



Beyond Testing: What Really Matters

Andreas Marx
CEO, AV-TEST GmbH

About AV-TEST GmbH

Innovations and Presentations

The AV-TEST Approach – Part 1 Protection, Performance,
Usability and Repair

Microsoft as Baseline

The AV-TEST Approach – Part 2

Summary

Q&A

**Decades of
experience in the
field of virus
research and
analysing
antivirus software**

We are a global acting and independent service provider in the field of IT security and antivirus research.

We have almost 20 years of experience in the field of malware and antivirus software.

We process more than 700 Terabyte testing data, including 120 million clean files and 150 million malware samples.

We feature more than 1 Petabyte storage space, with over 300 client and server systems.

We offer 30 employees and several students a secure, variable and interesting position.

ABOUT AV-TEST GMBH



The AV-TEST Institute in Magdeburg/Germany –
Hightech in historical ambience

WHERE IS MAGDEBURG?

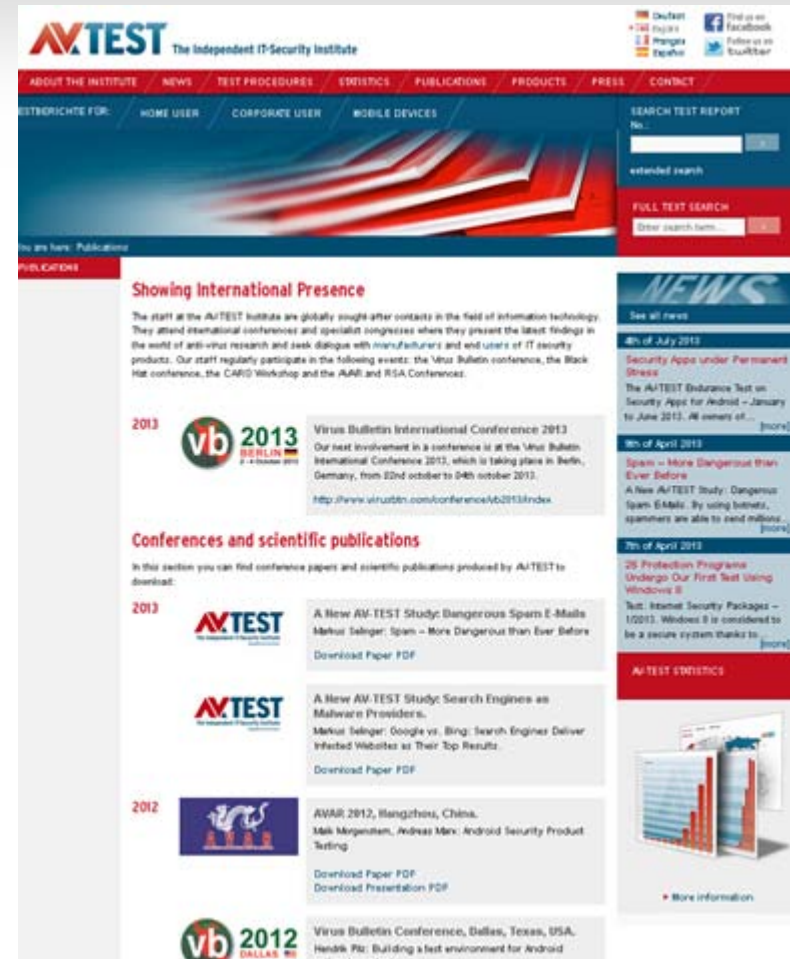


Our region is the
cultural heart of
Germany



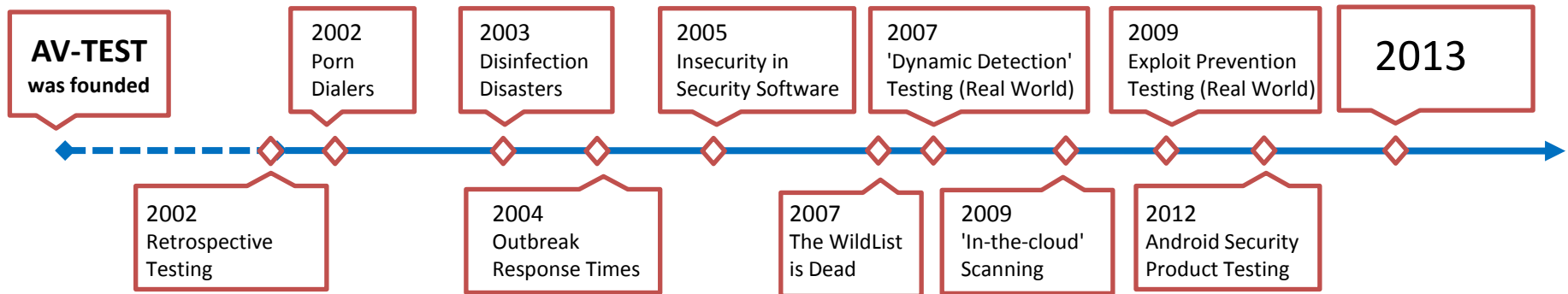
© IMG Investment and Marketing Corporation Saxony-Anhalt

