The Technoarete Transactions on Recent Advances in Cyber security and Digital Forensics Journal

Volume No. 12

Issue No. 1

January - April 2023

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

The Technoarete Transactions on Recent Advances in Cyber security and Digital Forensics Journal

Managing Editor

Mr. Amit Prasad

The Technoarete Transactions on Recent Advances in Cyber security and Digital Forensics Journal

(Volume No. 12,   Issue No. 1,   January - April 2023)

Contents

A Comparative Performance Analysis Of WOA Vs. SOA

Ashish Verma[1], Vikas Bhatnagar[2], Siddharth Jain[3]

[1,2,3]Assistant Professor, Department of Computer Science & Engineering, Anand International College of Engineering, Jaipur

ABSTRACT

Service-Oriented Architectures (SOA) is an Emerging approach that addresses the requirements of loosely coupled, standards-based, and protocol independent distributed computing. A Distributed Computing is always required a tight coupled relationship between all working services. Basically SOA provides a large number of objects that are working in modular services as reusable software components. Generally there are no any alternative for SOA to provide flexibility and reduction in the cost of services which are basically used in the IT field as reusable components. This functionality is provided by the Enterprise Service Bus (ESB) that is an integration platform that utilizes Web services standards to support a wide variety of communications patterns over multiple transport protocols and deliver value-added capabilities for SOA applications. But in this Context we are introducing the "WEB 2.0"which is used to provide reusable IT components dynamically. In this paper we will introduce the methodology of design WOA using the concept of SOA. The big picture will follow the existing SOA model. In particular, this WOA methodology comprises conceptual as well as realization issues and breaks WOA design down into three distinct phases.

Keywords: Design Methodology, Reusable Components and Web oriented Architecture.

1. Introduction

Web-oriented architecture (WOA) is a style of software architecture that extends service- oriented architecture (SOA) to web-based applications, and is sometimes considered to be a light-weight version of SOA. WOA is also aimed at maximizing the browser and server interactions by use of technologies such as REST and POX. In software engineering, a Service- Oriented Architecture (SOA) is a set of principles and methodologies for designing and developing software in the form of interoperable services. These services are well-defined business functionalities that are built as software components that can be reused for different purposes. WOA is simply a way of implementing SOA by creating services that are Restful resources, allowing any service or data to be accessed with a URI. (REST, by the way, stands for representational state transfer).

Long version: WOA is an architectural style that is a sub style of SOA based on the architecture of the www with the following additional constraints: globally linked, decentralized, and uniform intermediary processing of application state via self-describing messages.

Shorthand version: WOA = SOA + WWW + REST

Motivation: The main remainder for this paper is to develop a Restful Application for which we have to convince the users to use WOA as a REST because rest is not a piece of software or like software packages instead of these we have following questions which are only be answered by REST which is a set of principles or rules in WOA:

- Is there any service which is mainly concentrate on resources?
- Is there any service which works on Web methods?
- Is your web service works in remote distributed environment?
- What is lifetime of URI and URL's in Web services?
- How your web service looks like in enterprises?
- Is your web service highly consumable, efficient and interoperable with support of latest widgets and gadgets and embedded social apps.
- Is there any Simple tool to weave the Web of resources into new applications?
- What types of objects are supported in remote basic environment?
- Is any service is called by any object if yes can we define its scope of visibility or its lifecycle?
- How is the web services protected?
- How can we define any client interface in web services?

Existing System: Here we are presenting the algorithm in which previous applications or web services were developed before introducing the WOA (Restful Applications)

Input: Q. Any user Query
Output: Final WS*
Step 0: Begin
Step 1: Take User Input as Query (SOAP Message)
Step 2: Proceed this query to the UDDI (Discovery Mechanism)
Step 3: Now the Query will be search in the UDDI.
Step 4: Response if only Found and Iff Server gets registered in UDDI also.
Step 5: If Step 4 gets true results then describe its description by (WSDL)
Step 6: After Step 4 & 5 the user will get resultant WS*
Step 7: Then Transfer SOAP message to the Client from service provider in the form of XML.
Step 8: End

If we consider the above mentioned example of SOA based web service we can say that the main concentration of this service is on Functional Data it does not matter from where data is coming or from where that particular service is being analyzed this only consider the functional the functional part of the service.

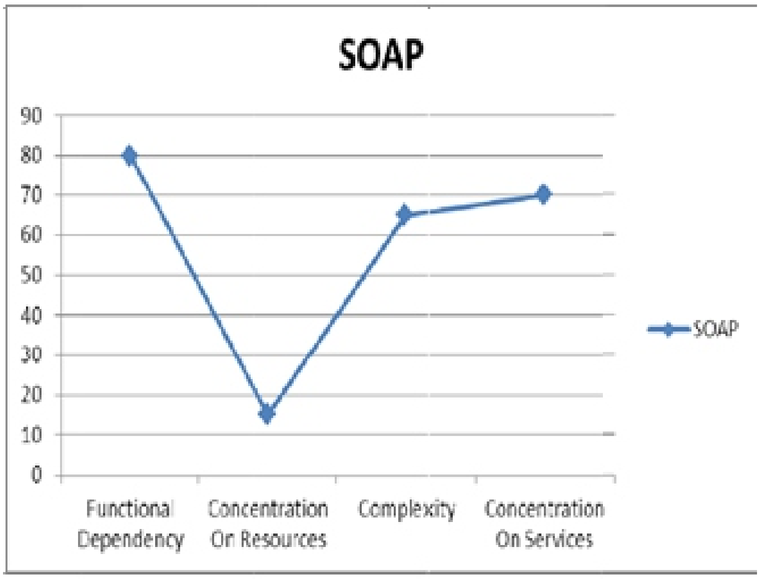


Figure 1.1: Resulting graph after applying previous algorithm

Proposed System

The proposed system is known as RESH Algorithm.

From the previous algorithm we are ready to develop SOA application in the SOAP environmentand now we are giving a proposed algorithm for developing any web services in WOAarchitecture with the use of REST architecture style:

RESH Algorithm for Developing RESTFul Application

Input: Q. Any Query from Client.

Output: Resultant WS*

Step 0: Begin

Step 1: Identify all the Resources by URI's

Step 2: Convert Verbs in to the Noun.

Step 3: After applying Step-2 get all Correct URI's

Step 4: Categorize them and apply suitable web method (GET, PUT, POST, and DELETE)

Step 5: Exposing them in URI Directory.

Step 6: Transfer them in XML Message Format.

Step 7: Repeat Step 2, 4 then and Perform Step 5 and 6.

Step 8: End

An Example Based On Resh Algorithm

Step 1: ET http://adduser?name=”Ashish”

After getting this URI we can see that it is not proper in terms of RESTful application so we need to convert it just need to convert all verbs in to the noun.

Step 2: GET http://user/Ashish

Step 3:GET/user/Ashish

host: myservertype:application/XML

Step 4: Now we need to apply web methods because the main purpose of this URI is to get data and ADD it in to the Database so we need to apply POST web method.

POST /User/HTTP 1.1

host :myserver

<?xml version=”1.0” ?>

<user>

<name>Ashish</name>

</user>

Step 5: Expose it if we have any other URI.

Step 6: Transfer them in to the XML format as per client Interface.

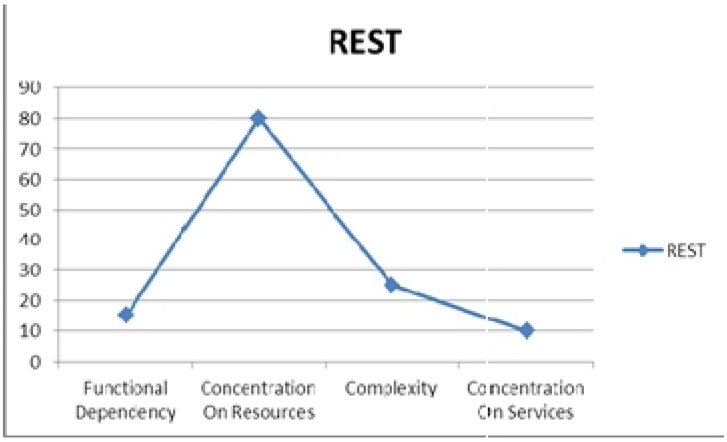


Figure 1.2: Resulting graph after applying RESH algorithm

The Comparative Analysis Between SO And WOA:

While comparing SOA and WOA we are taking some parameters one of them is round trip olatency in a network, latency, a synonym for delay, is for a packet of data to get from one designated point t n expression of how much time it takes much as la ency increases itdirectly impacts on the Quality of Service (QoS) and another parameter for comparing is them Ropacket size of eachequest / Response pair. So when we conducted these tests (Add, Update, getand Reove) comparison we found that total round trip time latency and average packet sizewas low in the case of Rest than the Traditional RPC based se vice or we can say serviceoriented Architecture. And it is clear that their REST equivalents.

