4. Die Methodik und das Verfahren des VTC

In diesem Kapitel wird eingehend das Verfahren des Virus Test Centers der Universität Hamburg zum Testen von Anti-Malware-Software dargestellt. Die in den folgenden Abschnitten enthaltenen Angaben zu der Testmethodik, der Aufgabenverteilung, dem Verfahren beim Testen, der Hardware und der Testmenge beziehen sich auf den Stand von Juni 2002. Ergänzende und detaillierte Angaben zum VTC-Projekt finden sich in Anhang A.

Die Projektarbeit im VTC ist am Fachbereich Informatik der Universität Hamburg im Hauptstudium in das Vertiefungsgebiet IT-Sicherheit bzw. das Studienprofil ISO in Form eines Projektseminars eingebunden[37]. Die Studenten arbeiten an den Tests mit, die durch den Projektleiter Prof. Brunnstein vorbereitet, geleitet und veröffentlicht werden. Dabei lernen die Studenten den Umgang mit Anti-Malware-Software, insbesondere die Installation, die richtige Einstellung der Optionen und die Analyse der Ergebnisse. Wissenschaftlich werden durch die kontinuierlichen Tests sowohl aktuelle (vgl. [VTC 2001-10a], [VTC 2001-04]) als auch langfristige Erkenntnisse (vgl. Kapitel 6 und 7, [Bjergstrom 2001], und [VTC 2001-10a], executive summary) über die Erkennung von Malware gewonnen. Diese Erkenntnisse beziehen sich nicht nur auf die Erkennung einzelner Produkte, sondern auch auf allgemein gültige Trends hinsichtlich der Erkennung bestimmter Arten von Viren, der Fähigkeit zur korrekten Säuberung von Malware durch die getesteten Produkte, der Erkennung ohne neueste Signaturen oder der Erkennung auf bestimmten Betriebssystemen (um nur einige wichtige Ergebnisse zu nennen), die aus den Testergebnissen abgeleitet werden können. Da bei den VTC-Tests eine große Zahl an Anti-Malware Herstellern ihre Produkte testen lässt, geben diese zusammenfassenden Ergebnisse auch Aufschluß über den Stand der Technologie der Erkennung bösartiger Software im allgemeinen. So lassen sich Aussagen über das "Hinterherhinken"[38] der Industrie gegenüber den Programmierern von Malware bezogen auf Klassen von Viren/Malware machen. Ferner wird der tatsächliche (und nicht der durch Marketing angepriesene) Schutz des Benutzers gegenüber verschiedenen Arten von Bedrohungen durch Malware sichtbar.

4.1 Grundlagen und Allgemeines zum Virus Test Center

Bevor in Abschnitt 4.2 die Vorgehensweise, die einzelnen Arbeitsschritte und Aufgaben sowie die Aufgabenverteilung im VTC erläutert werden, soll in diesem Abschnitt eine Einführung in die Arbeit des Virus Test Centers erfolgen. Im Virus Test Center werden

[37] Je nach gewählter Prüfungsordnung ist das Projekt eine Veranstaltung im Vertiefungsgebiet oder im Studienprofil; Diplomprüfungsordnung von 1985: Vertiefungsgebiet A6 "IT-Sicherheit und Datenschutz"; Diplomprüfungsordnung von 1998: Studienprofil ISO "Informatiksysteme in Organisationen"

[38] Da Antivirenprodukte Viren hauptsächlich durch Signaturen erkennen (vgl. Kapitel 1.2), kann kein Produkt jederzeit alle Viren erkennen, sondern nur die, für die bereits eine Analyse mit Signatur (zur Erkennung) und Reinigungsmethode (zur Säuberung) als Ergebnis durchgeführt wurde. Insofern "hinken" die Hersteller von Antivirenprodukten den Programmierern bösartiger Software hinterher.

Anti-Malware-Produkte hinsichtlich der Qualität ihrer Erkennung von bösartiger Software getestet. Das dabei angewandte Verfahren basiert auf den in den Abschnitten 3.3.1 und 3.3.2 beschriebenen Methoden.

In Abschnitt 4.1 werden die Grundsätze beim Umgang mit bösartiger Software beschrieben, die der Arbeit im Virus Test Center zugrunde liegen. Außerdem werden die Ziele (4.1.1), die Testumgebung (4.1.2), die Testmenge (4.1.3) und die getesteten Qualitätskriterien (4.1.4) der VTC-Tests geschildert.

4.1.1 Grundsätze des VTC und Ziele der Tests

Da die Arbeit mit bösartiger Software gewisse Gefahren mit sich bringt (etwa die der unabsichtlichen Verbreitung von Viren), hat das Virus Test Center Grundsätze zum Schutz der Außenwelt und der Projektmitglieder entwickelt. Diese Grundsätze sind im sogenannten "*code of conduct*" des Virus Test Centers schriftlich festgelegt und werden von jedem Projektmitglied vor Beginn der Mitarbeit im Projekt akzeptiert. Der vollständige *code of conduct* findet sich in Anhang A; die wesentlichen Inhalte sind:

- man muß sich beim Umgang mit bösartiger Software der Risiken bewußt sein und ständig auf Gegenmaßnahmen vorbereitet sein
- die Testumgebung muß abgeschottet von der Außenwelt sein
- Ein Experiment mit Viren darf niemals durchgeführt werden, wenn der Testrechner durch ein Netzwerk mit der Außenwelt verbunden ist (z.B. Internet)
- man soll immer vorsichtig mit Viren umgehen und alle Schritte bei der Durchführung von Experimenten dokumentieren
- Viren und andere Malware dürfen niemals ohne Genehmigung des Projektleiters das VTC-Labor verlassen
- es ist nicht erlaubt, mit Autoren von Viren oder Leuten, die Viren verbreiten oder austauschen, zusammenzuarbeiten

Das Ziel des Virus Test Centers ist es, mit Hilfe der Durchführung von Tests von Anti-Malware-Software wissenschaftliche Erkenntnisse über die Erkennung von bösartiger Software zu gewinnen. Diese Erkenntnisse sollen möglichst detailliert als auch umfangreich vorliegen, vollständig nachvollziehbar sein und unabhängig von Herstellern zu Stande gekommen sein. Durch die Tests hoffen die Mitglieder des VTC, den Herstellern bei der Qualitätsbeurteilung ihrer Produkte zu helfen und so diese langfristig zu verbessern. Auf diese Weise wird aus Sicht des VTC die Erkennung und damit letztlich auch die Bekämpfung bösartiger Software unterstützt.

Außerdem möchte das VTC mit seinen unabhängig durchgeführten Tests Benutzern bei der Beurteilung der Qualität von Anti-Malware-Software helfen. Da Benutzern oft das notwendige Fachwissen über bösartige Software fehlt, sind sie leicht durch Marketing-Aussagen der Hersteller irrezuleiten. Das VTC gibt den Benutzern eine

Entscheidungshilfe beim Erwerb von Anti-Malware-Software. Die Benutzer werden über den tatsächlich geleisteten Schutz von Anti-Malware-Software informiert und können sich entsprechend ihren Anforderungen das Programm mit dem besten Schutz auswählen.

Zusätzlich dienen die VTC-Tests der Ausbildung von Informatik-Studenten im Fachgebiet IT-Sicherheit. Durch die aktive Mitarbeit im Projekt lernen Studenten - entsprechend dem VTC *code of conduct* (s.o.) - den sicheren Umgang mit maliziöser Software. Außerdem können Studenten wissenschaftliche Methoden zum Testen von Anti-Malware-Software erlernen und im Rahmen der Projektarbeit sowie in Studien-, Baccalaureats- und Diplomarbeiten Erweiterungen und Verbesserungen für die im VTC angewandten Methoden entwickeln.

4.1.2 Getestete Kriterien

Bei den im VTC-Labor durchgeführten Tests von Anti-Malware-Software handelt es sich um Tests im *On-demand*-Modus quantitativer Kriterien. Es werden nur *On-demand*-Tests durchgeführt, da diese Art der Erkennung von Malware automatisiert ausführbar ist. Auf den großen Datenbanken im VTC wären aus Zeitgründen *On-access*-Tests auf den gesamten Datenbanken kaum denkbar. Die getesteten Produkte werden nur auf quantitative Kriterien getestet, da diese Kriterien als einzige mit einer objektiven Metrik messbar und wissenschaftlich aussagekräftig sind (vgl. Kapitel 2).

Im einzelnen werden folgende Kriterien getestet:

- Erkennungsrate[39]
- Erkennungsgenauigkeit
- Erkennungszuverlässigkeit
- Erkennung komprimierter Dateien
- Erkennung unterschiedlicher Dateiformate
- Häufigkeit von Falschmeldungen
- Reparatur infizierter Dateien (nur ART-Test)

Somit werden bis auf die Geschwindigkeit alle in Abschnitt 2.2 genannten quantitativen Kriterien getestet. Ferner gehen alle getesteten Kriterien (außer der Reparatur von infizierten Dateien, die bisher einmalig in einem Extra-Test, dem sogenannten ART-Test, getestet wurde) in die Bewertung der Testprodukte ein. Die Geschwindigkeit ist zwar in den Testprotokollen der Produkte ablesbar, doch aufgrund unterschiedlicher Hardware (siehe nächster Abschnitt) sind die Zeiten, die unterschiedliche Produkte zum Überprüfen der Datenbanken benötigen, nicht vergleichbar. Für einen Test und eine Bewertung der

[39] Die Erkennungsrate wird sowohl in Bezug auf erkannte Viren (mindestens eine Musterdatei des Virus wurde erkannt, *detected viruses*) als auch auf erkannte Musterdateien (insgesamt erkannte Objekte in einer Datenbank, *detected files*) angegeben. Soweit nicht anderweitig vermerkt, ist in dieser Arbeit mit der Erkennungsrate bei VTC-Ergebnissen die Anzahl erkannter Viren im Verhältnis zu der Anzahl von Viren in einer Testmenge gemeint.

Geschwindigkeit müßten alle Produkte auf identischer Hardware installiert sein, was aus organisatorischen Gründen im VTC nicht der Fall ist. Außerdem ist die Geschwindigkeit eines Produktes aus Sicht des Virus Test Centers nur sehr bedingt aussagekräftig für die Qualität der Erkennung. Aufgrund der schon in Abschnitt 2.2 angesprochenen gegenseitigen negativen Beeinflussung von Geschwindigkeit und Erkennung, führt eine hohe Geschwindigkeit oftmals zu einer geringen Erkennung. Die Erkennung bösartiger Software stellt aber die Hauptanforderung an Anti-Malware-Software dar.

Zusätzlich zu den getesteten Kriterien werden auch Aussagen über die Stabilität gemacht. Denn die Durchsuchung riesiger Datenbanken ist auch ein Test an die Stabilität der Produkte. Diese Eigenschaft wird aber nicht vordergründig getestet, negative Vorfälle (Abstürze und Fehlverhalten) der Programme werden lediglich vermerkt und im Testbericht veröffentlicht (in der sogenannten *problemslist*). Die Messung der einzelnen Werte erfolgt durch die Auswertung von Protokolldateien, die die Produkte beim Überprüfen der Datenbanken erzeugen. Das genaue Verfahren zur Auswertung wird in Abschnitt 4.2.2.5 ausführlich beschrieben.

4.1.3 Testumgebung

In Kapitel 3 wurden eine klar definierte Testumgebung und eine offene Beschreibung dieser als einige der Aspekte für nachvollziehbare Testergebnisse genannt. Um die Entstehung der Testergebnisse so transparent wie möglich zu gestalten, wird die Testumgebung bei den VTC-Tests im Testbericht genau dokumentiert (vgl. [VTC 2001-10a], 5PROTOCO.TXT), so daß klar ist, in welch einer Testumgebung die veröffentlichten Ergebnisse zustande gekommen sind.

Die Testumgebung ist auf die Durchführung von On-Demand-Erkennungstests quantitativer (also meßbarer) Qualitätskriterien abgestimmt. Sie besteht aus einem Client/Server-Netzwerk von Rechnern. Dieses Netzwerk von Rechnern steht im VTC-Labor im Arbeitsbereich AGN am Fachbereich Informatik der Universität Hamburg. Zu diesem physikalisch gesicherten Labor haben nur Mitglieder des Virus Test Centers zutritt. Eine physikalische Sicherung ist aufgrund der gefährlichen Malware, die im Labor gespeichert ist, notwendig. Es gibt einen Server, der zwei Funktionen im Netzwerk übernimmt:

- Bereitstellung der Malware-Datenbanken
- Speicherung der Testergebnisse und Datenaustausch

Zur Zeit (Stand Juni 2002) wird als Betriebssystem auf dem Server Windows NT Server (Service Pack 6) verwendet. Jede Datenbank ist als einzelnes Verzeichnis angelegt und im Netzwerk unter der entsprechenden Bezeichnung freigegeben. So können alle Rechner parallel auf die Datenbanken zugreifen, eventueller gleichzeitiger Zugriff auf eine Datei von verschiedenen Clients im Netzwerk aus wird vom Server-Betriebssystem geregelt (im Normalfall durch Serialisierung der Zugriffe). Der Zugriff der Client-Systeme auf die Datenbanken wird auf das Lesen von Dateien beschränkt, so daß von den Clients keine Veränderung an den Datenbanken durch Schreibzugriff (etwa durch Säuberung von Musterdateien durch Testprodukte, die auf den Clients laufen) erfolgen kann. Die

Testergebnisse werden in einem gesonderten Verzeichnis gespeichert, auf das alle Clients per Netzwerkfreigabe vollen Zugriff (inklusive Schreibzugriff) haben. Dieses Verzeichnis wird auch zum Datenaustausch zwischen den einzelnen Testern und der Auswertung benutzt (siehe Abschnitt 4.2.2).

Durch die beschriebene Struktur ist es möglich, auf den am Netzwerk angeschlossenen Client-Rechnern unterschiedliche Betriebssysteme und Anti-Malware-Software zum Testen zu installieren, On-Demand Scans der Datenbanken durchzuführen und die Ergebnisse (in Form von Protokolldateien der Produkte) auf dem Server abzulegen. Abbildung 4.A zeigt schematisch die Struktur der gesamten Testumgebung im VTC-Labor. Aus Sicherheitsgründen (vgl. 4.1.1.) handelt es sich um ein reines lokales Netzwerk (LAN) ohne Anschluß an externe Netzwerke. So kann gesichert werden, daß nicht versehentlich Malware der VTC-Datenbanken verbreitet werden.

