6. Erweiterungen und Verbesserungen für das Testverfahren des VTC

Nachdem in Kapitel 4 das Testverfahren des VTC ausführlich dargestellt wurde, sollen in diesem Kapitel einige Verbesserungsvorschläge und Erweiterungen für das Testverfahren aufgezeigt werden. Nach einer Betrachtung der Probleme des VTC-Testverfahrens (Abschnitt 6.1) werden mehrere Verbesserungen im Zusammenhang mit der Erstellung und Veröffentlichung des Testberichtes beschrieben. Der Ausgangspunkt ist dabei die Integration von Testergebnisdaten in eine Datenbank, mit der die Daten ausgewertet werden können (Abschnitt 6.2). Mithilfe dieser Datenbank können VTC-Testergebnisse auch grafisch dargestellt werden (Abschnitt 6.3). Diese Lösung verbessert allerdings nicht nur die Erstellung des Testberichtes, sondern ermöglicht neue Formen der Ergebnisanalyse. Abschnitt 6.4 befaßt sich in diesem Zusammenhang mit den Möglichkeiten der langfristigen Auswertung von Testergebnissen mehrerer Tests. Abschnitt 6.5 und 6.6 erläutern die automatische Bewertung der Testergebnisse und die automatische Erstellung von Teilen des Testberichtes.

6.1 Probleme und Schwächen des VTC-Verfahrens

Die Probleme beim vorgestellten Testverfahren des VTC (siehe Kapitel 4) lassen sich in unterschiedliche Kategorien unterteilen:

- Kapazitätsprobleme
- Organisatorische Probleme

Die beiden folgenden Abschnitte (6.1.1 und 6.1.2) widmen sich jeweils einer der beiden Kategorien von Problemen beim VTC-Verfahren. Mit Problemen sind verfahrensinterne Schwierigkeiten gemeint. Abschnitt 6.1.3 betrachtet die generellen Schwachpunkte des VTC-Testverfahrens im Vergleich zu anderen Tests von Anti-Malware-Software.

6.1.1 Kapazitätsprobleme

Das Fachgebiet IT-Sicherheit ist sehr schnellebig. Laufend tauchen neue Arten von Bedrohungen auf, und täglich werden neue Varianten bereits bekannter Bedrohungen entdeckt. So wie sich Anti-Malware-Produkte diesen sich ständig ändernden Bedrohungen anpassen müssen, so befindet sich auch ein Test von Anti-Malware-Software in einem ständigen Anpassungsprozeß. Da immer mehr Computerviren entdeckt werden, und die Entdeckung neuer Viren in immer kürzeren Abständen stattfindet, wächst die Zahl der in den VTC-Datenbanken enthaltenen Viren stetig. Diese wachsende Anzahl an Viren bereitet den VTC-Tests zunehmend Probleme.

Die mit jedem Test immer größer werdenden Virendatenbanken verursachen an mehreren Stellen im VTC-Verfahren Kapazitätsprobleme. Erstens wird das Einsortieren neu eintreffender Virenkollektionen und die Qualitätsüberprüfung der erhaltenen Musterdateien eine immer aufwendigere Aufgabe. Zweitens dauert der Scanvorgang zunehmend länger, und immer mehr Produkte stürzen auf den riesigen Datenbanken (insbesondere File-Zoo) ab, da diese Produkte anscheinend nicht in der Lage sind, so große Verzeichnisstrukturen und so viele Dateien zu überprüfen. Drittens wächst der Auswertungsaufwand überproportional zum Zuwachs der Datenbanken, da durch die häufigeren Produktabstürze die Anzahl der Nachscans und damit auch die Anzahl an Auswertungen von Protokollen schneller wächst als die eigentliche Vergrößerung der Datenbanken. Schließlich wird durch die immer häufiger auftretenden Probleme der Testprodukte auf den großen Datenbanken und die von Test zu Test größer werdenden Datenbanken auch die Aufgabe der Anfertigung und Veröffentlichung des Testberichtes immer aufwendiger.

Verstärkt werden die genannten Aspekte dadurch, daß im Rahmen der neuen Studienordnung des Fachbereichs Informatik an der Universität Hamburg[79] den Studenten immer weniger Zeit und Spielraum für intensive Projektarbeit eröffnet wird. Durch diese Entwicklung wird es immer schwieriger, Nachwuchs für das Projektteam zu finden. Und die Projektmitglieder, die nach neuer Studienordnung studieren, haben in der Regel weniger Zeit für das Projekt als Studenten der alten Studienordnung.

Desweiteren kommt hinzu, daß die Rechnerbeschaffung an einer Universität nur schwer mit der rasanten Hardwareentwicklung mithalten kann, wie sie in den letzten Jahren stattgefunden hat. So sind die meisten Rechner im VTC veraltet. Moderne Betriebsysteme lassen sich aufgrund ihrer Hardwareanforderungen nur auf einem Bruchteil der Rechner installieren, wodurch zwangsläufig Kapazitätsengpässe in der Testphase des Projektes entstehen, da sich viele Tester wenige Rechner teilen müssen. Da fast jährlich neue Microsoft Betriebssysteme auf den Markt gebracht werden, ergibt sich außerdem die Schwierigkeit, die Produkte auf den aktuellen und am meisten benutzten Betriebssystemen zu testen. Durch die Vielzahl der getesteten Betriebssysteme ist der Umfang der VTC-Tests in den letzten Jahren zusätzlich extrem stark angewachsen.

Die Gesamtentwicklung lässt sich anschaulich mit einer immer weiter auseinander klaffenden Schere vergleichen, bei der auf der einen Seite das stetig ansteigende Wachstum der Virendatenbanken steht und auf der anderen Seite die geminderte Zeit der Studenten und die Veralterung der eingesetzten Hardware. Zwei Lösungswege werden im VTC-Team verfolgt: einerseits die Einschränkung der Tests, andererseits die Verbesserung der Organisation, um die gegebenen Kapazitäten effizienter zu nutzen.

Folgende Entscheidungen sind während der letzten Tests getroffen worden, die eine Einschränkung bedeuten:

[79] Diplomprüfungsordnung des Fachbereichs Informatik der Universität Hamburg von 1998

- Es wird nur noch einmal im Jahr ein Test mit allen Datenbanken als Testmenge durchgeführt, alternierend dazu wird einmal im Jahr ein Test nur auf den Datenbanken der Makro- und Skriptviren sowie dazugehöriger Malware durchgeführt, so daß weiterhin zwei Tests pro Jahr durchgeführt werden
- Die Datenbanken VKit und Poly werden ab dem nächsten Test nicht mehr mitgetestet
- Das Betriebssystem Windows NT wird nicht mehr getestet (ersetzt durch Windows 2000)

Organisatorisch wurden folgende Maßnahmen ergriffen:

- Der Auswertungsprozeß wurde überarbeitet mit dem Ziel einer zeiteffizienteren und in höherem Maße automatisierten Auswertung (vgl. [Messerschmidt 2002a])
- Eine Betriebssystem-Neuinstallation wird durch das Anlegen von Partitions-Images erheblich beschleunigt
- Es wird versucht, die Kapazitätsauslastung der Rechner zu optimieren, indem die Tester zu unterschiedlichen Zeiten das Labor benutzen

Die vorgestellten Kapazitätsprobleme verlangsamen insbesondere die Testphasen Sortierung der Viren in Datenbanken, die Auswertung und das Erstellen des Testberichtes. In den Abschnitten 6.2 bis 6.6 wird gezeigt, wie sich die Erstellung und Veröffentlichung des Testberichtes mittels einer Ergebnisdatenbank beschleunigen und verbessern läßt. Der Auswertungsprozeß wurde bereits von Michel Messerschmidt grundlegend erneuert und optimiert (vgl. [Messerschmidt 2002a]). Das Einordnen der Viren in die Datenbanken und die Qualitätskontrolle der großen Datenbanken ist ein Teil des Testverfahrens, der in Zukunft ebenfalls grundlegend verbessert werden muß, um der beschriebenen Entwicklung entgegenzuwirken. Insgesamt muß in Zukunft die Kapazitätsausstattung des Virus Test Centers erheblich verbessert werden - sowohl durch Studenten als auch durch Hardware - um mit der schnellen Entwicklung mithalten zu können.