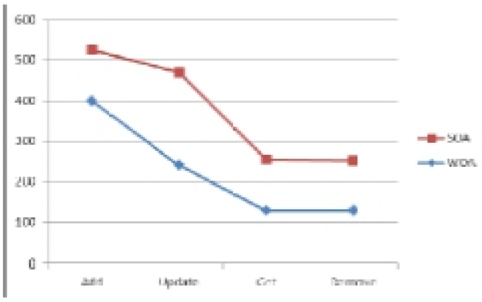


Fig 1.3:Benchmark Test/Add application Profile Service (Latency Gr ph)

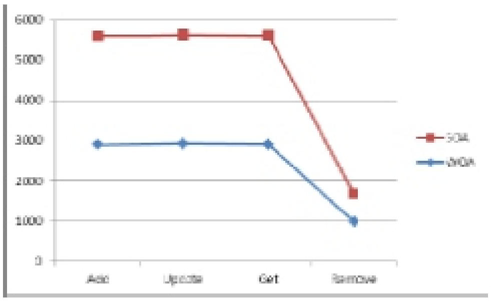


Fig 1.4: Benchmark Test/Add application Profile Service (Packet Size Graph)

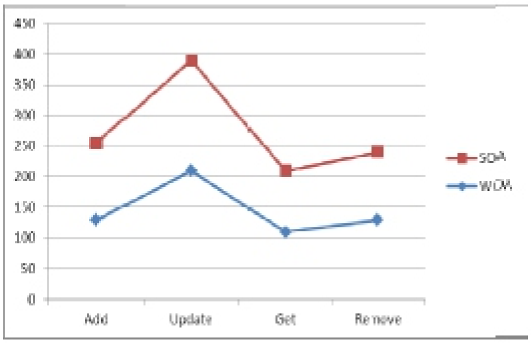


Fig 1.5:Benchmark Test/Device Profile Service (latency Graph)

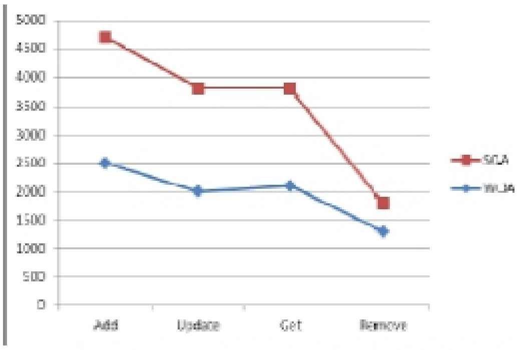


Fig 1.6: Benchmark Test/Device Profile Service (Packet Size Graph)

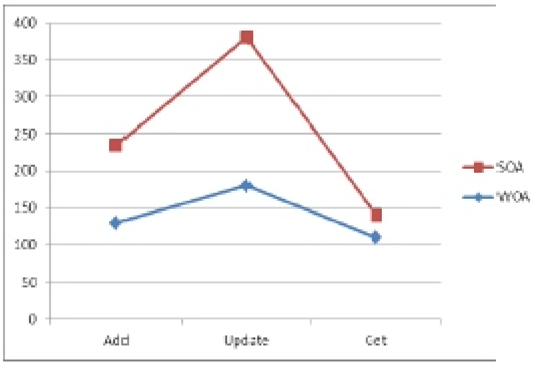


Fig 1.7: Benchmark Test/User Account (Laten(y Graph)

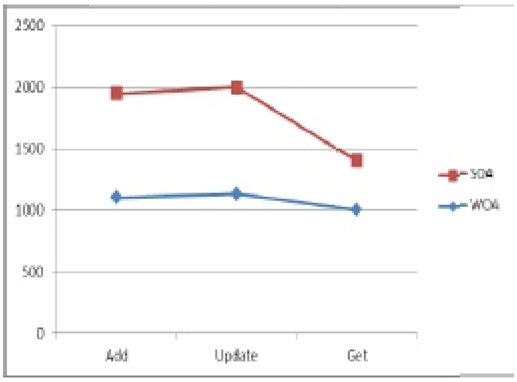


Fig 1.8: Benchmark Test/User Account (Packet Size Graph)

REST request and response messages add zero overhea to the messages being transmitted apartfrom the standard HTML headers which are used to route the packets through the network.SOAP, on the other hand, encloses each message payload within an additional SOAP 'envelope' set of XML tags and adds a few SOAP-related headers to the outbound HTTP packet. REST is able to take advantage of simplistic CRUD situations and execute them much more efficiently than the SOAP implementation. Some examples of this are the 'Remove' services for application profiles, device profiles, and user accounts. For these services, all that's required to delete anitem from the back-end data model is the item's ID number. Therefore, the ESTimplementation merely sends an HTTP DELETE command to the appropriate resource URLwith the ID number prepended. By doing so, it doesn't have to include any internal XMLpayload to represent this ID number and, thus, cuts down on its overall packet size.

SOA Versus WOA

- SOAP always transfer message in the form of its normal structure.
- On the other hand REST always transfers its message in the form of URI.
- Same from the response SOAP will always add some additional information on the payload of messages.
- But on the other hand REST simply transfers its message in the form of URI's.

References

[1].the Soa With Reach: Web-oriented Architecture By Dion Hinchcliffe.

[2]. Design And Development Of A Rest-based Web Service Platform For Applications Integration By Luis Oliva Felipe

[3]. A Guide To Designing And Building Restful Web Services With Wcf 3.5 Aaron Skonnard, Pluralsight October 2008

[4]. Web Services, Part 1: Soap Vs. Rest By Brennan Spies

[5]. Web Services, Part 1: Soap Vs. Rest By Brennan Spies, Olaf Zimmermann, Frank Leymann

[6]. Restful Web Services: A Reviewby Kurt Cagle In Reviews

[7]. Survey: Soa Delivering; Soap Out, Rest In By Joe Mckendrick For Service Oriented

[8]. Rest Battles Soap For The Future Of Information Services By John Newton [9].the Future Of Restful Drupal By Wscci Team

[10].restful Web Services By Leonard Richardson And Sam Ruby

[11]. Rest And Web Services: In Theory And In Practice By Paul Adamczyk, Patrick H. Smith, Ralph E. Johnson, And Munawar Hafiz [12]. Rest-an Introduction By Daniel Silva

[13]. The Design Of A Restful Web-service By Nadia Mohedano Troyano [14]. Designing Restful Web Applications By Ben Ramsey.

Development Of Predictive Simulation Models For Drug Dissolution Parameters Computing

Dr. H. B. Bhadka

Dean, Faculty of Computer Science, C. U. Shah Univeristy, Wadhwan. Gujarat (India)

ABSTRACT

Over recent years, drug dissolution computing has been the subject of intense and profitable scientific developments. Whenever a new batch profile is developed or produced, it is necessary to ensure that drug dissolution occurs in an appropriate manner. The quantitative analysis of the values obtained in dissolution tests is easier when statistical formulas that express the dissolution results as a function of some of the dosage forms parameters are used. In most of the cases the theoretical concept does not exist and some empirical equations have proved to be more appropriate.

Keywords *:Drug Dissolution, Drug release; Drug release models simulation; Parameters Computing.*

1. Introduction

Dissolution testing plays an important role in pharmaceutical quality control and in the development of solid, semi-solid, and transdermal pharmaceutical forms. The dissolution kinetic is reexamined under simulated physiological conditions, which are specified in both the U.S. Pharmacopeia (USP) and the European Pharmacopeia (EP) dissolution testing regulations. As such, these analytics are performed in a highly regulated Good Manufacturing Practice (GMP) environment, and present particular challenges for that facilitate those computations software applications. The role of computer based information systems has considerably increased in pharmacy as well as clinical practice in the last decade. However, the use of such systems in dissolution test is still not widespread. Several mathematical systems are commercially available for dissolution parameters calculations. In addition, comprehensive systems customized to specific needs have also been developed. The use of such systems in dissolution parameters calculations, questions regarding the role of structured data versus free text input, standardization of nomenclature, and compatibility with other systems, are hotly debated. This paper has in its scope the above stated considerations in the development of software for dissolution parameters calculations records, which attempts to resolve some of these issues. The model described herein is specifically designed to meet the requirements of the dissolution parameters calculations of a tertiary referral. An additional module has been included to allow modification and update of previously recorded data. A unique number assigned to each has been used as a primary identifier throughout the record.