Im Server befinden sich drei Netzwerkkarten: eine BNC-Karte mit 10 MBit/s Datenübertragungsgeschwindigkeit und zwei Ethernet-Karten mit 100 MBit/s Datenübertragungsgeschwindigkeit. Über die erstgenannte Karte sind alte Dos-Rechner mit dem Server verbunden. Die beiden schnelleren Netzwerkkarten dienen dem Anschluß von Rechnern mit Ethernet-Nertwerkkarten auf Clientseite. An einer dieser Karten ist ein 100 MBit/s Hub angeschlossen, an dem alle Rechner mit 100 MBit/s Netzwerkkarten angeschlossen sind. Alle anderen Rechner mit einer Datenübertragung von nur 10 MBit/s sind mittels zweier Switches an der anderen Netzwerkkarte am Server angeschlossen.

Es handelt sich um ein heterogenes Netzwerk mit verschiedenen Rechnern, auf welchen unterschiedliche Betriebssysteme installiert sind. Die Rechner im VTC-Labor sind zu unterschiedlichen Zeitpunkten erworben worden und variieren daher in Geschwindigkeit und Ausstattung. Die Rechner, auf denen DOS installiert ist (MS-DOS 6.22), hängen am (alten) BNC-Netzwerkstrang. Es handelt sich dabei um 15 Intel 486-50 Mhz Prozessoren mit 16 MB Arbeitsspeicher. Die Rechner an den anderen beiden Netzwerksträngen werden variabel für unterschiedliche Windows-Betriebssysteme und Linux eingesetzt. Diese Rechner unterscheiden sich in der Ausstattung. Der langsamste eingesetzte Rechner ist ein Pentium I mit 90 Mhz und 32 MB Arbeitsspeicher, der schnellste Rechner ist ein Pentium III mit 128 MB Arbeitsspeicher. Eine genaue Auflistung der einzelnen Rechner und der im VTC-Labor verwendeten Hardware ist in Anhang A zu finden.

4.1.4 Testmenge

Die Testmenge, auf deren Eingabe die Programme getestet werden, besteht beim Virus Test Center aus einer großen Sammlung von Musterdateien, welche mit Malware infiziert sind. Diese Musterdateien werden nach Kategorien sortiert; die Gesamtheit der Musterdateien einer bestimmten Kategorie heißt im VTC Datenbank.

Die Zuordnung der Malwareobjekte zu Datenbanken wird nach verschiedenen Gesichtspunkten vollzogen[40]. Die Haupteinteilung der vorhandenen Malware-Musterdateien erfolgt nach der Plattform, auf der die infizierte Datei ausführbar ist. Danach lassen sich folgende Kategorien von Malware (vgl. Kapitel 1) bilden[41]:

- Boot (Viren, die Bootsektoren von Disketten bzw. Festplatten infizieren)
- File (ausführbarer Code unter Windows oder Dos)
- Makro (Microsoft-Office Dokumente mit Visual Basic for Applications u.ä.)
- Skript (Visual-Basic-Skript Dateien, Java-Skript Dateien, u.ä.)
- Exot (seltene und andere Plattformen, z.B. OS/2, Linux, Java u.ä.)

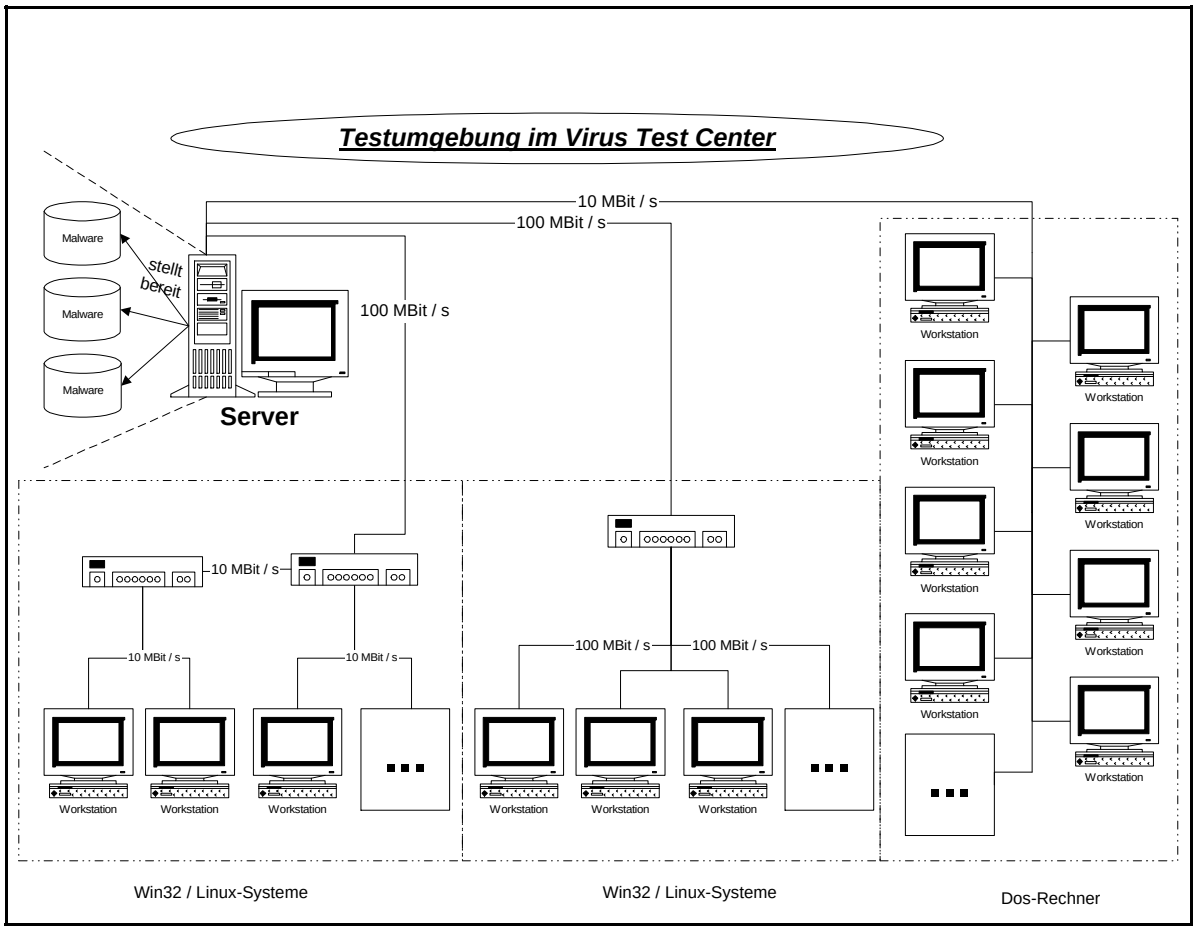

Abbildung 4.A: Testumgebung des VTC

[40] Zur genauen Beschreibung der Einsortierung von Malware in die hierarchische Verzeichnisstruktur der VTC-Datenbanken siehe Abschnitt 4.2.2.2

[41] Da es sich bei den VTC-Testberichten um internationale Veröffentlichungen handelt, sind die im VTC-Projekt verwendeten Bezeichnungen englische Fachbegriffe. In diesem Kapitel werden deshalb die Bezeichnungen von Datenbanken und anderen Fachbegriffen (teilweise) ebenso in englischer Schreibweise verwendet.

Mit *Exot* sind in diesem Zusammenhang exotische Plattformen gemeint (exotisch in dem Sinne, daß auf diesen Plattformen wenige Viren existieren). Je nach Anzahl der in einer Kategorie dieser Haupteinteilung vorhandenen Musterdateien werden die Datenbanken weiter unterteilt. Zum einen wird eine Unterteilung nach ITW und Zoo vorgenommen (vgl. Kapitel 1). Letztere Viren sind alle Viren einer Kategorie, die in der Sammlung des VTC enthalten sind. ITW-Viren sind eine Teilmenge der Zoo-Viren, und zwar die Viren, die in der Öffentlichkeit weit verbreitet sind[42]. Obwohl die ITW-Viren eine Teilmenge der Zoo-Viren bilden, werden sie zu Beginn eines Tests anhand der aktuellen *Wildlist* als Extra-Datenbank angelegt. Dies hat den Vorteil, daß bei Absturz eines Produktes auf den recht großen Zoo-Datenbanken die Erkennung von ITW-Viren dennoch gesondert gemessen werden kann.

Bei den meisten vorhandenen Musterdateien handelt es sich um Viren und Würmer (Definitionen vgl. Kapitel 1), die sich über die Plattform, auf der sie ausführbar bzw. lauffähig sind, einer Datenbank zuordnen lassen. Die restliche - nichtvirale und nichtreplizierende - Malware (z. B. *Trojanische Pferde* oder *Backdoors*) wird im VTC mit dem übergeordneten Begriff Malware bezeichnet[43]. Diese Musterdateien lassen sich ebenfalls nach Plattform unterteilen, analog zu der Vireneinteilung nach File-Malware, Makro-Malware und Skript-Malware.

Bei den Kategorien Boot und Exot sind die Malwareobjekte mit in den regulären Datenbanken enthalten, da hier aufgrund der geringen Anzahl der Objekte eine weitere Unterteilung nicht sinnvoll erscheint. Außerdem wird bei Exot auch keine Extra-Datenbank für ITW-Objekte angelegt.

Zusätzlich zu den beschriebenen Datenbanken werden für File und Makro jeweils Datenbanken mit komprimierten Malwareobjekten angelegt. Dazu werden aus den jeweiligen ITW-Datenbanken Musterdateien genommen und mit verschiedenen Komprimierungsverfahren (ZIP, LHA, ARJ, RAR, WinRAR, CAB) komprimiert. So ist ein Testen der Unterstützung dieser Komprimierungsformate gesondert möglich. Die Datenbanken heißen: File-Pack und Makro-Pack (Pack für gepackte, also komprimierte Dateien).

Desweiteren gibt es zwei gesonderte Datenbanken von File-Viren. Die eine Datenbank besteht aus verschiedenen Varianten sechs polymorpher Viren, die durch automatische Replikation erzeugt wurden. Für jede dieser sechs polymorphen Viren wurden 10.000 verschiedene Objekte erzeugt, so daß die Datenbank aus 60.000 maliziösen Musterdateien besteht. Die andere Datenbank besteht aus 104.640 maliziösen Objekten, die jeweils mit einem von 10.706 sogenannten *VKit*-Viren infiziert sind. VKit ist ein Programm, mit dem

[42] In den VTC-Datenbanken werden die Viren als ITW betrachtet, die laut der monatlich aktualisierten sogenannten *Wildlist* [Wildlist 2002] weit verbreitet sind. Somit bilden zu verschiedenen Zeitpunkten unterschiedliche Teilmengen der Zoo-Viren einer Plattform-Kategorie die ITW-Viren zu dieser Plattform.

[43] Diese Terminologie ist ungünstig gewählt, da der Begriff Malware auch als Oberbegriff für bösartige Software im allgemeinen steht (vgl. Abschnitt 1.2). Bei der Datenbankeinteilung im VTC bedeutet Malware allerdings nur eine Teilmenge aller bösartigen Software, nämlich nichtvirale - im Sinne von nicht selbst replizierende - Malware

nichtversierte Benutzer automatisch Viren erzeugen können. Durch unterschiedliche Optionen können sehr viele unterschiedliche Viren mit dem Programm erzeugt werden. Sowohl die große Testmenge an polymorphen Viren als auch die große Anzahl an *VKit*-Viren liefern aussagekräftige Aussagen über die Erkennung dieser Arten von Viren durch die Testprodukte. Die Datenbanken heißen entsprechend Poly und VKit.

Um die Anzahl an Falschmeldungen messen zu können, werden in die Datenbanken auch saubere, nichtmaliziöse Dateien integriert. Diese Dateien für den sogenannten *false positive* Test sollen von einem Anti-Malware-Produkt als saubere Dateien identifiziert werden. So wird getestet, ob ein Testprodukt nur genau die bösartigen Softwareobjekte der Testmenge identifiziert. Die Menge der so in die anderen Malware-Datenbanken eingefügten sauberen Dateien wird je nach Plattform der entsprechenden Dateien als File-FP oder Makro-FP (FP für *false positive*) bezeichnet. Die Dateien sind zwar nicht in einer gesonderten Datenbank gespeichert, dennoch stellt ihre Gesamtheit pro Plattform jeweils eine Datenbank mit Testobjekten dar und wird auch in der Auswertung als eine Extra-Testmenge ausgewertet.

Insgesamt besteht die Testmenge des VTC aus folgenden Datenbanken:

- Boot-Zoo, Boot-ITW
- File-Zoo, File-ITW, File-Pack, File-Mal, File-FP
- Makro-Zoo, Makro-ITW, Makro-Pack, Makro-Mal, Makro-FP
- Skript-Zoo, Skript-ITW, Skript-Mal
- Poly, VKit
- Exot

Abbildung 4.B zeigt die verwendeten Datenbanken des VTC-Tests 2001-04 ([VTC 2001-04], 5PROTOCO.TXT). Von Skript-Viren wurde in diesem Test nur eine Datenbank mit ITW-Viren angelegt[44]. Die Datenbanken sind mit den Verzeichnisnamen, in denen sie auf dem VTC-Server gespeichert sind, bezeichnet[45]. Diese Verzeichnisnamen sind maximal acht Buchstaben lang und weichen von der oben benutzen Bezeichnung ab: die letzen vier Buchstaben bezeichnen die Plattform, die vorderen Buchstaben konkretisieren die Viren der Plattform (ITW, Pac[k], FP, Zoo, Mal).

