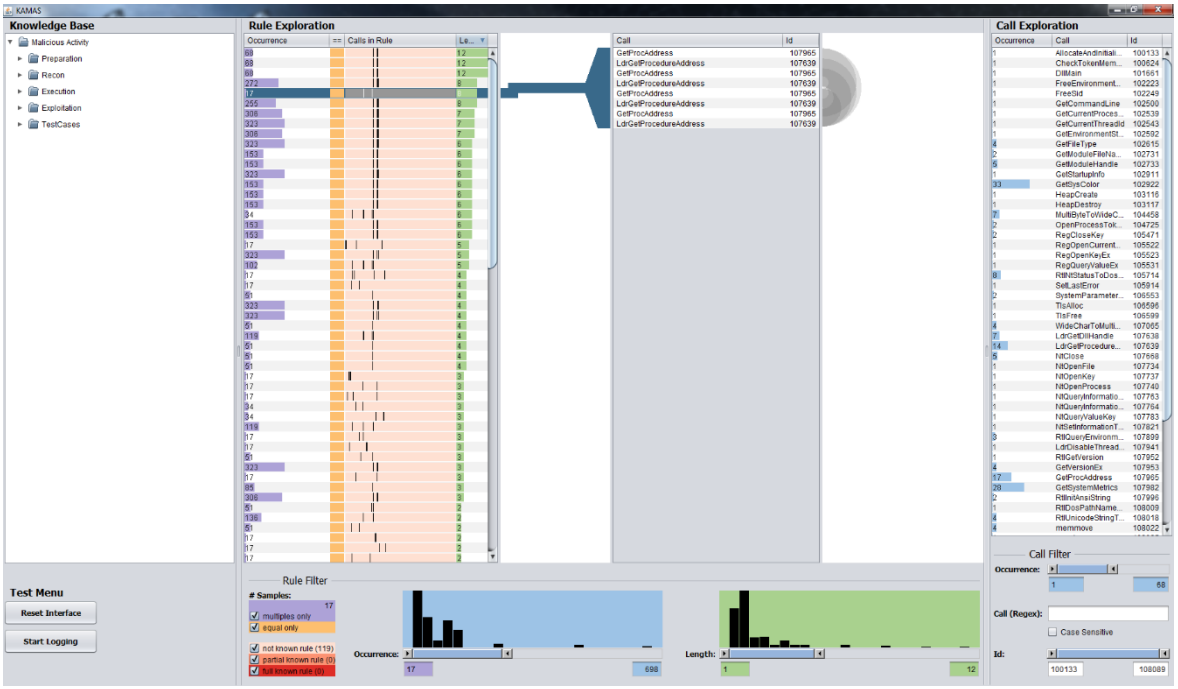
- 3x Graybird (Trojan)
- 1x IRCBot (Bot)
- 12x Koobface (Worm)



**Cluster C000-0073**

This cluster contains 17 malware samples and 195 rules:

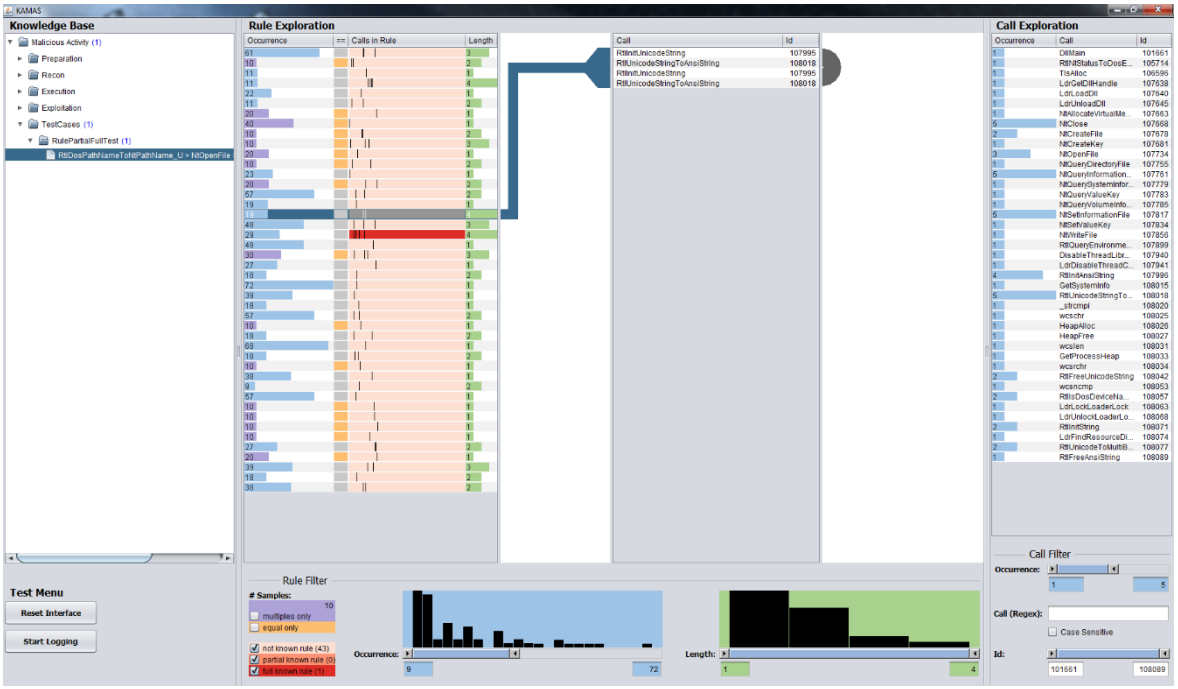
- 8x Bagle (Worm)
- 2x Bifrost (Trojan)
- 1x Brontok (Virus)
- 1x IRCBot (Bot)
- 5x Prorat (Trojan)



