



Heuristische Analyse – Die Erkennung unbekannter Viren

Antiviren-Software macht viel mehr als nur bekannte Viren zu entdecken; sie scannt proaktiv auch nach unbekanntem. Wie funktionieren solche Scanner eigentlich?

David Harley

Autor und Berater für Computersicherheit
und

Andrew Lee

Forschungsleiter
ESET LLC

Michael Dankert

deutsche Übersetzung

Inhaltsverzeichnis

Über die Autoren	2
<i>David Harley</i>	2
<i>Andrew Lee</i>	2
Einleitung.....	3
Detektive bei der Arbeit.....	4
<i>Viren</i>	4
<i>Würmer</i>	5
<i>Nicht-replizierende Malware</i>	5
Was bedeutet Heuristik nun wirklich?	7
<i>Scannen nach Signaturen</i>	8
<i>Das Gegenteil von Heuristik</i>	10
<i>Generische Antivirus-Methoden</i>	10
<i>Ich bin absolut sicher</i>	12
<i>Empfindlichkeit und Fehldiagnosen</i>	13
<i>Testprobleme</i>	15
Das Fazit: ein heuristisches Paradoxon.....	17
Literaturnachweise	20
Glossar.....	21

Über die Autoren

David Harley

David Harley forscht und schreibt bereits seit dem Ende der 80er Jahre über Schadsoftware und andere Sicherheitsprobleme. Seit 2001 arbeitete er im britischen staatlichen Gesundheitsdienst als 'National Infrastructure Security Manager', wo er sich auf die Abwehr von Bedrohungen und aller Formen von EMail-Missbrauch spezialisiert hatte sowie das 'Threat Assessment Centre' leitete. Seit April 2006 arbeitet er als unabhängiger Autor und Berater.

Er war Mitverfasser von 'Viruses Revealed' und hat nicht nur zahlreiche Kapitel zu vielen anderen Büchern von namhaften Verlagen über Computersicherheit beigesteuert, sondern auch eine Fülle von Artikeln und Beiträgen zu Konferenzen veröffentlicht.

SMALL BLUE-GREEN WORLD

8 Clay Hill House, Wey Hill, Haslemere, SURREY GU27 1DA

Telephone: +44 7813 346129

<http://smallblue-greenworld.co.uk>

Andrew Lee

Andrew Lee, CISSP (Certified Information Systems Security Professional), ist der Leiter der Forschungsabteilung der Firma ESET LLC. Er ist eines der Gründungsmitglieder des Anti-Virus Information Exchange Network (AVIEN) sowie seiner Schwesternvereinigung AVIEWS (AVIEN Information & Early Warning System), ist Mitglied von AVAR (Association of Anti Virus Asia Researchers) und Berichterstatter für die WildList-Organisation. Bis vor kurzem war er an vorderster Front bei der Abwehr von Schadsoftware als hochrangiger Sicherheitsadministrator in einer großen Regierungsbehörde tätig. Andrew ist Autor von zahllosen Artikeln zu Sicherheitsproblemen und tritt häufig auf Konferenzen und anderen Veranstaltungen, wie AVAR, Virus Bulletin und EICAR auf.

ESET, LLC

610 West Ash Street, Suite 1900, San Diego, California 92101, U.S.A.

Telephone: +1.619.876.5400

Fax: +1.619.876.5845

<http://www.eset.com>

Einleitung

„Nicht das, was du nicht kennst, bringt dich um, sondern das, was du zu kennen glaubst, das aber nicht so ist.“

Einige der hartnäckigsten Mythen rund um Computer beziehen sich auf Viren und Antivirus(AV)-Technologie. Die weit verbreitete Annahme, dass AV-Software nur bestimmte, bekannte Viren entdecken kann, hält sich schon seit den frühen Tagen der Antiviren-Forschung. Es war schon damals nicht vollkommen richtig; denn einige der ersten AV-Programme waren nicht dafür gedacht, spezifische Viren zu entdecken, sondern virusartiges Verhalten oder verdächtige Änderungen an Dateien zu erkennen bzw. zu verhindern. Heutzutage ist es ganz bestimmt nicht zutreffend.

Kommerzielle AV-Lösungen unterstützen das Scannen nach Signaturen durch eine Vielzahl eher generischer Ansätze, die oft unter dem Titel "Heuristische Analyse" zusammengefasst werden. Darüber hinaus sind moderne AV-Produkte in der Lage, eine große Auswahl an Malware (Malware ist eine Zusammenfassung der Worte "malicious - böse" und "software - Programm") zu erkennen, und nicht nur Viren. Dies kann noch mit anderen Sicherheitstechniken, wie der Erkennung von Spam und Phishing-Nachrichten kombiniert werden.

Ziel dieses Artikels ist es, das beim Verständnis von AV-Technologien herrschende Durcheinander etwas zu verringern und klarzustellen, was man realistischere vom Antivirenschutz und speziell von der heuristischen Analyse erwarten kann.

Die Spezifika des heuristischen Scannens werden später im Detail erörtert. Für den Moment beschreiben wir die heuristische Analyse einfach als eine Methode zur Abschätzung der Wahrscheinlichkeit, dass ein Programm, das nicht als eine bekannte Malware identifiziert wurde, trotzdem infektiös oder böse ist.

Detektive bei der Arbeit

Was erkennt ein Antiviren-Programm? Zufälligerweise eine ganze Menge, einschließlich einiger Dinge, die technisch betrachtet gar keine Viren sind. Das meiste, was Virus genannt wird, könnte man besser als Malware bezeichnen. Die Ironie dabei ist, dass viele spezialisierte Erkennungswerkzeuge (z.B. zum Aufspüren von Spyware oder Trojanern) als notwendig vermarktet werden, da AV nur Viren entdecken kann.

Das meiste, was Virus genannt wird, könnte man besser als Malware bezeichnen.

Tatsächlich erkennt kommerzielle AV-Software einen viel größeren Umfang an Malware als die meisten dieser spezialisierten Dienste. Ein spezifisches Programm kann mehr Bedrohungen auf seinem eigenen Spezialgebiet erkennen, aber das hängt nicht nur von den Fähigkeiten des Programms ab, charakteristische Bedrohungen und Bedrohungstypen zu finden, sondern auch von anderen Faktoren, wie:

- den generischen Erkennungsmöglichkeiten des Programms
- den benutzten Kriterien zur Unterscheidung zwischen Varianten der Malware
- den Mechanismen zum Verteilen von Virusproben zwischen den Herstellern (AV-Hersteller haben dafür besonders effektive und etablierte Wege, verglichen mit Anbietern auf anderen Gebieten der Malware-Erkennung.)

Die folgenden Abschnitte berücksichtigen nur drei Hauptformen von Malware. Eine komplette Systematik aller Malwareformen würde den Rahmen dieses Artikels sprengen.

Viren

Es ist sicher vernünftig, zu erwarten, dass AV-Software Viren erkennen kann und der über die Jahre gewachsene Erfolg der AV-Software bei ihrer Erkennung ist teilweise daran Schuld, dass ihre Fähigkeiten, andere Formen von Malware aufzuspüren, unterschätzt werden.

Während es viele Definitionen dafür gibt, was man unter einem Virus versteht, lautet eine von den meisten Forschern auf dem Gebiet der Malware akzeptierte Erklärung: "ein Computerprogramm, das andere Computerprogramme infizieren kann, indem es sie dadurch modifiziert, dass es eine (möglicherweise weiterentwickelte) Kopie von sich selbst einfügt" [1, 2].

Diese Definition deckt viele Virusarten ab, einschließlich:

- Infizierer von Boot- und/oder Partitionssektoren
- Datei-Infizierer (parasitische Viren)
- mehrteilige Viren
- Makro- und Skriptviren

Während einige dieser Virusformen heutzutage kaum noch in Erscheinung treten (z.B. Bootsektor- und Partitionssekturviren), erkennen AV-Programme üblicherweise alle bekannten Viren auf den Plattformen, auf denen sie vorkommen (und manchmal sogar auf anderen Plattformen). Im Allgemeinen sind sie auch ziemlich gut darin, neue und unbekanntere "echte" Viren heuristisch aufzuspüren.

Der über die Jahre gewachsene Erfolg der AV-Software ist teilweise daran Schuld, dass ihre Fähigkeiten, andere Formen von Malware aufzuspüren, unterschätzt werden.

Würmer

Die AV-Industrie hat sich bisher noch nicht darauf einigen können, ob Würmer wie Cohen feststellte, ein "Spezialfall von Viren" [1] sind, aber wie dem auch sei, AV-Software erkennt sie normalerweise trotzdem.

Es gibt mindestens genau so viele Definitionen für Würmer wie für Viren, aber die meistens AV-Forscher charakterisieren einen Wurm als ein Programm, das sich nicht-parasitisch repliziert, d.h. ohne sich an eine Wirtsdatei anzuhängen. Programme für Massen-Mails könnte man als einen speziellen Wurmtyp bezeichnen. Die meisten AV-Firmen bezeichnen diese Art von Email-basierter Malware als Wurm, aber einige Mailer und Massenmailer besitzen die Eigenschaften eines "reinen" Virus (Melissa war zum Beispiel eigentlich ein reiner Virus, ein Makrovirus, der sich wie ein Wurm verbreitet hat, während W32/Magistr ein Datei-Infizierer war).

