SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
Established in collaboration with MIT

Singapore, August 16th 2016

To: Committe for Doctoral Thesis Defence
Computer Science and Engineering
Brno University of Technology

Report for Ivan Homoliak's PhD Thesis "Intrusion Detection in Network Traffic"

Summary

This thesis deals with the application of supervised Machine Learning techniques for the detection of Buffer Overflow (BOF) attacks by analyzing several features of network traffic (such as packet sizes, frequency of packets, statistics on incoming and outcoming flows etc.). Different from previous work, this work proposes to adapt traditional supervised approaches to train on data containing *obfuscated* attacks. Since popular datasets containing normal traffic and attack traffic rely on standard attacks (as performed for instance using Metasploit), it is relatively easy to bypass the generated classifiers by deviating slightly from the exact way the attack was performed. In particular, this thesis explores *tunneling,* such as encapsulating malicious traffic within HTTP or HTTPS and randomly changing the frequency of packets, duplicating some packets, changing order etc. The results in this Thesis show that by training with obfuscated data, the performance of the classifier significantly increases and successfully hampers obfuscation-based attacks (for the obfuscation operators used in the training phase).

Strong points of the work

Given the sophistication of modern attacks and a professionalized malware industry, it is important to develop new techniques to identify and react to attacks. Obfuscation is a trend recently observed in Malware research, and that in my opinion will increase in sophistication and relevance in the future. Therefore, I think the topic of this work is very timely. The obfuscation operators considered are reasonable (specially tunneling) and there is a considerable amount of empirical validation and comparison with publicly available data-sets. The Thesis is overall well structured, the formalism used is clear.

Weak points of the work

Some of the points I'd expect to be explained in more depth during the presentation include the following. First, the title of the thesis is a bit too generic in my view, since Intrusion Detection is a much broader topic than just BOF detection. Although the focus on BOF is motivated on the work, I miss a discussion on the generality of the approach for other types of intrusion or anomalies. On the other hand, I appreciate the effort to come up with obfuscated *attacks,* but it is not so clear to me how much effort was put into come up with realistic *normal* traffic. I think this is a challenging issue with Machine Learning applications in general, since in many realistic situations the "normal" case (no attack) can be very heterogeneous and produce false positives in practice. Last, I missed a discussion on how the feature set would be impacted by choosing a different set of obfuscation operators, or how to generically come up with such a feature set based on

characteristics of the obfuscation operators.

Conclusion

I think that this thesis shows sufficient expertise in the selected area of dissertation and constitutes an interesting contribution to the state of the art in terms of empirical evaluation of network-based attack obfuscation. Therefore, I recommend the acceptance of this Thesis as part of the process of conferment of the PhD title to Ivan Homoliak.

Also in particular:

Is the topic appropriate to the particular area of dissertation and is it up-to-date from the viewpoint of the present level of knowledge?

The topic is appropriate and the document is to the best of my knowledge up-to-date with respect to the state of the art.

Is the work original and does it mean a contribution to the area - specify where the original contribution lies?

The originality of this work lies in my view in the in-depth empirical assessment of the impact of realistic network traffic obfuscation operators to malicious traffic containing BOF related payloads. The depth of this assessment is in my view an original contribution to the state of the art in this area.

Has the core of the doctoral thesis been published at an appropriate level?

I believe the core technical content has been sufficiently disseminated in Journals, Conferences, Workshops and Technical Reports. I think that the author could have been more ambitious in terms of the quality of the avenues for publication, given the interesting results of this work (such as higher ranked conferences and Journals).

Does the list of the candidate´s publications imply that he is a person with an outstanding research erudition?

The publications show that author possess sufficient knowledge in his field of expertise and ability for research.

Sincerely,

Martín Ochoa
Assistant Professor
Information Systems Technology and Design