2. Materials And Methods

2.1 Requirements:

The central objective of this initiative is to create a data model capable of accurately representing the calculations for dissolution parameters in a computer-suitable format. The main requirements include that the system should (a) be simple enough to be directly operated by the analytical scientist(s) in analytical research laboratory, (b) can easily run on personal computers, (c) allow comprehensive data entry conforming to accepted procedures which are currently carried out, (d) can generate a printed report, (e) allows modification and update of data, and (f) permit subsequent statistical analysis of records in a tabular format and displays the results of analysis. As dissolution data are to be handled by analytical scientists with minimal previous computer experience, an emphasis is laid on a user-friendly interface.

2.2 Software Construction:

The software has been developed in visual logway in Visual Basic. It has a set of two screens for data entry. One relating to (a) calibration curve and the other relates to (b) cumulative percentage release. Data flow is designed in two directions: (a) to a database, after appropriate coding, for storage and subsequent analysis at a later date, and (b) to the report generator. An additional module is included to allow modification and update of previously recorded data. Another module is designed for filling test reports on specimens obtained during the procedure, as and when these results became available. A unique identifier assigned to each test record is to be used as a primary identifier throughout the record.

2.3 Data Entry:

Modules are developed to allow easy user access and facilitate data entry. On completion of one module, automatic transfer to the subsequent module is envisaged. The basic module is structured as a large window, with smaller sub-windows appearing only on demand. The entire software is menu driven, with a simple and consistent hierarchical structure. As far as possible, all fields are structured, with the user allowed to choose one or more options from list of choices. These options included important and/or commonly observed conditions, and are chosen to cover majority of everyday findings after consulting experienced faculty members and reviewing previous records. A standard terminology developed for the structured items based on available literature and general consensus. The fixed choices are displayed either as searchable list boxes, check boxes, or as radio buttons. Free text is allowed in some fields, such as the information beyond the fixed choices available to the user. To allow complete data acquisition in each test, all data fields are marked mandatory, and the user is not allowed to proceed to a subsequent field without recording data in such fields. (Fig. 1 and 2)

2.4 Debugging and Modification:

After initial development, the software is tested over a four-week period by input of data. An attempt is made to rectify problems faced initially by the users. Opinion is sought from faculty members regarding possible modifications and improvements. Inconsistencies in the programming script, which gave rise to error messages during operation of software, are corrected. Finally the software is put to routine use.

2.5 Software Validation:

To evaluate the actual utility of the software, all consecutive test records entered using this computer software. Analytical scientists are asked to assess the overall quality of the reports and the content of information. After entry of data for 60 consecutive test procedures, these details are subjected to statistical analysis to evaluate the robustness of the database component.

2.6 Linearity or calibration curve [6]:

The linearity of an analytical procedure is its ability (within a given range) to obtain test results, which are directly proportional to the concentration (amount) of analyze in the sample. A linear relationship should be evaluated across the range of the analytical procedure. It may be demonstrated directly on the drug substance (by dilution of a standard stock solution) and/or separate weighing of synthetic mixtures of the drug product components, using the proposed procedure. The latter aspect can be studied during investigation of the range.

Linearity should be evaluated by visual inspection of a plot of signals as a function of analyze concentration or content. If there is a linear relationship, test results should be evaluated by appropriate statistical methods, for example, by calculation of a regression line by the method of least squares. In some cases, to obtain linearity between assays and sample concentrations, the test data may need to be subjected to a mathematical transformation prior to the regression analysis. Data from the regression line itself may be helpful to provide mathematical estimates of the degree of linearity. The correlation coefficient, y-intercept, slope of the regression line and residual sum of squares should be submitted for regulatory purpose. A plot of the data should be included. In addition, an analysis of the deviation of the actual data points from the regression line may also be helpful for evaluating linearity. For the establishment of linearity, a minimum of 5 concentrations is recommended. For the dissolution, concentrations of drug are calculated from the respective calibration curve (Fig. 3).

2.7 Dissolution study [1, 4-5]:

In vitro dissolution specifications are established to guarantee batch-to-batch consistency and to indicate potential bioavailability problems. For new drug products, dissolution specifications must be based on data obtained from the batch used in the bioavailability assay (bio-batch). For generic drugs, the dissolution specifications are generally the same of the reference drug product. These specifications are confirmed by testing the performance of the bio-batch dissolution. If the generic drug dissolution is substantially different from the reference drug product dissolution, and the in vivo study had proved the bio-equivalence between them, a different dissolution specification for the generic drug can be established, provided it is based upon a validated IVIVC. In that case, the specification must be fulfilled throughout the permanence of the generic drug in the market. The specifications must be based on the bio-batch dissolution characteristics. If the formulation developed for commercialization differs significantly from the bio-batch, the comparison of the dissolution profiles and the bio-equivalence study between these two formulations is recommended.

The dissolution tests must be undertaken under such conditions as: basket method at 50/100 rpm or paddle method at 50/75/100 rpm. To generate a dissolution profile, at least five sampling points must be obtained of which a minimum of three must correspond to percentage values of dissolved drug lower than 65% (when possible) and the last point must be relative to a sample period of time equal to, at least, the double of the former period of time. For drug products of rapid dissolution, samples at shorter intervals (5 or 10 minutes) may be necessary. For drug products with highly soluble drugs that present rapid dissolution (cases I and III of BCS), a dissolution test of a single point (60 minutes or less) that proves a dissolution of, at least, 85% is sufficient for batch to batch uniformity control. For drug products containing drugs poorly soluble in water, which dissolve very slowly (case II of BCS), a two points dissolution test, that is, one at 15 minutes and another at 30, 45 or 60 minutes, to ensure 85% of dissolution is recommended (Fig. 4 and 5).

2.8 Dissolution Efficiency [7]:

Khan suggested Dissolution Efficiency (D.E.) as a suitable parameter for the evaluation of in vitro dissolution data. D.E. is defined as the area under dissolution curve up to a certain time „t‟ expressed as percentage of the area of the rectangle described by 100% dissolution in the same time. The D.E. values are calculated from the dissolution data. (Fig. 6)

$$\text{Dissolution efficiency (D.E.)} = \frac{{}_0\int^t y.dt}{y100^t} x100$$

2.9 Comparison of dissolution profiles by similarity and dissimilarity factor [2-3, 8-9]:

To avoid the requirement of bioequivalence studies of the immediate release pharmaceutical forms of lower dosage, when several presentations with the same formulation exist, the dissolution profiles must be compared and must be identical among all dosages.

Until recently, single point dissolution tests and specifications have been employed to evaluate scale-up and post-registration changes. When minor alterations are carried out, the single point dissolution test may be adequate to ensure drug product quality and performance. For major alterations, the comparison of dissolution profiles obtained in identical conditions between the altered formulation and original one, is recommended. In this comparison, the curve is considered as a whole, in addition to each sampling point of the dissolution media, by means of independent model and dependent model methods. Independent model method employing the similarity factor. A simple independent model method employs a difference factor (f1, Fig. 7) and a similarity factor (f2, Fig. 8) to compare dissolution profiles. Factor f1 calculates the percentage difference between two the profiles at each sampling point and corresponds to a relative error measure between the profiles:

$$f_1 = \{ \left[\sum^{n}{}_{t-1} |R_t| \right] + \left[\sum^{n}{}_{t-1} R_t \right] \} x100$$

where:

n = number of sampling points

Rt = value dissolved in time t (percentage), obtained with the reference product or with the original formulation (before the alteration)

Tt = percentage value dissolved from the altered formulation, in time t. Factor f2 corresponds to a similarity measure between the two curves:

$$f_2 = 50 \times \log \left\{ \left[1 + \frac{1}{n} \sum_{t-1}^{n} |R_t - T_t| \right]^{-0.5} \times 100 \right\}$$

The procedure is described as follows:

- Determine the dissolution profile of products, test and reference, using twelve units of each.
- Calculate factors f1 and f2 using the equations presented previously.
- Criteria for two dissolution profiles to be considered similar.
- The nominal range of f1 and f2 values are 0 to 15 and 50 to 100, respectively.

3. Results

A software for drug dissolution parameter computation developed with Graphical User Interface (GUI). During execution, it takes for the drug concentration, instrument response and time data. After taking input it display list where user can opt for specific set of computations and can get the results for desired set of computation. The software supplements visualization along with computation. The user can opt for reports to be provided by the software. It generate calibration curve, cumulative percentage release, dissolution efficiency, comparisons of two products through similarity and dissimilarity factors. The software has various modules for input and modification of data, computation of various parameters and visualization with facilities to generate reports of dissolution parameters. The use of interface is designed for work with much ease in respecting. With little practice, scientists soon became adept at entering details correctly and quickly. The slightly increased time of data entry into the computer is more than made up by uniform and complete report generation. A user-friendly software providing computation and visualization parse drug dissolution parameters. The analytical scientists can utilize the software for intensive research as wide variety of parameter computation at simple key stroke.