Auch hier haben die Hersteller die Erkennung neuer Varianten ziemlich gut unter Kontrolle. Neue Massenmailer zum Beispiel werden von den Sicherheitsdienstleistern fast genauso schnell gekennzeichnet, wie sie erscheinen.

Nicht-replizierende Malware

Aus den oben genannten Definitionen folgt, dass ein böses Programm, das sich nicht repliziert, weder ein Virus noch ein Wurm sein kann. Das bedeutet aber nicht, dass es von AV-Software nicht gefunden wird oder dass es keinen Schaden anrichten kann.

Denken Sie daran, dass die Hersteller auch zu der Zeit, als sie noch gegen die Erkennung nicht-replizierender Objekte protestierten, da es sich ja nicht um Viren handelte, trotzdem einige dieser Objekte (manche davon nicht mal ausführbare Programme, geschweige denn böse) erkannten und als schädlich markierten [3]. Einige Beispiele:

- Intendeds (Viren, die sich auf Grund von Fehlern nicht replizieren können) und defekte Programme
- Dateien mit sinnlosem Inhalt
- Nicht infektiöse Programme aus dem Virenumfeld, wie Germs, Dropper und Virusgeneratoren
- Rechtmäßige Testprogramme wie die EICAR Testdatei [4]

Viele nicht-replizierende Objekte kursieren schon seit Jahren in schlecht gepflegten Virensammlungen.

Viele nicht-replizierende Objekte kursieren schon seit Jahren in schlecht gepflegten Virensammlungen, die von einigen Gutachtern zum Testen von AV-Software benutzt werden. Die meisten Hersteller gaben ihren Protest schon vor längerer Zeit auf und ergänzten ihre Datenbanken um Definitionen (Signaturen) für diese Objekte, in der Hoffnung, damit eine Benachteiligung für das Nichterkennen zu vermeiden. Unglücklicherweise hat die fortschreitende Verfeinerung der heuristischen Scanner kaum mit der Fähigkeit der AV-Tester Schritt halten können, neue und nicht immer angemessene Testmethoden zu entwickeln. Weiter unten werden wir uns kurz mit technisch akzeptablen Methoden zum Testen der heuristischen Fähigkeiten eines Produkts beschäftigen.

Die bekannteste nicht-replizierende Malware ist das Trojanische Pferd (kurz Trojaner genannt). Ein Trojaner ist "ein Programm, das behauptet, eine wünschenswerte oder notwendige Funktion zu erfüllen, was manchmal sogar stimmt, aber darüber hinaus eine bestimmte Funktion oder Funktionen ausführt, die derjenige, der das Programm gestartet hat, weder erwarten würde noch haben möchte." [5] Diese Definition umfaßt einen ganzen Bereich spezialisierter Malware, insbesondere:

- Dropper
- Keylogger
- Zerstörerische Trojaner
- Downloader
- Spyware
- Adware
- Rootkits und stealthkits
- Spaßprogramme (einige)
- Zombies (Bots, Trojaner für Fernzugriffe, Agenten für DDoS-Angriffe usw.)

Replikative Schadprogramme wie beispielsweise Viren können mitunter auch als Trojaner bezeichnet werden (oder als zum Trojaner geworden bzw. gemacht, was bedeutet, dass ein vorher legitimes Programm unterwandert, verändert oder ersetzt wurde, um damit irgendwie Schaden anzurichten), obwohl die meisten Leute die Verwendung des Begriffs in diesem Sinne wahrscheinlich eher verwirrend als hilfreich finden. Eine Erkennung aller Versionen nicht-replikativer Malware ist sogar noch schlechter zu erreichen als die aller Virusformen, da ein viel größerer Funktionsumfang getestet werden muss als nur die reine Fähigkeit, sich zu vermehren.

Die meisten Auseinandersetzungen darüber, was ein Trojaner (oder böseartig) ist oder nicht, stützen sich nicht auf die Funktion sondern vielmehr auf die Absichten. Ein Keylogger zum Beispiel ist kein Trojaner, wenn er rechtmäßig oder mit Einverständnis des Nutzers installiert wurde, und doch ist die Funktionsweise identisch. Dies führt zu Problemen bei der Erkennung, da Computer noch viel weniger als Menschen in der Lage sind, Absichten zu ermitteln.

Spyware und Adware werden neuerdings - vielleicht wegen des verstärkten Medieninteresses und der speziell auf sie zugeschnittenen Erkennungsprogramme - als Malware in ihre eigenen Unterklassen eingeteilt. Diese Unterscheidung ist hier allerdings meistens unnötig, obgleich argumentiert werden könnte (und oft wird), dass speziell Adware nicht immer auch Malware ist. Das gleiche Argument könnte man jedoch für fast alle Punkte dieser Liste geltend machen, denn ein Programm wird nicht durch das, was es tut, böseartig, sondern durch die Diskrepanz zwischen den unredlichen Absichten des Programmierers und den Erwartungen des Anwenders.

Was bedeutet Heuristik nun wirklich?

Heuristik bezieht sich auf den Vorgang beziehungsweise Prozess des Findens oder Entdeckens (“heurisko“ griech.: ich finde). Das Oxford-Wörterbuch definiert heuristisch als “eine Person zu befähigen, etwas selbständig zu entdecken oder zu lernen“ oder (im Zusammenhang mit Computern) “zu einer Lösung durch Versuch und Irrtum oder die Anwendung nur unscharf definierter Regeln zu kommen“ [6]. Das Merriam-Webster-Wörterbuch definiert es als “ein Hilfsmittel zum Lernen, Entdecken oder Problemlösen durch experimentelle Methoden, hier speziell Versuch und Irrtum“ oder (wiederum im Computer-Kontext) “bezogen auf Forschungsmethoden zur Problemlösung, die zur Verbesserung der Leistung selbstlernende Verfahren (in Auswertung von Feedback) benutzen“ [7].

Heuristische Programmierung wird im Allgemeinen als eine Anwendung von künstlicher Intelligenz und als Werkzeug zur Problemlösung angesehen. Heuristische Programmierung, wie sie in Expertensystemen eingesetzt wird, baut auf Erfahrungswerten auf und die von einem solchen System erzeugten Antworten werden besser, indem es aus weiteren Erfahrungen “lernt“ und seine Wissensbasis erweitert.

Bei der Anwendung zur Abwehr von Schädlingen (auch von Spam und ähnlichen Ärgernissen) hat die heuristische Analyse, obwohl eng verwandt mit Versuch und Irrtum oder Lernen aus Erfahrung, eine etwas eingeschränktere Bedeutung. Die heuristische Analyse benutzt hier einen regelbasierten Ansatz zum Erkennen einer potentiell gefährlichen Datei (oder Nachricht im Fall der Spam-Analyse). Während die Analyseeinheit ihre Regeln abarbeitet und die Nachricht mit Merkmalen möglicher Schadsoftware vergleicht, vergibt sie Punktwerte für jeden Treffer. Erreichen oder überschreiten die Punkte einen Schwellwert [8], wird die Datei als verdächtig (oder potentiell gefährlich bzw. als Spam) markiert und entsprechend weiter verarbeitet.

In gewissem Sinne versucht heuristische Anti-Malware den Prozess der menschlichen Analyse auf ein Objekt anzuwenden. Genauso wie ein Schädlingsanalytiker aus Fleisch und Blut versuchen würde, Ablauf und Auswirkungen eines vorgegebenen Programms zu bestimmen, führt die heuristische Analyse den gleichen intelligenten Vorgang der Entscheidungsfindung aus und handelt damit wie ein virtueller Malware-Forscher. Während der menschliche Malware-Analytiker mehr von und über neu entstehende Bedrohungen lernt, kann er oder sie dieses Wissen durch Programmierung in den heuristischen Analysator einbringen und damit zukünftige Erkennungsraten verbessern.

Die heuristische Programmierung spielt bei der Funktion von AV-Lösungen eine Doppelrolle: Geschwindigkeit und Erkennungsrate. Tatsächlich wird der Begriff Heuristik auf anderen Wissensgebieten [9] in einem sehr ähnlichen Sinn angewandt, nämlich zum Zweck der Leistungssteigerung (speziell des Durchsatzes), indem nicht das genaueste Ergebnis gesucht wird, sondern nur ein Ergebnis, das “gut genug“ ist. Mit der ständigen Erhöhung der Anzahl bekannter Viren ist auch das Bedürfnis nach Steigerung der Erkennungsgeschwindigkeit gewachsen. Anderenfalls würde die weiter ansteigende Zeit zum Scannen nach einer immer größeren Zahl von Schädlingen das System praktisch unbrauchbar machen.

Trotz der stark verbesserten Leistungsfähigkeit mancher aktueller Heuristik-Einheiten besteht die Gefahr, dass die Belastung durch heuristisches (und selbst nicht heuristisches) Scannen als schwerwiegender eingeschätzt wird als die Vorteile der verbesserten Erkennung. So gibt es die verbreitete Vorstellung, dass heuristische Scanner generell langsamer als statische sind, was aber ab einem gewissen Entwicklungsstand nicht mehr gilt.

Selbst die nur mit einfacher Mustererkennung arbeitenden heuristischen Scanner aus den Anfangstagen profitierten von Optimierungsmethoden, die nur diejenigen Teile eines Objekts durchsuchten, in denen man das

Die heuristische Analyse benutzt einen regelbasierten Ansatz zum Erkennen einer potentiell gefährlichen Datei (oder Nachricht im Fall der Spam-Analyse).