6.1.2 Organisatorische Probleme

Die überwiegende Anzahl der Probleme im VTC sind Kapazitätsprobleme. Dennoch gibt es auch organisatorische Probleme, die den Testablauf negativ beeinflussen. Diese Probleme sind häufig bedingt durch die mangelnde Kommunikation der Teammitglieder untereinander. Die Schnittstelle zwischen Testern und Auswertung ist verbesserungsfähig, so daß unnötige Nachscans vermieden werden können[80].

[80] unnötige Nachscans können durch Mißverständnisse zwischen Auswertung und Test der Produkte entstehen, außerdem werden Nachscans oft durch falsche Produkteinstellungen notwendig, die bei besserem Informationsaustausch der Tester untereinander hätten vermieden werden können

Durch die neue Studienordnung am Fachbereich Informatik, die eine kürzere Studienzeit der Studenten bewirkt[81], sind die Studenten kürzere Zeit Mitglieder im VTC-Team als noch vor einigen Semestern. Dadurch herrscht eine höhere Fluktuation der Testmitglieder, die folgende Probleme mit sich bringt:

- weniger Zeit für die Einarbeitung in das VTC-Verfahren
- größere Anzahl an Projektmitgliedern mit jeweils geringerer Zeit, dadurch höhere Anforderungen an den Informationsfluß zwischen den Projektmitgliedern
- größere Notwendigkeit zur Dokumentation von Verfahren und Wissen

Diese organisatorischen Probleme werden durch die angesprochenen Kapazitätsprobleme noch verstärkt, da der zu dokumentierende Testumfang stetig wächst. Den organisatorischen Problemen kann nur durch eine Strukturierung des Testverfahrens und insbesondere durch eine Verbesserung des Informationsflusses im Projekt begegnet werden. Deshalb wurden einige Verfahren, beispielsweise die Verzeichnisstruktur zur Übergabe der Reportdateien von den Testern an die Auswertung, überarbeitet und um Verzeichnisse zur Dokumentation der Produktinstallation erweitert. Die Übermittlung produktspezifischer Informationen (zum Beispiel spezielle Installationsvorgehensweisen, Einstellungen, usw.) ist um so wichtiger, je mehr Personen sich das Testen und Auswerten der Produkte teilen. Diese strukturierte Speicherung spezifischen Wissens verringert die genannten organisatorischen Probleme, stellt aber auch einen höheren Aufwand für die Testteilnehmer dar, der nicht direkt in die Bearbeitung des Tests eingeht, sondern sich erst langfristig bezahlt macht.

6.1.3 Schwächen des VTC-Verfahrens

Vergleicht man das VTC-Verfahren mit anderen Tests von Anti-Malware-Software, so fällt auf, daß das Testverfahren des VTC einzig auf das Testen von quantitativen Qualitätskriterien fokussiert ist und diese Kriterien nur im *On-Demand*-Modus getestet werden. Eine solche Fokussierung ist einerseits sinnvoll und notwendig, da bereits das ausführliche und wissenschaftliche Testen dieser Kriterien die Kapazitäten des VTC voll ausschöpft. Andererseits stellt sich die Frage, ob nicht ein Test im *On-Access*-Modus sinnvoll wäre, weil der Großteil der Anwender heutzutage Anti-Malware-Software in diesem Modus einsetzt. Die Möglichkeit eines solchen Testes wird derzeit im VTC untersucht (vgl. [Siekierski 2002]).

Außerdem besteht die Möglichkeit, die Kapazitäten des VTC-Teams anders zu verteilen, indem andere Kriterien getestet werden und dafür die bisher getesteten Kriterien nicht so intensiv. So ist beispielsweise vorstellbar, nur noch einmal pro Jahr (statt zweimal pro Jahr) einen Test der bisher benutzten Kriterien durchzuführen. Die dadurch gewonnenen Kapazitäten könnten für einen neuartigen Test bisher nicht betrachteter Qualitätskriterien von Anti-Malware-Software eingesetzt werden, so zum Beispiel der Geschwindigkeit. Auch ein

[81] sowohl durch eine verringerte Regelstudienzeit als auch durch die Zwangsexmatrikulation nach 6 Semestern ohne Vordiplom

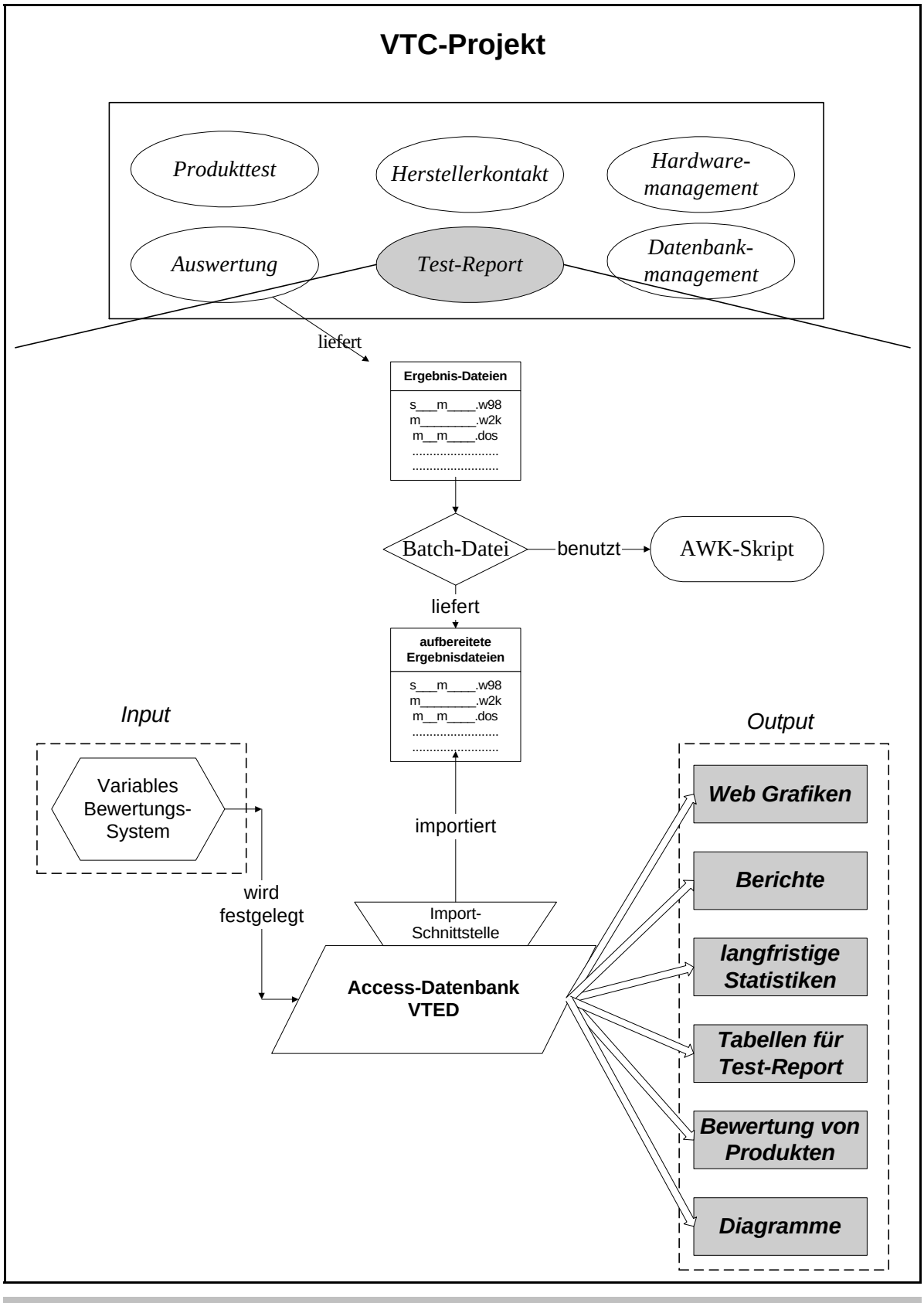

Abbildung 6.A: Integration der Datenbank VTED in die Erstellung des Testberichtes

Test der reinen Heuristikfunktion der Produkte (also ohne oder mit extrem veralteter Signatur

und mit speziellen Heuristikeinstellungen) wäre denkbar.