The computer software currently used has two modules for data input: (a) calibration curve, and (b) cumulative percentage release. The data is linked to a MS-SQL Server having a set of two tables related to (a) calibration curve, and (b) cumulative percentage release and their reports. The two tables are linked to each other using the unique number. Another module deals with screen preview of reports and generation of printed reports.

In the calibration curve module, the number identifies each test record uniquely. The date of procedure is automatically derived from the system date maintained by the computer clock, but can be changed manually. The user has to enter the number of observations of concentration and instrumental response. After completion of calibration curve test record, the user is transferred directly to the „cumulative percentage release" module. The possible locations in the dissolution parameters tree are represented by a cascading hierarchy of tables. An additional table listing the appropriate divisions/segments appears.

On completion of data entry, the user is transferred to the print module, where he can preview the report prior to printing. The printed report contains all the information entered in the database. It also contains a standard set post- procedure instruction for the test, and also has space for signatures for the analytical scientist carrying out the procedure.

Problems initially faced by users are primarily related to data entry. Scientists, not having any working knowledge of computers, encountered problems such as a slow speed of data entry and failure to enter data in mandatory fields (with a consequent error message that did not allow the user to proceed further without rectifying the mistake). With little practice, they became adept at entering details correctly and quickly. Almost all the analytical scientists reported a slightly increased time of data entry into the computer, in comparison to writing reports on a standard printed proforma. However, all agreed that the report and data generated through the software are uniformly complete, and more than made up for the extra time spent. The new report has a uniform and easily understood structure, and is free of any inadvertent omissions.

The database component is evaluated by analyzing 60 consecutive records entered over a 4-month period. Data access and analysis are easily and quickly performed. Data are found to have been completely transferred from data entry screens to the database and no missing values are encountered.

4. Discussion

Structured input and free-text input represent two fundamentally different ways of entering data into a computer. Initial reports of test databases relied heavily on text based tools. Such input facilitates personalized style and flexibility in description of test records, and generates a well readable report. However, free-text input weakens the utility of the database, as it is not suited to subsequent analysis. Structured input and the resulting categorical data offer an important advantage in this regard. Data thus entered is more likely to be complete and is well suited for research and analysis, as well as for the generation of analytical reports and for quality control. It has been estimated that use of computerized test records improves completeness of data entry by more than 50 percent. However, a major trade-off for structure is flexibility. We therefore used a basic structured data entry protocol, supplemented by use of free text only under special situations. Besides operator related factors, it is related to the amount of free text entered and the number of tables accessed during structured data entry. However, the additional effort is rewarded by a more comprehensive, precise and accurately documented report.

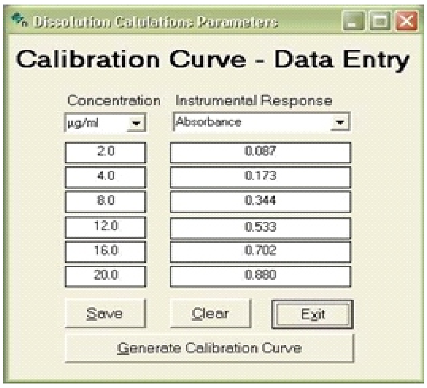
A major feature of the software is the powerful database component. This portion of the software has been built as a set of two interrelated database in MS-SQL Server, which can easily handle large database and also offers a wide range of analytical tools through a versatile query system. We have evaluated the robustness of this module of the software through an analysis of 60 consecutive test records.

Although such analysis requires some working knowledge of the database system, it is easy of learn. No data is lost and statistical analysis could be easily performed. Both user-friendliness of the software and completeness of data entry are critical to the success and acceptance of such software. This software allows easy integration of buttons, text boxes, check boxes and fields for free text to achieve this end. The format for data input is optimized through continuous interaction between scientists and the programmer. Scientists and other faculties are involved early and frequently during the development of the software, so that they are able to contribute ideas and advice. The software has been under routine use, and has performed well in areas of data entry, report generation and data analysis. Successful development and routine application of the database is, however, only a short-term achievement. The system is adaptable and capable of keeping pace with new technological advances.

Reference:

1. "Dissolution testing", United State Pharmacopoeia, XXIII, NF XVIII, The USP convention, Inc, Rockville, MD, (1995).
2. M. C. Gohel and M. K. Panchal, "Refinement of lower acceptance value of the similarity factor f2 in comparison of dissolution profile". Dissolution Technol., vol. 2, pp. 18-22, 2003.
3. M. C. Gohel, K. G. Sarvaiya, N. R. Mehta, C. D. Soni, V. U. Vyas and R. K. Dave, "Assessment of similarity factor using different weighting approaches". Dissolution Technol., vol. 11, pp. 22-27, 2005.
4. Guidance for industry: Dissolution testing of immediate release solid oral dosage forms. US food and drug administration, Rockville, MD, USA, 1997.
5. Guidance for industry: Immediate release solid oral dosage forms, scale-up and post- approval changes: chemistry, manufacturing and controls, in vitro dissolution testing, and in vivo bioequivalence documentation. US food and drug administration, Rockville, MD, USA, 1995.
6. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, ICH harmonised tripartite guideline validation of analytical procedures: Text and methodology Q2 (R1), 2005.
7. K. A. Khan and C. T. Rhodes, "Concept of dissolution efficiency". J. Pharm. Pharmacol, vol. 27, pp. 48-49, 1975.
8. J. W. Moore and H. H. Flanner, "Mathematical comparison of dissolution profiles".
Pharm. Technol., vol. 20, pp. 64-74, 1996.
9. J. E. Polli, G. S. Rekhi, L. L. Augsburger and V. P. Shah, "Methods to compare dissolution profiles and a rationale for wide dissolution specification for metoprolol tartrate tablets", J. Pharm. Sci., vol. 86(6), pp. 690-700, 1997.

Fig. 1 Input form of calibration curve – Data entry



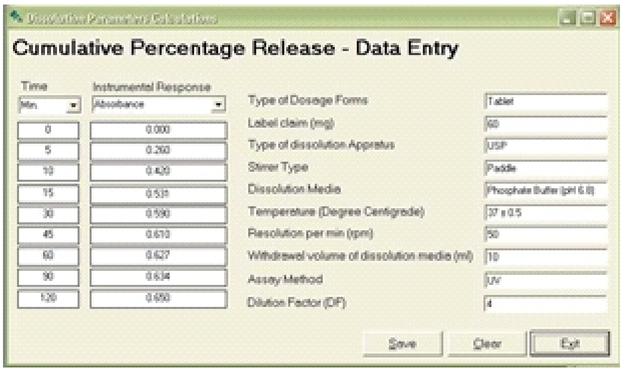
Dissolution Calculations Parameters

Calibration Curve - Data Entry

Concentration	Instrumental Response
µg/ml	Absorbance
2.0	0.087
4.0	0.173
8.0	0.344
12.0	0.533
16.0	0.702
20.0	0.880

Save
Clear
Exit
Generate Calibration Curve

Fig. 2 Input form of Cumulative percentage release – Data entry

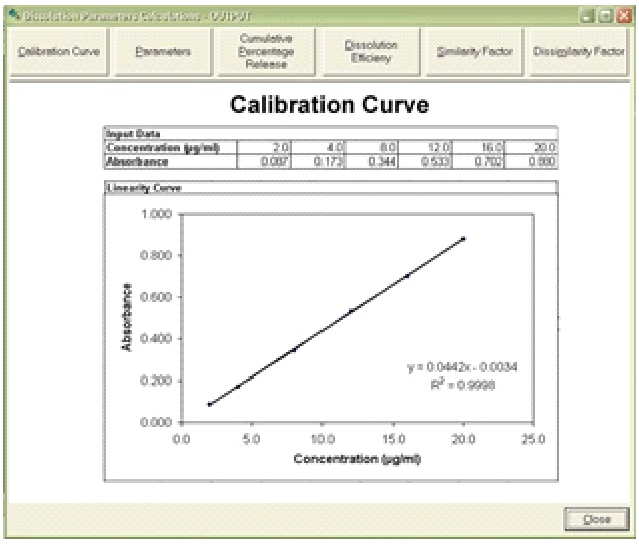


Cumulative Percentage Release - Data Entry

Time	Instrumental Response
Min.	Absorbance
0	0.000
5	0.260
10	0.420
15	0.531
30	0.590
45	0.610
60	0.627
90	0.634
120	0.650