Auftreten eines bestimmten Virus erwarten konnte. (Ein simples Beispiel - es hat keinen Zweck, eine komplette Datei nach einer Virensignatur zu durchsuchen, wenn der Virus seine entscheidenden Teile immer am Anfang oder am Ende einer infizierten Datei speichert.) Das reduziert den Aufwand für das Scannen und verringert das Risiko falsch positiver Treffer.

Die unzutreffende Erkennung einer Virensignatur an einer Stelle, an der das Virus unter normalen Umständen niemals auftreten würde, ist nicht nur eine Begleiterscheinung einer mangelhaften Erkennungsmethodik, sondern auch ein Merkmal unzureichend gestalteter Erkennungstests. So haben einige Prüfer beispielsweise versucht, die Fähigkeiten eines AV-Programms zu testen, indem sie Viruscode wahllos in eine Datei oder ein anderes infizierbares Objekt eingefügt haben. Ebenso kann ein bestimmtes Objekt, etwa eine Datei oder ein Bootsektor, selektiv nur nach den Malware-Typen gescannt werden, die man realistischerweise in dieser Art Objekt erwarten würde, ein Vorgang, der manchmal als "Filterung" bezeichnet wird. Schließlich gibt es keinen Grund, in einem Bootsektor nach einem Makrovirus zu suchen.

Mit der ständigen Erhöhung der Anzahl bekannter Viren ist auch das Bedürfnis nach Steigerung der Erkennungsgeschwindigkeit gewachsen.

Allerdings ist die korrekte Erkennung des Dateityps kein konkreter Beweis dafür, dass die Datei nicht infiziert ist. Beispielsweise waren Microsoft-Word-Dokumente, die eingebettete Schadprogramme enthielten, lange Zeit ein bevorzugtes Angriffsziel für Informationsdiebstahl und Industriespionage. In gleicher Weise suchen die Autoren von Malware ständig nach Angriffsmöglichkeiten auf Objekte, die normalerweise nicht in der Lage sind, Code auszuführen, bei denen man aber zum Beispiel durch Ändern der Ablaufumgebung die Ausführung ermöglichen kann. Der Virus W32/Perrun hingte sich zum Beispiel an .JPG- und .TXT-Dateien an, konnte aber nicht ausgeführt werden, sofern nicht bestimmte Änderungen an der Systemkonfiguration vorgenommen wurden, die den Perrun-Code extrahierten und zur Ausführung brachten.

Scannen nach Signaturen

Scannen nach Signaturen ist durch ziemlich geradlinige Algorithmen zur Mustererkennung gekennzeichnet, die nach einer für jeden Virus oder seine Varianten in der Virusdatenbank des Scanners charakteristischen Bytefolge (einer Zeichenkette) suchen, deren Wahrscheinlichkeit für das zufällige Auftreten in einer nicht infizierten Datei aber gering ist. Einige AV-Forscher haben versucht, die Verwendung des Begriffs "Signatur" zu Gunsten von "Suchstring" oder "Scanstring" aufzugeben [2], aber das erscheint zwecklos, wenn selbst AV-Firmen diesen Ausdruck routinemäßig verwenden.

Tatsächlich können viele Viren nicht durch Suche nach einer statischen Zeichenkette identifiziert werden.

Ein Einwand gegen diesen Ausdruck ist, dass er eine überholte Vorstellung von der Arbeitsweise der Scanner weiterführt, obwohl man das Gleiche auch von den Alternativen behaupten könnte.

Die tatsächlichen Schwierigkeiten beim Gebrauch des Begriffs "Signatur-Scannen" sind, dass er:

den Mythos aufrechterhält, dass dies die einzige Erkennungsmethode der AV-Scanner ist. Tatsächlich können viele Viren nicht einfach durch Suche nach einer statischen Zeichenkette identifiziert werden.

suggeriert, dass es in jedem Virus eine einzige Bytefolge gibt, die von allen Scannern zur Erkennung verwendet wird.

Einige Quellen [10] haben die Angelegenheit noch weiter verwirrt, da sie den Eindruck erweckt haben, Scanner würden nur nach einfachen Zeichenketten suchen statt nach Bytefolgen. Eine derartige Praktik ist generell unzuverlässig, bei vielen Typen von Malware wirkungslos und aus Sicht der Programmierung unwirtschaftlich. Außerdem ist sie durch Virenschreiber - eigentlich durch jeden, der eine Datei editieren kann - leicht auszunutzen und durch die Möglichkeit, zahlreiche falsch positive Treffer hervorzurufen, gefährlich.

Platzhalter und UNIX-typische reguläre Ausdrücke erlauben eine höhere Flexibilität bei der Zeichenkettenuche. Anstatt nach einer statischen Zeichenkette (oder einer festen Bytefolge) zu suchen, erkennt der Scanner eine zu einem Virus gehörende Zeichenfolge selbst dann, wenn andere Bytes oder Bytefolgen (Störbytes) zwischen die Elemente der Zeichenfolge eingeschoben werden. Ein simples Beispiel für ein Störbyte ist das Einfügen eines NOP(No Operation)-Befehls, der keine weitere Funktion erfüllt als Prozessorzeit zu verbrauchen, ohne eine tatsächliche Aktion auszuführen.

Diese Verbesserungen gegenüber dem einfachen Scannen nach Zeichenketten erlauben die Erkennung einiger verschlüsselter und polymorphischer Viren [8]. Dennoch ist das Scannen nach Strings selbst mit dieser Erweiterung nicht besonders effizient, wenn es darum geht, nach einer Vielzahl von Viren zu suchen und das Aufkommen komplexer polymorphischer Viren hat sogar einige Scanner zur Aufgabe gezwungen, die nicht in der Lage waren, auf fortschrittlichere Erkennungsmethoden umzusteigen [8, 11].

Das algorithmische, virus-spezifische Scannen der aktuellen AV-Technologie beruht oft auf interpretiertem Code innerhalb einer virtuellen Maschine. Virtualisierung und Emulation können zum Beispiel dazu genutzt werden, zufällige oder vorsätzliche Verschleierungstechniken wie Packen, Komprimierung oder Verschlüsselung zu entfernen. Sobald die Datei unkomprimiert oder unverschlüsselt vorliegt, kann sie von einem AV-Scanprozess algorithmisch - oder heuristisch - analysiert werden.

Virtuelle Maschinen spielen ebenfalls eine Hauptrolle bei der Implementierung der heuristischen Analyse und können dabei sehr erfolgreich sein, trotz der vielen Probleme, die mit der Emulation einer so komplexen Umgebung verbunden sind, wie sie ein modernes Window™-System darstellt [12]. (Man muss dabei jedoch wissen, dass eine Emulation nicht perfekt ist und die Latenzzeit (gestiegene Verarbeitungszeit) erheblich sein kann und in Abhängigkeit von der einzelnen getesteten Datei variiert.)

Das Aufkommen komplexer polymorphischer Viren hat sogar einige Scanner zur Aufgabe gezwungen, die nicht in der Lage waren, auf fortschrittlichere Erkennungsmethoden umzusteigen.

Das Gegenteil von Heuristik

Das Gerücht, dass kommerzielle AV-Lösungen nur bekannte Exemplare, Varianten und Untervarianten von bekannter Malware erkennen können, ist vielleicht nicht mehr ganz so verbreitet wie früher. Es wurde jedoch teilweise durch das unbedeutendere Gerücht ersetzt, dass viren-spezifische und heuristische Scanner zwei völlig verschiedene Typen darstellen. Tatsächlich wird die heuristische Analyse, wie wir sie heute kennen, schon seit mehr als einem Jahrzehnt benutzt, aber heuristische Methoden zur Optimierung der Virenabwehr werden schon viel länger in Scannern nach "bekannten Viren" verwendet. Sie hatten ebenfalls ihren Platz in ähnlichen Gegenmaßnahmen wie Programmen zum Überwachen und Blockieren schädlichen Verhaltens sowie Integritätsprüfern.

In gewisser Weise ist das Gegenteil der heuristischen Analyse bei AV-Produkten nicht das Scannen nach Signaturen, sondern das algorithmische Scannen, wovon das Scannen nach Signaturen einen Spezialfall darstellt. Algorithmisches Scannen beruht, wie andere Formen von algorithmischer Programmierung, auf mathematisch beweisbaren Prozeduren [13]. Was man in der Antiviren-Industrie als algorithmisches Scannen bezeichnet, basiert normalerweise auf einem Algorithmus (im Gegensatz zum einfachen Scannen nach einer statischen Zeichenkette oder festen Bytefolge), der speziell auf den Virus zugeschnitten ist, den er erkennen soll.

In der Realität wird die heuristische Analyse, wie sie oben beschrieben wurde, natürlich ebenfalls als algorithmisch im weiteren Sinne angesehen. Allerdings wird der Begriff "algorithmisch" im spezialisierten, viren-spezifischen (und deshalb etwas irreführenden) Sinne innerhalb der Industrie schon in zu weitem Umfang gebraucht [12], als dass man ihn ignorieren könnte. Heuristik ist normalerweise dadurch gekennzeichnet, dass sie mit einem speziellen Bewertungsalgorithmus die Wahrscheinlichkeit bestimmt, dass ein gescanntes Objekt bösartig ist, anstatt ein bestimmtes Schadprogramm zweifelsfrei zu identifizieren.

In gewisser Weise ist das Gegenteil der heuristischen Analyse bei AV-Produkten nicht das Scannen nach Signaturen, sondern das algorithmische Scannen, wovon das Scannen nach Signaturen einen Spezialfall darstellt.