6.2 Die Datenbank VTED und erweiterte Möglichkeiten der Ergebnisanalyse

Die Datenbank VTED (Virus Test Ergebnis Datenbank) wurde von Jens Gallion, Thomas Buck und Jan Seedorf im Rahmen des VTC-Projektes entwickelt. Ursprüngliches Ziel der Datenbank war, die Ergebnisse der Tests zu importieren und daraus Grafiken zu erstellen. Die Grafiken sollten die Testergebnisse auf der Webseite des VTC visualisieren. Seit Test 2000-04 werden die Datenbank VTED zum Erstellen von Ergebnisgrafiken benutzt und die Grafiken auf der Webseite als Teil des Testberichtes veröffentlicht.

6.2.1 Vorteile einer Datenbank zur Ergebnisanalyse

Die Integration der Testergebnisse bietet jedoch viel mehr Möglichkeiten als nur die grafische Darstellung der Ergebnisse. Durch kontinuierlichen Import der Testergebnisse kann eine Datenbank als Speicherstelle aller historischen Testergebnisse dienen. So können alte Testergebnisse zentral verwaltet werden und bestimmte Testergebnisse durch Datenbankabfragen schnell gefunden werden. Durch Abfragen auf den Daten können Analysen vorgenommen und Statistiken erstellt werden. Auch spezielle, langfristige Statistiken, die die Ergebnisse mehrerer Tests beinhalten und vergleichen, können so realisiert werden.

Außerdem kann eine Datenbank mit Testergebnissen hilfreich bei der Erstellung des Testberichtes sein. So können die Produkte anhand ihrer Testergebnisse und festgelegter Kriterien automatisch bewertet werden (siehe Abschnitt 6.5). Teile des Testberichtes, zum Beispiel die integrierten Ergebnistabellen, können automatisch erstellt werden (siehe Abschnitt 6.6.). Auf diese Art kann eine Datenbank die Organisation des Testes verbessern und die Durchführung einzelner Arbeitsschritte beschleunigen.

Zusammenfassend betrachtet hat die Speicherung der Testergebnisse in einer Datenbank folgende Vorteile:

- grafische Darstellung der Ergebnisse
- zentrale Speicherung aller Ergebnisdaten sämtlicher VTC-Tests
- umfangreiche Analysemöglichkeiten der Testergebnisse durch Abfragen
- Möglichkeit des Vergleichs der Ergebnisse mehrerer Tests
- automatische Bewertung der Testergebnisse nach festgelegten Kriterien
- automatische Generierung von Teilen des Testberichtes

Abbildung 6.A zeigt die Integration der Datenbank VTED in das VTC-Projekt und die Erstellung des Testberichtes. Die Auswertung liefert Ergebnisdateien, die nach einer

Anpassung durch ein AWK-Skript[82] in die Datenbank importiert werden. Die Datenbank erzeugt daraus Web-Grafiken, Berichte, langfristige Statistiken, Tabellen für den Testreport und eine Bewertung von Produkten.

Die Ausgaben der Datenbank werden in den Abschnitten 6.3 bis 6.6 näher erläutert. Der Import der von der Auswertung gelieferten Ergebnisse wird in Abschnitt 6.2.3 dargestellt.

6.2.2 Der Aufbau der Datenbank VTED

Bei der Datenbank VTED handelt es sich um eine relationale Datenbank[83]. Die Datenbank ist in Microsoft Access 97 entworfen und realisiert. Den Kern der Datenbank bilden die folgenden Relationen:

- Ergebnis
- Version
- Betriebssystem
- Testart
- Scanner

Die Datenbank[84] befindet sich in der 2. Normalform, das heißt in allen Relationen sind alle Nichtschlüsselattribute voll funktional abhängig vom Primarschlüssel. Die Datenbank befindet sich nicht in der 3. Normalform, da teilweise transitive Abhängigkeiten zwischen Nichtschlüsselattributen bestehen. Abbildung 6.B zeigt die Beziehungen zwischen den genannten Relationen.

[82] AWK ist eine Skriptsprache, die im VTC zur Auswertung der Reportdateien benutzt wird (vg. Abschnitt 4.2.2.5)

[83] Es wird bei der Beschreibung der Datenbank davon ausgegangen, daß der Leser mit grundlegenden Datenbankkonzepten und insbesondere relationalen Datenbanken vertraut ist. Eine Einführung in Datenbankmodelle und das relationale Datenbankmodell findet sich in [Informatik-Duden 1993] (157ff), [Vossen 1999] (Kapitel 4 - 6) und [Stahlknecht 1995] (S 189-212).

[84] In diesem Kapitel wird die Bezeichnung Datenbank sowohl (im eigentlichen Sinne) für die relationale Datenbank VTED, als auch für die sogenannten Datenbanken des VTC, bei denen es sich um in Verzeichnisse sortierte Objekte bösartiger Software handelt, verwendet. Dem Leser sollte jedoch klar sein, wann mit Datenbank die relationale Datenbank VTED oder eine Datenbank als Sammlung von Malware zum Testen gemeint ist.