4.2 Die Testmethodik

Zum Testen von Anti-Malware-Software hat das Virus Test Center über die Jahre eine Testmethodik entwickelt, die zum Ziel eine möglichst umfangreiche, objektive, exakte, nachvollziehbare und wissenschaftliche Evaluierung der Erkennung von bösartiger Software durch die getesteten Produkte hat. Getestet werden quantitative Qualitätskriterien von Anti-Malware-Software (s. 4.1.4) in einer klar definierten Testumgebung (s. 4.1.2) auf einer

[44] Da zum Zeitpunkt des Tests noch nicht viele Skript-Viren existierten.
[45] Ausnahme: False Positive Testmengen, die in die anderen Datenbanken integriert sind.

möglichst umfangreichen und so weit wie möglich vollständigen Testmenge von Malwareobjekten (s. 4.1.3).

```
This file contains the content of VTCs testbeds (see A3TstBed.zip)
used in Test 2001-04:

    1) In-The-Wild Testbeds:
    ------------------------
    ITW-BOOT.VTC    content of ITW boot virus testbed
    ITW-FILE.VTC    content of ITW file virus testbed
    ITW-MACR.VTC    content of ITW macro virus testbed
    ITW-SCRI.VTC    content of ITW script virus testbed
    PAC-FILE.VTC    content of packed ITW file virus testbed
    PAC-MACR.VTC    content of packed ITW macro virus testbed
    FP-FILE.VTC     content of File virus FalsePositive Testbed
    FP-MACR.VTC     content of Macro virus FalsePositive Testbed

    2) Zoo (=full collection) Testbeds:
    -----------------------------------
    ZOO-BOOT.VTC    content of full boot virus testbed
    ZOO-FILE.VTC    content of full file virus testbed
    ZOO-MACR.VTC    content of full macro virus testbed
    ZOO-POLY.VTC    content of polymorphic file virus testbed
    ZOO-VKIT.VTC    content of VKIT-generated virus testbed
    MAL-FILE.VTC    content of file malware testbed
    MAL-MACR.VTC    content of macro malware testbed (O:)
    EXOTIC.VTC      content of "exotic" viruses&trojans
                            (platforms: Java, Linux, OS2, ...)
```

Abbildung 4.B: Verwendete Datenbanken im VTC-Test 2001-04

Abschnitt 4.2.1 gibt einen Überblick über die Testmethodik und die Arbeitsschritte. Die einzelnen Arbeitsschritte und Aufgaben werden in Abschnitt 4.2.2 ausführlich beschrieben. Abschnitt 4.2.3 befaßt sich mit der Aufgabenverteilung im VTC-Projekt und Abschnitt 4.2.4 geht auf die Sicherstellung der Qualität der veröffentlichten Ergebnisse ein.

4.2.1 Überblick über die Testmethodik

Der Testablauf im Virus Test Center bei der Durchführung eines Tests von Anti-Malware-Software lässt sich grob in acht Arbeitsschritte einteilen:

- Virenkollektionen zusammentragen
- Viren sortieren
- Testprodukte anfordern
- Testen
- Auswerten
- Endauswertung

- Testbericht erstellen
- Testergebnisse veröffentlichen

Diese Arbeitsschritte stellen in der obigen Auflistung von oben nach unten zwar eine chronologische Reihenfolge dar, werden in der Regel aber nicht strikt sequentiell abgearbeitet sondern teilweise parallel und überlappend. Abbildung 4.C zeigt die Abhängigkeiten zwischen den einzelnen Schritten als Kausaldiagramm. Sofern mit einem Arbeitsschritt schon begonnen werden kann, bevor der vorherige abgeschlossen ist, wird dies aus Effizienzgründen auch getan. Dies ist im Diagramm durch diagonale Linien mit Pfeil in der Mitte gekennzeichnet. Es wird deutlich, daß der zweite Arbeitsschritt in diesem Fall erst leicht zeitversetzt begonnen werden kann, da Teilergebnisse des ersten Arbeitsschrittes vorliegen müssen. Besteht eine kausale Abhängigkeit zwischen zwei Arbeitsschritten, so kann der zweite Arbeitsschritt erst nach Beendigung des ersten angefangen werden (im Diagramm durch horizontale und vertikale Pfeile dargestellt).

Das Diagramm (Abbildung 4.C) zeigt lediglich den zeitlichen Fortschritt des Tests. Die Arbeitsschritte beanspruchen unterschiedlich viel Zeit, so sind das Testen und das Auswerten erfahrungsgemäß sehr zeitintensiv, während das Anfordern der Testprodukte recht einfach per e-mail und Download erledigt werden kann. Die angegebenen Arbeitsschritte umfassen jeweils mehrere Aufgaben, die Grafik zeigt den groben Ablauf beim Testen. Eine genaue Beschreibung der einzelnen Aufgaben erfolgt im nächsten Abschnitt.

4.2.2 Beschreibung der Aufgaben und Arbeitsschritte

In diesem Abschnitt werden die einzelnen Aufgaben, die bei der Durchführung eines Tests anfallen, ausführlich beschrieben. Die Aufgaben sind in den einzelnen Unterabschnitten jeweils den groben Arbeitsschritten - entsprechend Abbildung 4.C - zugeordnet.

4.2.2.1 Virenkollektionen zusammentragen

Um eine gute Aussage über die Erkennung von Anti-Malware-Produkten machen zu können, ist eine möglichst große Datenbasis an Musterdateien vonnöten. Je umfangreicher diese Menge an Viren und anderer bösartiger Software ausfällt, desto objektiver und aussagekräftiger ist das Ergebnis, da eine Übertragung der Ergebnisse auf die Realität angenommen werden kann.

Beim VTC arbeitet der Projektleiter, Prof. Brunnstein, in der internationalen Antivirus-Organisation *CARO* (Computer Antivirus Research Organization) mit. Desweiteren pflegt er einen engen Kontakt zu den Herstellern von Antivirensoftware, von denen einige dem VTC regelmäßig Musterdateien von Viren zuschicken. Dadurch ist eine Vielfalt an unterschiedlichen Quellen vorhanden, welche die Testmenge in zweierlei Hinsicht verbessert.

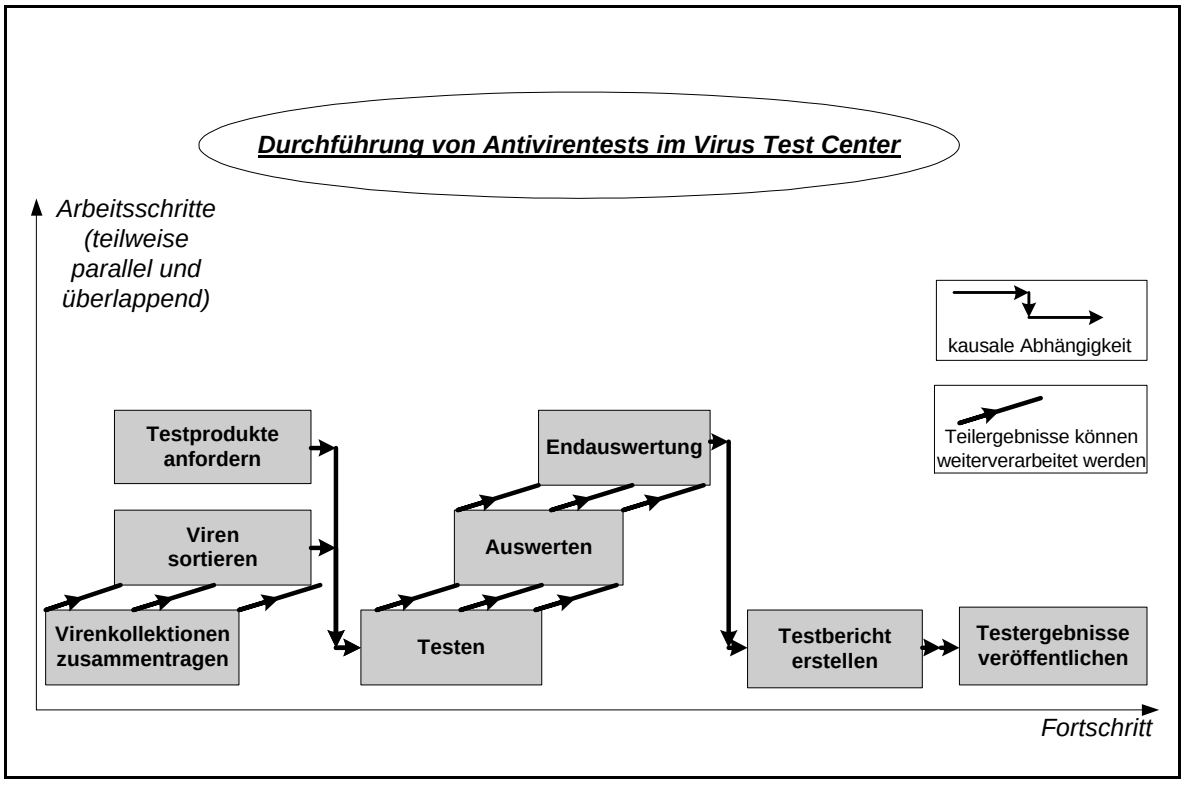

Abbildung 4.C: Kausaldiagramm der Arbeitsschritte beim Durchführen eines Antivirentests im Virus Test Center

Erstens gelangen durch diese Vielzahl an Quellen auch seltene Viren und Exemplare anderer Malware in das Labor des VTC. Durch diese Vielseitigkeit entsteht eine *Breite* der Testmenge. Zweitens erhält das VTC von den bekannteren und öfter vorkommenden Viren eine große Menge an Sampledateien. Durch diese *Tiefe* der Testmenge können Erkennungsgenauigkeit und -zuverlässigkeit (vgl. Abschnitt 2.2.2) umfangreicher und differenzierter gemessen werden.

Von folgenden Herstellern bekommt das VTC monatlich Kollektionen mit maliziösen Dateien zugeschickt:

- Kaspersky Labs
- Network Associates
- Symantec
- Frisk Software

Bei entsprechendem Anlaß wird auch neu entdeckte Malware über CARO-Kollektionen zugesandt. Zusätzlich erhält das VTC auch Musterdateien von Viren und Malware durch den am Arbeitsbereich AGN von Sönke Freitag entwickelten *Malware-Crawler* (siehe [Freitag 2000]), der das Internet automatisch nach bösartiger Software durchsucht und dabei entdeckte Objekte auch an das VTC weiterleitet.

4.2.2.2 Viren sortieren

Durch die große Anzahl an Quellen erhält das VTC eine Vielzahl an Malware-Musterdateien jeden Monat zugeschickt. Diese müssen nicht nur in vorhandene Datenbanken einsortiert werden, sondern auch daraufhin geprüft werden, ob es sich tatsächlich um Viren handelt. Denn es ist nicht auszuschließen, daß bei der Vielzahl der eingesandten Dateien die eine oder andere Datei nichtmaliziös ist (vgl. 3.2.2).

Die Prüfung auf die tatsächliche Verseuchung mit bösartiger Software der erhaltenen Sampledateien erfolgt im Virus Test Center dadurch, daß die eingegangenen Kollektionen von drei bekanntermaßen recht zuverlässigen Virenscannern, die in vergangenen Tests bei der Erkennung von Malware sehr gut abgeschnitten haben, gescannt werden. Diese drei Produkte sind[46]: Antiviral Toolkit Pro von Kaspersky Labs, McAfee Virusscan von Network Associates und F-Prot Antivirus von Frisk Software. Durch diese mehrfache Überprüfung wird die Qualität der Testmenge sichergestellt. Die Dateien, die von keiner der drei genannten Produkte als bösartig eingestuft werden, werden zu einer genauen Inspektion an die Hersteller gesendet. Findet auch dort keiner der genannten drei Hersteller maliziöse Elemente in einer Datei[47], so wird diese aus der Testmenge entfernt. Dieser Prozess der Qualitätssicherung der Datenbanken wird im VTC-Projekt als *Pre-Test* bezeichnet, da vor dem eigentlichen Test die Testmenge auf ihre Qualität hin getestet wird.

Eine Schwierigkeit beim Einsortieren der Dateien in die Verzeichnisstruktur der VTC-Datenbanken stellt die einheitliche Benamung der eingesandten Dateien und das Feststellen des tatsächlichen Dateityps von Musterdateien dar, denn die Dateien sind in der Regel nicht nach gleichem Schema benannt. Da die Dateien aus unterschiedlichen Quellen stammen, ist solch eine Diskrepanz verständlich und unvermeidbar. Viele Dateien, die beim VTC eintreffen, haben eine fortlaufende Dateiendung (z.B. .001, .002, ...). Konvention in den VTC-Datenbanken ist aber, daß die Dateien gemäß ihrem Dateityp enden (also z.B. .doc für Word-Dateien oder .vbs für Visual-Basic-Skript-Dateien). Dafür werden die eigentlichen Dateinamen pro Verzeichnis eines Malwaretyps in der VTC-Datenbankstruktur fortlaufend durchnummeriert (z.B. JVS_000_.JS, JVS_001_.JS, JVS_002_.JS für drei Javaskript-Samples desselben Virus im selben Verzeichnis). Die Änderungen der Dateiendungen werden von Hand vorgenommen, die Namensanpassung beim Einsortieren der Musterdateien in die entsprechenden Verzeichnisse erfolgt automatisch durch Skriptverarbeitung.

Eine weitere Arbeit beim Präparieren der Test-Datenbanken ist das absichtliche Einstreuen nicht verseuchter Dateien, um später die Falschmeldung (*false positive detection*) von Antivirensoftware testen zu können. Hierzu werden von öffentlich verfügbaren CDs, die mit großer Sicherheit als virenfrei angesehen werden können (etwa *Microsoft Windows*-CDs, Compiler-CDs oder AOL Installations-CDs) Dateien kopiert und an zufälliger Stelle in die

[46] Stand: Juni 2002
[47] Genauer: die Datei muß auch einwandfrei replizieren (sofern es sich um selbstreplizierende Malware handelt, vgl. auch Definition von *Virus* in Kapitel 1), um in den VTC-Datenbanken zu bleiben.

Datenbanken eingefügt[48]. Das Einfügen der sauberen Dateien muß protokolliert werden, damit diese Dateien bei der Auswertung der Scan-Protokolle richtig berücksichtigt werden.