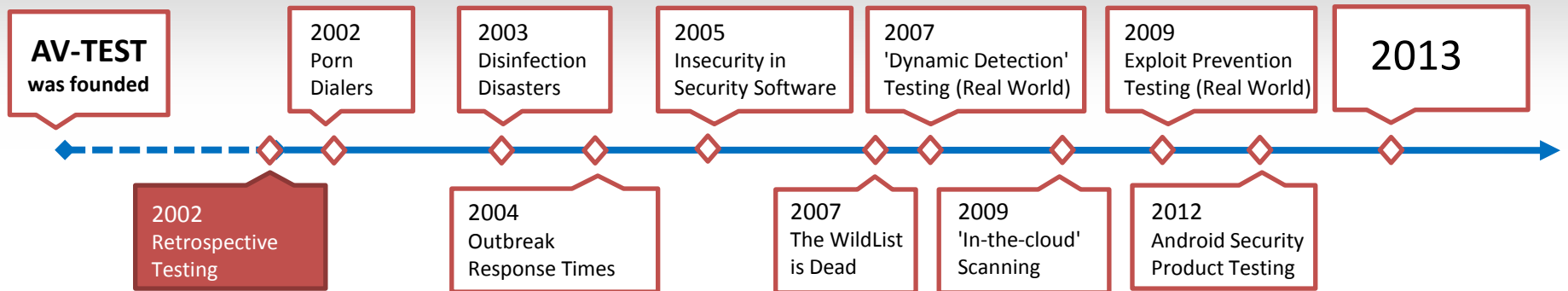
Find our research papers and
conference presentations on
www.av-test.org/en/publications



The screenshot shows the AVTEST website's 'PUBLICATIONS' page. The header includes the AVTEST logo and navigation links: ABOUT THE INSTITUTE, NEWS, TEST PROCEDURES, STATISTICS, PUBLICATIONS, PRODUCTS, PAGES, CONTACT. Below the header, there are search options for 'SEARCH TEST REPORT' and 'FULL TEST SEARCH'. The main content area is titled 'Showing International Presence' and features a list of publications. The first entry is for the 'Virus Bulletin International Conference 2013' in Berlin, Germany, with a 'Download Paper PDF' link. The second entry is 'A New AV-TEST Study: Dangerous Spam E-Mails' by Markus Seliger, with a 'Download Paper PDF' link. The third entry is 'A New AV-TEST Study: Search Engines as Malware Providers' by Markus Seliger, with a 'Download Paper PDF' link. The fourth entry is 'AVAR 2012, Hangzhou, China' by Mal Meningham and Andrea Mei, with 'Download Paper PDF' and 'Download Presentation PDF' links. The fifth entry is for the 'Virus Bulletin Conference, Dallas, Texas, USA' by Hendrik Pfo, with a 'Download Paper PDF' link. On the right side, there is a 'NEWS' section with a 'See all news' link and a 'PUBLICATIONS' section with a 'More information' link.