Type of Dosage Forms	Tablet
Label claim (mg)	60
Type of dissolution Appratus	USP
Stirrer Type	Paddle
Dissolution Media	Phosphate Buffer (pH 6.8)
Temperature (Degree Centigrade)	37 ± 0.5
Resolution per min (rpm)	50
Withdrawal volume of dissolution media (ml)	10
Assay Method	UV
Dilution Factor (DF)	4

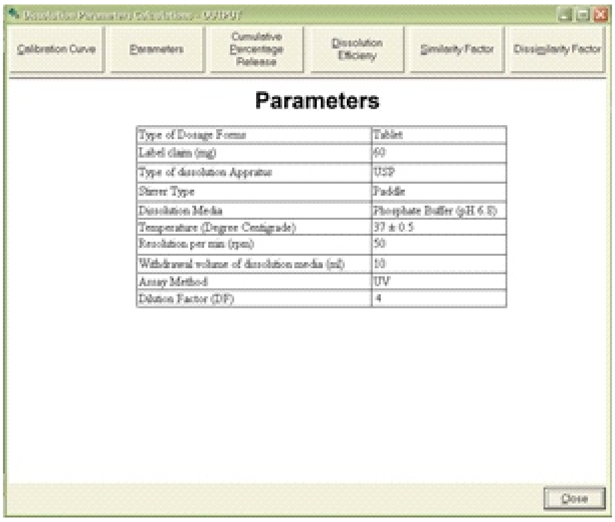
Fig. 3 Output report of calibration curve



Calibration Curve

Input Data						
Concentration (µg/ml)	2.0	4.0	8.0	12.0	16.0	20.0
Absorbance	0.087	0.173	0.344	0.533	0.702	0.880

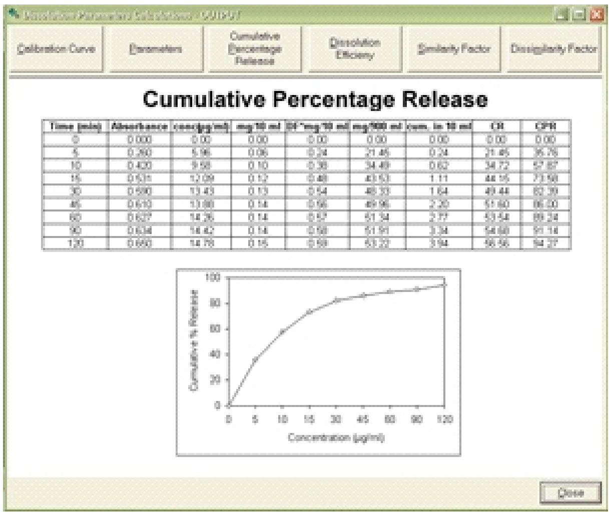
Fig. 4 Output report of parameters



Parameters

Type of Dosage Forms	Tablet
Label claim (mg)	60
Type of dissolution Appratus	USP
Stirrer Type	Paddle
Dissolution Media	Phosphate Buffer (pH 6.8)
Temperature (Degree Centigrade)	37 ± 0.5
Resolution per min (rpm)	50
Withdrawal volume of dissolution media (ml)	10
Assay Method	UV
Dilution Factor (DF)	4

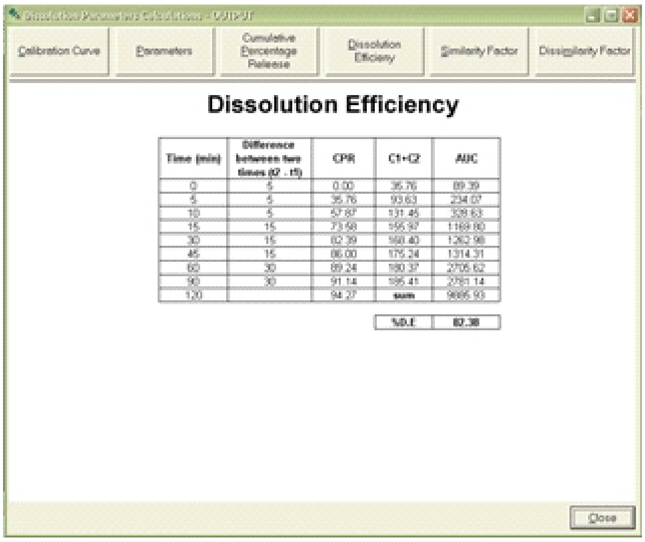
Fig. 5 Output report of cumulative percentage release



Cumulative Percentage Release

Time (min)	Absorbance	conc(µg/ml)	mg/10 ml	DF*mg/10 ml	mg/900 ml	cum. in 10 ml	CR	CPR
0	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.260	5.96	0.06	0.24	21.45	0.24	21.45	35.76
10	0.420	9.58	0.10	0.38	34.49	0.62	34.72	57.87
15	0.531	12.09	0.12	0.48	43.53	1.11	44.15	73.58
30	0.590	13.43	0.13	0.54	48.33	1.64	49.44	82.39
45	0.610	13.88	0.14	0.56	49.96	2.20	51.60	86.00
60	0.627	14.26	0.14	0.57	51.34	2.77	53.54	89.24
90	0.634	14.42	0.14	0.58	51.91	3.34	54.68	91.14
120	0.650	14.78	0.15	0.59	53.22	3.94	56.56	94.27

Fig. 6 Output report of dissolution efficiency

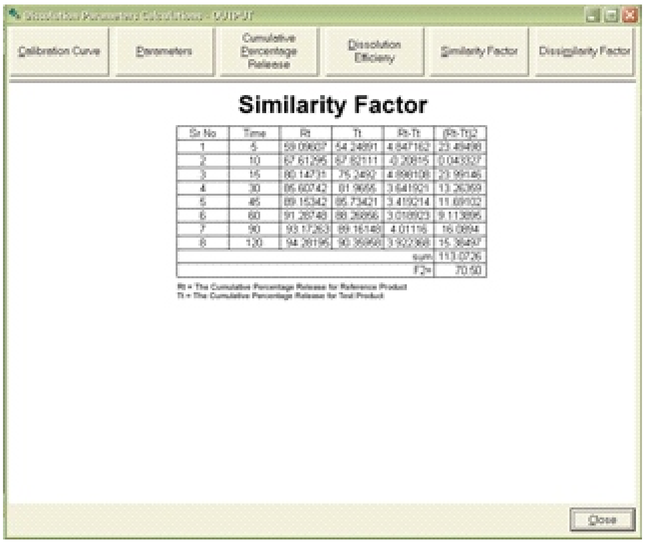


Dissolution Efficiency

Time (min)	Difference between two times (t2 - t1)	CPR	C1+C2	AUC
0	5	0.00	35.76	89.39
5	5	35.76	93.63	234.07
10	5	57.87	131.45	328.63
15	15	73.58	155.97	1169.80
30	15	82.39	168.40	1262.98
45	15	86.00	175.24	1314.31
60	30	89.24	180.37	2705.62
90	30	91.14	185.41	2781.14
120		94.27	sum	9885.93

%D.E	82.38

Fig. 7 Output report of similarity factor

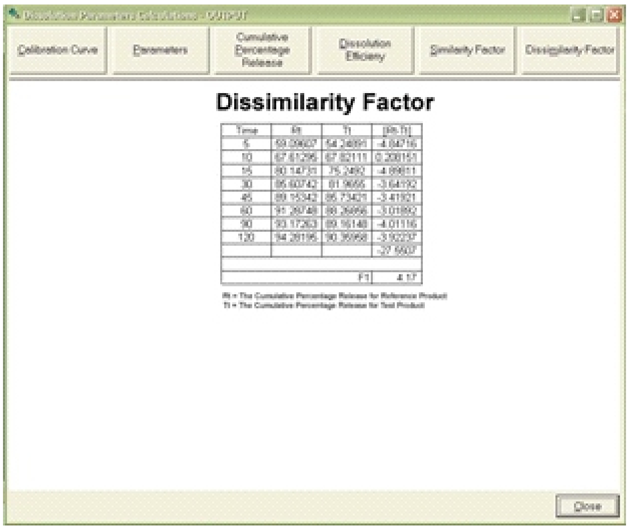


Similarity Factor

Sr No	Time	Rt	Tt	Rt-Tt	(Rt-Tt)2
1	5	59.09607	54.24891	4.847162	23.49498
2	10	67.61295	67.82111	-0.20815	0.043327
3	15	80.14731	75.2492	4.898108	23.99146
4	30	85.60742	81.9655	3.641921	13.26359
5	45	89.15342	85.73421	3.419214	11.69102
6	60	91.28748	88.26856	3.018923	9.113895
7	90	93.17263	89.16148	4.01116	16.0894
8	120	94.28195	90.35958	3.922368	15.38497
				sum	113.0726
				F2=	70.50