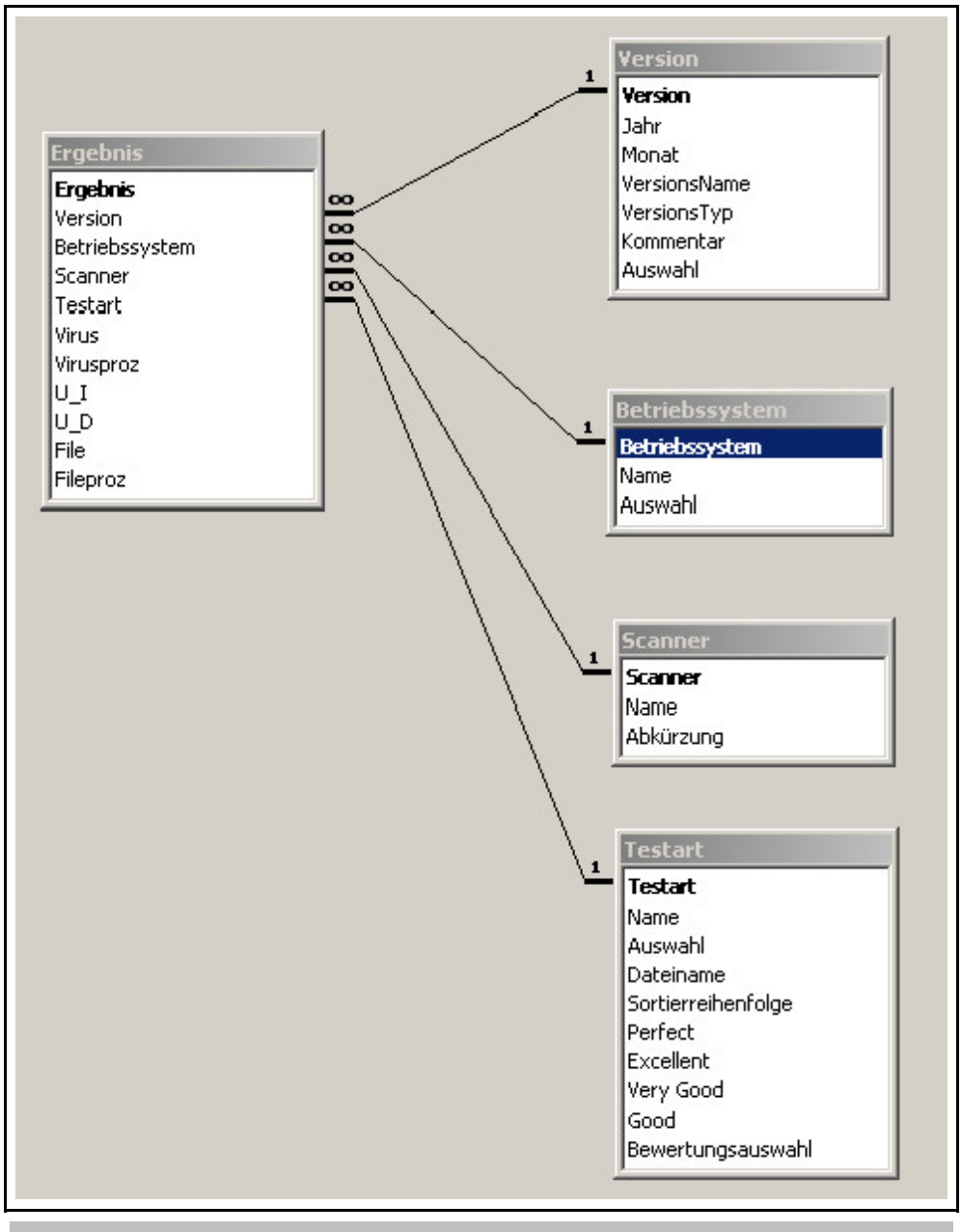

Abbildung 6.B: Beziehungen in der Datenbank VTED

In der Relation *Ergebnis* werden die importierten Testergebnisse gespeichert. Für *Version*, *Betriebssystem*, *Scanner* und *Testart* werden in der Relation *Ergebnis* jeweils nur datenbankinterne, numerische Werte als Fremdschlüssel gespeichert. Über diese Fremdschlüssel stehen die anderen Relationen in 1:n Beziehungen zur Relation *Ergebnis*. Das bedeutet, daß die Fremdschlüsselwerte in der Relation *Ergebnis* mehrfach vorkommen können, jeder Wert ist aber nur genau einmal in der entsprechend über den Fremdschlüssel verknüpften Relation definiert (durch Attributausprägungen in den Relationen). Den Werten der Fremdschlüssel sind in den einzelnen Relationen *Version*, *Betriebssystem*, *Scanner* und *Testart* andere Attribute (z. B. Name) zugeordnet.

In der Relation *Ergebnis* haben die gespeicherten Attributwerte folgende Bedeutung:

Attribut	***Bedeutung***
Ergebnis	fortlaufender Zahlenwert
Version	Fremdschlüsselwert für Testversion
Betriebssystem	Fremdschlüsselwert für Betriebssystem
Scanner	Fremdschlüsselwert für Scanner
Testart	Fremdschlüsselwert für Datenbank
Virus	erkannte Viren in Datenbank
Virusproz	erkannte Viren in Datenbank in Prozent
U_I	Anzahl Objekte "unreliably identified"
U_D	Anzahl Objekte "unreliably detected"
File	erkannte Objekte in Datenbank
Fileproz	erkannte Objekte in Datenbank in Prozent

Tabelle 6.C: Bedeutung der Attribute in der Tabelle Ergebnis

In der Relation *Version* sind die durchgeführten VTC-Tests[85] jeweils als eine Testversion aufgeführt. In den Relationen *Betriebssystem*, *Scanner* und *Testart* sind entsprechend alle Betriebssysteme, Scanner und Datenbanken aufgeführt. In der Datenbank wird eine Testversion aus der Relation *Version* als sogenannte "*aktuelle Testversion*" markiert. Dies ist sinnvoll, da der Hauptnutzen der Datenbank in der Bearbeitung der Testergebnisse des jeweils aktuellen Tests besteht. Durch die Markierung der aktuellen Version als *aktuelle Testversion* können sich viele Datenbankabfragen automatisch auf die Ergebnisse des so markierten Tests beziehen. Sollen Ergebnisse eines alten Tests betrachtet oder durch Abfragen ausgewertet werden, so ist dies durch Änderung der Markierung *aktuelle Testversion* einfach möglich. Weitere Angaben zur Benutzung und Wartung der Datenbank VTED finden sich in Anhang B.

6.2.3 Der Datenimport

Der Import von Testergebnissen in die Datenbank VTED erfolgt in zwei Schritten. Als erstes werden die von der Auswertung erzeugten Ergebnistabellen (pro Kombination aus Betriebssystem und Datenbank je eine Tabelle) so angepasst, daß sie in die Datenbank importiert werden können. Im zweiten Schritt werden die veränderten Ergebnistabellen dann in die Datenbank importiert.

Die Ergebnistabellen der Auswertung liegen als ASCII-Textdateien vor. In den Dateien sind die einzelnen Werte durch Leerzeichen oder Tab getrennt. Um in die Datenbank importiert werden zu können, werden die Werte der Ergebnistabellen durch ein AWK-Skript in eine

[85] (1997-02 bis 2002-03)

neue Datei geschrieben, in der die einzelnen Werte durch Semikola getrennt sind. Außerdem werden Zeilen, die keine Ergebniswerte, sondern nur Überschriften oder Kommentare enthalten, bei der Anpassung unterdrückt.

Abbildung 6.D zeigt den Anfang einer Ergebnistabelle vor der Anpassung, Abbildung 6.E zeigt die Tabelle nach der Anpassung. Über eine Batchdatei können so unter Benutzung eines AWK-Skriptes alle Ergebnisdateien eines Tests automatisch für den Datenimport angepaßt werden.

```
                          This  includes
           Viruses     ---- unreliably ----     Files
Scanner    detected    identified   detected   detected
-----------------------------------------------------------
Testbed   17561 100.0%         %           %  132576 100.0%
-----------------------------------------------------------
ANT       15310  87.2   690   3.9  1242   7.1 113426  85.6
AVA       17107  97.4   631   3.6   134   0.8 129634  97.8
AVG       15283  87.0   521   3.0   322   1.8 118702  89.5
...
```

Abbildung 6.D: Ergebnistabelle vor Anpassung an Datenimport

```
ANT;15310;87.2;690;3.9;1242;7.1;113426;85.6
AVA;17107;97.4;631;3.6;134;0.8;129634;97.8
AVG;15283;87.0;521;3.0;322;1.8;118702;89.5
...
```

Abbildung 6.E: Angepaßte Ergebnistabelle mit Werten durch Semikola getrennt

Die so angepaßten Ergebnisse können in die Datenbank importiert werden. Abbildung 6.F zeigt die Benutzeroberfläche zum Importieren von Testergebnissen in die Datenbank[86]. Dabei ist zu beachten, daß vom Benutzer die richtigen Parameter ausgewählt werden, damit die importierten Testergebnisse in der Datenbank richtig zugeordnet werden. Der Benutzer gibt an, in welchem Pfad (unterstes Feld) sich welche Testergebnisse befinden (obere drei Felder). Diese importierten Daten werden dann entsprechend den Angaben in der Datenbank in der Relation *Ergebnis* gespeichert.

[86] Der zugehörige Quellcode zum Datenimport ist in Anhang B zu finden (Abschnitt B.3).

Abbildung 6.F: Die Benutzeroberfläche beim Datenimport

6.3 Grafische Darstellung der Ergebnisse

Die Datenbank VTED bietet dem Benutzer die Möglichkeit, die enthaltenen Testergebnisse zu visualisieren. Dabei können Testversion, Betriebssystem und Datenbanken ausgewählt werden, und es können sämtliche, in die Datenbank importierte Testergebnisse benutzt werden.