The story so far

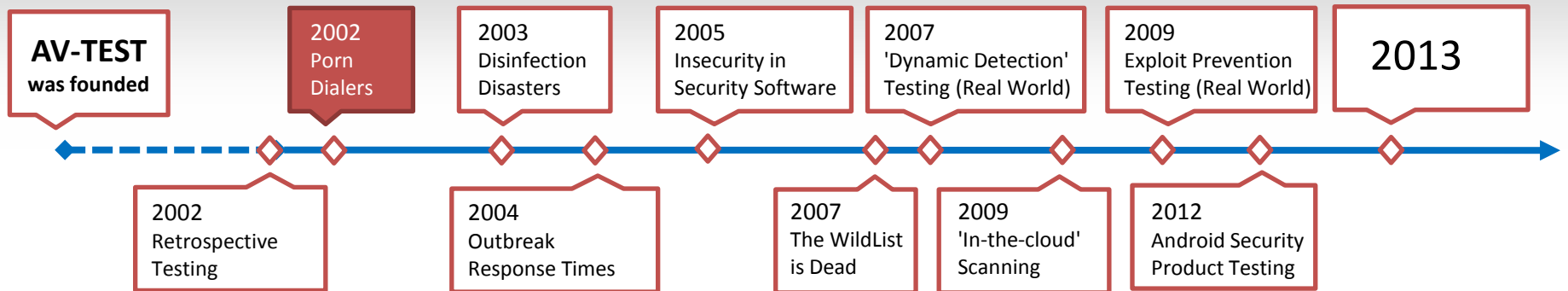




Retrospective Testing - How Good Heuristics Really Work

VB Conference 2002

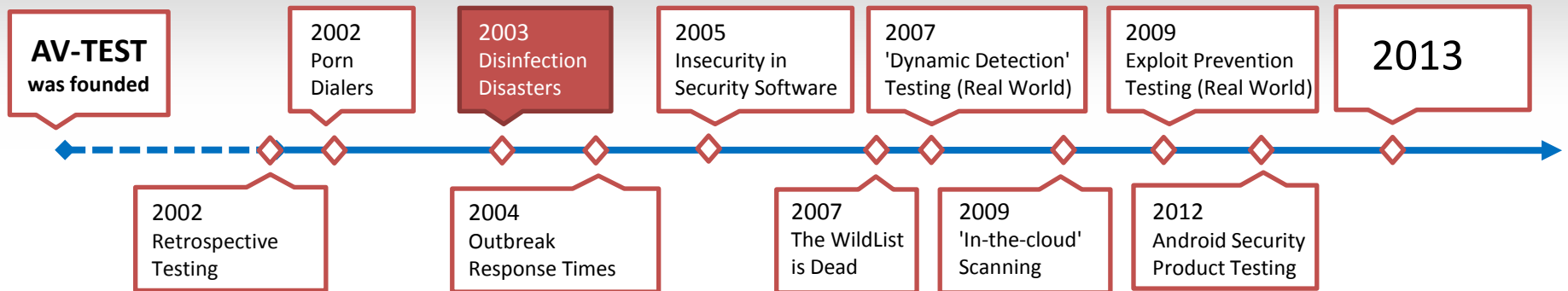
At the time of writing, this was a state-of-the-art single-feature test but such tests are now obsolete, as on-demand tests are outdated and you cannot "freeze" AV updates anymore and cloud access should not be limited, and not single features should be tested anymore.



(Porn) Dialers - Another Class of Malware?

VB Magazine 12/2002

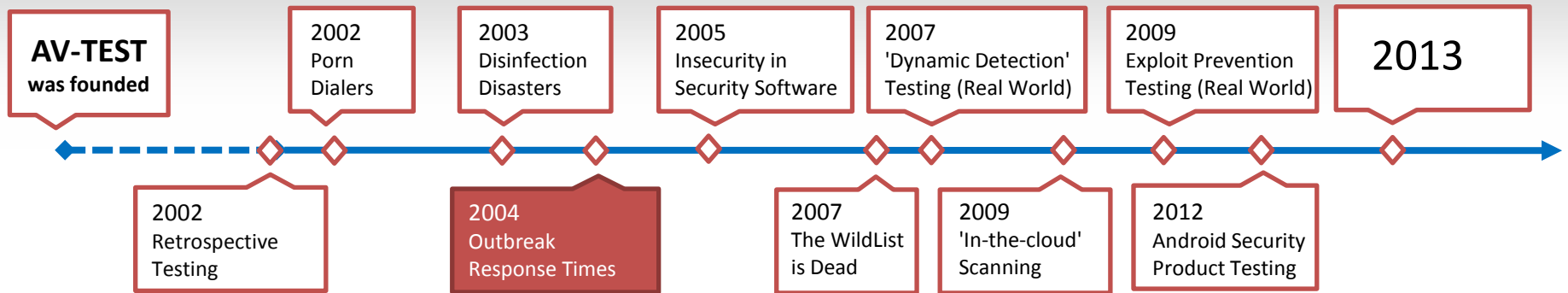
At that time, dialers were a heavy problem to the Windows world, now the problem has shifted to mobile phones (especially in the Android space), calling expensive numbers or sending out text messages



The Sober Effect: Disinfection Disasters

VB Magazine 12/2003

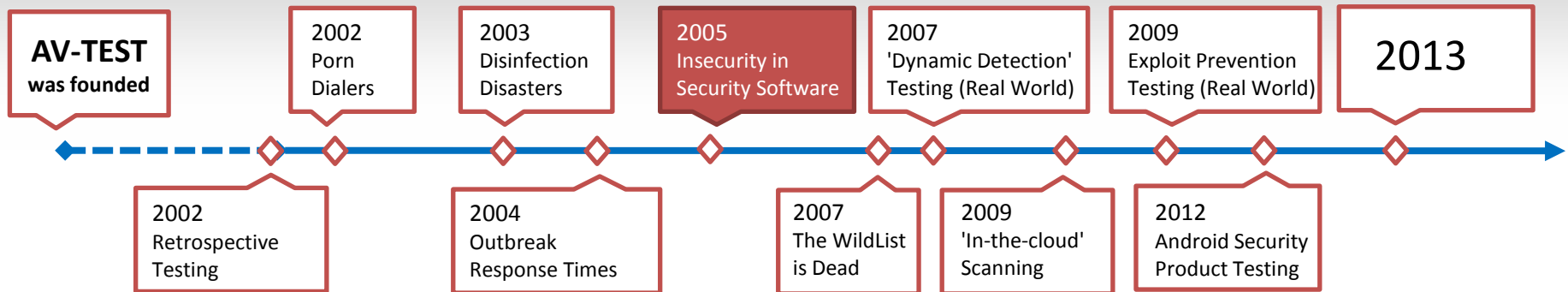
- Products had problems to effectively clean-up infected systems, e.g. due to the self-protection of malware (tasks cannot easily be killed)
- Fact: Repair is still one of the most challenging things these days
- Many more papers and presentations by us followed, still disinfection is often not tested at all or not tested properly