Das Einsortieren der Viren in Datenbanken und Unterverzeichnisse wird durch entsprechende Skripte weitestgehend automatisch vollzogen. Die Verzeichnisstruktur im VTC ordnet die Musterdateien hierarchisch:

<Datenbank> \ <Plattform> \ <Bezeichnung> \ <Variante> \ <Objekt>

Also z.B. „R:\W97M\CRONO\A\W97_002_.DOC" für eine Musterdatei (W97_002_.DOC) der Variante A des Word-97-Makroviruses (W97M) „Crono" (Bezeichnung) auf der Datenbank Makro (in diesem Beispiel Festplattenpartition R:). Die Datenbank bezeichnet dabei eine Kategorie von Plattformen (zum Beispiel Makro); für jede Plattform dieser Kategorie (z.B. Word97 oder Access97) existiert ein eigenes Unterverzeichnis, in dem die entsprechenden Musterdateien weiter durch entsprechende Unterverzeichnisse nach Bezeichnung und Variante differenziert werden[49]. In einem Verzeichnis der untersten Hierarchieebene (also einer bestimmten Virus- oder Malware-Variante) können sich mehrere Musterdateien befinden. Mehrere Musterdateien pro Virus und Variante ermöglichen das Testen der zuverlässigen und genauen Erkennung (vgl. Kapitel 2, Abschnitt 2.2.2). Ist für eine Musterdatei nicht die genaue Variante des Virus bekannt, so liegt diese Datei im Verzeichnis für den Virus, auch wenn dort noch Unterverzeichnisse für unterschiedliche Varianten existieren.

4.2.2.3 Testprodukte anfordern

Die im Virus Test Center getesteten Produkte gelangen auf unterschiedliche Arten ins VTC-Labor:

- Download
- Zusendung via e-mail
- Zusendung per Post

Steht ein neuer Test bevor, so wird im Rahmen der Testplanung ein Termin festgelegt, bis zu welchem Testprodukte spätestens eingeschickt werden können (*submission deadline*). Dieser Termin garantiert die Gleichbehandlung aller am Test teilnehmenden Produkte, denn so hat kein Hersteller länger Zeit als die anderen, Signaturen zur Virenerkennung in seine Software zu integrieren. Ist der Termin festgelegt, werden die Hersteller der Produkte, die am letzten Test teilgenommen haben und die Hersteller, die Interesse für einen Test angemeldet haben, über Inhalt und Ziel des anstehenden Tests sowie die submission deadline informiert. Die Hersteller antworten daraufhin (in der Regel via e-mail), wie sie die Produkte zu liefern gedenken. Entweder schicken sie die Produkte via e-mail oder Postversand zu, oder sie geben

[48] Gerade diese "Standard"-CDs sollten von keinem Testprodukt fälschlicherweise als maliziös gemeldet werden, da sie von vielen Benutzern eingesetzt werden.
[49] In einigen Datenbanken bestehen noch weitere Hierarchieebenen, die die Musterdateien noch differenzierter unterteilen. Das beschriebene Schema stellt das grundlegende Verfahren der hierarchischen Sortierung von Musterdateien in den VTC-Datenbanken dar.

eine Internet-Adresse (ftp oder http) inklusive zum Download berechtigender Zugangsdaten an.

Das Anfordern der Testprodukte beinhaltet also folgende Aufgaben:

- Hersteller über anstehenden Test und submission deadline informieren
- gegebenenfalls Produkte downloaden
- erhaltene und downgeloadete Produkte an das Labor übergeben

Produkte, die an einem VTC-Test teilnehmen möchten, müssen Teilnahmebedingungen (*test conditions*) erfüllen, die in jedem Testreport aufgelistet sind (siehe z.B. [VTC 2001-10a], 4Testcon.txt). Diese Bedingungen sind deshalb notwendig, damit sichergestellt ist, daß sich die Produkte unter dem Testverfahren des VTC testen lassen. Dazu müssen die Produkte möglichst automatisierbar ausführbar und die erzeugten Protokolldateien automatisiert mit dem VTC-Verfahren (siehe 4.1.3.5) auswertbar sein. Eine genaue Beschreibung der Teilnahmebedingungen für Produkte findet sich in Anhang A; im wesentlichen müssen folgende Bedingungen erfüllt sein, damit ein Produkt im Virus Test Center getestet werden kann:

- Die Parameter und Optionen für eine optimale Erkennung müssen verfügbar sein
- Der Scanvorgang muß nach annehmbarer Zeit beendet sein
- Es muß eine Protokolldatei erzeugt werden können, in der der komplette Pfad der durchsuchten Dateien gespeichert wird und deren Größe unbegrenzt ist (außer durch verfügbaren Festplattenspeicherplatz)
- Es muß möglich sein, Dateien auf Malware zu untersuchen, ohne daß bei jeder Erkennung eine Benutzereingabe gefordert wird, ein Piepton ertönt oder die Desinfektion der Datei eingeleitet wird ("*scan-only*" Modus); erkannte Dateien müssen unberührt bleiben

Diese Testkriterien werden leider nicht immer ganz genau von den Testkandidaten eingehalten, was zeitaufwendige, manuelle Anpassungsarbeiten an die Testumgebung und das Testverfahren nach sich zieht (vgl. [VTC 2001-10a], 4Testcon.txt: "Several of the scanners in this test did NOT conform to those conditions. Very few had to be withdrawn from the test, whereas several required manual support. The task to test such non-conforming scanners is very difficult and time-consuming").

4.2.2.4 Testen

Sind alle Produkte im VTC-Labor angelangt, kann mit dem Testen begonnen werden. Das Testen der Produkte erfolgt prinzipiell nach dem in Abschnitt 3.3.2 vorgestellten Verfahren zum Testen im On-Demand Modus. Im einzelnen werden folgende Tätigkeiten beim Testen eines Produktes nacheinander vorgenommen:

1. Betriebssystem installieren / Betriebssystem-Image aufspielen
2. Produkt installieren
3. Produkt anpassen / Optionen einstellen
4. Probescan anwerfen
5. Probeprotokolldatei überprüfen
6. Gegebenenfalls Änderungen an Installation vornehmen und ab Schritt 4 wiederholen
7. Scan auf Datenbank anwerfen
8. Schritt 7 wiederholen, bis alle Datenbanken getestet sind
9. Reportdateien in Auswertungsverzeichnis kopieren
10. Schritt 1 bis 8 wiederholen, bis das entsprechende Produkt auf allen Betriebsystemen getestet ist

Abbildung 4.D veranschaulicht das Vorgehen beim Testen eines Produktes auf einem ausgewählten Betriebssystem. Innerhalb der Testumgebung wird ein freier Rechner (auf dem gerade keine Testläufe durchgeführt werden) für die Installation eines Betriebssystems und Produktes gewählt. Da Interdependenzen zwischen einzelnen Produkten nicht ausgeschlossen werden können und Deinstallationsroutinen - sowohl von Betriebssystemen als auch von Programmen bereitgestellte - nicht gesichert alle Programmdateien eines Produktes entfernen, muß vor jedem Testen eines jeden Produktes ein "frisches" Betriebssystem installiert werden. Nur so können gleiche Testbedingungen für alle Testkandidaten garantiert werden. Da das Installieren von Betriebssystemen mitunter recht zeitaufwendig sein kann (die Installation von Windows NT auf einem Pentium II Rechner dauert ca. eine Stunde), kann Zeit gespart werden, indem von einem Betriebssystem gleich nach der Installation und Integration in das Netzwerk ein sogenanntes Image (etwa mit Programmen wie *Norton Ghost* oder *Powerquest Drive Image*) angelegt wird, vor der Installation von Programmen. Dieses Image kann dann in kurzer Zeit auf den entsprechenden Rechner wieder aufgespielt werden, mit demselben Effekt einer komplett neuen Betriebssysteminstallation. Aufgrund unterschiedlicher Hardware muß für jeden Rechner und jedes Betriebssystem ein separates Image angelegt werden. Wird die Erkennung auf einem Betriebssystem in einem Test zum ersten Mal getestet (z.B. Windows XP im kommenden Test) oder ist ein altes Image einer Betriebssysteminstallation aufgrund von Updates (z.B. Windows NT oder Windows 2000 Service und Security Packs) nicht mehr aktuell, muß das Betriebssystem neu installiert werden, bevor ein neues, wiederverwendbares Image angelegt wird.

Die Installation eines Produktes erfolgt entsprechend der mitgelieferten Anleitung bzw. nach mitgesandten Angaben der Hersteller. Eine Installation inklusive Kommandozeilenversion (CLI, *command line interface*) des Produktes wird nach Möglichkeit bevorzugt, da durch den Aufruf per Batch-Datei auf Kommandozeilenebene Testdurchläufe auf mehreren Datenbanken mit unterschiedlichen Protokollen automatisiert auf einmal gestartet werden können.

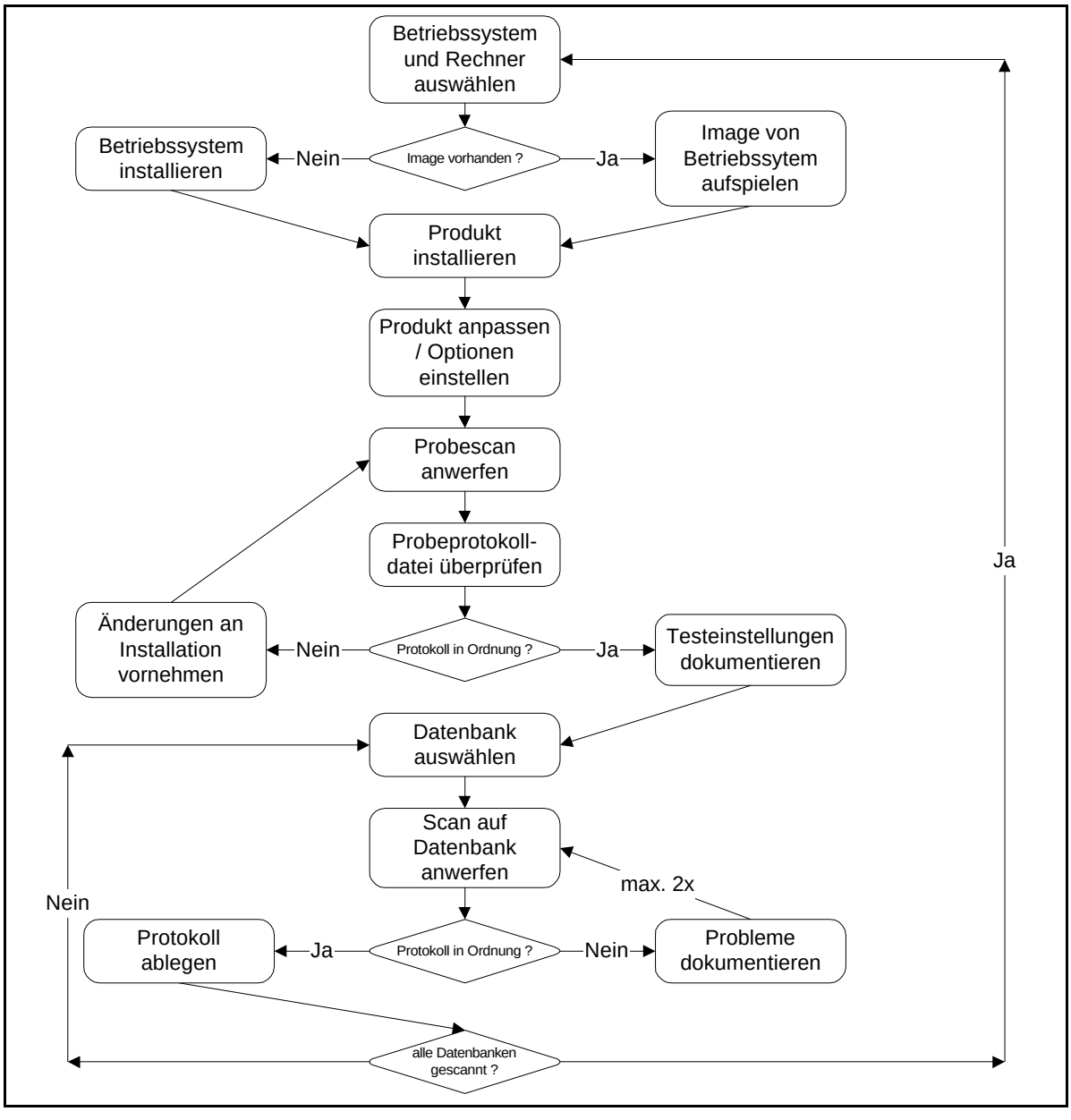

Abbildung 4.D: Vorgehen beim Testen eines Produktes im VTC

Die Anpassung eines jeden Produktes umfasst folgende Teilaufgaben zur Sicherstellung einer einwandfreien und korrekten Installation des jeweiligen Produktes:

- Datum von Signatur und Engine auf Einhaltung der *submission deadline* (siehe 4.1.3.3) prüfen
- e-mails vom Hersteller lesen, um Empfehlungen und zusätzliche Informationen zum Programm zu erhalten
- Gegebenenfalls (vor *submission deadline* datierte) Signaturupdates durchführen

- Gegebenenfalls Programm durch vom Hersteller erhaltenen Schlüssel (*key*) freischalten
- Optionen entsprechend Herstellervorgaben bzw. optimal für Test einstellen:
  - maximale Heuristik aktivieren
  - Durchsuchen aller Dateien aktivieren (nicht nur bestimmte Dateiendungen)
  - Durchsuchen komprimierter Dateien aktivieren
  - Speicherung von Protokolldatei aktivieren (Pfad von Protokolldatei angeben und Speicherung aller Informationen aktivieren)
  - Reparatur oder Quarantänefunktion deaktivieren
  - Soundfunktion deaktivieren
  - Benutzerinteraktion bei Erkennung deaktivieren
  - Restliche Optionen untersuchen und optimal für Test einstellen
- Probescans durchführen, analysieren und gegebenenfalls Änderungen an der Installation vornehmen
- Dokumentation von Version, Signaturdatum und Einstellungen des Produktes speichern

Zur Überprüfung der gewählten Einstellungen wird ein Probescan (in der Regel auf der kleinsten Datenbank) durchgeführt und danach die erzeugte Reportdatei betrachtet. Ist der Scan einwandfrei verlaufen (ohne Abstürze, Benutzerinteraktion, Warnmeldungen und ähnliches) und entspricht die Reportdatei den zur automatisierten Auswertung nötigen Anforderungen des VTC (vgl. Anhang A, A.2: Bedingungen für am Test teilnehmende Softwareprodukte), werden die gewählten Einstellungen und Anpassungen gespeichert. Dies ist für die Nachvollziehbarkeit und Dokumentation der späteren Testergebnisse wichtig. So können Hersteller exakt nachvollziehen, unter welchen Einstellungen die Ergebnisse erzielt wurden.

