Session 1: General Testing Issues

Test Strategies & Common Mistakes

Andreas Marx, CEO, AV-Test GmbH
email: amarx@av-test.de

Abstract: *To be sent*

Exploiting the Testing System

Viorel Canja, Head of BitDefender Labs, Bitdefender
email: vcanja@bitdefender.com

Abstract: *High detection rates, fast response time, low false positives rates have become increasingly important as the number of threats has grown continuously and the nature of those threats has changed. Control over a certain computing device can now be turned intomoney and money is a pretty good incentive for the professional malware writers.*

To measure the exact values of those parameters (detection rates, response time) one would need to collect an immense amount of data. As this is not feasible, tests are conducted against a set of malware which is considerably smaller than what exists in the wild.

This paper describes some of the weaknesses of current testing procedures and some techniques that are already used to exploit those weaknesses. While those techniques help in getting better test results they sometimes have a negative impact on real world performance.

Session 2: Not so straightforward Testing

Testing Heuristic Detections

Andrew Lee, Chief Research Officer, ESET LLC
email: aj@eset.com

Abstract: *Testing the heuristic capabilities of an anti-malware is both controversial and problematic.This presentation will examine the problems that such testing encounters, discuss some possible ways forward, and open some issues for further discussion around the issues, that may lead towards a sound and scientific methodology for testing heuristic and behavior based products.*

The Importance of Re-creating In-the-wild Infection Conditions for Testing Multi-Layered Security Products

Mark Kennedy, Distinguished Engineer, Symantec
email: mkennedy@symantec.com

Abstract: *The threat landscape has changed much over the past several years. Viruses lead to worms lead to blended threats lead to spam bot distributed Trojans. In the past, a good signature engine and plenty of definitions was all that was needed to protect from most threats. As packers and other obfuscators entered the scene, it became trivial for attackers to create a multitude of threats very quickly. Moreover, given that downloaders and spam bots could return to a central server to get the next incarnation of a threat it is now possible to create virtually unique threats. All this seeks to undermine the signature based detection systems.*

As the threats grew more complex, so too did the security solutions which detect and block them. Firewalls, host based IPS, behavior blockers were all added to the arsenal. Threats can be detected at many layers of the system. Signatures offered a high certainty of positive identification. The newer heuristic systems offer better protection against unknown threats, but at a loss – to some degree or another – of that certainty.

*In order to reduce false positives heuristic systems are highly tuned to detect threats ***as they actually exist in the wild**. This includes tracking how they arrive on the target machine, how they are initially launched, how they persist, and what actions they take. Any change in any of these items can influence how a heuristic will score a potential threat. As reviewers and testers change their methods – as they must – it is important that they test the threats in as realistic environment as possible. Otherwise, they risk introducing a "lab bias" into the testing which can skew the results, and changing them from what a real user would experience with a real new threat. This presentation will cover what steps testers and reviews could follow to help limit this lab bias.*

Session 3: Creating and Maintaining Collections for Testing

Maintaining a Malware Collection

Dr. Vesselin Bontchev, Antivirus Researcher, FRISK Software
e-mail: bontchev@complex.is

***Abstract**: A well-maintained malware library, or as it is often called, a malware collection, is an important tool to the anti-virus researcher. It can be used to test anti-virus software, to systemize the knowledge about the hundreds of thousands of currently existing malicious programs, as a basis of information exchange with other anti-virus researchers and so on. However, the creation of such a collection and its maintenance in a clean and well-ordered state is not a trivial task, especially with the huge amount of currently known viruses and Trojan horses and new ones appearing at the rate of several thousands per month. This paper describes the major guidelines and procedures used by the author to maintain such collections during the two decades of his career as an anti-virus researcher.*

Determining & Sorting out the Trash

Michael St. Neitzel, Senior Antivirus Architect, FRISK Software
email: mike@f-prot.com

***Abstract:** Sorting out the trash of malware collections is a essential task to gain accurate testing results. It's often underestimated and not quite that simple as it looks. The presentation shows which tools are needed, which knowledge is essential and how to determine and classify trash. It also demonstrates that you cannot fully rely on automated systems for this task. Furthermore it gives a brief overview on the tasks to perform and points out important facts and tricks.*

Building & Leveraging White Database for Antivirus Testing

Mario Vuksan, Director, Knowledgebase Services, Bit9
email: mario@bit9.com

***Abstract:** Did you ever find a false positive and wonder how many files in its family could be potentially affected? What products shipped these files? How to find these products? Are other languages or OS-es affected? Do you ever worry that today's signatures will flag down harmless new software of the future? So you decided to build your ultimate white listed repository of known software to improve the accuracy of your blacklist. How to size the software universe? What should you consider?*

Session 4: Testing in Hindsight

The Difference between Track and Testing Performance

Roel Schouwenberg, Senior Antivirus Researcher, Kaspersky Lab Benelux
email: Roel.Schouwenberg@bnl.kaspersky.com

Abstract: *Real life situations show that there can be significant differences between them and AV testing results. In this presentation, I will go over some of the flaws that result from the current ways of anti-virus testing. I will also attempt to present some patches for these flaws in an effort to make testing and real life performance come closer to each other.*

The VTC experience

Prof. Dr. Klaus Brunnstein, University of Hamburg, Germany
email: brunnstein@informatik.uni-hamburg.de

Abstract: *Established in 1987, Virus Test Center at Hamburg University was the first lab where students learned how to analyse security threats esp. related to malicious software and prepare software solutions to counter related threats (later, other labs worked about chipcard security, biometrics and incident reponse methods). After initial projects (including Morton Swimmer´s ANTIJERU), Vesselin Bontchev (coming from the virus lab of the Bulgarian Academy, Sofia) joined VTC in 1992 and started his AntiVirus test suit; Vesselin was probably the first ever to systematically organise AV tests, and his experiences taught several AV experts and their companies how to improve their products. When Vesselin left (for Iceland), a series of student projects were started where students could learn to organise and maintain a malware datase, prepare testbeds, develop criteria for testing, perform AV/AM tests with special emphasis on detection quality of AntiVirus and AntiMalware products. VTC results were sometimes controversially recognized, esp. when the author announced that product tests would also adress detection of non-replicating malware (aka trojans); at that time, some AV producers withdrew their product from the test (some of which joined later, after having been convinced that AntiVirus-only tests are too restrictive).*

The paper describes methods used by VTC in maintaining testbeds and how tests were performed, esp. also adressing problems found in testing. After the principal investigator finished his teaching carreer (in fall 2004), VTC was closed because of lack of students devoting time to test procedures.