Antivirus Outbreak Response Testing and Impact

VB Conference 2004

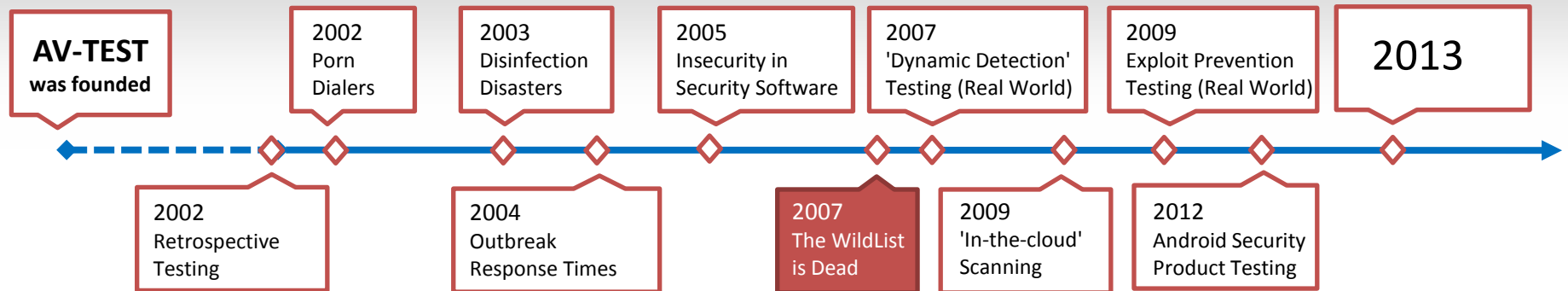
- “How long does it take until signature updates are publicly available in cases of major worm outbreaks?”
- Still a valid question, but replace the word “outbreaks” with “background noise malware”, something around 200,000 unique samples per day
- Ideal protection is when the malware is blocked at the time it arrives at the system (it doesn't matter if this is an hour or just a minute before, as long as the system is not compromised)



Insecurity in Security Software

VB Conference 2005

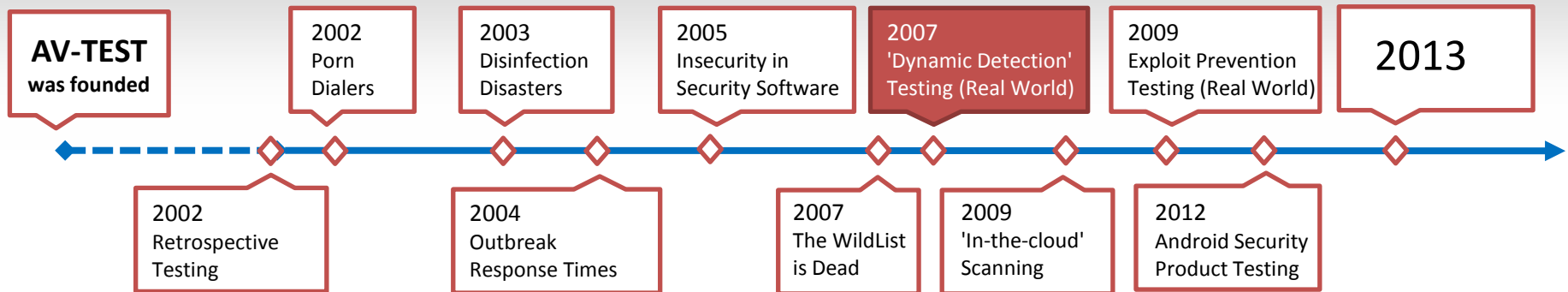
- The paradox: Security software is meant to secure the system, but nowadays it introduces new security holes. Every error could be security relevant when it happens in security software!
- Trustworthy computing development lifecycle:
 - Secure by design, Secure by default, Secure in deployment, Communications



The WildList is Dead, Long Live the WildList!

VB Conference 2007

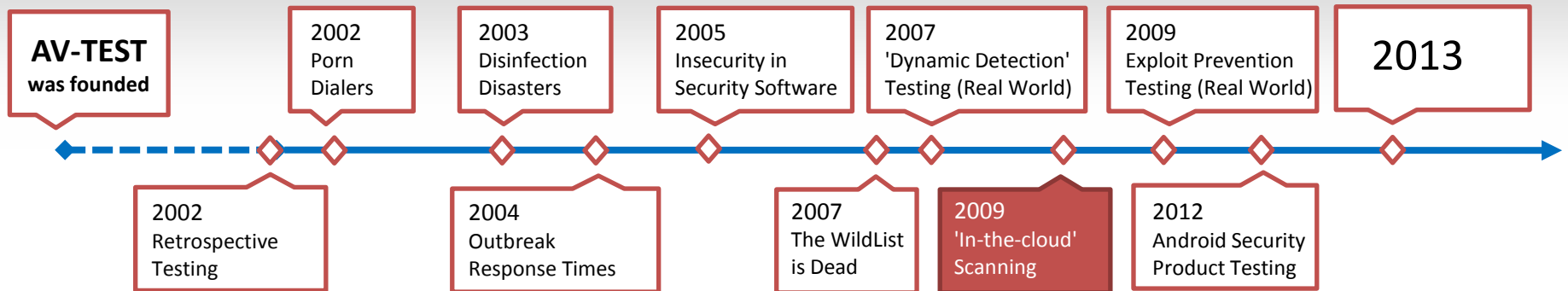
- Problems at this time: The Changing Threat Landscape, Number of Malware Samples, Nobody Wants to Report, Outdated WildList
- Problems today: The Changing Threat Landscape, Number of Malware Samples, Nobody Wants to Report, Outdated WildList
- Quite a lot of suggestions have been made “to make it better”
- Main issue: WildLists tests are easy to pass (you know the test set in advance), they are good for marketing purposes, but doesn't tell you anything about the real capabilities of AV programs