## Cluster C000-00134

This cluster contains 10 malware samples and 61 rules:

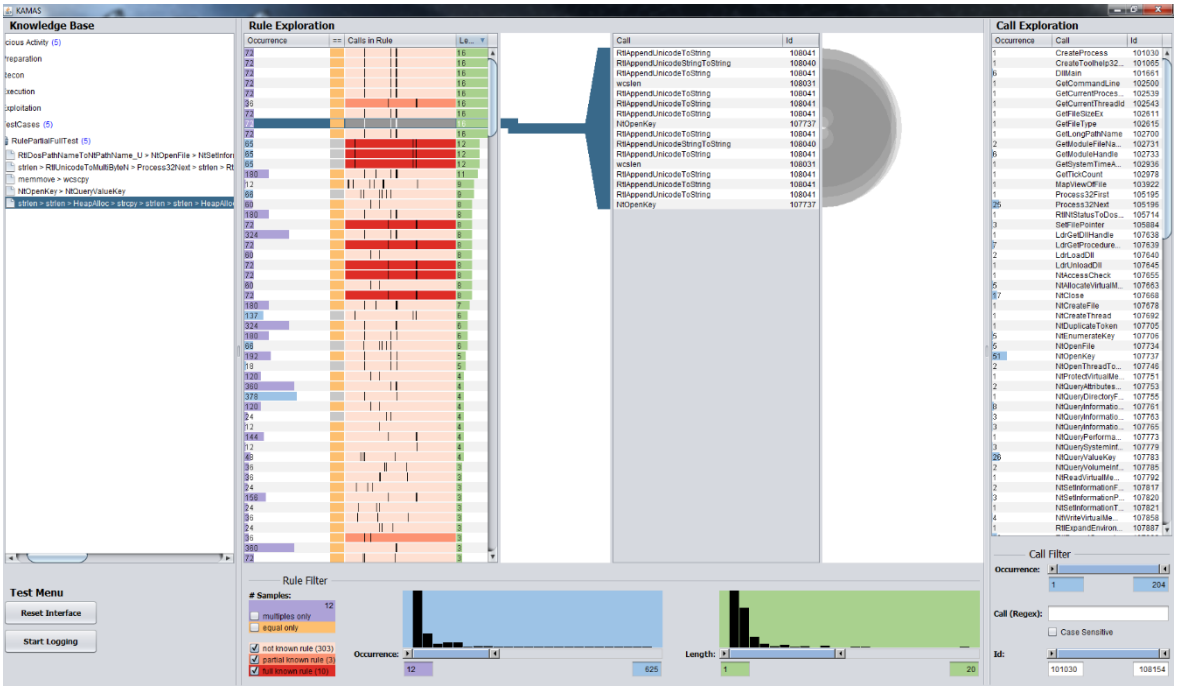
- 10x Hybris (Worm)



## Cluster C000-00166

This cluster contains 12 malware samples and 332 rules:

- 12x Badtrans (Worm)



## Cluster C000-00199

This cluster contains 15 malware samples and 748 rules:

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