Grundsätzlich stehen dem Benutzer drei verschiedene Auswahlmöglichkeiten für die Erstellung von Grafiken zur Verfügung:

- Ein Betriebssystem und eine Datenbank
- Ein Betriebssystem und mehrere Datenbanken
- Langfristige Statistiken über mehrere Testversionen

Bei den ersten beiden Auswahlmöglichkeiten werden immer die Ergebnisse der aktuellen Testversion angezeigt, bei der dritten Option können mehrere Testversionen ausgewählt werden. Abbildung 6.G zeigt die Auswahlmaske zur Erstellung einer Grafik der Ergebnisse eines Betriebssystems und mehrerer Datenbanken für den in der Datenbank als aktuelle Testversion markierten Test. Der Benutzer wählt ein Betriebsystem und die gewünschten

Datenbanken aus und klickt rechts auf die Schaltfläche "*Diagramm erstellen*". Abbildung 6.H zeigt die erzeugte Grafik bei entsprechender Auswahl in Abbildung 6.G. Die erzeugten Grafiken können gespeichert werden. Dabei wird zum gewählten Dateinamen automatisch die datenbankinterne Nummer von Betriebssystem und Testversion angehängt, damit am Dateinamen eindeutig die darin enthaltenen Testergebnisse abgelesen werden können. Die Auswahl für Grafiken mit nur einer Datenbank verläuft ähnlich, mit dem Unterschied, daß nur eine Datenbank ausgewählt werden kann und nur Ergebnisse dieser Datenbank in der erzeugten Grafik angezeigt werden.

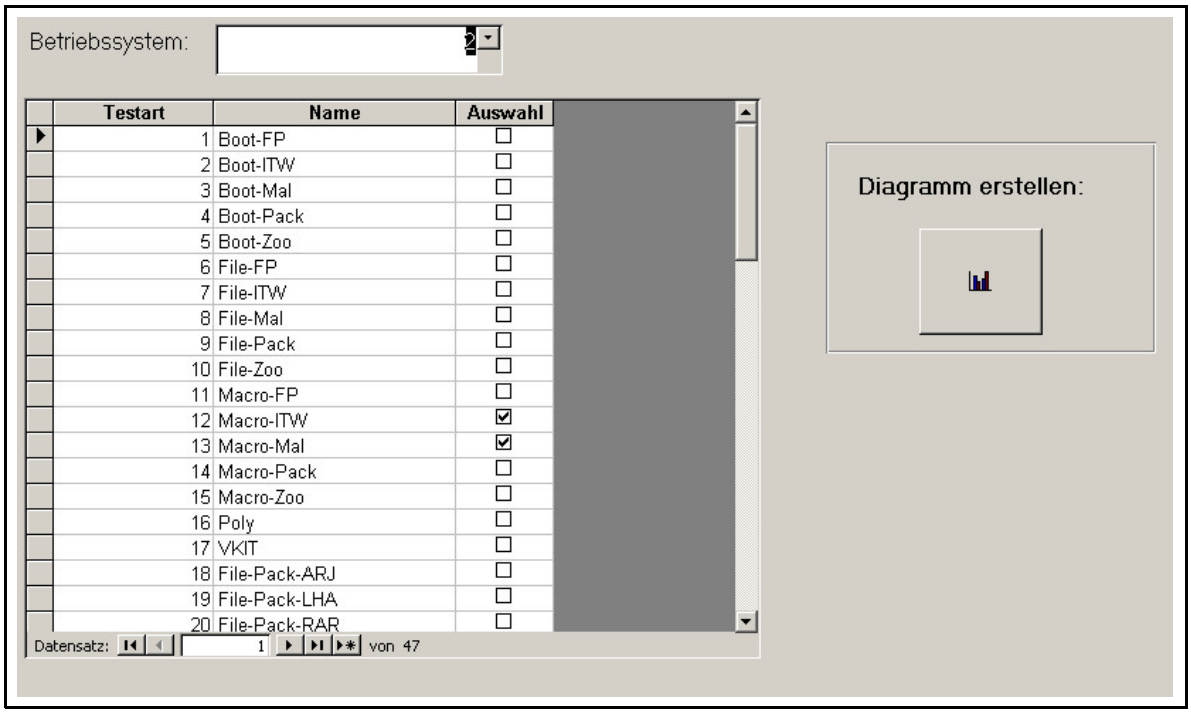

Abbildung 6.G: Auswahl von Betriebssystem und Datenbanken zur Erstellung einer Ergebnisgrafik

Die Benutzeroberfläche zur Auswahl der Daten für langfristige Statistiken ähnelt der in Abbildung 6.G. Zusätzlich können hier noch ein oder mehrere Testversionen ausgewählt werden, und es besteht die Möglichkeit, nicht nur Balken-, sondern auch Liniendiagramme zu erstellen. Ein Beispiel für ein Liniendiagramm findet sich in Abbildung 6.I. Diese Grafik veranschaulicht die Entwicklung der durchschnittlichen Erkennung aller getesteten Produkte für Malware (File-Mal und Macro-Mal) unter Windows NT in den Jahren 1999 bis 2001[87]. Ähnliche Grafiken können für andere Testversionen, Betriebssysteme und Datenbanken erstellt werden. Bei den Grafiken langfristiger Statistiken werden die durchschnittlichen

[87] Die Linien zwischen den jeweiligen Durchschnittswerten der Tests approximieren linear die Entwicklung der durchschnittlichen Erkennungsrate im Zeitraum zwischen den entsprechenden VTC-Tests.

Ergebnisse pro Testversion angezeigt. Es ist aber auch denkbar, das jeweilige Maximum oder Minimum auszugeben.

Die grafische Darstellung der Testergebnisse ermöglicht den Lesern des Testberichtes und den Besuchern der VTC-Webseite einen schnellen Überblick über das Abschneiden der getesteten Produkte. Außerdem eignen sich die Ergebnisgrafiken zur Darstellung der VTC Testergebnisse auf Messen oder Ausstellungen. Durch die Datenbank VTED und die beschriebenen Auswahlmöglichkeiten können nach dem Import der Testergebnisse in die Datenbank beliebige Grafiken flexibel und automatisch erstellt werden. Ein Teil des Quellcodes der Datenbank VTED zum Export von Grafiken ist in Anhang B abgebildet.