Generische Antivirus-Methoden

Die heuristische Analyse wird oft als ein generischer und nicht als ein viren-spezifischer Vorgang zur Viruserkennung angesehen. Dabei wird nicht immer berücksichtigt, dass auch die Umkehrung stimmt: generische Lösungen benutzen heuristische Regelsätze als Teil des Diagnosevorgangs.

Zum Beispiel:

- Filter auf Mail Gateways benutzen Regeln, um festzulegen, welche Dateitypen und -namen als Anhänge erlaubt sind. Solche Filter sind sehr gut zur Abwehr offensichtlicher Bedrohungen geeignet, wie beispielsweise Dateien mit Erweiterungen wie .LNK, oder .JPG und .EXE, können aber ziemlich unflexibel sein, da sie ganze Klassen von ausführbaren Dateien zurückweisen.¹ Andere Filter verwenden fortgeschrittenere Methoden, indem sie zum Beispiel prüfen, ob die Header der gescannten Dateien mit der Dateierweiterung übereinstimmen. Dies kann das Risiko falsch positiver (ebenso wie falsch negativer)

¹ Wieso sind das offensichtliche Bedrohungen? Im ersten Fall, weil die Endung .LNK eine Programmverknüpfung kennzeichnet, die üblicherweise als E-Mail-Anhang keinen Sinn macht, da es keine direkte Verbindung zwischen der Verknüpfung und dem Programm, auf das sie verweisen soll, geben kann: dagegen ist eine Verknüpfung als Anhang oft nur ein ausführbares Windows-Programm, das umbenannt wurde, um Filter zum Sperren ausführbarer Anhänge zu umgehen. Im zweiten Fall deutet die doppelte Erweiterung auf einen Versuch hin, eine ausführbare Datei als eine nicht-ausführbare (Graphik)-Datei auszugeben, ein beliebiger Trick von Virenschreibern.

Treffer deutlich verringern.

- Programmteile zum Erkennen von Veränderungen verwenden die Regel, dass ein Objekt als verdächtig zu behandeln ist, wenn sich seine Eigenschaften verändert haben. Da aber viele Umstände vorstellbar sind, unter denen sich die Prüfsumme eines Binärprogramms zu Recht ändert (wie bei selbst-modifizierendem Code, neu übersetztem Code, bei einer Neukonfiguration, durch Laufzeit-Komprimierung, Patches oder Updates), kann so ein simples Kriterium zur Erkennung von Änderungen (d.h. die Datei ist verändert, also muss sie infiziert sein) eine hohe Rate von falsch positiven Treffern erzeugen. Dagegen kann die Erkennung von Änderungen gut mit virus-spezifischem Scannen zusammenarbeiten. Dabei ist es eine bewährte Methode, das Objekt mit seiner Prüfsumme zu vergleichen und nur dann einen kompletten Scan durchzuführen, wenn sich die vorher berechnete Prüfsumme geändert hat, wodurch die Zeit zum Prüfen einer unveränderten Datei reduziert wird. Deshalb kann bei manchen AV-Programmen der erste Scan-Durchlauf länger dauern als die folgenden.
- Programme zur Überwachung und zum Blockieren schädlichen Verhaltens, die das Verhalten von Anwendungen beurteilen und danach handeln, gehören zu den frühesten Formen von AV-Software. Deren Ansatz passt gut zur Heuristik, da sie die Leistungsfähigkeit der Blockier-Software steigern und die falsch Positiven verringern kann. Klassische AV-Programme zur Überwachung des Programmverhaltens prüfen gewöhnlich auf zwei Verhaltensweisen: Replikation und potentiell Verursachen von Schäden.
 - Sich replizierender Code deutet schon per Definition stark auf die Anwesenheit eines Virus hin (oder eines Wurms, abhängig von der Art des Codes und der Definition, die Sie bevorzugen). Dieser Ansatz hat seine Vorteile darin, dass die Systemaufrufe, die auf replizierenden Code schließen lassen, vom Programm vergleichsweise leicht zu identifizieren sind, besonders dort, wo der Code nicht übermäßig verschleiert wurde. Die Erkennung eines Virus ist allerdings einfacher, wenn er sich durch Schreiben einer direkten Kopie von sich selbst repliziert (d.h. ein nicht polymorphischer Virus), anstatt eine weiterentwickelte Kopie zu erzeugen.
 - Potentiell schädlicher Code spiegelt die Wahrscheinlichkeit einer bösartigen Nutzlast wider. Dieser Ansatz ist dort unwirksam, wo es keine Nutzlast gibt oder wo sie nicht offensichtlich schädlich ist. Manche Schadensformen, wie das Löschen von Dateien, sind vom Programm leichter zu erkennen als andere, wie die unerwünschte und möglicherweise verwirrende Anzeige von anstößigen Mitteilungen oder Bildern. Andererseits hat die erfolgreiche Erkennung anhand der Nutzlast einen Vorteil beim Aufspüren nicht-replizierender Malware (wie Trojaner und anderer nicht infektiöse Programme). Dennoch muss man Vorsicht walten lassen. So ist beispielsweise das Löschen einer Datei an sich ein unzuverlässiger Hinweis auf Böswilligkeit, da viele Programme routinemäßig und völlig legitim Dateien löschen oder überschreiben, wie etwa nicht mehr benötigte Konfigurations- oder Datendateien.

Generische Lösungen
benutzen heuristische
Regelsätze als Teil des
Diagnosevorgangs.

Ich bin absolut sicher

Die Identifizierung von Viren ist eine Balance zwischen zwei Notwendigkeiten: das Vermeiden falsch negativer (der Misserfolg, eine existierende Infektion zu entdecken) und falsch positiver (Erkennung eines Virus, wo keiner vorhanden ist) Treffer. Wie durch die Häufung von Problemen mit falsch Positiven in etlichen Scannern führender Hersteller in den ersten paar Monaten 2006 demonstriert wurde, haben die Fortschritte bei der Optimierung der Scannertechnik das Risiko falsch positiver Treffer nicht beseitigt.

Der Ausschluss falsch Positiver ist auch bei Anwendung von Heuristik nicht immer möglich, da sie definitionsgemäß ein gewisses Maß an Versuch und Irrtum enthält. Wie bereits weiter oben ausgeführt wurde, ist das Ziel der heuristischen Programmierung weniger das "perfekte" Ergebnis zu erreichen, als ein einheitliches Ergebnis, das "gut genug" ist. Wo ist also das Problem? Die "sicherste" Methode, einen bekannten Virus zu identifizieren, ist die Suche nach jedem einzelnen Byte des Viruscodes, der in einem infizierten Objekt vorhanden sein sollte, indem eine Prüfsumme über jedes konstante Bit des Viruskörpers erzeugt wird. Dieser Vorgang wird oft als "exakte Identifizierung" bezeichnet.

Die Identifizierung ist ein Maß für die Fähigkeit von AV-Software, in einer Virenprobe einen bestimmten Virus oder eine seiner Variante aufzuspüren und als solchen zu erkennen. Als exakte Identifikation bezeichnet man deshalb einen Grad an Genauigkeit, bei dem jedes unveränderliche Byte des Virencodes berücksichtigt wird. Während es wünschenswert scheint, diese Genauigkeit bei jedem Virenskan anzuwenden, kommt sie doch in der Realität kaum vor, einerseits wegen der möglichen Auswirkungen auf die Scanzeit und des Verbrauchs an Systemressourcen und zum anderen, weil dieser Genauigkeitsgrad nur selten benötigt wird.

Der Ausdruck "fast genaue Identifikation" wird angewandt, wenn die Erkennung "gut genug ist, um sicherzustellen, dass ein Versuch, den Virus zu entfernen, keinen Schaden am infizierten Objekt durch Anwendung einer ungeeigneten Desinfektionsmethode hervorruft" [2]. Erkennung und Beseitigung werfen nicht immer die gleichen Probleme auf. Einige AV-Firmen haben schon lange dafür plädiert, infizierte Binärprogramme zu ersetzen statt sie zu säubern und sich lieber auf die Erkennung zu konzentrieren. Auch gibt es Szenarien (Rootkits und Stealthkits sind gute Beispiele dafür), wo der Austausch eines legitimen Programms durch einen Trojaner bedeutet, dass die Sicherheits-Software sowieso nur noch löschen und nicht säubern kann. In so einem Fall ist es üblicherweise notwendig, dass entweder der Administrator oder der Anwender das ursprüngliche Programm wiederherstellt; eine automatische Wiederherstellung ist entweder nicht möglich oder zu unsicher.

Malware hat sich in den vergangenen Jahren weg von der klassischen, parasitischen Infektion von Dateien, hin zur Manipulation der Arbeitsumgebung (z.B. Veränderung der Registry) bewegt. Dies kann es viel schwieriger machen, alle Spuren der Malware zu beseitigen, sobald sie sich einmal eingenistet hat. Eine unvollständige (oder fehlerhafte) Beseitigung kann Schäden am System hervorrufen oder es sogar unbrauchbar machen, was manchmal radikale Maßnahmen erfordert, wie die Neuinstallation des Betriebssystems und der Anwendungssoftware sowie die Wiederherstellung von Daten aus den Sicherungskopien.

Wo Malware jedoch proaktiv mit heuristischen oder generischen Methoden aufgefunden gemacht wird (z.B. bevor sie die Chance hatte, sich auf dem Zielsystem zu installieren), tritt dieses Problem im Allgemeinen nicht auf, es sei denn, das schädliche (infektiöse oder vom Trojaner unterwanderte) Objekt wird in seiner nicht infizierten Form gebraucht, weil es beispielsweise Daten enthält.