Treten Fehler beim Scannen auf, oder ist die Reportdatei nicht in Ordnung, müssen die gewählten Einstellungen überprüft und geändert werden, bevor ein weiterer Probedurchlauf gestartet wird. Bei unerwartet großen Problemen bei der Anpassung von Produkten werden die Hersteller um Hilfe gebeten. Läßt sich ein Produkt auch nach Rücksprache mit dem Hersteller nicht richtig oder nur mit nicht vertretbarem Aufwand einstellen, scheidet es aus dem Test aus. Der Grund für das Ausscheiden wird ebenfalls dokumentiert und später im Testbericht veröffentlicht.

Ist die Installation eines Scanners eingerichtet, kann dieser auf einer Datenbank zum Scannen gestartet werden. Hierbei gibt es zwei grundsätzlich verschiedene Ausführungsarten. Beim Scannen über eine grafische Benutzeroberfläche (GUI, *graphical user interface*), erfolgen Auswahl von zu scannenden Dateien und der Start des Scans durch Markieren und Klicken mit der Maus. Für jede zu scannende Datenbank müssen also vom Tester erneut die zu scannenden Dateien markiert werden und der Scan manuell gestartet werden. Zusätzlich muß bei den meisten Scannern der Name der Reportdatei in den Einstellungen geändert werden, damit nicht die Reportdatei vom vorherigen Scan überschrieben wird.

Beim Scannen über die Kommandozeile (CLI, *command line interface*) werden die Scanner über eine Befehlszeile aufgerufen (zum Beispiel den DOS-Prompt unter Microsoft Windows-Systemen). Dabei werden die gewählten Einstellungen und Optionen dem Scanner über Parameter mitgeteilt. Diese Art des Scannens hat den Vorteil, daß das Durchsuchen mehrerer Datenbanken mit unterschiedlich benannten Reportdateien mittels einer Stapeldatei (engl. *batch*) gestartet werden kann. In einer Batch-Datei sind mehrere Befehlszeilen untereinander aufgelistet, die bei Aufruf der Datei vom Computer sequentiell ohne Benutzerinteraktion ausgeführt werden. Auf diese Art kann das Scannen mehrerer Datenbanken effizient automatisiert werden. Leider verfügen nicht alle Anti-Malware-Produkte über eine Kommandozeilenversion.

Nach dem Testen einer Datenbank (bei GUI Modus) oder nach dem Testen aller Datenbanken (bei CLI Modus) mit einem Produkt unter einem Betriebssystem müssen die erzeugten Reportdateien ins Auswertungsverzeichnis kopiert werden. Dadurch wissen die Auswerter und anderen Tester, daß das Scannen erfolgreich fertiggestellt wurde. Vor dem Einsortieren werden die Reportdateien durch den Tester noch einmal überprüft und gegebenenfalls angepaßt. Einige Produkte erzeugen Protokolldateien nicht in dem von der im VTC bei der Auswertung verwendeten ASCII-Format, so daß die Dateien durch entsprechende Konvertierungsprogramme umgewandelt werden müssen. Ferner müssen die Reportdateien, damit sie ausgewertet werden können, folgender Namenskonvention genügen:

<Scannerabkürzung>.ful (z.B. SCN.ful)

Die Scannerabkürzung ist dabei die im gesamten Test verwendete eindeutige Abkürzung jedes getesteten Produktes auf 3 Zeichen[50]. Die Endung .ful ist für die Verwendung in Auswertungsskripten notwendig. Die fertigen Reportdateien werden auf dem Server der Testumgebung nach folgendem Schema einsortiert:

P:\ <Test> \ <Datenbank> \ <Betriebssystem> \ <Reportdatei>

(z.B. P:\2001-10\H\Win2000\AVP.ful[51])

Die erzeugten und im richtigen Verzeichnis abgelegten Reportdateien bilden die Schnittstelle zwischen der Auswertung und dem Testen der Produkte. Nach der Ablage der Reportdateien ist das Testen abgeschlossen (für das entsprechende Produkt unter dem entsprechenden Betriebssystem, ausgenommen Nachscans, s.u.), die Auswertung kümmert sich um die Verarbeitung der Reportdateien zu Testergebnissen.

[50] eine Liste mit den in VTC-Tests verwendeten Produktabkürzungen findet sich in Anhang A.
[51] wobei auf Laufwerk H für diesen Test eine Datenbank installiert war

4.2.2.5 Auswerten

Die Auswertung hat mehrere Aufgaben zu erfüllen (vgl. [Messerschmidt 2002b], Folie 6):

- Testergebnisse aus Protokolldateien ermitteln
- Qualitätssicherung der Ergebnisse
- Verzeichnisse und Testergebnisse verwalten

Hauptaufgabe der Auswertung ist es, die beim Testen erzeugten Reportdateien der Testprodukte so auszuwerten, daß die Ergebniswerte der getesteten Kriterien ermittelt werden. Zur automatischen Auswertung der von den Testprodukten erzeugten Protokolldateien werden im VTC Test die Skriptsprachen *AWK* und *Perl* verwendet. Mithilfe dieser Sprachen lassen sich sogenannte *Skripte* erzeugen, mit denen Dateien auf bestimmte Merkmale (z.B. Wortfolgen wie "virus found") untersucht werden können. Durch die Benutzung von Stapeldateien (*Batch-Dateien*) kann ein Skript nacheinander auf mehrere Protokolldateien angewendet werden. Die Verarbeitung einer Protokolldatei[52] erfolgt in mehreren Schritten:

1. Protokolldatei in einheitliches Format umwandeln
2. Protokolldatei aufteilen
3. Erkennungsrate und andere Kriterien ermitteln
4. Überprüfung des Protokolls

Protokolldateien in einheitliches Format umwandeln

Um die Protokolldateien der Scanner automatisch verarbeiten zu können, müssen diese in ein einheitliches Format umgewandelt werden. Jeder Scanner benutzt eine andere Nomenklatur zur Beschreibung der Ereignisse beim Überprüfen einer Verzeichnisstruktur. So kann zum Beispiel ein Produkt "suspicious code found: ..." bei der Erkennung von Malware melden, während andere Produkte "virus found: ..." und wieder andere "found the ... virus" in ihren Protokollen vermerken. Um für alle Produkte die Protokolldateien einheitlich mit Skripten auswerten zu können, müssen deshalb die Protokolle der einzelnen Produkte im Rahmen der Auswertung standardisiert werden. Dafür gibt es für jedes Produkt ein spezifisches AWK-Skript (sog. *scanner key*), in welchem speziell für dieses Produkt die in den Protokolldateien benutzten Bezeichnungen vermerkt sind und dementsprechend in ein einheitliches Format umgewandelt werden können. Diese *scanner keys* müssen oftmals aktualisiert werden, da die Produkte ihr Ausgabeformat und ihre Meldungen recht häufig ändern. Eine standardisierte Protokolldatei beinhaltet durch Semikola getrennte Zeilen mit:

- Pfad des getesteten Objektes
- Meldung des Scanners
- Malwarebezeichnung des Scanners für getestetes Objekt

[52] Der Begriff Protokolldatei wird in dieser Arbeit als Synonym für Reportdatei verwendet (vgl. Kapitel 1)

Protokolldateien aufteilen

Die standardisierten Reportdateien werden in drei Dateien geteilt:

- infiziert gemeldete Dateien
- nicht infiziert gemeldete Dateien
- übrige Zeilen

Die Protokolldatei mit den als infiziert gemeldeten Objekten wird zur Berechnung der prozentualen Erkennungsrate und anderer Kriterien benutzt. Die nicht infiziert gemeldeten Dateien und übrigen Zeilen dienen der Qualitätskontrolle. So kann zum Beispiel durch Addition der Zeilen in den Protokolldateien der infizierten und nichtinfizierten Dateien festgestellt werden, wieviele Objekte gescannt wurden. Durch Vergleich mit der Anzahl der Objekte in der Datenbank kann auf diese Weise festgestellt werden, ob alle Objekte in der Datenbank überprüft wurden.

Erkennungsrate und andere Kriterien ermitteln

Aus der Datei mit den als infiziert gemeldeten Malwareobjekten können mittels eines Skriptes durch Vergleich mit der Dateiliste der gescannten Datenbank die Erkennungsrate in Prozent sowie die falsch als Malware identifizierten Objekte (*false positives*) ermittelt werden. Insgesamt werden ermittelt:

- die Anzahl der erkannten Viren (mindestens ein Objekt pro Virus)
- die Anzahl der erkannten Objekte
- die Anzahl der Falschmeldungen (*false positives*)
- die Anzahl der ungenau identifizierten Objekte (*unreliable identification*)
- die Anzahl der unzuverlässig erkannten Objekte *(unreliable detection)*

Überprüfung der Protokolldateien

Zusätzlich werden die Reportdateien überprüft. Diese Überprüfung kann als Qualitätskontrolle des Testens angesehen werden. Entspricht eine Reportdatei nicht den Anforderungen für die Auswertung, oder hat zum Beispiel ein Testprodukt nicht alle Dateien in der Datenbank überprüft, so wird eine erneute Überprüfung der Datenbank durch das Testprodukt angeordnet. Im einzelnen werden folgende Punkte überprüft:

- hat der Scanner alle Dateien in der Datenbank inspiziert ?
- sind die richtigen Optionen beim Scannen gewählt worden ?
- ist das Datum der Virensignaturen vor der *submission deadline* datiert ?

Hat ein Scanner nicht alle Dateien einer Datenbank überprüft, so wird von der Auswertung ein sogenannter *Nachscan* für das entsprechende Produkt und die entsprechende Datenbank veranlasst. *Nachscannen* bedeutet, daß ein Produkt auf einer Datenbank wiederholt zum Scannen von Dateien angeworfen wird. Dabei übermittelt die Auswertung den Testern, ob die gesamte Datenbank (z. B. bei unleserlichem Reportformat oder falsch gewählten Optionen) oder nur einzelne Unterverzeichnisse (z. B. bei geringerer Anzahl gescannter Dateien als in der gescannten Datenbank vorhanden oder Absturz des Scanners) erneut gescannt werden

sollen. Bei Fehlverhalten des Scanners (nicht bei falschen Einstellungen durch Tester) werden maximal 2 Nachscans pro Produkt und Datenbank durchgeführt. Danach werden die inklusive des letzten Nachscans erreichten Ergebnisse gewertet.

Ein Nachscan wird auch dann veranlasst, wenn ein Produkt unzuverlässige Erkennung von bestimmten Viren zeigt, das heißt, wenn alle bis auf einige Musterdateien eines bestimmten Virus erkannt werden. Diese nicht erkannten Musterdateien werden von der Auswertung als Liste zum wiederholten Scannen an die Tester weitergereicht. Dadurch soll die tatsächliche Nichterkennung der Objekte verifiziert werden und ausgeschlossen werden, daß einzelne Nichterkennungen auf Dateizugriffsfehler im Netzwerk zurückzuführen sind.

Neben der beschriebenen Ermittlung der Testergebnisse und der Qualitätskontrolle der Protokolldateien ist die Auswertung für die Archivierung der Protokolldateien zuständig und überwacht den Verlauf des Testens und den Stand der Nachscans. Dies geschieht durch entsprechende Verzeichnisse, in denen die Protokolldateien von den Testern (siehe 4.2.2.4) und Informationen von zu erledigenden Nachscans von den Auswertern abgelegt werden.

4.2.2.6 Endauswertung

Sind für ein Betriebssystem alle Produkte auf allen Datenbanken getestet und sämtliche Nachscans durchgeführt wurden, so kann für dieses Betriebssystem die Endauswertung vorgenommen werden. Es werden sämtliche Auswertungen der Protokolldateien zu Übersichtstabellen zusammengestellt. Dies geschieht skriptgesteuert und automatisch, indem die Ergebnisse jedes Produktes als Zeile in der entsprechenden Übersichtstabelle für jede Datenbank eingefügt werden. Außerdem werden andere Informationen für den Testbericht zusammengestellt (siehe 4.2.2.7) und an den Projektleiter weitergeleitet.

Erst im Rahmen der Endauswertung werden die Gesamttestergebnisse eines Betriebssystems berechnet. Diese abschließende Auswertung der Ergebnisse eines Betriebssystems kann erst erfolgen, wenn keine Nachscans mehr durchzuführen sind**.**

4.2.2.7 Testbericht erstellen

Der Projektleiter, der den Testbericht anfertigt, bekommt von der Auswertung folgende Daten als Grundlage für den Testbericht:

- für jedes Betriebssystem und jede auf diesem Betriebssystem getestete Datenbank eine Übersichtstabelle mit den erzielten Ergebnissen aller Scanner, inklusive Übersichtstabellen zu Falschmeldungen (*false positive detection*) und Detailtabellen für die Erkennung einzelner Komprimierungsformate
- Sämtliche Protokolldateien von Produkten des gesamten Tests
- Dokumentationen der aufgetauchten Probleme

- Dokumentationen von Version, Engine, Signaturdatum und benutzter Einstellungen aller getesteten Produkte (teilweise unterschiedliche Einstellungen pro Betriebssystem**)**
- Eine Liste der Verzeichnisstruktur aller Datenbanken, auf denen getestet wurde

Jede Übersichtstabelle enthält die Ergebnisse der Erkennung aller Scanner unter einem Betriebssystem auf einer bestimmten Datenbank. Abbildung 4.E zeigt als Beispiel die Ergebnistabelle der Erkennung von Viren aus Macro-Zoo (Datenbank) unter Windows NT (Betriebssystem) aller darauf getesteten Produkte (aus [VTC 2001-10a], 6GWNT.TXT).