Testing of 'Dynamic Detection'

AVAR Conference 2007

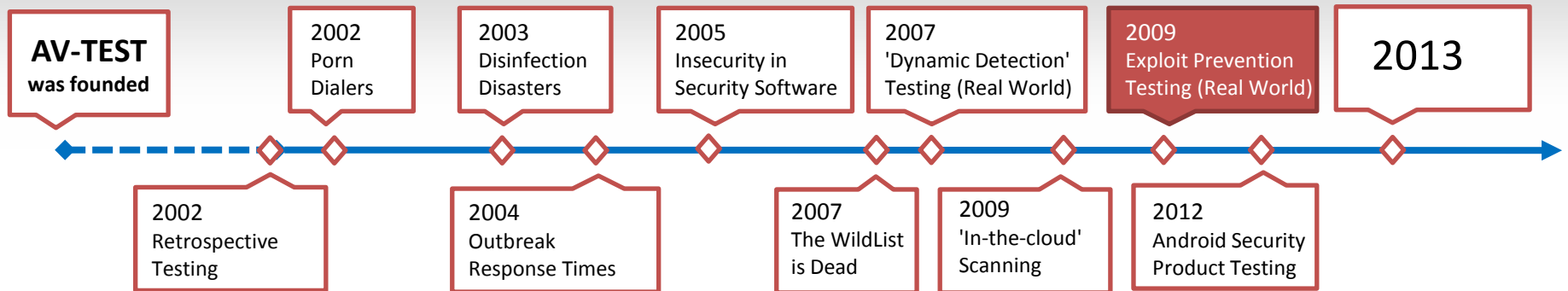
- Historic: Static detection as part of the "traditional" way of AV testing
- Newly introduced: Dynamic detection -- and we demonstrated how to test it
- First full-feature "Real World" test description presented more than 6 years ago (the first "Real World" tests have started earlier in the year 2007)
- "Ideal setup": real (not virtualized) hardware, base system with recent operating system and patch level, default settings of products under test, high volume and many different malware types, use the appropriate introduction vector (e.g. e-mail, web, download, P2P, USB key, network port), "Record the impact of the security software and compare the result to the actions of the malware on the clean base system", check for detection, reporting and blocking
- With some minor changes, most parts of the setup are still valid



Why 'In-the-cloud' Scanning is not a Solution

VB Conference 2009

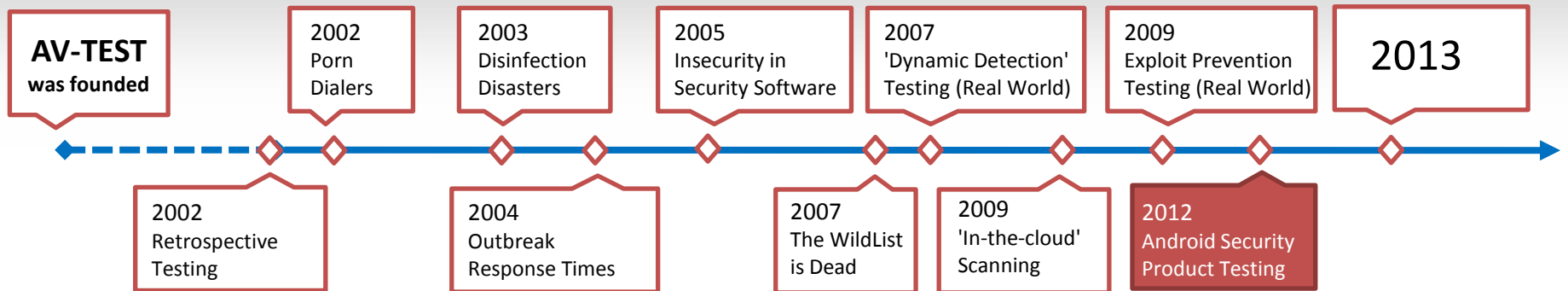
- 'In-the-cloud' scanning is helping the vendors to get their static detections rate up
- With reputation systems and further statistical analysis, those approaches can help even further in detecting malware
- But: 'In-the-cloud' scanning is still only a part of a whole security infrastructure (and not every product can access the cloud, e.g. in critical infrastructures)
- New (much better!) developments these days: reputation services instead of "pure" blacklisting and whitelisting