Die Identifizierung von Viren ist eine Balance zwischen zwei Notwendigkeiten: das Vermeiden falsch negativer (der Misserfolg, eine existierende Infektion zu entdecken) und falsch positiver Treffer (Erkennung eines Virus, wo keiner vorhanden ist).

“Generische Erkennung” ist ein Ausdruck für die Suche nach einer Anzahl bekannter Varianten mittels eines einzigen Suchstrings, mit dem sich all diese Varianten erkennen lassen. Während man damit auch eine gegenwärtig unbekannte Variante finden kann, in welcher der gleiche Suchstring vorkommt, ist es nur dann eine heuristische Erkennung, wenn dabei eine Bewertungsmethode zum Einsatz kommt. Anderenfalls ist es tatsächlich nur ein Spezialfall der viren-spezifischen Erkennung. Einige Systeme verwenden einen Hybridansatz, bei dem die generischen Fähigkeiten um ein Bewertungssystem ergänzt werden, mit dem die Wahrscheinlichkeit der Abweichung oder der Zugehörigkeit zu einer Virenfamilie mit unterschiedlicher Zuverlässigkeit bestimmt wird. Wenn die Ähnlichkeit beispielsweise groß genug ist, könnte der Scanner “eine Variante von x” melden, ist er sich weniger sicher, könnte es “vermutlich eine Variante von x” heißen.

"Generische Erkennung" ist ein Ausdruck für die Suche nach einer Anzahl bekannter Varianten mittels eines einzigen Suchstrings, mit dem man all diese Varianten erkennen kann.

Empfindlichkeit und Fehldiagnosen

Die Treffgenauigkeit der heuristischen Analyse hängt davon ab, wie aggressiv die Bewertungskriterien eingestellt werden. Wenn der zu findende Schädling neu für den Scanner ist, kann man die Bewertung des Analyseergebnisses nicht anhand einer simplen Ja/Nein-Entscheidung vornehmen (entweder “ja, das ist ein bekannter Virus namens XXX” oder “nein, das ist kein bekannter Schädling”). Stattdessen liegt die Ansprechschwelle eines Scanners auf einer kontinuierlichen Skala zwischen den Endwerten Hoch (was die Anzahl der falsch Positiven so niedrig wie möglich hält) und Niedrig (erlaubt die Erkennung möglichst vieler neuer Viren). Ein aggressives Ansprechverhalten des Scanners priorisiert die Erkennung möglicher Viren gegenüber dem Risiko falsch positiver Treffer, wogegen eine nicht-aggressive Reaktion dort geeigneter ist, wo die nachteilige Wirkung von Fehlalarmen nicht akzeptabel erscheint.

Es ist heutzutage nicht unüblich, dass ein Produkt die Auswahl zwischen einer Standardeinstellung (Heuristik ausgeschaltet) und einer Einstellung mit Heuristik anbietet. (Da wir schon darauf hingewiesen haben, dass alle Scanner zu einem gewissen Grad heuristisch arbeiten, dürfte es vielleicht zutreffender sein, die Standardeinstellung mit “einfache Heuristik aktiviert” zu bezeichnen.) Einige Hersteller unterscheiden auch zwischen passiver und aktiver Heuristik. In beiden Fällen wird der Code nach verdächtigen Eigenschaften durchsucht, aber im aktiven Modus benutzt der Scanner eine Emulationsumgebung, um den Code auszuführen und zu verfolgen. Im passiven Modus führt er nur eine statische Überprüfung des Codes durch.

Die Ansprechschwelle eines Scanners liegt auf einer kontinuierlichen Skala zwischen den Endwerten Hoch (was die Anzahl der falsch Positiven so niedrig wie möglich hält) und Niedrig (erlaubt die Erkennung möglichst vieler neuer Viren).

Eine mögliche Abbildung der Scanner-Technologien auf den Bereich der Schwellwertskala könnte folgendermaßen aussehen:

Schwellwert	Zugehöriger Grad der Heuristik
Höchster	Nur für exakte (oder nahezu exakte) Identifikation; Heuristik wird nicht benutzt oder minimal gehalten
Normal	Sowohl Erkennung bekannter Viren durch algorithmisches Scannen und Emulation, wo angebracht, als auch exakte (oder nahezu exakte) Identifikation, wo erforderlich. Wohl auch mit generischen Signaturen, um relativ nahe Varianten zu identifizieren.
Heuristik-Modus	Mittlere Heuristikstufe, verbesserte Erkennung; ziemlich geringes Risiko falsch Positiver, Verwendung von passiver Analyse statt auf Emulation beruhender Heuristik.
Niedrigster	Höchste (fortgeschrittene oder empfindlichste) Heuristik, einschließlich bestimmter Formen der Emulation. Ein hoher Anteil neuer Malware wird erkannt, aber das Risiko falsch Positiver ist erhöht.

Weder besitzen alle Scanner diese ganzen Empfindlichkeitsstufen, noch erlauben sie die manuelle Einstellung der Schwellwerte oder eine Änderung ihrer Konfiguration, und diejenigen, die tatsächlich Empfindlichkeitsstufen unterstützen, dokumentieren sie möglicherweise nicht. Es sollte ebenfalls betont werden, dass gewisse Formen von Emulation überall auf der obigen Skala im Einsatz sein könnten.

Anbieter von AV-Software, die ihre fortgeschrittene Heuristik standardmäßig deaktivieren, versuchen damit nicht nur, das Risiko falsch Positiver zu verringern, sondern sie könnten tatsächlich damit beabsichtigen, die empfundene Scan-Geschwindigkeit ihres Produkts zu verbessern. Alle Stufen der heuristischen Analyse erhöhen die Scanzeit durch zusätzliche Verarbeitungsschritte und bei einigen Produkten kann die dadurch verringerte Leistungsfähigkeit allzu offensichtlich werden. Wie wir jedoch bereits erwähnt haben, können die Auswirkungen auf den heute üblichen leistungsfähigen Computern mit gut implementierten Programmroutinen, selbst wenn sich die Anzahl der bekannten Schädlinge weiter erhöht, auf ein handhabbares Maß reduziert werden. Tatsächlich gibt es bezüglich der Geschwindigkeitseinbußen eine große Schwankungsbreite zwischen den Scannern unterschiedlicher Hersteller. Ein richtig implementiertes Heuristikmodul sollte nur einen minimalen Einfluß auf die Systemleistung haben.

Die Empfindlichkeit der Heuristik ist nicht nur ein technisches Problem, das die Genauigkeit betrifft, mit der die Anwesenheit eines vorher unbekanntes Virus festgestellt wird. Es ist ebenfalls eine psychologische Streitfrage; wie sollten wir den Anwender auf einen möglichen Virus hinweisen und was sollten wir ihm empfehlen zu tun?

Die Art und Weise, wie auf einen möglichen Virus hingewiesen wird, erzählt dem Kunden eine Menge über den Anbieter der AV-Software. Einige Produkte sind dabei übervorsichtig und benutzen Mitteilungen, die im Kern aussagen, dass es sich um eine Variante von Virus X handeln könnte, sie sich aber nicht vollkommen sicher sind. Dies schaltet für den Hersteller das Risiko falsch Positiver aus, überlässt aber dem Kunden die endgültige Diagnose und die Wahl der Gegenmaßnahme.

Alle Stufen der heuristischen Analyse erhöhen die Scanzeit durch zusätzliche Verarbeitungsschritte und bei einigen Produkten kann die dadurch verringerte Leistungsfähigkeit allzu offensichtlich werden.

Die meisten Kunden würden aber sicher bevorzugen, dass die Diagnose vom Scanner gestellt wird. Auch könnten sich die Anwender bei dem Gedanken unbehaglich fühlen, die Software hätte möglicherweise Unrecht, was die Technologie unzuverlässiger erscheinen lassen könnte als sie tatsächlich ist.

Andere Hersteller zeigen durch mehr Details beeindruckende Meldungen an, etwa in der Art "Schädling XXX gefunden und blockiert" oder "W32/schrecklichertrojaner erkannt und entfernt". Das klingt großartig und der Kunde kann seine gebührende Dankbarkeit über die Erkennung und Neutralisierung des Schädlings zeigen, er weiß aber meist nicht, dass es sich bei diesen Bezeichnungen einfach um generische Namen handelt, die auf eine heuristische Erkennung möglicher Malware hinweisen und kein Beispiel für einen speziellen Virus sind.

Unglücklicherweise gibt es keine zuverlässige Statistik darüber, wie viele rechtmäßige Programme, E-Mails usw. von einem zu selbstsicheren Scanner als Schädling verleumdet worden sind.

Einige Hersteller empfehlen, die fortgeschrittene Heuristik nur in solchen Umgebungen zu aktivieren, wo die Anwesenheit von Schädlingen vermutet wird oder am wahrscheinlichsten ist, zum Beispiel bei Scannern auf E-Mail Gateways. Das verringert die Irritationen, die durch falsch positive Treffer auf Arbeitsplatzrechnern hervorgerufen werden, erhöht aber das Risiko falsch Negativer, wenn das Scannen an den Schnittstellen nach Außen fehlschlägt.