Table WNT.M1: "MacroVirus 1": Results of "full" zoo test for macro viruses under Windows NT:

Scanner	Viruses detected		This includes ---- unreliably ---- identified		detected		Files detected	
Testbed	6762	100.0%					21677	100.0%
ANT	6566	97.1	185	2.7	56	0.8	20811	96.0
AVA	6604	97.7	42	0.6	32	0.5	21202	97.8
AVG	6651	98.4	55	0.8	13	0.2	21387	98.7
AVK	6762	100.0	118	1.7	1	0.0	21674	100~
AVP	6761	100.0	118	1.7	1	0.0	21673	100~
AVX	6703	99.1	134	2.0	8	0.1	21474	99.1
CMD	6760	100.0	93	1.4	1	0.0	21672	100~
DRW	6725	99.5	81	1.2	14	0.2	21574	99.5
FPR	6760	100.0	29	0.4	1	0.0	21672	100~
FPW	6760	100.0	29	0.4	1	0.0	21672	100~
FSE	6762	100.0	120	1.8	0	0.0	21677	100.0
IKA	6451	95.4	421	6.2	170	2.5	20723	95.6
INO	6755	99.9	110	1.6	6	0.1	21651	99.9
MR2	44	0.7	4	0.1	4	0.1	156	0.7
NAV	6726	99.5	110	1.6	17	0.3	21495	99.2
NVC	6751	99.8	89	1.3	12	0.2	21622	99.7
PAV	6762	100.0	118	1.7	1	0.0	21674	100~
QHL	0	0.0	0	0.0	0	0.0	0	0.0
RAD	6726	99.5	346	5.1	16	0.2	21546	99.4
RAV	6726	99.5	348	5.1	15	0.2	21545	99.4
SCN	6762	100.0	66	1.0	0	0.0	21677	100.0
VSP	1	0.0	0	0.0	1	0.0	1	0.0

Abbildung 4.E: Beispiel einer Ergebnistabelle
(Ergebnisse unter Windows NT auf der Datenbank Macro-Zoo)

Beim Schreiben des Testberichtes werden die einzelnen Ergebnistabellen zusammengefügt und um Kommentare ergänzt. Desweiteren analysiert der Projektleiter die Ergebnisse manuell und schreibt daraus gewonnene Erkenntnisse in den Testbericht. Die Vielzahl der Daten durch die vielen Ergebnistabellen (pro Datenbank und Betriebssystem eine Tabelle, sowie Tabellen zur *false positive* Erkennung und der Erkennung komprimierter Malwaredateien) wird in zwei Schritten verdichtet, dabei entstehen drei Ebenen von Ergebnisdaten:

Ebene 1: Pro Betriebssystem eine Datei mit allen Ergebnistabellen

Ebene 2: Evaluation der Ergebnisse pro Betriebssystem, für jedes Betriebssystem eine Datei mit zusammengefassten Ergebnissen sowie Kommentaren und Analysen zu den Ergebnissen, außerdem eine Bewertung der unter diesem Betriebssystem getesteten Produkte

Ebene 3: *Executive Summary*, Zusammenfassung der Ergebnisse und Bewertungen des gesamten Tests, Angabe der aus dem Test gewonnenen (Meta-)Erkenntnisse[53]

Diese Verdichtung der Daten ist zum einen zur sinnvollen Analyse der großen Menge an Testergebnisdaten nötig, denn nur durch Verdichtung können die Allgemeinheit betreffende Erkenntnisse gewonnen werden (Meta-Erkenntnisse). Zum anderen interessieren sich die Leser des Testberichtes für unterschiedlich verdichtete Ergebnisdaten. Anwender möchten eine Bewertung der Produkte (Ebene 2), die Wissenschaft generelle Erkenntnisse zur Erkennung von Viren (Ebene 3) und die Hersteller genaue Details der Erkennung ihres Produktes zu dessen Verbesserung (Ebene 1) als Testergebnisse erhalten.

Die Bewertung der Produkte wird im VTC nach einem klaren Schema vollzogen. Ein *perfektes* Anti-Virus bzw. Anti-Malware-Produkt muß die Kriterien in Abbildung 2.A bzw. 2.B (siehe Kapitel 2) voll erfüllen. Da dies in der Regel kaum ein Produkt tut, wird für einige, wichtige Kategorien jeweils eine Erkennungsrate festgelegt, ab der das Produkt in der entsprechenden Kategorie als *perfekt* oder *exzellent* gilt. Die Kategorien sind[54]:

- Erkennung unter File-Zoo
- Erkennung unter Makro-Zoo
- Erkennung unter Skript-Zoo
- Erkennung unter Poly
- Erkennung unter Vkit
- ITW-Erkennung (aller Plattformen)
- Erkennung komprimierter Dateien (sowohl File-Pack als auch Makro-Pack)
- Anzahl von Falschmeldungen
- Erkennung von (nichtreplizierender) Malware

```
Test category:        "Perfect"              "Excellent"
  ------------------------------------------------------------
```

[53] Mit Meta-Erkenntnissen sind in diesem Zusammenhang Erkenntnisse gemeint, die als generelle Aussagen aus der Analyse der verdichteten Teilergebnisse gewonnen werden können.
[54] (vgl. [VTC 2001-04], z.B. 7evalwnt.txt)

```
    WNT zoo file test:       ---               AVP,PAV,AVK,SCN

    WNT zoo macro test: CMD,FPR,FPW,FSE,PAV,SCN      ---
    WNT zoo script test:       FSE, SCN              ---
    WNT zoo Poly test:      AVG,AVP,FSE,RAV          ---
    WNT zoo VKit test:         FSE, SCN              ---
    WNT ITW tests:           AVK,AVP,SCN     AVG,CMD,FPR,FPW,FSE,
                                                 INO,NAV,NVC
    WNT pack-tests:          AVK,SCN               ---
    WNT FP avoidance: AV3,AVG,AVK,INO,PRO,SCN      ---
    WNT Malware Test:        ---             FSE,SCN,AVP,PAV,AVK,
                                  FPR,FPW,CMD
    ------------------------------------------------------------

In order to support the race for more customer protection, we
evaluate the order of performance in this WNT test with a simple
algorithm, by counting the majority of places (weighing "perfect"
twice and "excellent" once), for the first places:

    ************************************************************
    "Perfect" WNT AntiVirus product: =NONE=
    "Excellent" WNT AV products:
        1st place:                    SCN          (13 points)
        2nd place:                    FSE          ( 9 points)
        3rd place:                    AVK          ( 7 points)
        4th place:                 AVG, AVP        ( 5 points)
        6th place:          CMD,FPR,FPW,INO,PAV    ( 3 points)
       11th place:               AV3,PRO,RAV       ( 2 points)
    ************************************************************
    "Perfect" WNT AntiMalware product: =NONE=
    "Excellent" WNT AntiMalware product:
        1st place:                    SCN          (14 points)
        2nd place:                    FSE          (10 points)
        3rd place:                    AVK          ( 8 points)
        4th place:                    AVP          ( 6 points)
        5th place:              CMD,FPR,FPW,PAV    ( 4 points)
    ************************************************************
```

Abbildung 4.F: Beispiel einer Produktbewertung für ein Betriebssystem (WindowsNT im VTC Test 2001-04)

Perfekte Erkennung bedeutet eine Erkennungsrate von 100% (außer bei Falschmeldungen: dort 0%)[55]. *Exzellente* Erkennung bedeutet eine Erkennungsrate von mindestens 99% (außer bei Malware: dort über 90% und bei Falschmeldungen: dort unter 0,5%). So kann eine Gesamtbewertung der Produkte pro Betriebssystem vollzogen werden, indem jedes Produkt pro Kategorie mit *perfekter* Erkennung 2 Punkte und pro Kategorie mit *exzellenter* Erkennung 1 Punkt erhält. Die Gesamtzahl der Punkte ermöglicht eine Gesamtbewertung aller Testprodukte eines Betriebssystems: das Produkt mit den meisten Punkten hat am besten abgeschnitten. Abbildung 4.F zeigt die Produktbewertung im Test 2001-04 unter WindowsNT.[56]

[55] Für die Bewertung "*perfekt*" wird vom VTC zusätzlich auch eine Erkennung aller Objekte (100% *erkannte Objekte*) sowie eine durchweg zuverlässige (0% *unreliably detected*) und genaue (0% *unreliably identified*) Erkennung gefordert.
[56] (vgl. [VTC 2001-04], 7evalwnt.txt)

Der fertige Testbericht besteht im einzelnen aus folgenden Teilen (in Klammern die Bezeichnung entsprechend der tatsächlichen Nummerierung und Bezeichnung der Dateien in der Veröffentlichung, vgl. [VTC 2001-10a]), dabei entsprechen Dateien mit einer 6 am Anfang Datenverdichtungsebene 1 (s.o.), Dateien mit einer 7 am Anfang Ebene 2 und der *Executive Summary* der Ebene 3:

- Zusammenfassung *(executive summary*) und letzte Meldungen
  (0README.1ST, 0XECSUM.TXT)
- Übersicht über die einzelnen Dokumente
  (1CONTENT.TXT)
- Vorwort
  (2PROLOG.TXT, 3INTRO.TXT)
- Bedingungen für teilnehmende Produkte
  (4TESTCON.TXT)
- Beschreibung der Testmenge und Überblick über das Verfahren
  (5PROTOCO.TXT)
- Detaillierte Ergebnisse pro Datenbank und Betriebssystem aller Scanner (pro Betreibssystem ein Dokument, zusätzlich eine Datei zum Vergleich der Windows32-Erkennung)[57]
  (6DDOSMAC.TXT, 6FW98.TXT, 6GWNT.TXT, 6IW2K.TXT, 6MCMP32.TXT, 6XLinux.TXT)
- Für jedes Betriebssystem und die Windows32-Erkennung eine Zusammenfassung und Bewertung der Ergebnisse
  (7EVAL-DOS.TXT, 7EVAL-WNT.TXT, 7EVAL-W98.TXT, 7EVAL-W2k.TXT, 7EVAL-CMP.TXT, 7EVAL-LIN.TXT)
- Liste mit während des Tests aufgetauchten Problemen
  (8PROBLMS.TXT)
- Dokumentationen zu Version, Engine, Signatur und benutzter Einstellungen pro Scanner und Betriebssystem
  (A2SCANLS.TXT)
- Zusammenfassung und Ausblick
  (9EPILOG.TXT)
- Copyright des Testberichtes
  (DISCLAIM.TXT)
- Liste der ITW-Viren (laut Wildlist.org)
  (A1ITW00b.TXT)
- Liste mit der im Testbericht benutzten Abkürzungen der Produktnamen
  (A5CODNAM.TXT)
- Beschreibung der benutzten Testmenge als Liste der Verzeichnisstruktur aller Datenbanken, auf denen getestet wurde
  (A4TSTDIR.TXT, A3TSTBED.ZIP)
- Sämtliche Reportdateien aller Scanner (in komprimierter Form)
  (SCAN-RES)

[57] Windows32-Erkennung: Vergleich der erzielten Ergebnisse pro Scanner unter unterschiedlichen Windows-32-bit-Systemen (Windows 98, WindowsNT, Windows 2000)

Durch diese im Testbericht enthaltenen Informationen ist die Nachvollziehbarkeit der Testergebnisse gegeben, da die Testumgebung dokumentiert wird, die Testmenge detailliert beschrieben ist, die benutzten Einstellungen dokumentiert sind und sämtliche Reportdateien veröffentlicht werden. So kann jeder Leser des Testberichtes nachvollziehen, wie die Resultate der einzelnen Scanner entstanden sind. Ferner ist in den Zusammenfassungen genau beschrieben, aus welchen Ergebnissen sich die Gesamtbewertung zusammensetzt. Das einzige, was im Rahmen des Testberichtes nicht veröffentlicht wird, sind die maliziösen Dateien selber. Stattdessen wird eine detaillierte Verzeichnisliste aller in den Datenbanken (also der Testmenge) vorhandenen Malwareobjekte angegeben. Eine Veröffentlichung der Malwareobjekte selber ist deshalb nicht zulässig, weil auf diese Weise die gesamte Malware der umfangreichen Testmenge des Virus Test Centers öffentlich verfügbar gemacht würde. Ein Mißbrauch der bösartigen Software (zum Beispiel durch Verbreitung) wäre nicht auszuschließen und zu kontrollieren. Da das Virus Test Center aber die Bekämpfung von bösartiger Software unterstützen möchte (auch deshalb wird die Erkennung in Tests gemessen), ist eine Veröffentlichung bösartiger Software nicht mit den Grundsätzen des Virus Test Centers vereinbar.

4.2.2.8 Testbericht veröffentlichen

Bevor der Testbericht veröffentlicht wird, wird er dem Testteam intern vorgelegt. Die Mitglieder des Testteams überprüfen den Testbericht und haben dann vor der Veröffentlichung Zeit, Anmerkungen und Fehlerkorrekturen anzugeben, die in den Testbericht eingefügt werden. Da der gesamte Testbericht aus ASCII-Dateien besteht, ist die Veröffentlichung einfach. Die Dateien werden auf dem eigenen Webserver des Arbeitsbereichs AGN (in den das Virus Test Center eingegliedert ist) des Fachbereichs Informatik der Universität Hamburg durch Kopieren und Verlinkung auf der Webpage des VTC der Öffentlichkeit verfügbar gemacht.

Zusätzlich werden die von der Auswertung erzeugten Ergebnistabellen in eine Datenbank[58] importiert, die von Thomas Buck, Jens Gallion und Jan Seedorf im Rahmen des VTC-Projektes entwickelt wurde. In der Datenbank werden die Ergebnisdaten so aufbereitet, daß daraus Grafiken zur Veranschaulichung der Testergebnisse erzeugt werden können. Diese Grafiken werden aus der Datenbank erstellt[59] und durch Kopieren auf den AGN-Webserver (s.o.) und Verlinkung auf der VTC-Webpage veröffentlicht.

Der Viren-Datenbankmanager erhält von den Auswertern zu jedem Testprodukt eine Liste aller in diesem Test von dem jeweiligen Produkt nicht erkannten maliziösen Objekte. Er kann dann für jedes Produkt eine Kollektion aus Kopien der nichterkannten Dateien zusammenstellen. Diese als *missed samples* bezeichneten Dateien werden den Herstellern zugeschickt. Alle Testteilnehmer bekommen *missed samples* von den ITW-Datenbanken. Die jeweils nichterkannten Musterdateien der anderen Datenbanken erhalten nur Hersteller, von denen ein kompetenter und vertraulicher Umgang mit der bösartigen Software erwartet werden kann. Die Notwendigkeit dieser Vertrauensbeziehung zwischen Hersteller und Virus

[58] es handelt sich um die Microsoft Access 97 Datenbank "*VTED*", siehe Kapitel 6.2
[59] mittels Code in *Visual Basic for Applications* und unter Benutzung von *Microsoft Graph*, siehe Kapitel 6.3

Test Center zum Erhalt von Musterdateien ergibt sich aus den Grundsätzen des VTC (*code of conduct*), die den vorsichtigen Umgang mit bösartiger Software sicherstellen. Das Zusammenstellen der *missed samples* erfolgt skriptgesteuert, so daß bei Eingabe einer von der Auswertung gelieferten Liste nicht erkannter Dateien automatisch Kopien aller dieser Dateien aus den Datenbanken angefertigt werden können.