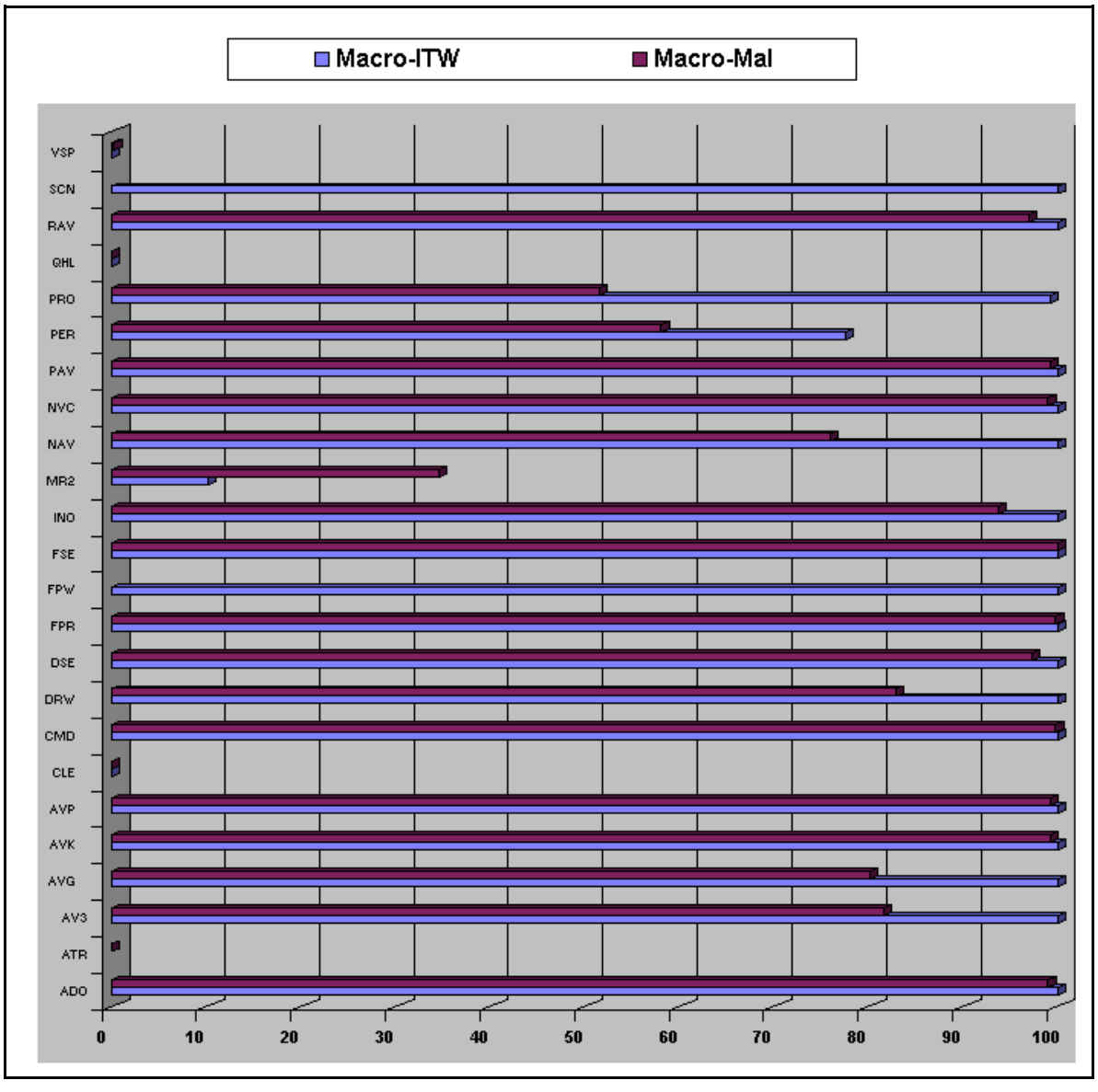

Abbildung 6.H: Grafische Darstellung der Ergebnisse mehrerer Datenbanken eines Betriebssystems

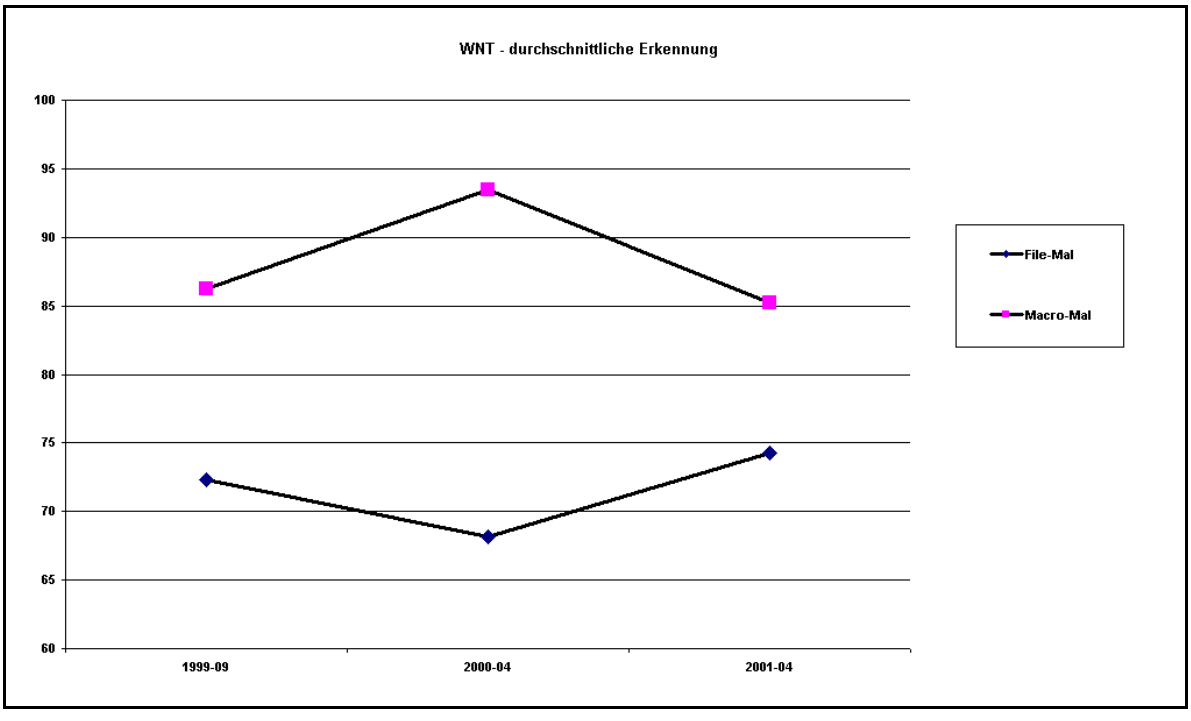

Abbildung 6.I: Beispiel einer langfristigen Ergebnisgrafik (durchschnittliche Erkennungsrate von Malware in Prozent unter Windows NT 1999-2001)

6.4 Untersuchung der langfristigen Entwicklung der Erkennung

Im Rahmen des Testberichtes werden nur die jeweiligen Ergebnisse eines Tests veröffentlicht und untersucht. Einzige Ausnahme sind die Zoo-Datenbanken (File-Zoo, Makro-Zoo und Skript-Zoo), für die auf Ebene 2 des Testberichtes für jedes Betriebssystem eine langfristige Tabelle mit den Ergebnissen aller getesteten Produkte für den aktuellen und die vergangenen Tests aufgelistet wird.

Durch die Integration aller historischen Testergebnisse in der Datenbank VTED können mittels Datenbankabfragen und Selektionen beliebige langfristige Statistiken aller Datenbanken, Betriebssysteme und getesteten Produkte programmiert werden. So kann zum Beispiel die Entwicklung der durchschnittlichen Erkennung aller Produkte auf einer oder mehreren Datenbanken untersucht werden. Oder es kann die Entwicklung der getesteten Qualitätskriterien eines bestimmten Produktes verfolgt werden. Diese Abfragen lassen sich sowohl als Tabelle als auch als Grafiken (vgl. 6.3) darstellen und veröffentlichen.

Folgende Arten von langfristigen Statistiken können zur Analyse der Testergebnisse über mehrere Tests benutzt werden und mit Hilfe der Datenbank VTED erstellt werden:

- Abweichung der Erkennung unter den einzelnen Windows-32-Systemen pro Test[88]
- automatische Erstellung der Tabellen der historischen Zoo-Erkennung der getesteten Produkte für den Testreport
- Erstellung von Tabellen der historischen Entwicklung der Erkennung beliebiger Datenbanken (für alle Produkte)
- Verlauf der durchschnittlichen Erkennung von Malware, komprimierter Malware oder anderer Datenbanken über die Tests
- Schwankungen der Erkennungsrate pro Produkt über mehrere Tests pro Datenbank (welcher Scanner zeigt über mehrere Tests konstant gute Ergebnisse)

Betriebssystem. Name	Testart. Name	Abkürzung	1999-03	2000-04	2001-04	Mittelwert von Virusproz
DOS	File-Mal	AVA	66,50	56,50	51,00	58,00
DOS	File-Mal	AVG	65,30		50,60	57,95
DOS	File-Mal	AVK	94,80		96,00	95,40
DOS	File-Mal	AVP	88,30	83,20	96,40	89,30
DOS	File-Mal	CMD		92,80		92,80
DOS	File-Mal	DRW	74,60			74,60
DOS	File-Mal	DSS	97,50			97,50
DOS	File-Mal	FPR	89,20	95,30	94,00	92,83
DOS	File-Mal	FSE	88,70	94,60		91,65
DOS	File-Mal	INO	75,30	74,70	47,90	65,97
DOS	File-Mal	IRS	43,50			43,50
DOS	File-Mal	itm	44,80			44,80
DOS	File-Mal	MR2			41,90	41,90
DOS	File-Mal	NAV	76,80	73,40	45,10	65,10
DOS	File-Mal	NOD	63,40	77,60		70,50
DOS	File-Mal	NVC	69,40	64,90		67,15
DOS	File-Mal	PAV	94,80	90,80	96,20	93,93
DOS	File-Mal	PRO	12,00			12,00
DOS	File-Mal	SCN	97,20	93,40	90,40	93,67
DOS	File-Mal	SWP		78,30		78,30
DOS	File-Mal	TSC	57,20			57,20
DOS	File-Mal	VET	43,80			43,80
DOS	File-Mal	VIT		6,30		6,30
DOS	File-Mal	VSP	69,30	50,50	43,70	54,50

Tabelle 6.J: Analyse der Erkennungsrate über mehrere Tests

[88] Mittlerweile werden pro VTC-Test drei Windows-32-Betriebssysteme getestet (Windows 98, Windows 2000 und Windows XP). Es gilt zu untersuchen, inwieweit die Produkte unterschiedliches Verhalten unter den einzelnen Windows-32-Betriebssystemen zeigen. Bei statistisch geringen Abweichungen könnte zum Beispiel in Zukunft nur noch ein Referenz-Betriebssystem dieser Kategorie zum Testen der Produkte (mit Verweis auf das statistisch gemessene Verhalten der Produkte in vorangegangenen Tests) benutzt werden.