Testing Exploit-Prevention Mechanisms in Anti-Malware Products

CARO Workshop 2009

- Extension to the "Real World" testing methodology from 2007 to cover drive-by attacks etc.
- Testing needs to reflect these additional protection mechanisms: Whole product evaluation instead of only testing (possibly misleading) on-demand scanning capabilities



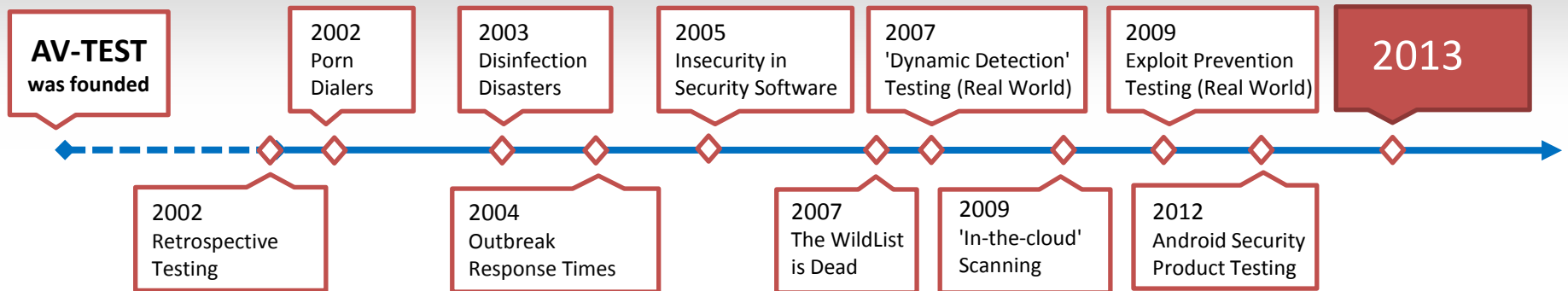
Android Security Product Testing

AVAR Conference 2012

- Wrong focus in past: Are malware detection and all the other technical features really the most important items?
- Problem: Results don't help the user to choose the "right" product, according to his or her needs

What really matters: What happens when I lose my phone?

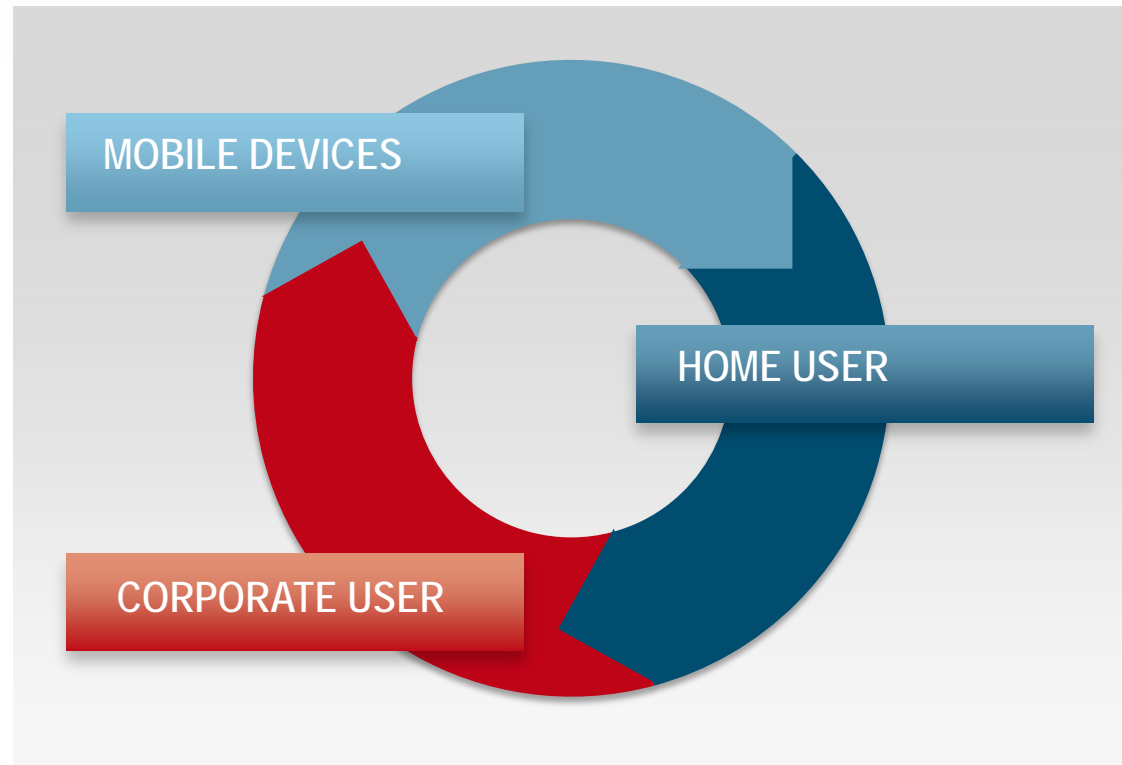
- Can I get it back? → Anti-Theft (Locate Device)
- Is my data safe? → Remote Wipe, Remote Lock, Encryption
- Can I get my data back? → Online Backup
- Is my privacy ensured? → Which apps spy on me and can security software tell me and protect me?
- Is malware or adware a problem for me? → Malware and PUA Detection rates
- I want to protect my child from inappropriate content on the phone. → Parental Control



A wide range of other security-related areas are covered, too, e.g.