4.2.3 Aufgabenverteilung

Die in 4.2 beschriebenen Aufgaben lassen sich vier unterschiedlichen Verantwortungsbereichen zuteilen:

- Projektleiter
- Viren-Datenbank-Manager
- Tester
- Auswerter

Durch diese Verantwortungsbereiche, im folgenden als Funktionen bezeichnet, läßt sich die Arbeit im Projektteam aufteilen. Jedes Mitglied im Team wird einer Funktion zugeordnet und hat dadurch einen definierten Aufgabenbereich. Die drei Funktionen außer dem Projektleiter sind dabei jeweils für eine umfangreiche Kernaufgabe bei der Durchführung des Tests verantwortlich. Der Projektleiter übernimmt die Projektplanung und Vorbereitung sowie die (nicht weniger zeitintensive) Veröffentlichung der Ergebnisse. Er vertritt das Projekt nach außen.

Je nach Umfang der Aufgaben werden ein oder mehrere Personen einer jeweiligen Funktion zur Durchführung des Tests benötigt. Die Einteilung nach Funktionen ist jedoch nicht strikt, so daß im Bedarfsfall Teilnehmer im Projekt auch andere Aufgaben übernehmen können. Eine solche Flexibilität der Aufgabenteilung ist in einem Projekt vonnöten, in dem einige Funktionen nur durch wenige Personen übernommen werden (Viren-Datenbank-Manager, Auswerter). Nur so kann der Ausfall (etwa durch Krankheit oder Prüfungsvorbereitung) eines Mitglieds des Testteams kompensiert werden.

Der Projektleiter steht am Anfang und Ende des Testprozesses. Er hält den Kontakt zu Herstellern von Anti-Malware Produkten, die ihre Produkte zum Testen einschicken und Musterdateien von Malware liefern. Diese Musterdateien übergibt er dem Viren-Datenbank-Manager. Am Ende des Testberichtes analysiert er die Ergebnisse und schreibt den zusammenfassenden Testbericht. Zusätzlich überwacht der Projektleiter den gesamten Prozess und ist Ansprechpartner bei auftretenden Problemen.

Der Viren-Datenbankmanager pflegt die Malware-Datenbanken des VTC. Dazu gehören das Einsortieren neuer Kollektionen von Musterdateien und die Qualitätsüberprüfung der Malwareobjekte in den Datenbanken. Am Ende des Tests stellt er Kollektionen mit nichterkannten Musterdateien (*missed samples*) pro Testprodukt zusammen. Er übergibt diese Zusammenstellungen dem Projektleiter, damit dieser sie an die Hersteller weiterleiten kann.

Den Kern des Testverfahrens, sowohl inhaltlich als auch von der zeitlichen Abfolge betrachtet, bildet das eigentliche Testen der Produkte. Dieses wird von mehreren Testern vorgenommen, da die Aufgabe aufgrund der großen Anzahl an Produkten und der Größe der Virendatenbank nur so bewältigt werden kann. Das Testen der Produkte umfaßt im wesentlichen das Installieren der Betriebssysteme und Produkte, die Anpassung der Produkte an die Testumgebung (Einstellungen vornehmen) und das eigentliche Scannen der Datenbanken mit den Testprodukten (vgl. 4.2.2.4). Teilweise führen Tester auch den Download von teilnehmenden Produkten durch.

Die Auswerter sind für die Auswertung der Produkte (entsprechend Abschnitt 4.2.2.5) zuständig. Die Aufgabe des Auswertens kann bereits mit den ersten vorliegenden Protokolldateien begonnen werden. Es werden nicht so viele Auswerter wie Tester benötigt, obwohl trotz Automation auch die Auswertung zeitaufwendig ist. Dies liegt an der häufig nötigen Anpassung der *scanner keys* (vgl. 4.2.2.5) aufgrund von Änderungen der Testprodukte von Test zu Test und an der Verwaltung und dem Austausch von Protokolldateien und Testergebnissen an zwei Schnittstellen (Tester und Testbericht). Bei der Überprüfung des Testberichtes vor der Veröffentlichung können die Auswerter am besten die im Testbericht enthaltenen Testergebnisse überprüfen, da die Tester in der Regel die Protokolldateien nur erzeugen, aber nicht inspizieren, während die Auswerter die Ergebnistabellen erzeugt haben und so Fehler (zum Beispiel falsche Bezeichnungen) entdecken können.

Tabelle 4.H zeigt die Zuordnung von Aufgaben zu Funktionen und die benötigte Anzahl an Personen pro Funktion. Abbildung 4.G zeigt das gesamte Verfahren des VTC inklusive Aufgabenverteilung. Links in der Abbildung sind untereinander die Arbeitsschritte in chronologischer Reihenfolge abgebildet. Die kleineren Rechtecke schräg rechts unter der jeweiligen Arbeitsschrittbezeichnung stellen die Ergebnisse dar, die am Ende des betreffenden Arbeitsschrittes vorliegen. Rechts außen stehen zu jedem Arbeitsschritt die verantwortlichen Funktionen. Zwischen den Verantwortlichen und den Zwischenergebnissen sind die Tätigkeiten dargestellt, die im jeweiligen Arbeitsschritt ausgeführt werden müssen.

4.2.4 Sicherstellung der Qualität

Durch die automatische Datenverarbeitung und die vielen Schnittstellen zwischen den Arbeitsschritten im Projekt ist es notwendig, die Richtigkeit der berechneten Testergebnisse zu überprüfen. Die Sicherstellung der Qualität der VTC-Testberichte erfolgt an mehreren Stellen:

- beim Sortieren der maliziösen Objekte
- im Auswertungsprozeß
- vor Veröffentlichung des Testberichtes

Die Qualitätssicherung beim Sortieren der maliziösen Objekte erfolgt durch Überprüfung der Musterdateien mit ausgewählten Scannern (dem sogenannten *Pre-Test*), wie in Abschnitt 4.2.2.2 beschrieben.

Da sämtliche Testergebnisse während der Auswertung automatisch erstellt werden, ist es wichtig, die erzeugten Testergebnisse zu überprüfen. Diese Qualitätssicherstellung ist im VTC-Projekt Teil des Auswertungsprozesses und erfolgt durch die Generierung von Kontrolldateien und Analyse dieser Kontrolldateien. In den Kontrolldateien werden zusammenfassende Angaben über den Auswertungsprozeß bzw. den Ablauf eines Auswertungsskriptes gespeichert. Diese Daten können dann exakt mit den tatsächlichen Werten verglichen oder auf Plausibilität überprüft werden. Ersteres kann automatisch durch Skripte erfolgen, letzteres geschieht durch Begutachtung der Kontrolldatei durch den Auswerter.[60]

Funktion	***Aufgabenbereich***	***Anzahl***
Projektleiter	• Virenkollektionen zusammentragen • Testprodukte anfordern • Testbericht erstellen • Testergebnisse veröffentlichen	1
Viren-Datenbank-Manager	• Viren sortieren • Testergebnisse veröffentlichen	1-2
Tester	• Testen • Testprodukte anfordern • Testergebnisse veröffentlichen	5-10
Auswerter	• Auswerten • Endauswertung • Testergebnisse veröffentlichen	2

Tabelle 4.H: Funktionen und Aufgaben im Virus Test Center

[60] Die Qualitätssicherung während der Auswertung ist ausführlich in [Messerschmidt 2002a] beschrieben.

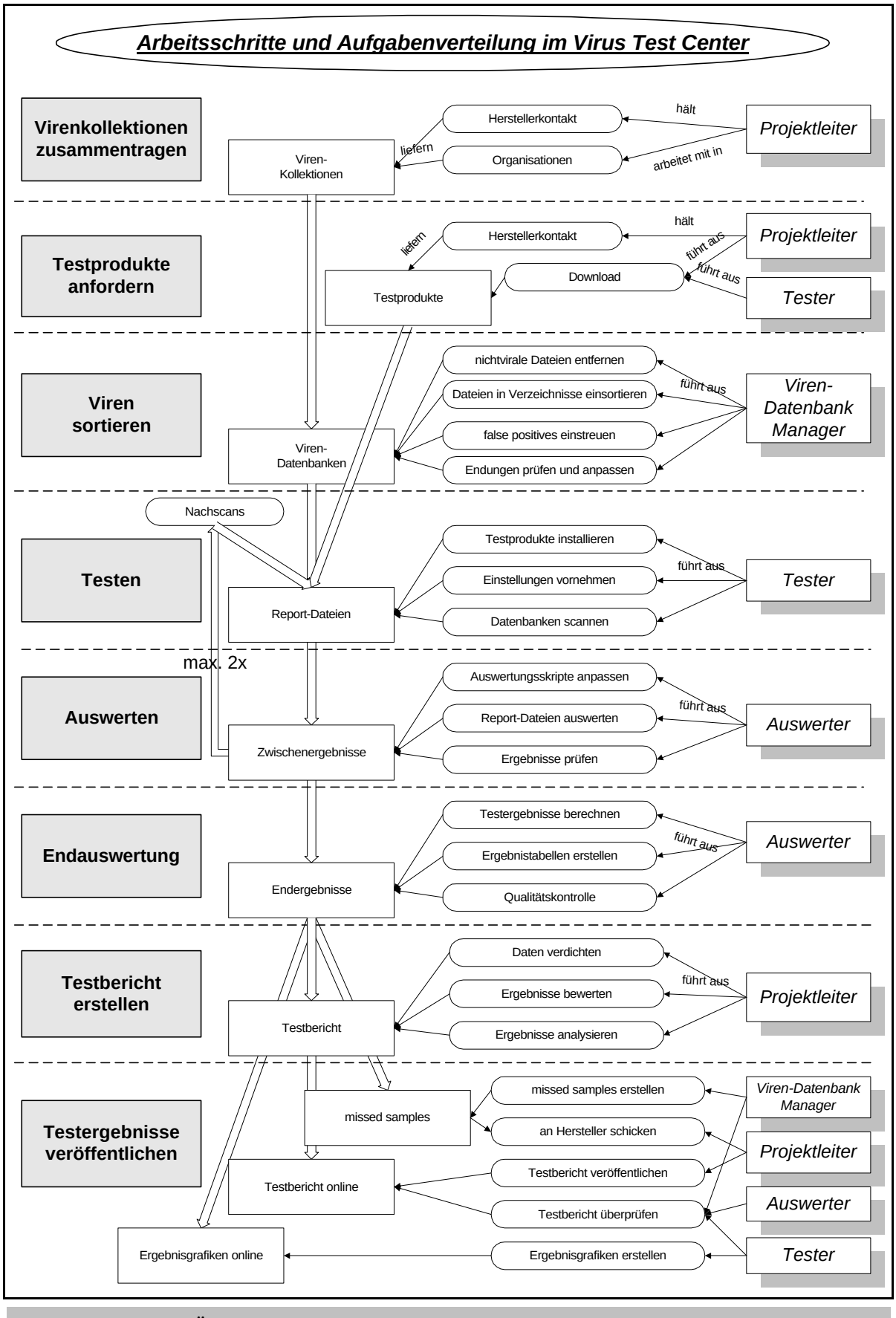

Abbildung 4.G: Übersicht über das VTC-Testverfahren inklusive Aufgabenverteilung

Ein Beispiel ist die Aufteilung der Protokolldateien der Scanner. Während des Auswertungsprozesses werden die Protokolldateien der Scanner in mehrere Dateien aufgeteilt (siehe 4.2.2.5). Da dabei keine Zeilen entfernt oder hinzugefügt werden, kann automatisch die Zahl der Zeilen der Ausgangsdatei mit der Summe der Zeilen in den drei Ausgabedateien verglichen werden. Stimmen diese Zahlen nicht überein, so ist ein Fehler während der Ausführung des Auswertungsskriptes aufgetreten (z.B. sind durch einen Fehler im Skript nicht alle Zeilen in den Dateien verarbeitet worden). Solch ein Fehler wird bemerkt und kann dann entsprechend korrigiert werden.

Vor der Veröffentlichung des Testberichtes wird dieser allen Mitgliedern des Testteams zur Revision vorgelegt. Die einzelnen Gruppen (Funktionen) des Testteams überprüfen zur Qualitätssicherung die Ergebnisse im Testbericht auf Plausibilität und vergleichen Aussagen kritisch mit den persönlich gemachten Erfahrungen. So können beispielsweise die Tester die Liste mit Problemen beim Testen auf Richtigkeit überprüfen. Der Viren-Datenbankmanager überprüft die Angaben zur Testumgebung und zu den Datenbanken. Die Auswerter überprüfen die veröffentlichten Ergebnistabellen. Sämtliche Testmitglieder durchsuchen den Testbericht auf Rechtschreibfehler. So wird abschließend die Qualität des Testberichtes noch einmal von den Inhabern der einzelnen Funktionen im Testprozeß sichergestellt.

4.3 Durchgeführte Anti-Malware-Tests

Im Virus Test Center werden in regelmäßigen Abständen (zweimal pro Jahr) mit der in den beiden vorangegangenen Abschnitten vorgestellten Testmethodik Tests von Anti-Malware-Software durchgeführt. Wenn es sinnvoll erscheint, andere als die standardmäßigen Verhaltensweisen der Produkte zu testen, werden aber auch andere, spezielle Tests durchgeführt. Diese Tests haben das Testen bestimmter Qualitätsmerkmale zum Ziel, die mit den regelmäßigen Tests gar nicht - oder nur in geringem Maße - betrachtet werden.