Rt = The Cumulative Percentage Release for Reference Product
Tt = The Cumulative Percentage Release for Test Product

Fig. 8 Output report of dissimilarity factor



Dissimilarity Factor

Time	Rt	Ti	[Rt-Ti]
5	59.09607	54.24891	-4.84716
10	67.61295	67.82111	0.208151
15	80.14731	75.2492	-4.89811
30	85.60742	81.9655	-3.64192
45	89.15342	85.73421	-3.41921
60	91.28748	88.26856	-3.01892
90	93.17263	89.16148	-4.01116
120	94.28195	90.35958	-3.92237
			-27.5507
		F1	4.17

Rt = The Cumulative Percentage Release for Reference Product
Ti = The Cumulative Percentage Release for Test Product

Internet As A Growing And Dynamic Network: An Economic View

Prof. Jayesh Kumar Jha

ABSTRACT

The past few decades have witnessed renewed interest and research efforts on the part of the scientific community. After spending decades to disassemble nature, focusing the attention on its components, scientists have shifted their attention on complex networks. These basic structures constitute a wide range of systems in nature and society, but their design is irregular, evolves dy- namically over time and their components can fit in a large multiplicity of alternative ways. Nev- ertheless, the most recent studies of networks have made remarkable progresses by investigating some critical issues of structure and dynamics, thereby improving the understanding of the to- pology and the growth processes of complex networks. From an economic point of view, networks are especially interesting because they can be considered as a problem of allocation of a critical resource, information, under multiple constraints. They can also be viewed as forms of poliarchies that reproduce, for many aspects, the market paradigm, with surprising properties of self-organ- ization and resilience, which go much beyond the characteristics that are generally attributed to general equilibrium structures. In this paper we first address the major results achieved in the study of complex network and then focus our attention on two specific, highly dynamic and com- plex networks: Internet and the World Wide Web.

Keywords: *Network Analysis, Internet, Social Networks*

1. Introduction

Since the 60's, the desire to understand the properties of the networks has prompted scientists from different fields to investigate the mechanisms that determine the topology of a variety of systems, ranging from biology to the structure of social relations. In the past two decades the availability of large databases on the topology of various real networks and the increased availability of computing power have offered scientists the chance to investigate networks of millions of nodes. Motivated by these circumstances, many new and important concepts about the topology of the interactions between different components have been proposed.

“The greatest challenge today not just in cell biology and ecology but in all of science, is the accurate and complete description of complex systems. Scientists have broken down many kinds of systems. They think they know most of the elements and forces. The next task is to reassemble them, at least in mathematical models, that capture the key properties of the entire ensemble.”

In our analysis, we first introduce the basic framework for the treatment of complex networks and then inves- tigate the mechanism determining the structural properties and topologies of real networks.

2. Graph Theory

Graph Theory is the natural framework for the treatment of complex networks, since a graph can be considered the natural representation of the topology of a complex network Leonhard Euler, one of greatest mathemati- cians of all time, pioneered graph analysis and made important discoveries both from the point of view of its theory and its applications. Euler spent most of his life in the city of Konigsberg, Prussia, which was set on the Pragel River and included two large islands, connected to each other and the mainland by seven bridges. In 1776 Euler solved the famous Konigsberg Bridge problem, consisting in finding a path that traversed each of the sev- en bridge of the Prussian city of Konigsberg exactly once and returned to the starting point. He offered a rigor- ous mathematical proof that the problem had no solution This result is considered the first theorem of graph theory, specifically of "planar graph theory". Since then, graph theory has undergone many more interesting developments and today represents the basis for the thinking about real networks.

In order to better understand graph theory, we must define some of its basic concepts. A graph is composed by a pair of sets G = (V,E), where V is a set of n vertices (also called points or nodes) and E is a set of K edges (links or lines) which connects two elements of the V. The graphs are constituted by a set of dots, which represents the nodes, and where two dots may be joined together by a line representing a link. In graph theory how these dots and lines are drawn is irrelevant, the only thing that matters is which pairs of nodes form a link and which do not. While it is clear that a graph may represent a network as a set of nodes and links, concepts such as arcs, paths and a number of other characteristics have been developed and superimposed, so to say, to the definition of the original definition of a graph. In this regard, a graph can be considered also a representation of a "structure", a more general concept, which more naturally lends itself to be described by additional, more abstract properties.

Graphs are also representations of market structures, in that they may depict trade flows, or other systems of individuals and linkages that have economic relevance. Contractual relations, for example, can be represented as connections among contracting partners, and can be analyzed through graphs that represent not only bilateral obligations, but also the interdependencies that are created, as a consequence of bilateral deals, in a system of production and exchange.

3. Hierarchical Clustering

Lin (1999) noticed that hierarchical position and network location facilitate the access to embedded resources such as wealth, status and power of social ties. As an alternative to community networks, hierarchical clustering appears to be also a powerful basis of organization and power in social networks according to two distinct strategies: 1) agglomerative clustering and 2) divisive clustering. The agglomerative model provides for a series of fusions of the single nodes into groups. It can be defined as the result of a “bottom up” approach, where each node starts in its own cluster and then pairs of clusters are merged as one moves up the hierarchy. Divisive methods, on the other hand, separate the single objects and correspond to a “top down approach”, where all nodes start in one cluster and then splits are performed as one by one they move down the hierarchy. The hierarchical structure of clusters can be graphically represented by dendograms, or hierarchical trees, which are often used to display the clusters which are produced at each step of agglomeration.

The importance of clusters for economic theory arises from Coase's approach to the theory of the firm. Ac- cording to Coase (1988) in fact, the existence of the firm is the attempt to reorganize contracts of exchange in alternative to the market, by economizing on transaction costs. The value of the firm thus derives from a pe- culiar configuration of “rights”. This depends on its “dedicated hierarchical nature”, that is, its specialized clus- ter structure, which is the consequence of both agglomerative and divisive clustering, and the assignment of dif- ferent “rights” to its various stakeholders with ownership and control embedded into residual “rights” of shareholders. Because most economic activities can be interpreted as “enterprises”, i.e. as business ventures of a sort, the Coasian approach and the further developments of the neo-institutional school imply that the cluster paradigm (the enterprise as a cluster of contracts) may be applied to a wide variety of situations and, in particu- lar, to the configuration of different actors presented by Internet and WWW. The concept of the enterprise as a cluster of contracts and of the parties involved as stakeholders has forced economists to face the issue of the plurality and heterogeneity of economic agents, especially in the new forms of "user based" enterprises.

4. Short Path Lengths

While the random graph model is the basis for the study of the formation of long-range connectivity in random systems (which is studied by percolation theory), it does not consider the random growth of complex structures, consisting of different points and connections of different types to obtain more complex real growth processes. In its original formulation furthermore, it only considers static networks in which the number of vertices is fixed, thus neglecting the fact that in reality many networks evolve with the continuous addition of new ele- ments to the system. For any given network, however, the random graph model has the merit of identifying an underlying random structure which can be very useful to establish some basic properties that do not depend on complexity and/or on growth.

The so called “small world property” is the most important network property in this respect. It was discovered by Milgram (1967) who proposed an experiment, where randomly selected people in

Nebraska would be asked to send letters to a distant individual in Boston, identified only by name, occupation and rough location, so that the letter could only be sent to someone presumably closer to him. Milgram was tracking the letters and found a surprising result. The average number of links needed to find the targeted person was found to be only six. This result is called “the six degrees of separation” and is the consequence of the fact that two individuals who don't know each other may nevertheless be linked by a common acquaintance.

The small world property is important because it reveals properties of the performance of a given action that depend on the underlying structure rather than on any special procedure of optimization. It also shows that local information may be conducive to global success and that social networks, regardless of their size and complexity, exhibit two key characteristics: 1) a plurality of short path lengths and, 2) a structure that enables individuals to find short path lengths even in absence of a global knowledge of the network. These aspects are very important not only for social systems, but also for the World Wide Web, the traffic way-finding in a city, and the transport of information packets on the Internet and the diffusion of signaling molecules in biological cells The small world property also suggests that whereas complexity may make more difficult to comprehend the properties of a network, it carries with itself an increase in connectivity that makes easier to cover seeming longer distances with relatively short paths. This property is the basis, for example, of the so called strength of weak ties. This argument, put forward by Granovetter (1973) asserts that “Our acquaintances (weak ties) are less likely to be socially involved with one another than are our close friends (strong ties). Thus the set of people made up of any individual I and his or her acquaintances comprises a low-density network (one in which many of the possi- ble relational lines are absent) whereas the set consisting of the same individual and his or her close friends will be densely knit (many of the possible lines are present)”. Complex networks can thus be conceived as sets of simpler closely connected networks of strong ties (such as cliques) loosely connected by weak ties. The small world property would be the result of this local-global structure, whereby weak ties would be the means to bridge the gaps between two or more densely connected “strongly tied” networks.