Viele weitere Arten von Abfragen sind denkbar und können bei Bedarf durch den Datenbankadministrator erstellt werden. Die Datenbank VTED dient somit als Werkzeug zur Betrachtung und Auswertung aller über die verschiedenen durchgeführten Tests gesammelten Ergebnisse. Tabelle 6.J zeigt exemplarisch die Entwicklung der Erkennungsrate pro getestetem Produkt unter DOS für die Malware-Datenbank *File-Mal* während der Tests 1999-03 bis 2001-04 (leere Zellen bedeuten, daß das Produkt an jeweiligem Test nicht teilgenommen hat).

6.5 Automatische Bewertung der Testergebnisse

Ein weiterer Vorteil der Verarbeitung der Testergebnisse in einer Datenbank ist die automatische Berechnung der Bewertung der Produkte unter den einzelnen Betriebsystemen anhand festgelegter Kriterien. In einem Formular können für jede Malware-Datenbank die Werte für die Erkennungsrate eingegeben werden, ab deren Erreichung das Produkt mit dieser Rate als "*perfekt*", "*exzellent*", "*gut*" oder "*sehr gut*" gilt. Abbildung 6.K zeigt dieses Formular zur Festlegung der Bewertungskriterien[89]. Der Benutzer der Datenbank kann somit genau festlegen, was ein *perfektes* oder *exzellentes* Produkt für Merkmale besitzen sollte. Auf diese Art lassen sich die Bewertungskriterien des Virus Test Centers für die getesteten Produkte innerhalb der Datenbank VTED festlegen. Die Kriterien können von Test zu Test geänderten Umständen angepasst werden[90], ohne daß die automatische Berechnung der den Bewertungskriterien entsprechenden Produkte beeinflußt wird.

In die Spalten *Perfect*, *Excellent*, *Very Good* und *Good* werden die *Mindest*-Erkennungsraten an Viren für die entsprechende Bewertung eingetragen. Ausnahme ist die *false positive* Erkennung (z.B. File-FP). Hier bedeutet der eingetragene Wert, daß der Scanner *höchstens* eine solche Rate an Falschmeldungen liefert. Dies wird durch Markierung des Feldes in der Spalte "Höchstwert" der Datenbank angegeben. In den zusätzlichen drei Spalten werden zusätzliche Nebenbedingungen für Werte von Kriterien für die Bewertung "*perfekt*" spezifiziert[91] ("Erkannte Objekte in %" als Mindestwert, "unreliably identified" und "unreliably detected" als Höchstwert).

Sind die Bewertungskriterien festgelegt, können automatisch für jedes Betriebssystem und jede Datenbank des aktuellen Tests die auf dieser Datenbank *perfekten* oder *exzellenten* Produkte angezeigt werden. Der Benutzer braucht also lediglich die Kriterien im Formular (Abb. 6.K) festzulegen. Abbildung 6.L zeigt den Anfang der Ausgabe der *perfekten* Produkte

[89] Die Bewertungen sind in der Datenbank VTED in Anlehnung an die bisherige Darstellung in Testberichten in Englisch.
[90] Es kann beispielsweise die Staffelung der notwendigen Erkennungsrate bei einer neuen Art von Malware (wie Skript-Viren im Test 2000-04) für die einzelnen Bewertungsstufen mit gewissem Abstand vorgenommen werden. Einige Tests später jedoch können die Bewertungskriterien verschärft werden , da eine Anpassung der Anti-Malware-Produkte an die Entwicklung zunehmend erwartet werden kann.
[91] Für die Bewertung "perfekt" wird vom VTC auch eine Erkennung aller Objekte (100% *erkannte Objekte*) sowie eine durchweg zuverlässige (0% *unreliably detected*) und genaue (0% *unreliably identified*) Erkennung gefordert.

pro Betriebssystem und Datenbank. Die Ergebnisse können exportiert werden (zum Beispiel für den Testbericht); als Format dafür stehen in Microsoft Access 97 unter anderem *Rich Text Format* (.rtf), *ASCII*-Text (.txt) oder *HTML*-Seiten (.htm) zur Verfügung.

Für jeden VTC-Test wird als Entscheidungshilfe für die Benutzer zu jedem getesteten Betriebssystem eine Rangliste der besten Testprodukte unter dem jeweiligen Betriebssystem veröffentlicht. Zur Ermittlung dieser Rangliste wird ein Punktesystem verwendet, bei dem jedes Produkt zwei Punkte für die Auszeichnung *perfekt* und einen Punkt für die Auszeichnung *exzellent* auf einer Datenbank erhält (vgl. Kapitel 4, Abschnitt 4.2.2.7). Die Scanner werden nach Punkten aufgelistet; der Scanner mit den meisten Punkten ist Testsieger auf dem jeweiligen Betriebssystem. In die Berechnung dieser Rangliste gehen nur die Zoo-, Itw-, Pack- und Malware-Datenbanken und die Datenbanken Poly und VKit sowie die Anzahl von Falschmeldungen (*false positive*s) mit ein.

Testart	Name	Erkennungsrate für Bewertung					weitere Kriterien für Perfect			Bewertung
		Perfect	Excellent	Very Good	Good	Höchst-wert	Erkannte Objekte in %	unreliably identified	unreliably detected	
1	Boot-FP	0	0	0	0	☐				☐
2	Boot-ITW	0	0	0	0	☐				☐
3	Boot-Mal	0	0	0	0	☐				☐
4	Boot-Pack	0	0	0	0	☐				☐
5	Boot-Zoo	0	0	0	0	☐				☐
6	File-FP	0	0,5	2,5	5	☑	0	0	0	☑
7	File-ITW	100	99	95	90	☐	100	0	0	☑
8	File-Mal	100	90	80	60	☐	100	0	0	☑
26	File-Mal-FP	0	0	0	0	☐				☐
9	File-Pack	100	83,3	66,66	50	☐	100	0	0	☑
18	File-Pack-ARJ	0	0	0	0	☐				☐
46	File-Pack-CAB	0	0	0	0	☐				☐
19	File-Pack-LHA	0	0	0	0	☐				☐
20	File-Pack-RAR	0	0	0	0	☐				☐
47	File-Pack-WRAR	0	0	0	0	☐				☐
21	File-Pack-ZIP	0	0	0	0	☐				☐
10	File-Zoo	99,9	99	95	90	☐	100	0	0	☑
11	Macro-FP	0	0,5	2,5	5	☑	0	0	0	☑
12	Macro-ITW	100	99	95	90	☐	100	0	0	☑

Abbildung 6.K: Formular zum Festlegen der Bewertungskriterien pro getestete Datenbank

In der Datenbank VTED können die in die beschriebene Gesamtbewertung pro Betriebssystem eingehenden Datenbanken ausgewählt werden (Auswahl in der rechten Spalte

in Abb. 6.K). Nur die Testergebnisse dieser im Formular ausgewählten Datenbanken werden zur Berechnung der Punktzahl pro Scanner benutzt. Absteigend nach Punktzahl werden dann die Produkte ausgegeben, zusätzlich wird jeweils die Anzahl der Datenbanken mit *perfektem* und *exzellentem* Abschneiden angegeben. Abbildung 6.M zeigt exemplarisch eine solche Auflistung für das Betriebssystem Windows 98.