Testprobleme

Das Testen der Virens Scanner auf ihre Erkennungsleistung war schon immer eine strittige Angelegenheit [14], und nur sehr wenige Tester und Testeinrichtungen werden von den anderen Mitgliedern der AV-Forschungsgemeinschaft als Sachverständige auf diesem Gebiet anerkannt.

Folgende Testorganisationen werden im Allgemeinen als auf diesem Gebiet kompetent angesehen:

- AV Comparatives (<http://www.av-comparatives.org/>)
- AV-Test.org (<http://www.av-test.org/>)
- ICSA Labs (<http://www.icsalabs.com/>)
- SC Magazine/West Coast Labs (<http://www.westcoastlabs.org/>)
- Virus Bulletin (<http://www.virusbtn.com/>)
- Virus Research Unit, Universität von Tampere (<http://www.uta.fi/aitokset/virus>)
- Virus Test Center, Universität Hamburg (<http://agn-www.informatik.uni-hamburg.de/vtc/naveng.htm>.)

(Beachten Sie, dass die letzten beiden Einrichtungen in jüngster Zeit nicht besonders aktiv beim Testen waren.)

Nur sehr wenige Tester und Testeinrichtungen werden von den anderen Mitgliedern der AV-Forschungsgemeinschaft als hochkompetent auf diesem Gebiet anerkannt.

Anders als Testern ohne Verbindungen zur AV-Forschergemeinde wird diesen Organisationen von der Gemeinschaft generell das Vertrauen entgegen gebracht - obwohl nicht notwendigerweise von allen Mitgliedern des Verbands - dass sie kompetent, sicher und ethisch korrekt testen und dabei unabhängig bleiben. Diese Vertrauensbasis bedeutet, dass sie oft Zugriff auf bestätigte Virusproben erhalten, wie sie von der WildList International Organization (<http://www.wildlist.org/>) gesammelt, getestet und beglaubigt werden, einer Gruppe gemeinsam

tätiger Forscher, die die meisten größeren AV-Hersteller sowie eine Anzahl großer Firmen und Bildungseinrichtungen repräsentieren.

Die AV-Gemeinschaft behauptet, dass die meisten anderen Tests, die außerhalb der Gruppe der von der Industrie gebilligten Tester durchgeführt werden, potentiell ungültig oder auf andere Weise ungeeignet sind, da:

- die Kompetenz der Tester nicht festgestellt werden kann und deshalb auch nicht:
 - die Eignung der Testmethoden
 - die Einhaltung sicherer Praktiken, ethischer Codes der Industrie und von Standards

Wegen dieser Probleme können die Mitglieder der AV-Forschungsgemeinde mit den Testern, denen sie nicht voll vertrauen, aus ethischen Gründen keine Virusproben austauschen. Deshalb können auch die Herkunft und die Echtheit der Proben, gegen die man die Produkte testet, nicht bestätigt werden. Oft versuchen Tester, die keinen Zugang zu Beispielsammlungen der AV-Gemeinschaft haben, die Beispiele durch solche von Virenaustausch-Seiten und anderen (möglicherweise zweifelhaften) Quellen zu ersetzen, die alle möglichen Sorten von nicht-infektiösen Proben (Dateien mit sinnlosem Inhalt, unfertige Viren, fehlerhafte Beispiele usw.) enthalten können. Einige dieser Probleme könnten überwunden werden, wenn die Einrichtung, die die Bewertung vornimmt, die Tests durch eine anerkannte Organisation durchführen lassen würde. (Beispielsweise führt AV-Test verschiedene Tests für die Besprechung in Zeitschriften durch.)

Seltsamerweise haben diese Schwierigkeiten zu einer Situation geführt (sie aber nicht verursacht), wo die über die Leistungsfähigkeit eines bestimmten Scanners gegenüber unbekanntem Viren besorgten Tester mittels Heuristik getestet haben, noch bevor diese Technologie überhaupt ihren Namen und ihre Fähigkeiten für das 21. Jahrhundert hatte, und zwar durch das Erzeugen von Varianten. Leider ging das normalerweise einher mit der Verwendung unzuverlässiger Virusgeneratoren, unbedeutender Virussimulatoren, dem willkürlichen Einfügen oder Ausschneiden von Viruscode bzw. Zeichenketten und Ähnlichem [15].

Selbstverständlich ist das Testen der heuristischen Fähigkeiten eines Scanners eine absolut zulässige Zielsetzung (besonders jetzt, da die Scanner heuristische Fähigkeiten besitzen). Es ist jedoch wichtig, dass solche Tests genauso kompetent und sicher durchgeführt werden, wie die Tests der Erkennung bekannter Viren. Durch das Fehlen eines sachkundig verwalteten grundlegenden Testdatensatzes gibt es keine Garantie dafür, dass die Scanner gegen gültige und funktionsfähige Viren getestet werden. Tester, deren Sachkunde bereits fragwürdig ist, da sie keinen direkten Kontakt zur AV-Forschergemeinde haben, schaffen sich und denen, die sich auf ihre Tests verlassen, nur weitere Schwierigkeiten, wenn sie keine Informationen über ihre Testmethodik und speziell über die Validierung ihrer Proben veröffentlichen.

Durch das Fehlen eines sachkundig verwalteten, grundlegenden Testdatensatzes gibt es keine Garantie dafür, dass die Scanner gegen gültige und funktionsfähige Viren getestet werden.

Unter Validierung verstehen wir, ob der zu prüfende Code tatsächlich schädlich ist - d.h. ein Virus muss die Fähigkeit zur Replikation besitzen, Würmer müssen in der Lage sein, sich zu verbreiten usw. Wenn die Tester ihre Prüfungen ohne Validierung durchführen, wird oftmals später entdeckt, dass viele der Codeteile gar nicht schädlich waren, sondern es sich nur um beschädigte oder seriöse Dateien handelte, die irrtümlich benutzt wurden.

In einem aktuellen Beispiel [16], wurde indirekt behauptet, dass die mit den Tests beauftragte Gruppe Virusgeneratoren benutzte. Das ließ die AV-Forscher sofort an der Kompetenz der Tester zweifeln, da Virusgeneratoren offenkundig unzuverlässig sind, wenn es darum geht, brauchbare Viren zu erzeugen. Da sie außerdem keine

brauchbaren Details ihrer Testmethodik beschrieben haben, war nicht bekannt, ob und wie die Teststichproben verifiziert wurden. Die Möglichkeit, dass einige oder alle Proben nicht ansteckend waren, macht die Tests von AV-Scannern ungültig, wenn davon ausgegangen wurde, dass die Beispiele infektiös waren. Wenn das der Fall ist, dann entspricht die höchste Erkennungsrate nicht notwendigerweise der besten Leistung, da sie eine große Anzahl von falsch Positiven enthalten könnte [15], selbst unter der Annahme, dass alle getesteten Scanner einheitlich konfiguriert waren.

Die AV-Industrie ist zurückhaltend, wenn es darum geht, die Erzeugung neuer Malware oder von infektiösem Code zu dulden, nicht einmal zu Testzwecken. Für diese Haltung gibt es viele Gründe: das Befolgen eines bindenden ethischen Codes durch viele Forscher, Sorge über Sicherheitsprobleme bei der Handhabung neuer Viren durch unerfahrene Tester, Schwierigkeiten bei der Validierung und vieles mehr. Allerdings ist es auch nicht notwendig, dass irgendjemand Viren erzeugt, um damit die Heuristik zu testen.

“Rückwirkendes Testen“ bedeutet, einen Scanner, der über einen bestimmten Zeitraum (drei Monate ist eine übliche Dauer) nicht aktualisiert wurde, mit einem bestätigten Schädling zu testen, der erst nach der letzten Aktualisierung des Scanners aufgetreten ist. Das bietet ausreichende Gewissheit, dass die heuristische Fähigkeit getestet wird und nicht die Erkennung bekannter Viren mit einem viren-spezifischen Algorithmus. Solche Überprüfung vermindert keineswegs die Notwendigkeit sachkundigen Testens, aber sie vermeidet die ethischen und praktischen Schwierigkeiten, die mit der Erzeugung neuer Viren zu Testzwecken verbunden sind. Sie beseitigt jedoch nicht die Notwendigkeit der Validierung von Proben oder der sorgfältigen Zusammenstellung sinnvoller Tests.

“Rückwirkendes Testen“ bedeutet, einen Scanner, der über einen bestimmten Zeitraum nicht aktualisiert wurde, mit einem bestätigten Schädling zu testen, der erst nach der letzten Aktualisierung des Scanners aufgetreten ist.

Fast alle der größeren AV-Hersteller liefern tägliche (oder noch häufigere) Aktualisierungen der Datenbasis, weshalb des Testens eines Scanners, wenn er bereits drei Monate veraltet ist, nicht sehr viel über seine aktuellen Erkennungsfähigkeiten aussagt. Ein zuverlässigerer Ansatz könnte darin bestehen, die Fähigkeiten zu verschiedenen Zeitpunkten zu testen oder mit einem speziellen Virus den Zeitpunkt festzustellen, an dem er das erste Mal erkannt wird. Auf jeden Fall ist es aber erwähnenswert, wenn ein Scanner in der Lage war, eine Malware zu erkennen, bevor ihre Existenz allgemein bekannt wurde.

Das Fazit: ein heuristisches Paradoxon

Trotzdem die Heuristik-Technologie heutzutage technisch viel ausgereifter ist als in den 90er Jahren, sind die Erkennungsraten insgesamt dramatisch gesunken, obgleich diejenige für Malware der “alten Schule” (Makroviren, Massenmailer usw.) weiterhin beeindruckend hoch ist.