5. Small World Networks

The small world property (SWP) was the object of different analysis on the structure of real networks, especially in biological and technological networks. Watts and Strogatz (1998) extensively analyzed SWP in their “Collective dynamics of small-world networks”, where they found the existence of a relationship between the “small-world” networks and a high value of the clustering coefficient. Analytically, the clustering coefficient is a parameter introduced by the authors in 1999, to characterize the structure of complex networks. It represents a measure of the degree to which nodes in a graph tend to cluster together. Two versions of this measure exist: a local and a global clustering coefficient. The local clustering coefficient gives an indication of the clustering around a single node inside the network, while the global clustering coefficient is used to define an overall indi- cation of the clustering inside the network.

The small-world property is also important for economics, where technological development, trade growth and the so called globalization phenomenon can be interpreted, as a consequence of “global clustering”, as re- ducing the distance between people and make the world smaller. Goyal, van der Leij and Moraga-Gonzalez (2005) studied the evolution of social distance among economists who publish in journals in the period from 1997 to 2000, to show that, despite the fact that the number of economists has more than doubled in this period, the distance between any two of them has declined.

6. Post-Structuralism and Hypertexts

Complex networks find their most recent and egregious incarnation in Internet and the World Wide Web, two constructs whose complexity, because of their continuous, never ending growth, appear boundless. It is an ex- treme level of present and expected convexity that, paradoxically, stimulates the search for paradigm that cut across the intricacy and multiplicity of links, to discover drastic simplifications. The small world property is one of these, but a more pervasive idea is that of the personal classification embedded in the so called hypertext. This and other ideas appear to incarnate many of the intuitions of the so called post-structuralism, a strain of thought concentrating on a set of themes on the transmission of meaning through language, the role of networks of signification, and the perpetuation of power. These themes parallel and to some extent predict hypertexts and some other features of Internet and the World Wide Web and, in particular, share with the originators of hyper- text and the Web the notions about the structure and workings of text, and of the network as the coordinating principle behind the transmission of meaning through texts. Post-structuralist theory challenges the assumption that organizing structures can be imposed on information in a neutral and objective fashion. This is a similar mi- strust to discrete set approaches to information organization and retrieval influencing the innovators of hypertext and the Web.

Post structuralism economics may also be interpreted as an extension of Coase's theory (1988) which sees the power of the firm arising from its capacity to tie its stakeholders in a multiplicity of explicit and implicit contractual knots. As in a post-structuralist Foucault, 1980 in the case of language and power, the neo-in- stitutional school that emerged after Coase's work interprets economic power as the cause, rather than the con- sequence of the economic structure. Thus, for example, as Coase firmly establishes, in the presence of transac- tion costs, the distribution of property rights, by empowering one particular set of stakeholders rather than another one, determines the ensuing structure of the market and, as such, the particular clustering of contracts that characterizes the firms and their relations. This approach has also caused a “new theory of corporations” to emerge in the school of law and legal analysis, whereby corporations are considered networks that lock in equity investors' initial capital contributions by making it far more difficult for those investors to subsequently withdraw assets from the firm.

These characteristics of economic enterprises and their networks appear also important to understand the post- structuralist nature of much of the World Wide Web. Foucault, as a key theorist of the post-structuralist movement, describes the properties of an interdependent system by exploring the role of power within a special category, which he calls “discourse”. Discourse for Foucault is a framework through which knowledge is trans- mitted and exploited, and, what is more, a framework regulated by power relations. Those power relationships are evident both between individuals, and more importantly between groups. Through language, customs, classi- fications and other more subtle means that impose a structure on knowledge, discourse therefore manifests its power by delimiting what it is acceptable and even possible to say about given subjects at given times. Contrary to Bacon who believed that knowledge was essentially empowering, Foucault argued that power defines what can be considered knowledge.

As a hypertext defined by essentially free associative relations, which can be traced by constructing a highly personalized and unpredictable chain of links among texts of different levels, the WWW appears an apparently successful attempt to overcome the dictatorship of an exogenously established discourse, which determines the extent and the nature of the knowledge that can be gained. The WWW is free from the arbitrary and tendentious nature of the classifications used to index and navigate the system of traditional texts and lends itself to be explored, without having to use the scaffolding of analogical categories that are the base of all dictionaries, encyc- lopedias and library classification systems.

The economic side of this analysis resides in the nature of the World Wide Web as a system of information management, which arose from Enquire, a personal information retrieval system developed by Tim Berners-Lee, who recognized its potential as a global information system from the outset. Berners-Lee attempted to overcome the formal hierarchical structures imposed on information management solutions, because of their es- sential sub-optimality in retrieving and organizing information. The basic idea of the new system, which has important economic implications, is very remindful of the emergence of Coase's enterprise, and consists in the intuition that self-organizing clusters of knowledge would come more efficiently from textual networks con- nected by semantic and associative relationships. Contrary to the formal structures dominating traditional text indexing and retrieval algorithms, hypertext and the Web could thus progressively emerge from an underlying loser structure of random networks, by creating dynamic clusters of associative relationships emerging from the texts of an information collection. This would in turn give rise to self-organizing associative networks of infor- mation, which would dynamically optimize information search and retrieval.

7. Scale-Free Networks

The year 1999 can be considered a turning point in the analysis of complex networks because scientists found that networks don't show static scale-free graphs but expand continuously by the addition of new vertices. The network models discussed by Erdos and Renyi and by Watts and Strogatz assume that the number of vertices in- side the network remains fixed. In this way, static scale-free graphs are models in which growth or aging processes do not play a dominant role in determining the structural properties of the network. In reality many real networks are ruled by the dynamical evolution of the whole system. In this respect, Barabasi and Albert observed that most real networks are open systems which grow by the continuous addition of new nodes.

The Barabasi and Albert (BA) model was inspired by the topology structure of the World Wide Web that constitutes a network in continuous evolution and where the number of sites increases dynamically. By explor- ing several large databases describing the topology of large networks, AB found that, for most large networks, the degree distribution deviates from the Poisson law and that, in most of cases, it follows a power-law for large K. Since power-laws are independent of the unit of measure, these networks are called “scale-free” This topological characteristic is determined by two mechanisms that interact inside the network: growth and preferential attachment. In contrast with the static models, the scale-free model describes a dynamic system which grows by the continuous addition of new vertices, as for example does the World Wide Web, which grows by the continuous addition of new Web pages. So, growth means that the number of nodes increases over time.

The algorithm of this mechanism can be represented as an algorithm which starts with a small number of nodes m0 and at each time step, adds new nodes with $m \leq m0$ edges, with each new node being linked to the m nodes that are already present in the system. The algorithm is also non-random in the connectivity for a node inside the network, and dependent on the node's degree. This means that, when choosing the vertices to which the new node connects, the probability that a new node will be connected to node i depends on the degree Ki (the number of nodes already connected) of this node. New vertices attach preferentially to already well connect ones. An example of preferential attachment is represented by the hyperlinks of the Web page that will have a higher probability to include links to the more popular documents than to less-known ones. New pages link pre- ferentially to hubs, very well-known sites such as Google, rather than to less-known pages. In this way, older vertices increase their connectivity, leading to a rich-gets-richer phenomenon that can easily be identified inside real networks.

Growth and preferential attachment represent two important mechanisms of the networks, and both lead to the discovery of the networks with a power-law degree distribution.

$$P (K) = AK -Y \qquad (1)$$

The exponent takes different values with respect to different networks, within a relatively narrow range (2.1 to 4): for example for the World Wide Web the value is approximately 3. K stands for the average degree of a node i, that is the number of edges incident with the node, while P stands for the probability that a node chosen at random has degree K. BA investigated two different variants of the model: one with growth and no preferential attachment and one with preferential attachment without growth. In both cases no scale free structure emerged. Thus, both properties are needed to empower the network to self-organize according to a stationary power law distribution. The scale-free nature of networks, which has been widely accepted by most scientists, forces us to acknowledge that networks constantly change over time. The evidence comes from better maps and data sets but also from the agreement between the empirical data and the analytical models that predict the network structure. In his book "Linked" (2002), Barabasi states that "power-laws are at the heart of some of the most stunning conceptual advances in the second half of the twentieth century, emerging in fields like chaos, fractals and phase transitions. Spotting them in networks signaled unsuspected links to other natural phenomena and placed networks at the forefront of our understanding of complex systems in general. The fact that the networks behind the Web, Hol- lywood, scientists, the cell, and many other complex systems all obey to a power law allowed us to paraphrase Pareto and claim for the first time that perhaps there were laws behind complex networks." It was the well know Italian economist Wilfredo Pareto who, at the end of the nineteenth century, noticed that a few quantities in nature follow a power law. As a careful observer of economic inequalities, Pareto noticed that 80 per cent of the money is earned by 20 per cent of the population and also that 80 per cent of his peas were produced by only 20 per cent of the peapods. Pareto's rule is a power-law degree distribution and appears to approximately hold for many networks, including the World Wide Web, where around 80% of the links on the Web point to only 15% of the Webpages. As Barabasi (2000) puts it, "power laws mathematically for- mulate the fact that in most real networks the majority of the nodes coexist with a few big hubs, nodes with an anomalously number of links. The few links are not sufficient to connect the entire network, but this function is secured by the rare hubs."