Perfekte Scanner

Betriebssystem.Name **DOS**

Testart.Name *File-ITW*

Abkürzung	***Virusproz***
AVA	100
AVG	100
AVK	100
AVP	100
DRW	100
FPR	100
INO	100
NAV	100
NVC	100
PAV	100
SCN	100

Testart.Name *Macro-ITW*

Abkürzung	***Virusproz***
AVG	100
AVK	100
AVP	100
DRW	100
FPR	100
NVC	100
PAV	100
SCN	100

. . .

. . .

. . .

Abbildung 6.L: Ausgabe der perfekten Scanner pro Betriebssystem und Datenbank

Grading W98 gesamt

Gesamtpunktzahl	*Scanner*	*Count perfect*	*Count excellent*
14	FSE	6	2
14	SCN	6	2
12	AVK	4	4
12	AVP	4	4
11	PAV	3	5
10	CMD	4	2
10	FPR	4	2
9	FPW	4	1
8	INO	3	2
8	NVC	3	2
6	NAV	3	
6	AVG	3	
6	RAV	2	2
5	ADO		5
4	AV3	2	
3	PRO	1	1
3	DSE		3
0	ATR		
0	AVA		
0	QHL		
0	DRW		
0	PER		
0	VSP		
0	MR2		
0	CLE		

Abbildung 6.M: Automatische Berechnung der Gesamtbewertung pro Betriebssystem

6.6 Automatische Erstellung des Testberichtes

Der Testbericht für die VTC-Tests besteht aus drei Ebenen (vgl. Kapitel 4):

Ebene 1: Pro Betriebssystem eine Datei mit allen Ergebnistabellen

Ebene 2: Evaluation der Ergebnisse pro Betriebssystem; Bewertung der unter diesem Betriebssystem getesteten Produkte

Ebene 3: *Executive Summary*, Zusammenfassung der Ergebnisse und Bewertungen; Angabe der aus dem Test gewonnenen Erkenntnisse

Mit höherer Ebene werden die Testergebnisdaten immer weiter verdichtet. Zusätzlich werden ab Ebene 2 die Testdaten um gewonnene Erkenntnisse aus dem Test ergänzt. Da die Dokumente der höheren Ebenen zu immer größerem Anteil aus Analysen bestehen, lassen sie sich schwer automatisch erstellen. Denn die Analyse der Testergebnisse geschieht durch genaue Betrachtung der Testergebnisse und kann nicht automatisiert werden. Die Dateien aus Ebene 1 bestehen hingegen nur aus Testergebnissen, ohne deren genauere Betrachtung, und können deshalb sehr gut automatisch erstellt werden.

Testbericht - alle Testergebnisse

Jahr 2001
Monat 4
OS **DOS**
Datenbank *Boot-ITW*

Scanner	*Virus%*	*Virus*	*U_I*	*U_D*	*File*	*File%*
AVG	100,00	22	2	0	300	100,00
AVK	100,00	22	1	0	300	100,00
AVP	100,00	22	1	0	300	100,00
DRW	100,00	22	1	0	300	100,00
FPR	100,00	22	0	0	300	100,00
INO	100,00	22	0	0	300	100,00
NAV	100,00	22	1	0	300	100,00
NVC	100,00	22	1	0	300	100,00
PAV	100,00	22	1	0	300	100,00
SCN	100,00	22	1	0	300	100,00

Datenbank *Boot-Zoo*

Scanner	*Virus%*	*Virus*	*U_I*	*U_D*	*File*	*File%*
AVG	85,20	1117	25	21	4958	86,60
AVK	99,10	1299	62	4	5697	99,50
AVP	99,20	1301	62	4	5699	99,60
FPR	97,20	1274	4	7	5627	98,30
INO	97,20	1274	0	118	5402	94,40
NAV	95,90	1257	600	145	5021	87,70
NVC	98,20	1287	31	5	5662	98,90
PAV	99,40	1303	63	3	5705	99,70
SCN	82,70	1084	21	15	5017	87,70

. . .

. . .

. . .

Abbildung 6.N: Export aller Testergebnisse als Textdatei im RTF-Format

Sämtliche Testergebnisse eines Testes können - sortiert und gruppiert nach Betriebssystemen und Datenbanken - in einer Datei ausgegeben werden, die als Teil des Testberichtes exportiert

werden kann. Dieser Teil entspricht der Ebene 1 des Testberichtes. Als Ausgabeformat stehen *Rich Text Format* (.rtf), *ASCII*-Text (.txt) oder *HTML*-Seiten (.htm) zur Verfügung. Eventuelle Kommentare können leicht in den exportierten Text eingefügt werden. Somit läßt sich die Ebene 1 des Testberichtes automatisch und schnell erstellen. Abbildung 6.N zeigt den Anfang einer exportierten Datei im Rich Text Format mit der Auflistung aller Testergebnisse eines Testes beispielhaft an Testversion 2001-04.

Der Testbericht auf Ebene 2 kann teilweise automatisch erstellt werden, da die Berechnung der Produktbewertung pro Betriebssystem automatisch vollzogen werden kann (siehe Abschnitt 6.5). Die Analyse der Testergebnisse und insbesondere der Erkennung pro Betriebsystem müssen aber durch menschliche Begutachtung der Daten erfolgen.

Der *executive summary* (Ebene 3) läßt sich durch automatische Berechnungen über Daten in der Datenbank zwar unterstützen, aber die Erstellung des *executive summary* kann nicht automatisiert werden. Denn auf dieser Ebene werden nur Erkenntnisse veröffentlicht, die durch die Analyse der Testergebnisse durch das Testteam erlangt wurden. Die Erkenntnisse werden zwar aus den Testergebnissen, die auch in der Datenbank enthalten sind, erlangt. Doch läßt sich eine solche Analyse nicht durch Datenbankabfragen automatisieren, da dem Betrachter unterschiedliche Zusammenhänge auffallen.[92]

Abbildung 6.O zeigt die Benutzeroberfläche zum Export der Testergebnisse als Grundlage für den Testbericht. Die ersten drei Schaltflächen geben die gesamten Testergebnisse aus, allerdings nach unterschiedlichen Gesichtspunkten sortiert und gruppiert. Üblicherweise werden die Testergebnisse nach Betriebssystem, Datenbank und Scannernamen sortiert veröffentlicht (dies ist auch die Sortierung in Abbildung 6.N). Durch die Speicherung in einer Datenbank ist aber auch möglich, die Testergebnisse nach Datenbank oder Scannerabkürzung sortiert auszugeben. Die Sortierung nach Datenbank eignet sich gut für den übersichtlichen Vergleich des Abschneidens der Testprodukte auf bestimmten Datenbanken unter unterschiedlichen Betriebssystemen. Die Sortierung nach Produktabkürzung ermöglicht die gezielte Zusammenstellung aller Ergebnisse pro Produkt.

Die unterste Schaltfläche führt in ein Menü für die in Abschnitt 6.5 bereits beschriebene Bewertung der Produkte pro Datenbank und Gesamtbewertung pro Betriebssystem. Der Benutzer kann die Rangliste pro Betriebssystem und die Auflistung der perfekten und exzellenten Produkte pro Datenbank exportieren und als Grundlage für den Testbericht der Ebene 2 nutzen. Die Ergebnisse müssen allerdings noch um Kommentare und Analysen ergänzt werden.

[92] Erweiterte Techniken der Datenbankanalyse wie *data mining* ermöglichen auch eine computerisierte Suche nach Zusammenhängen in großen Datenmengen. Durch Datenbankabfragen lassen sich jedoch solche Ergebnisse nicht erzielen.

Abbildung 6.O: Benutzeroberfläche zum Export von Testergebnissen