- "Spam – More Dangerous than Ever Before"
- "Google vs. Bing: Search Engines Deliver Infected Websites as Their Top Results"

More than 3,000
individual and
comparative tests
per year



PROTECTION

Tested monthly



PERFORMANCE

Tested every second month



USABILITY

Tested every second month



REPAIR

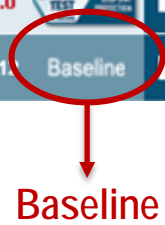
Tested bimonthly during the last years,
now 1 to 2 times a year



THE AV-TEST APPROACH – PART 1

Windows

Testreport	Producer: Product		PROTECTION	PERFORMANCE	USABILITY	Platform	Date
131594	F-Secure: Client Security 10.00					Windows XP	04-2013
131524	Fortinet: FortiClient 5.0					Windows XP	04-2013
131537	Kaspersky: Endpoint Security 10.1					Windows XP	04-2013
131577	McAfee: VirusScan Enterprise with EPO 8.8					Windows XP	04-2013
131581	Sophos: Endpoint Security and Control 10.2					Windows XP	04-2013
131502	Symantec: Endpoint Protection 12.1					Windows XP	04-2013
131516	Trend Micro: Office Scan 10.6					Windows XP	04-2013
131567	Webroot: SecureAnywhere Endpoint Protection 8.0					Windows XP	04-2013
131564	Microsoft: System Center Endpoint Protection 2012					Windows XP	04-2013



All results are presented together and are comparable against the industry average and Microsoft as baseline.



Products are tested having these items in mind:

- Home user products: use default settings only
- Corporate products: test with settings as provided (suggested) by the vendor
- Always most current publicly available version of all products
- Can update themselves at any time and query their 'in-the-cloud' services
- Products have to demonstrate their capabilities using all components and protection layers (nothing is deactivated)
- Vendor review (feedback) phase after the end of testing phase (usually around two weeks)

PROTECTION

Protection against malware infections (such as viruses, worms or Trojan horses)

- Protection against 0-day malware attacks, inclusive of web and e-mail threats (“Real World” Testing)
 - more than 100 samples per test-run and product
 - manual work with automation in place for preparation and tracking of system changes on the file system, registry, processes, threads, network; no replay tests
 - all items are reviewed (with different IP addresses) at (almost) the same time for all products, using the real internet (in a safe way for other internet users)



PROTECTION

- Detection of widespread and prevalent malware discovered in the last 4 weeks (the AV-TEST reference set)
 - Also using a full “Real World” approach: all 10,000 to 15,000 samples are not only scanned, but all misses are executed, too!
 - Only prevalent samples are used (based on telemetry data received from many sources, incl. Microsoft)
 - Products without an on-demand scanner are testable using the same criteria, e.g. whitelisting solutions can be tested in the same test as ‘traditional’ AV products



PERFORMANCE

Average influence of the product on computer speed in daily usage

- Use cases: visiting websites, downloading software, installing and running programs and copying data
 - “What a normal user does”, some items are not covered, e.g. system start-up time as many products are “cheating” here and it's not clear when the system is really “up” and the guard is fully working
 - Low ‘Performance’ impact is important (as this is what users are experiencing every day)



USABILITY

Impact of the security software on the usability of the whole computer

- False warnings or blockages when visiting websites
 - False detections of legitimate software as malware during a system scan
 - False warnings concerning certain actions carried out whilst installing and using legitimate software
 - False blockages of certain actions carried out whilst installing and using legitimate software
-
- This includes the impact of the suite on the 'Usability' of the whole system, all kind of "noise" messages only a human tester will see and details which might be missed if a fully automated testing system is in place
 - All tests are static and dynamic, simulating the whole user experience, using a "Real World" test approach



REPAIR

Cleaning and repair of a malware-infected computer

- Detection of actively running widespread malware (including Rootkits and stealth malware)
- Removal of all active components of widespread malware (including Rootkits and stealth malware)
- Removal of further malicious components and remediation of critical system modifications

In case of an infection, 'Repair' gets important. We test it less frequently now, but with more samples and covering more test criteria, including:

- Malware is activated before the security product is installed (and not only after the AV/ISS software is running)
- All kind of offered rescue options are used, including the product itself, stand-alone cleaning utilities and rescue CDs
- Not all products and OS can be tested equally (e.g. Microsoft Windows 8 comes with Defender out-of-box)



Microsoft Baseline: the 'out-of-box' protection

- Windows 8 includes Windows Defender as full-feature anti-virus program
- For older OS, Microsoft Security Essentials is available free-of-charge from the OS vendor
- Question: Why should I install a different AV when a baseline protection is available for free?
 - Free AV: might want user data, display ads
 - Paid AV: why pay, what extras can I get?
- What can 3rd party tools do better than the baseline? Let's focus on this!



How good is the Microsoft baseline in the Windows world?