4.3.1 Periodische Tests der Erkennung

Das in diesem Kapitel vorgestellte VTC-Testverfahren beschreibt die periodisch durchgeführten VTC-Tests quantitativer Qualitätskriterien von Anti-Malware-Produkten im *On-Demand*-Modus. Die Tests werden halbjährlich durchgeführt, eine höhere Frequenz ist aufgrund der Kapazitäten nicht möglich. Bis zum Test 2000-04 wurden stets sämtliche vorhandenen Datenbanken getestet, seit Test 2000-08 werden nur noch einmal pro Jahr alle Datenbanken des VTC getestet. Der zweite Test im Jahr spezialisiert sich auf Makro- und Skriptmalware. Dieses Vorgehen ist vom VTC-Team aufgrund von kapazitären Problemen durch stetig wachsende Datenbanken und damit wachsendem Testaufwand beschlossen worden (siehe Abschnitt 6.1).

Die kontinuierliche Durchführung von Tests mit der gleichen Methodik ermöglicht die Analyse der Entwicklung des Verhaltens der getesteten Produkte und der Erkennung über die Zeit. Durch konstante Anwendung der gleichen Testprinzipien und eines im wesentlichen gleich gebliebenen Testverfahrens können die Testergebnisse einzelner Produkte oder der Gesamtdurchschnitt der Erkennung pro Datenbank verglichen werden (vgl. auch Abschnitte 6.4 und 7.3).

4.3.2 Antivirus Repairtest (ART-Test)

Der Antivirus-Repairtest (abgekürzt ART-Test, [RetschTode 2000]) ist im Rahmen des VTC-Projektes entstanden und wurde von Martin Retsch und Stefan Tode 2000 durchgeführt und in ihrer Diplomarbeit beschrieben. Beim ART-Test wird nicht nur die Erkennung von Viren getestet, sondern insbesondere die Entfernung des maliziösen Codes aus den befallenen Dateien. Dieses ist eine äußerst sinnvolle Untersuchung, da ein Benutzer häufig wertvolle Daten in von Malware befallenen Dateien hat, auf die er - trotz zum Beispiel Virusbefall - weiter zugreifen möchte. Man denke nur an eine Firma, die wichtige Unternehmens-Daten (z.B. Umsätze, Preis-Kalkulationen, Gehälter oder kalkulatorische Abschreibungen) in einer Excel-Tabelle kalkuliert. Wird diese Datei wird von einem Excel-Makro-Virus infiziert, so möchte die Firma die Daten trotzdem weiter nutzen können. Eine Reparatur und Entfernung des Virus durch Anti-Malware-Software ist also wünschenswert.

In ihrer Diplomarbeit ([RetschTode 2000]) beschreiben Retsch und Tode ausführlich die von Ihnen aufgestellten Kriterien und das verwendete Testverfahren. Getestet wurde die Reparatur von mit Makro-Viren infizierten Microsoft-Office Dokumenten. Automatisch werden eine Infektion, die Reparatur und die Bewertung der Reparatur von Office-Dokumenten mittels einer virtuellen Maschine gesteuert. Die Kriterien für die Reparatur von infizierten Dokumenten wurden folgendermaßen festgelegt (vgl. [VTC 2000-11]):

- Das Dokument ist nicht mehr infiziert
- Gereinigte Dokumente enthalten nur die vor der Infektion vorhandenen Makros
- Gereinigte Dokumente können geöffnet und gespeichert werden
- Der Visual-Basic-Editor[61] kann geöffnet werden
- Makros im gereinigten Dokument funktionieren
- Keine Warnmeldungen erscheinen bei der Bearbeitung eines gereinigten Dokumentes
- Word-Dokumente sind nach der Reinigung keine Vorlagen
- Die Makro-Warnung von Microsoft-Office erscheint nicht beim Öffnen eines Dokumentes

Eine Übersicht über die Testergebnisse und die Bewertung der Produkte findet sich unter [VTC 2000-11].

[61] zum Erstellen und Bearbeiten von Makros unter Microsoft Office

4.3.3 Heureka-Test

Bei den standardmäßig durchgeführten Tests von Anti-Malware Produkten im VTC werden die Datenbanken immer vor der Einsendung der Produkte "eingefroren"[62], das bedeutet, in den Datenbanken des Virus Test Centers befinden sich für den jeweiligen Test nur Viren, die vor der Einsendung der Testprodukte beim VTC eingegangen sind (vgl. 4.2.2.1, Virenkollektionen zusammentragen). Die Hersteller haben in der Regel nach dem *Einfrieren* der Datenbanken mehrere Wochen Zeit, ihre Produkte mit den aktuellen Engines und Signaturen einzuschicken. Daher werden bei den regulären VTC Tests die in 4.1.4 genannten Kriterien bezogen auf Datenbanken mit bereits bekannten Viren getestet.

Viele Benutzer (und auch Sicherheitsadministratoren in Firmen) führen aber nicht regelmäßig Updates ihrer Antivirenprodukte durch. Dies kann viele Gründe haben, einige davon sind:

- Nachlässigkeit der Benutzer
- falsche Einstellungen in der Software
- Lizenzen, die nur eine bestimmte Anzahl an Signaturupdates beinhalten
- Unwissenheit über die Wichtigkeit von regelmäßigen Updates
- unbemerkte Fehlfunktion der Software (etwa durch Absturz)

Um die Wichtigkeit von regelmäßigen Updates zu untersuchen, erschien es dem VTC-Team sinnvoll, die Testkriterien an Produkten mit veralteten Engines und Signaturen zu testen. Da die meisten Viren, bevor sie als ITW eingestuft werden, bereits bekannt sind, ist auch ohne aktuelles Update eine gute bis sehr gute ITW-Erkennung vorstellbar, zumindest, wenn das entsprechende Produkt zum Zeitpunkt des Signaturupdates eine gute Zoo-Erkennung besaß.

Der Test verläuft prinzipiell genauso, wie die in Abschnitt 4.2.1 beschriebenen regelmäßig durchgeführten Tests quantitativer Kriterien im On-Demand Modus. Die Unterschiede sind, daß die Produkte und deren Signaturen vom letzten durchgeführten Test genommen werden, obwohl die Malware-Datenbanken seither aktualisiert wurden. Außerdem werden nur Makro- und Skriptviren getestet. Insgesamt wird auf zwei mal 6 Datenbanken getestet, nämlich die Datenbanken Makro-Zoo, Macro-Itw, Macro-Mal, Skript-Zoo, Skript-Itw und Skript-Mal jeweils auf dem Stand 3 Monate und 6 Monate nach dem *Einfrieren* der Datenbanken beim letzten Test. Desweiteren wird der Heureka-Test aus pragmatischen und Kapazitätsgründen nur auf Windows NT durchgeführt.

Die Unterschiede zu den sonst durchgeführten VTC-Tests sind also:

[62] Mit „Einfrieren" ist die Festlegung eines Datums gemeint, bis zu welchem die Datenbanken eines Tests erhaltene Malware enthalten. An diesem Datum werden die Datenbanken „eingefroren" und danach neu aufgetauchte Malware oder zugesandte Objekte alter Malwaretypen werden nicht mehr in die Datenbanken für den Test integriert (vgl. 1.2).

- Produkte mit alten Signaturen vom vorherigen Test aber Datenbanken mit neu aufgetauchter Malware bis jeweils drei und sechs Monate nach Datenbankstand des vorherigen Tests
- Test nur auf den Datenbanken zu Skript und Makro
- Test nur unter WindowsNT

Die Benamung Heureka[63]-Test ist deshalb gewählt, weil im eigentlichen Sinne nicht die Heuristik der Scanner getestet wird, obwohl die Erkennung von Malware ohne aktuelle Signatur im Mittelpunkt des Tests steht. Die Heuristik eines Scanners kann durch entsprechende Optionen auf Maximum gesetzt werden. Diese Optionen werden aber beim Heureka-Testverfahren nicht zwangsläufig gesetzt. Stattdessen werden die gleichen - von den Herstellern angegebenen - Optionen je Produkt wie beim vorherigen, "normalen" VTC-Test genommen. Diese Optionen sind auf eine maximale On-Demand Erkennung durch Signaturen optimiert.[64] So können die Testergebnisse und die Abnahme der Erkennung im Heureka-Test mit den Testergebnissen des vorangegangenen VTC-Test verglichen werden.

Bisher wurden zwei Heureka-Tests mit identischem Verfahren (dadurch wird ein langfristiger Vergleich der Ergebnisse möglich) veröffentlicht, und zwar Heureka-I im Juli 2001 und Heureka-II im März 2002. Zur näheren Beschreibung und den erzielten Ergebnissen siehe [VTC 2001-07] und [VTC 2002-03].

4.3.4 Überblick über durchgeführte Tests

Zum Abschluß dieses Kapitels soll ein Überblick über die in der Vergangenheit durchgeführten Tests von Anti-Malware-Software gegeben werden. Tabelle 4.I zeigt die durchgeführten Tests chronologisch und gibt an, was für ein Test (entsprechend den Abschnitten 4.2.1 bis 4.2.3) jeweils durchgeführt wurde, ergänzt durch Bemerkungen[65]. Tabelle 4.J zeigt für die durchgeführten Tests im Detail, welche Datenbanken als Testmenge benutzt wurden[66]. Es wird deutlich, wie die Zahl der Datenbanken im Laufe der Jahre gewachsen ist. Dies liegt an neuen Arten von Malware (z.B. Skriptviren) und an einer besseren Differenzierung der Malware in den Datenbanken des VTC (z.B. Trennung nach Makro-Malware und File-Malware). In beiden Tabellen sind die Tests ab 1997 aufgelistet, da erst ab dieser Zeit regelmäßig im Internet Testberichte des VTC veröffentlicht werden. Vor 1997 wurden allerdings auch Tests von Anti-Malware-Produkten im VTC durchgeführt und veröffentlicht.

---

[63] griechisch "Heureka" bedeutet: "ich habe gefunden"
[64] Normalerweise kann ein Scanner ohne Signatur einen Virus nur durch eine gute Heuristik erkennen.
[65] Außer den aufgeführten Tests wurden mehrere "kleinere" Tests durchgeführt:
Test für „Computer Bild” (1998-06), Test für „c’t”, Test für „Chip”
[66] In Tabelle 4.I ist der Test 2000-11 (ART-Test) nicht berücksichtigt, da es sich um einen Sondertest handelt.

Test	Bezeichnung	Test-Art	Bemerkungen
2002-03	"Heureka-2" Scanner test March 2002	Heureka (Macro/Script)	Macro/Skript 3 und 6 Monate
2001-10	Scanner test October 2001	Macro/Script	
2001-07	"Heureka(-1)" Scanner test July 2001	Heureka (Macro/Script)	Macro/Skript 3 und 6 Monate
2001-04	Scanner test April 2001	File/Macro/Script/Boot	erstmals Linux
2000-11	AntiVirus Repair Test (ART 2000-11)	ART-Test (Macro)	Spezieller Test der Reparatur inf. Dateien
2000-08	Scanner test August 2000	Macro/Script	erstmals Windows 2000
2000-04	Scanner test April 2000	File/Macro/Boot	
1999-09	Scanner test September 1999	File/Macro/Boot	
1999-03	Scanner test March 1999	File/Macro/Boot	erstmals Poly und VKit
1998-10	Scanner test October 1998	File/Macro/Boot	
1998-02	Scanner test February 1998	File/Macro/Boot	
1997-07	Scanner test July 1997	File/Macro/Boot	erstmals Windows 95 und Windows NT
1997-02	Scanner test February 1997	File/Macro/Boot	nur DOS

Tabelle 4.I: Chronologie von VTC Tests ab 1997

Test	OS	Datenbanken																
		Boot		***File***						***Macro***				***Skript***				***Exot***
		ITW	***Zoo***	***ITW***	***Zoo***	***Mal***	***Pack***	***Poly***	***VKit***	***ITW***	***Zoo***	***Mal***	***Pack***	***ITW***	***Zoo***	***Mal***	***Pack***	
1997-02	DOS	■	■	■	■					■	■	■						
	W95																	
	WNT																	
	W2K																	
	LIN																	
1997-07	DOS	■	■	■	■	■				■	■	■						
	W95			■	■					■	■							
	WNT			■	■					■	■							
	W2K																	
	LIN																	
1998-02	DOS	■	■	■	■	■	■			■	■	■	■					
	W95			■	■	■	■			■	■	■	■					
	WNT			■	■	■	■			■	■	■	■					
	W2K																	
	LIN																	
1998-10	DOS	■	■	■	■	■	■			■	■	■	■					
	W95			■	■	■	■			■	■	■	■					
	WNT			■	■	■	■			■	■	■	■					
	W2K																	
	LIN																	
1999-03	DOS	■	■	■	■	■	■	■	■	■	■	■	■					
	W98			■	■	■	■	■	■	■	■	■	■					
	WNT			■	■	■	■	■	■	■	■	■	■					
	W2K																	
	LIN																	
1999-09	DOS	■	■	■	■	■	■	■	■	■	■	■	■					
	W98			■	■	■	■	■	■	■	■	■	■					
	WNT			■	■	■	■	■	■	■	■	■	■					
	W2K																	
	LIN																	
2000-04	DOS	■	■	■	■	■	■	■	■	■	■	■	■					
	W98			■	■	■	■	■	■	■	■	■	■					
	WNT			■	■	■	■	■	■	■	■	■	■					
	W2K																	
	LIN																	
2000-08	DOS									■	■	■	■		■			
	W98									■	■	■	■		■			
	WNT									■	■	■	■		■			
	W2K									■	■	■	■		■			
	LIN																	
2001-04	DOS	■	■	■	■	■	■	■	■	■	■	■	■	■	■			■
	W98			■	■	■	■	■	■	■	■	■	■	■	■			■
	WNT			■	■	■	■	■	■	■	■	■	■	■	■			■
	W2K			■	■	■	■	■	■	■	■	■	■	■	■			■
	LIN			■	■	■	■	■	■	■	■	■	■	■	■			■
2001-10	DOS									■	■	■	■	■	■	■		
	W98									■	■	■	■	■	■	■		
	WNT									■	■	■	■	■	■	■		
	W2K									■	■	■	■	■	■	■		
	LIN									■	■	■	■	■	■	■		

Tabelle 4.J: Getestete Datenbanken in VTC-Tests ab 1997 (Legende auf der folgenden Seite)

Legende zu Tabelle 4.J:

OS:	Betriebssystem
W95:	Windows 95
W98:	Windows 98
WNT:	Windows NT
W2K:	Windows 2000
LIN:	Linux
ITW:	"in-the-wild"-Viren
Zoo:	Zoo-Viren
Exot:	exotische Viren