Während manchmal behauptet wird, diese globale Verschlechterung wäre auf die Erfolglosigkeit der AV-Industrie oder ihren Wunsch, an einem viren-spezifischen Erkennungsmodell festzuhalten, zurückzuführen, ist das in Wirklichkeit nicht so. Ein entscheidender Faktor ist dagegen die gestiegene Erfahrung der Virenschreiber, die eine breite Palette an Methoden entwickelt haben, um die Anfälligkeit ihrer Produkte für die heuristische Erkennung zu minimieren und die die Effektivität dieser Methoden gegen entsprechend aktualisierte und konfigurierte Scanner testen. Die Problematik ist heute komplizierter

Die Virenschreiber haben eine breite Palette an Methoden entwickelt, um die Anfälligkeit ihrer Produkte für die heuristische Erkennung zu minimieren.

als noch vor ein paar Jahren, als es (zumindest von den Herstellern) als eine Art Zugabe betrachtet wurde, wenn ein AV-Produkt etwas anderes außer Viren erkennen konnte.

Heutzutage stellen Viren (d.h. Programme mit einer erkennbaren replikativen Funktion) einen viel kleineren Anteil aller Schadprogramme dar [17]. In gewissem Sinne macht das die Arbeit für die heuristischen Scanner viel schwieriger; es ist konzeptionell einfach, einen Virus heuristisch zu erkennen, wenn man den Code nur so weit entwirren kann, dass die Absicht zur Replikation erkennbar wird, obwohl es nicht immer technisch möglich ist, ein replikatives Programm aufzuspüren. Automatisch zu entscheiden, ob ein Programm ein Bot oder eine Art Trojaner ist oder einfach nur, dass es bösartige Absichten verfolgt, ist eine viel größere Herausforderung [5].

Betrachten wir ein klassisches Beispiel: ein Programm, das die Festplatte neu formatiert, ist nicht per Definition schädlich - tatsächlich kann das sein offenkundiger und einziger Zweck sein. Wenn es jedoch ausgeführt wird, weil man dem Anwender vorgespiegelt hat, dass es einen Film abspielen oder seinen Internetzugang verbessern würde, kann man es sinnvollerweise als schädlich ansehen. In so einem Fall liegt das tatsächliche Problem darin, einen Algorithmus aufzustellen, der nicht anhand der Programmeigenschaften unterscheidet, sondern danach, was der Anwender als Zweck des Programms ansieht und was der Programmierer beabsichtigt hat.

Wenn wir auch keine zuverlässige Heuristik für Arglist oder Vorsatz schaffen können, so können wir doch andere Heuristiken anwenden und einem Programm einen entsprechenden Punktwert zuweisen. Eine große Ähnlichkeit eines Programms mit bekannter Malware führt dann zu einer höheren Bewertung. Es gibt viele andere Verhaltensweisen, die die Alarmglocken klingeln lassen, wie beispielsweise, abhängig von der Umgebung, das Öffnen von SMTP- bzw. IRC-Kanälen oder der Start einer Dateiübertragung. Die Analyse von ausführbaren Dateien kann viele Merkwürdigkeiten in der Programmierung aufdecken, wie zum Beispiel verdächtige Patches und Wertekombinationen, widersprüchliche Merkmale der Dateiheders, Differenzen bei Längenangaben und vieles mehr. Die gesamte Ablaufumgebung, in der ein möglicherweise schädliches Programm entdeckt wird, kann ebenfalls wertvolle Hinweise zu seiner Beschaffenheit liefern. Die Analyse von Meldungen kann auf Ähnlichkeiten mit einem bekannten Massenmailer oder über E-Mail verbreiteten Trojaner hinweisen und könnte sogar nützliche Informationen, wie das Passwort für ein verschlüsseltes Archiv liefern.

Obwohl einige Scanner diese Fähigkeit besitzen, wäre es doch zu optimistisch, von einem heuristischen Scanner zu erwarten, dass er automatisch nach einer Passphrase sucht, besonders in Nachrichten mit einem hohen Prozentsatz an graphischem Inhalt. Die Chance, eine Passphrase zu finden, könnte besser sein, wenn die Nachricht anderen schädlichen Nachrichten ähnelt. E-Mail-Nachrichten, die Schadprogramme oder böswillige URLs transportieren, können wiederum anderen Arten von schädlichem Nachrichtenverkehr wie Phishing und Spam ähneln - Virenschreiber und Verbreiter von Spam tauschen ihre Techniken schon seit vielen Jahren gegenseitig aus; Beweise deuten auf wachsende gemeinsame Interessen der einst unvereinbaren Gruppen. Von E-Mail-Scannern wird oft erwartet, dass sie diese und andere Formen des E-Mail-Missbrauchs ebenso wie reine Malware erkennen. Die Analyse des Netzwerkverkehrs kann mit schädlichen Aktivitäten verbundene Muster aufzeigen, wie sie beispielsweise von Massenmailern, durch Botnetze erzeugtem Spam und Scams, und vielem mehr hervorgerufen werden. Aus diesen Gründen kann das Scannen nach Spam (heuristisch oder auf andere Weise) an den Schnittstellen nach Außen die Effektivität der Malware-Erkennung erheblich steigern.

Es ist jedoch auf keinen Fall sicher, dass wir in der vorhersehbaren Zukunft die gleichen hohen Anteile proaktiver Erkennungen wie in den frühen Tagen des heuristischen Scannens haben werden, obgleich das sowohl von den Anwendern als auch von den Herstellern begrüßt werden würde. Die Urheber von Malware haben aber andere Prioritäten. Statt mit der Schrotflintenmethode (maximale Verbreitung einer einzigen Variante) zu arbeiten, setzen sie den Schwerpunkt auf häufige, aber kurze Angriffe mit einem bestimmten Schädlingsexemplar, das auch noch auf bestimmte Personen oder Gruppen zugeschnitten sein kann. Selbst einfache Änderungen, wie eine Reihe schnell zusammengestrickter und modifizierter Laufzeitkomprimierer zur Verschleierung seiner Spuren, können die Erkennungsrate (sowohl heuristisch als auch nicht-heuristisch) reduzieren und die Möglichkeiten selbst der größten Anti-Malware-Labors über Gebühr beanspruchen. Bestimmte Schädlingsformen, die häufigen Gebrauch

von Botnetz-Techniken machen, um sich selbst zu aktualisieren und zu modifizieren, sobald sie auf einem kompromittierten Rechner installiert sind, können sehr schwer zu entdecken sein.

Allerdings gibt es keinen Grund zur Panik - wir leben schon seit etlichen Jahren mit diesen Problemen. Auch das Benutzen des gesunden Menschenverstandes bei der Arbeit am Computer, das Einspielen aller relevanten Patches und die regelmäßige Aktualisierung der AV-Programme bieten einen ziemlich guten Schutz. Nicht nur das, sondern auch weiter verfeinerte Virtualisierungs- und Emulationstechniken, verbunden mit heuristischer Analyse bleiben ein starker und sich ständig weiterentwickelnder Bestandteil im Arsenal der Anbieter von Sicherheitslösungen. Jedoch können weder die AV-Hersteller noch die Befürworter von alternativen "Technologien des Monats" wahrheitsgetreu behaupten, alle zukünftigen Bedrohungen proaktiv erkennen zu können.

Der Trick dabei ist, mit den Erwartungen realistisch zu bleiben.

Literaturnachweise

- [1] "A Short Course on Computer Viruses 2nd Edition", Seiten 2, 49 (Dr Frederick B Cohen): Wiley, 1994.
- [2] "VIRUS-L/comp.virus Frequently Asked Questions (FAQ) v2.00" (N. FitzGerald et al., 1995): <http://www.faqs.org/faqs/computer-virus/faq/> (Zugriff am 12.01.2007)
- [3] "Analysis and Maintenance of a Clean Virus Library" (Dr. V. Bontchev): <http://www.people.frisk-software.com/~bontchev/papers/virlib.html> (Zugriff am 12.01.2007)
- [4] "The Anti-Virus or Anti-Malware Test File": http://www.eicar.org/anti_virus_test_file.htm
- [5] "Trojans" (Harley), in "Maximum Security 4th Edition" (ed. Anonymous): SAMS, 2003
- [6] Oxford Compact English Dictionary, Oxford University Press: <http://www.askoxford.com/> (Zugriff am 12.01.2007)
- [7] Merriam-Webster Online: <http://www.m-w.com/> (Zugriff am 12.01.2007)
- [8] "Viruses Revealed" (Harley, Slade, Gattiker) Seiten 158-159: Osborne 2001
- [9] "Evolution Discussion Group Fall 1996 Phylogenies and Evolution, Useful Terms" - University of British Columbia Zoology Department: www.bcu.ubc.ca/~otto/EvolDisc/Glossary.html (Zugriff am 12.01.2007)
- [10] "Virus Proof" (P. Schmauder), Seite 187: Prima Tech (2000)
- [11] Dr. Solomon's Virus Encyclopaedia (Solomon, Gryaznov), Seiten 30-31: S&S International (1995).
- [12] The Art of Computer Virus Research and Defense (Szor), S. 441, Seiten 451-466: Addison-Wesley (2005).
- [13] "Heuristic Programming": http://www.webopedia.com/TERM/h/heuristic_programming.html (Zugriff am 12.01.2007)
- [14] Anti-virus programs: testing and evaluation (Lee): in "The AVIEN Guide to Malware Defense in the Enterprise" (Ed. Harley): Syngress (2007, in Vorbereitung).
- [15] "AV Testing SANS Virus Creation" (Harley): Virus Bulletin Seiten 6-7, October 2006
- [16] "Consumer Reports Creating Viruses?" (Sullivan): http://redtape.msnbc.com/2006/08/consumer_report.html (Zugriff am 12.01.2006).
- [17] "Email Threats and Vulnerabilities" (Harley). In "The Handbook of Computer Networks" (Ed. Bidgoli): Wiley (2007 – im Druck).