- **Protection:** Scores usually in the area of 85-95% in our tests, not very good, "leaves room for improvements"
- **Performance:** Good to very good, many 3rd party products are slower
- **Usability:** Excellent, we can rarely see any false positive or misleading warning messages
- **Repair:** Overall good results, much better than the industry average



WHY MICROSOFT AS BASELINE?

How good is the Microsoft baseline (85-95% protection) compared with other OS' out-of-box protections?

As tested by AV-TEST in 2013:

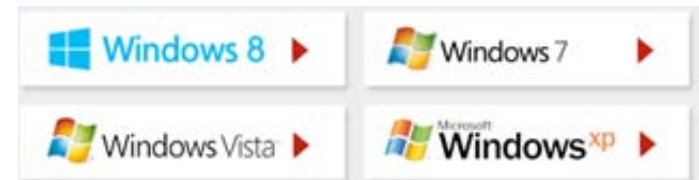
- **Google Android** (App Verification): cloud service, only available for Android 4.2 and up, not enabled by default, hard to test (only a few valid responses until the system will switch to a random mode), but protection against Android malware is less than 50% (more than 900,000 Android malware samples exist)
- **Apple** (OS X Gatekeeper): blocks less than 50% of OS X malware threats (we know about 2,000 different samples for the Apple platform), but the protection can be configured in more secure way, e.g. to deny all apps with no valid developer ID from Apple
- **Linux** (ClamAV?): blocks about 60% of Linux-related malware (about 5,000 samples are known to us), good response times to new threats, acceptable scan speed, extremely high number of false positives for certain file types (e.g. Windows EXE files)



© Think Linux

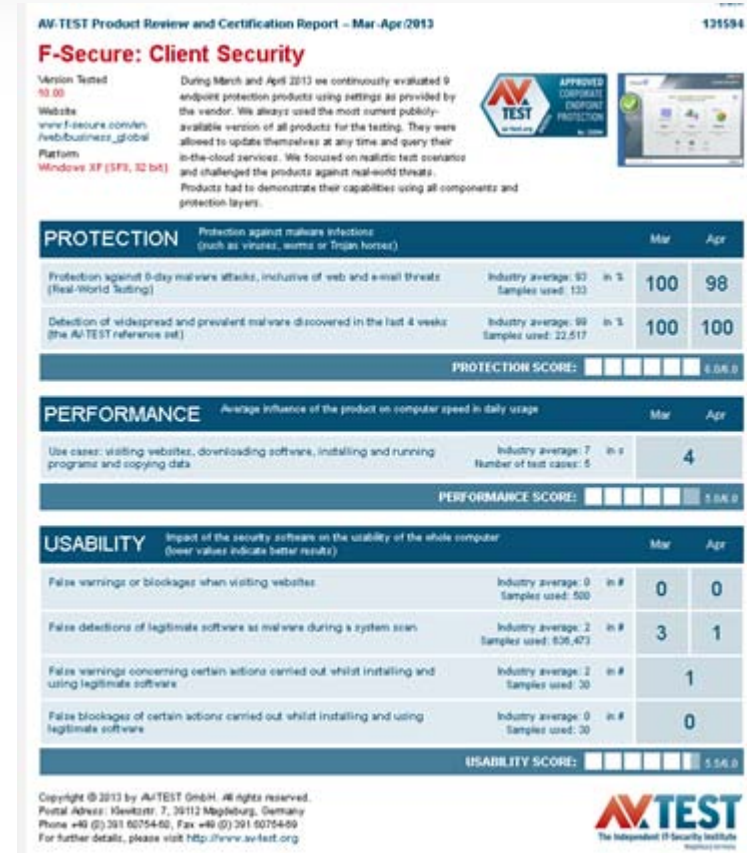
What is AV-TEST doing?

- Test as many products as possible
- Test as many aspects of the products as possible
- Test as many scenarios/samples as possible
- Perform the tests as often as possible
- Use all common OS platforms (32 and 64 bit)
- High quality in methodology and sample selection



That is only half of the story

- Generating all the data is necessary but not enough
- No user could dig through all the raw data
- Interpretation of the data, according to certain real scenarios has to be done
- From the different testing criteria it will be possible to derive answers for users questions
- Different user groups and their different demands can be considered



Ultimately providing three different answers

- AV-TEST certifies products that perform well overall and are a good choice for most of the scenarios
- Using AV-TEST data you can find the best fitting products for certain demands
- Yearly AV-TEST AWARDS for the best-performing products in each tested category



Advantages for Vendors

Quality assurance through reliable and independent tests.

Certification as a proof of high quality and as marketing tool.

Advantages for Home User

Test results as a decision aid.

Find solutions for PC and mobile devices.

Advantages for Corporate User

Decision guidance for the right anti virus software solution.

Costs and time can be saved.

AV products have developed significantly over time (additional layers of protection).

Our testing methodologies have been implemented and adjusted to cover these changes.

Many testers haven't evolved over time, and are still too focused on "old" ideas.

We presented our approach and of course, we're highly interested in all kind of feedback.



Thank you for your kind attention!
Are there any questions?