avast | blog

English

Protecting over 400 million PCs, Macs, & Mobiles – more than any other antivirus.

Ondrej Vlcek
3 December 2012

A New Toy in the Avast Research Lab

The Avast Research Lab is where some of the Avast's brightest brains create new ways of detecting malware.

These are either features inside the product (such as FileRep and autosandboxing, including all of its recent development) as well as components that run on our backend – i.e. things that users don't necessarily see but that are equally important for the overall quality of the product.

In fact, working on the backend stuff takes up more of their time these days, as more and more intelligence in Avast is moving to the cloud and/or is being delivered in almost real time via the avast! streaming update technology.

The Avast backend classifiers use a number of techniques, but the two hot ones that the team has been working on hard recently are things that we call ***Malware Similarity Search*** and ***Evo-Gen***.

Avast Malware Similarity Search

Malware Similarity Search is an important feature that allows us to pretty much instantly categorize a big amount of incoming samples. That is, for any file, it is able to say whether the file looks similar to an already seen malware file (or a whole cluster of malware files) as well as whether it's similar to a known clean file (or a cluster of these). This may sound like an easy problem to solve, but in practice this is actually pretty difficult. Of course, the secret sauce here is how you actually define the metric (to be able to talk about similarity) and what all you take into account when representing a file. In Avast we take into account both static properties of the file as well as the outcome of a dynamic analysis (i.e. basically logs gathered during the execution of the file).

Now, a technology like this is obviously very valuable as it allows us to make fast decisions about files that we have never seen before. For example, if a file is very similar to a cluster of known malware samples, and at the same time it is not similar to any clean files, we categorize it immediately as malware. Believe it or not, we're seeing thousands of files like this every day.

Avast Evo-Gen

The second technology I mentioned, *Evo-Gen*, is somewhat similar but a bit subtler in nature. This is about finding as short and generic descriptions of large sets of malware samples as possible. Say you take a set of 1,000,000 malware samples (and 1,000,000 clean files) and give the algorithm the following task: find as few, and as brief descriptions of as many samples in the malware set, without describing any file in the clean set. Evo-Gen is a genetic algorithm that we have developed just for that. It often happens to find some real gems for us – e.g. a description of an apparently random set of tens of thousands of malware files scattered somewhat randomly across our virus sets. And the size of the description? 8 bytes.

Now, if you think about this for a while, you will find out that both of these algorithms have something in common. I mean, for both of them it's necessary to have super-fast access to our vast sets of clean and malware files. Forget about sequential access (or any kind of processing of the files one by one). Even reading the samples off the disks takes hours.

The Real Breakthrough

For this purpose, the team has developed another **great piece of technology that we call *MDE***. It's basically an in-memory database that works on top of indexed data and allows heavily parallel access.

Traditionally, we have been running these things on classic server hardware. For the most part, we use standard Dell servers based on Intel Xeon CPUs. However, the performance has never been great and we always thought we should be doing better.

The real breakthrough came when we started experimenting with the GPUs. For starters, modern GPUs (both from NVidia and AMD) are not limited to high-end graphics or gaming. The good thing about them is that they can be massively parallelized – while today's high-end Intel CPUs contain 6, 8 or maybe 10 cores, the high-end gaming GPUs contain thousands of cores. True, each of them is not that powerful, but if you can unleash their potential with some good parallel algorithms, the resulting power is insane.

So, with MDE, we're now in the process of transitioning to a GPU-based "supercomputing" farm.

It's not a rackmount server – but a workstation instead. A hell of a workstation, I should say though. With Intel i7 E3820 4C 3.6GHz CPU and 32 GB DDR3 RAM, it's not a bad start, but what's really cool about the box is the 4 NVidia GPU-based graphics cards, each with 3 GB of RAM and connected to each other by a hose for external water cooling. The whole beast is powered by a 1,500W power supply but in case it's not enough, we are ready to add one more.

While we haven't put these systems in production yet, we will likely do so soon. And I'm truly looking forward to that – as doing so will allow us to serve you, our users, even better. You never know - if this proves to be as useful as we think it will be, we may end up building something like the Titan one day...

(Now, my job in the meantime will be to keep the gamers off the server room :-)).

Threat Research, Security News

RejZoR

12/21/2013 1:48:15 PM

Is any of this already being used for malware processing or is still planned for avast! 8 release and it will be fired up then for all the new goodies that are probably coming with the v8 release?

---

spywar
12/22/2012, 6:08:52 PM

Hi Vlk,

Atm, if I correctly understand, AV analysts at avast! receive tons of samples everydays from Honeypots, and mainly from community IQ right ?

And to classify them really fast, they use these amazing automated analysis systems, so only a few samples need to be manually processed.

But do you plan something in v8 like : Any suspicious file that an avast protected pc sees is sent directly to these servers for automated analysis if it's classified as malware you update DB from cloud, if not manual scan is done. Btw, I'm sure you guys are preparing well new stuffs/improvements for v8.

---

vlk
12/29/2012, 3:14:41 PM

@RejZoR Only in a limited mode. Most of the good stuff will be gradually deployed during Jan/Feb.

@spywar This is, in a way, it works. But it's not complately real-time (i.e. there's still some delay). That is, if you're the first person in the world who hits a completely new nasty virus (and the heuristics doesn't detect it, as well as the dynamic detection in the autosandbox) you will probably still get infected. But the rest of the 170m+ avast community will benefit from you getting infected.

This is no different from biological viruses, by the way.

More of the good stuff is coming in 2013 - we have a bunch of cool new things that will be rolling out in the first half of the year. Stay tuned!

---

joe
11/24/2014, 12:26:00 PM

So then why does Avast find Eye-Gen as a 'threat?' ???

RSS feed

More blog topics

Security News (1064)

Tips (297)

Threat Research (287)

Mobile Security (277)

Corporate News (91)

SMB/Business (76)

Tag Cloud

March 2016 August 2016

SMB/Business February 2016

Tips Mobile Security

Security News

Threat Research

January 2016

Corporate News

July 2016

September 2016

October 2016

June 2016

Tweets

Follow @avast_antivirus

Avast Software
@avast_antivirus

All your #iMessage contacts belong to Apple bit.ly/2dpgOgG via @IonutArghire @SecurityWeek

3h

Avast Software
@avast_antivirus

Facebook's Messenger Lite lights up old #Android phones cnet.co/2dKFS5f via @CNET

Facebook's Messen…
The social network's …
cnet.com

6h

Avast Software
@avast_antivirus

Sensitive FDA Systems at risk of #cyberattacks bit.ly/2dpgXAP via @EduardKovacs @SecurityWeek

9h

Avast Software
@avast_antivirus

How hard is it to #hack the average DVR? Sadly, not hard at all bit.ly/2dsNMwH via @dangoodin001

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.