Glossar

Adware	Ein Programm, das eine bestimmte Aktion ausführt (wie das Anzeigen eines Popup-Fensters oder das Umleiten des Browsers auf eine Webseite), mit der die Aufmerksamkeit des Anwenders auf eine Werbung/ein Produkt gerichtet werden soll. Oft als Trojaner angesehen, wenn es ohne Wissen oder Zustimmung des Anwenders installiert wurde.
Fast exakte Identifikation	Erkennung eines Virus, wobei die Identifikation nur ausreichend ist, um sicherzustellen, dass ein Versuch, den Virus zu entfernen, keinen Schaden am infizierten Objekt durch Anwendung einer ungeeigneten Desinfektionsmethode hervorruft. Nicht jeder Abschnitt der unveränderlichen Teile des Virenkörpers wird eindeutig identifiziert.
Prüfsumme	In diesem Zusammenhang ist die Prüfsumme ein berechneter Wert, der vom Inhalt der jeweiligen Datei abhängt. Wenn sich der Inhalt der Datei ändert, so ändert sich auch die Prüfsumme. (Einige Berechnungsmethoden sind anfällig für Kollisionen - d.h. es kann eine Datei mit der gleichen Prüfsumme wie eine andere Datei erzeugt werden, aber in der Mehrzahl der Fälle auf eine einzelne Datei angewandt, wird eine Änderung in der Datei auch die berechnete Prüfsumme betreffen - das reicht für die meisten Zwecke der Integritätsprüfung.)
Corruption	Ein Schaden, der eine Änderung oder Beeinträchtigung der Funktionalität bzw. den Verlust der Lebensfähigkeit (in diesem Kontext speziell für Viren) verursacht.
DDoS Distributed Denial of Service	(Verteilter Angriff zur Behinderung einer Webseite). Typischerweise benutzt ein entfernter Angreifer Zombies oder andere Client-Software, die in bössartiger Absicht auf einem Netzwerk von Computern installiert wurden, um andere Systeme so anzugreifen, dass ihre Funktionalität gestört wird.
Destruktiver Trojaner	Ein Trojaner, der (üblicherweise vorsätzlich) direkten Schaden verursacht, im Gegensatz zu weniger schädlichen Aktionen, wie dem Stehlen von Passwörtern oder anderen Daten.
Dropper	Ein Programm (normalerweise nicht infektiös), das ein anderes Schadprogramm, wie z.B. einen Wurm oder Virus installiert.
EICAR Testdatei	Eine eindeutig formatierte Programmdatei, welche die meisten AV-Programme als Testdatei erkennen und auf die sie auf die gleiche Art reagieren wie auf einen Virus. Die EICAR-Datei ist kein Virus und stellt keine Bedrohung dar: wenn sie ausgeführt wird, zeigt sie einen Text an, mit dem sie sich als Testdatei zu erkennen gibt.

Exakte Identifikation	Erkennung eines Virus, wobei jeder Abschnitt der unveränderlichen Teile des Virenkörpers eindeutig identifiziert wird.
Falsch Negativ	Beschreibt den Fall, dass ein Anti-Malware-Scanner bei der Erkennung eines aktuellen Schädlings versagt.
Falsch Positiv	Beschreibt den Fall, dass ein Anti-Malware-Scanner unzutreffenderweise einen Schädling erkennt, wo es keinen gibt.
Datei mit sinnlosem Inhalt	Gilt in der AV-Forschung nicht als Schädling, ist aber in schlecht gewarteten Beispielsammlungen enthalten, als wäre es einer.
Generisch	Beschreibt Sicherheitsprogramme, die keine bestimmten Bedrohungen erkennen, sondern mit einer Methode schützen, die eine ganze Klasse (oder Klassen) von Bedrohungen blockiert. Eine generische Signatur ist ein Spezialfall davon; eine ganze Reihe von Varianten wird von einer einzigen Signatur erkannt und verarbeitet, statt individuelle Signaturen für jede Variante zu benutzen. Gegenteil von "viren-spezifisch".
Germ	(Keim) Ein Virus der "nullten Generation", der noch nichts infiziert hat (z.B. eine Datei, die nur aus Virencode besteht, statt einer infizierten Programmdatei).
Heuristische Erkennung/ Heuristisches Scannen	Erkennung eines Objekts, das genügend infektiöse oder bösartige Eigenschaften hat, um möglicherweise ein Virus oder ein anderer Schädling sein zu können.
Intendeds	Viren (oder seltener andere Schadprogramme) die aus dem einen oder anderen Grund nicht funktionieren, meist weil sie von ihrem Autor nur unzureichend getestet wurden.
Spaßprogramm	Ein Programm, das irgendeine unerwartete Aktion ausführt, die lästig sein kann, aber nicht wirklich schädlich ist. Die Grenze zwischen einem Spaßprogramm und einen Trojaner kann mitunter sehr schmal sein.
Keyloggers	Ein Programm, das die Tastaturanschläge aufzeichnet, oft zu böswilligen oder kriminellen Zwecken, wie Passwortdiebstahl, installiert.
Scannen nach bekannten Viren, viren-spezifisches Scannen	Scannen nach bekannten Viren mit dem Ergebnis, dass der Name des Virus in der gescannten Umgebung identifiziert wird.
Negative Heuristik	Eine Regel oder ein Kriterium, die falls sie erfüllt werden, die Wahrscheinlichkeit verringern, dass das analysierte Objekt infektiös oder schädlich ist.
Passphrase	Im Gegensatz zum Passwort, das üblicherweise nur aus einem einzelnen Wort oder einer Zeichenkette besteht, ist eine Passphrase normalerweise eine größere Gruppe von Wörtern, die als eine sicherere Form von Passwort genutzt wird.

Positive Heuristik	Eine Regel oder ein Kriterium, die falls sie erfüllt werden, die Wahrscheinlichkeit vergrößert, dass das analysierte Objekt infektiös oder schädlich ist.
Rückwirkendes Testen	Eine Methode zum Testen der heuristischen Fähigkeiten eines oder mehrerer Scanner, indem er über einen bestimmten Zeitraum nicht aktualisiert wird, um anschließend nach Malware zu scannen, die erst nach der letzten Aktualisierung des Scanners aufgetreten ist.
Rootkit	Ein Programm oder eine Reihe von Programmen, die heimlich installiert werden, um unbefugten privilegierten Zugang zu einem System zu erhalten. Manchmal wird auch der Ausdruck "Stealthkit" verwendet, aber das kann auch unbefugter, aber unprivilegiertes Zugang bedeuten. [Siehe auch "Die Wurzel allen Übels? - Rootkits enttarnt" von David Harley & Andrew Lee - http://www.eset.com/download/whitepapers.php]
Scanstring, Suchstring	Eine Folge von Bytes, die in einem bekannten Virus gefunden wurde, die aber nicht in einem seriösen Programm vorkommt. Der Ausdruck ist nicht auf statische Zeichenketten beschränkt, sondern kann auch Platzhalter und reguläre Ausdrücke, sowie die Verwendung anderer viren-spezifischer Erkennungsalgorithmen einschließen. Manchmal auch als "Scan-Signatur" bekannt.
Selbststartend	Ein Ausdruck zur Beschreibung von Schadsoftware, die keinerlei Aktivität auf Seiten des Opfers benötigt, um sich entweder auszubreiten oder zu starten oder beides.
Signatur	Ein Synonym für "Scanstring". Kann man auf eine statische Zeichenkette anwenden, sollte aber am Besten ganz vermieden werden, zumal es die Leute oft zu dem Irrglauben verleitet, dass es eine einzige Bytefolge gibt, mit der alle Virens Scanner einen Virus oder seine Varianten erkennen.
Spyware	Ein Programm, das heimlich Informationen über den Nutzer des Computers sammelt und an interessierte Dritte weitergibt. Umfasst auch einige Formen von Adware.
Virusgenerator	Ein Programm, das selbst kein Virus ist, aber Viren erzeugen kann. Auch als "Viruskit" bezeichnet.
Viren-spezifische Erkennung	Erkennung bekannter Viren mittels spezieller Suchstrings für genau diese Viren oder ihre Varianten.
Platzhalter	Ein Zeichen, das ein anderes Zeichen oder eine Bytefolge ersetzen kann oder das in speziellen Formen regulärer Ausdrücke vorkommt.
Zombie	Ein Programm mit Hintertüren, das auf einem kompromittierten System läuft, auf Befehle von einem entfernten Rechner oder vom befallenen PC selbst wartet und diese ausführt.



www.eset.de

Exklusiv-Distribution Deutschland

DATSEC® Data Security e. K.

Talstr. 84, 07743 Jena, Germany

Tel.: +49 (0) 3641 / 63 78 - 3 Fax: +49 (0) 3641 / 63 78 - 59

E-Mail: info@datsec.de Web: <http://www.datsec.de>

Übersetzung aus dem Englischen: Michael Dankert