8. Conclusion

The analysis of complex and dynamic networks is at the heart of several new fields of scientific inquiry and the basis of an interpretation of reality that cuts across several disciplines. As a method to understand Internet and its economic significance, modern network theory appears especially relevant, even though most of its discipli- nary and interdisciplinary connections are yet to be discovered. The Social Accounting Matrix and the input output systems are an early application of network theory, even though their development in economics has been autonomous and mostly centered on the quantitative implications of impact analysis. On a different front, as in- formation management systems, both Internet and the Web are environments of enterprise creation that recall Coase's original theories and the subsequent outgrowth of institutional economics. The basic idea here is that markets can be viewed as a loose network of connections with the ensuing emergence of denser sub-networks as hierarchical clusters of contracts and other types of relationships. In this respect, both the Internet and the Web enlarge the horizons of Coase's original theory much beyond the classical concept of the firm to the idea of an enterprise that can be recursively and completely defined in terms of its internal and external relations, and whose organization and production is largely dependent on the contribution of a plurality of users/stakeholders.

A second important element of Internet as a social and economic system, which isrelated to its clustered nature, is the fact that it is a system of small worlds, or, to cite a phrase that has become popular also in other con- texts, a system of strong and weak ties that make possible communications of different types and intensities within and across communities. While the determinant factor of strong and weak ties for internet is built in its physical configuration, the small world characteristics of the web depend on its nature of a dynamic clustering system and the scale free property of the distribution of its links. This property is most intriguing, because it seems to denote a form of accumulation of “network capital”, whose distribution is based on a more than pro- portional connection reward to the nodes that already have a higher number of connections. The Web seems thus characterized by increasing benefits of accumulating information in a few privileged hubs, without correspon- dent increases in congestion costs. As in Coase's model, this property may be itself the consequence of the ten- dency of self-organizing clusters to reduce transaction costs.

As an information management tool, the Web presents itself as a network of relations clustering around the principle of cognitive gain from free association. The hypertext results from the possibility to navigate among different texts without the limitations imposed by external classifications. As such, it is a source of allocative ef- ficiencies that deserves further analysis. In principle, not only it allows exploiting more fully the information contained in the texts examined, but it also frees the reader from the dictatorship of the framework superimposed by any existing authority, which may effectively assert its power by limiting the extent and the form of knowledge that can be acquired

9. References

[1.] Strogatz, S.H. (2001) Exploring Complex Networks. Nature, 410, 268-276. http://dx.doi.org/10.1038/35065725

[2.] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.-U. (2006) Complex Networks: Structure and Dy- namics. Physics Reports, 424, 175-308.

[3.] Barabasi, A.L. (2002) Linked: the New Science of Networks. Perseus Books Group, New York City. [4]L i n , N. (1999) Building a Network Theory of Social Capital. Connections, 22, 28-51.

[4.] Coase, R. (1988) The Firm, the Market, and the Law. University of Chicago Press, Chicago.

[5.] Ackerman, B.A. and Alstott, A. (1999) The Stakeholder Society. Yale University Press, New Haven. [7] Erdos, P. and Renyi, A. (1959) On Random Graphs. Publicationes Mathematicae (Debrecen), 6, 290. [8] Milgram, S. (1967) The Small World Problem. Psychology Today, 160.

[6.] Granovetter, M.S. (1973) The Strength of Weak Ties. Sociol., 78, 1360.

[7.] Watts, D.J. and Strogatz, S.H. (1998) Collective Dynamics of Small-World Networks. Nature, 393, 440-442.

[8.] Goyal, S., Van der Leij, M. and Moraga-Gonzàlez, J.L. (2005) Economics: An Emerging Small-World. Journal of Po- litical Economy, 114, 403-412.

[9.] Foucault, M. (1980) War in the Filigree of Peace Course Summary. Oxford Literary Review, 4, 15-19. [13] Stout, L. (2004) On the Nature of Corporations. Deakin Law Review, 9, 775-789.

[10.] Berners Lee, T.J. and Fiaschetti, M. (1999) Weaving the WEB, Harper, San Francisco.

[11.] Albert, R. and Barabasi, A.L. (2002) Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science and Everyday Life. Plumes, New York City.

[12.] Albert, R. and Barabasi, A.L. (2002) Statistical Mechanics of Complex Networks. Review of Modern Physics, 74, 47-http://dx.doi.org/ 10.1103/RevModPhys.74.47

[13.] Barabasi, A.L. and Albert, R. (1999) Emergence of Scale in Random Networks. Science, 286, 509.http://dx.doi.org/10.1126/science. 286.5439.509

[14.] Barabasi, A.L., Albert, R. and Jeong, H. (1999) Scale Free Characyeristics of Random Networks: The Topology of the World Wide Web. Physica A, 272, 173.

[15.] Albert, R. and Barabasi, A.L. (2000) Topology of Evolving Networks: Local Events and Universalities. Physical Re- view Letters, 85, 5234. http://dx.doi.org/10.1103/PhysRevLett.85.5234

ATM Safety & Security

Krishan Tuli[1], Gurpreet Kaur[2]

[1,2]Lecturer, Chandigarh Group of Colleges, Landran

ABSTRACT

Security in the ATM Network is very necessary because it is widely spread in all areas such as financial, network administration and other important parts of financial network which requires very sensitive handling transmission of data. Manipulating the transmitted data, spoofing and misuse of ATM channels would be very fatal in accounting system. Mostly the ATM transactions should rely on the integrity of secure Crypto-processor (Dedicated processor for carrying Crypto- graphic operations). Therefore the ATM Forum, developed new standards & specification called ATM security specification 1.0 in 1998 and then in March 2001, the ATM Forum developed a new Technical Committee called ATM security & specification version 1.1. This paper presents the safety & security of the Automated Teller Machine which includes the basic introduction, threats to an ATM network, network security framework.

Keywords: *Automated Teller Machine, Confidentiality, Data Integrity, Eavesdropping, Masquerade, Denial of Service, Traffic Analysis, Corruption of information, Forgery, Controlled Access, Authentication, Logging, Reporting, Audit, Recovery.*

1. Introduction

The security specification includes the security services that are needed and necessary for the users to protect their ATM cards from being misused. Confidentiality, Data Integrity, Accountability, Correct Functionality, Availability and Access Control are the main objectives for ATM. Principal functional security requirements can be identified to deal with the generic threats. They are:

- AF-SEC-1: Verification of Identities;
- AF-SEC-2: Controlled Access and Authorization;
- AF-SEC-3: Protection of Confidentiality;
- AF-SEC-4: Protection of Data Integrity;
- AF-SEC-5: Strong Accountability;
- AF-SEC-6: Activity Logging;
- AF-SEC-7: Alarm Reporting;
- AF-SEC-8: Audit;
- AF-SEC-9: Security Recovery;
- AF-SEC-10: Management of Security. These functions are from AF-SEC-1 to AF-SEC-10.

Threats To An ATM Network

ATM network will suffers a lot of threats. Few of the network threats are

Eavesdropping

Eavesdropping is a threat in which attacker connects into the transmission media and gain unauthorized access to the data. It is one of the most common attacks to the network.

Masquerade

Masquerade is a threat in which one person pretends to be someone else and by doing that, tries to gain access to information.

Service Denial

Service Denial occurs when on entity fails to perform its work and prevents other entities to perform its work.

Traffic Analysis

Traffic analysis refers to a threat that the hacker can get information by collecting and analyzing the information like the volume, timing and the communication parties of a VC (Virtual Channels). Volume and timing can reveal a lot of information to the hacker even though the data is encrypted, because encryption won't affect the volume and timing of information.

Corruption of Information

The transmitted data is altered, deleted, changed and delayed by an entity with a proper authorization.

Forgery

Forgery refers to, when the fake data is sent and is claimed to have been received. The authenticated person's information must be changed to do this authentically.

Security Services

The ATM Security Framework

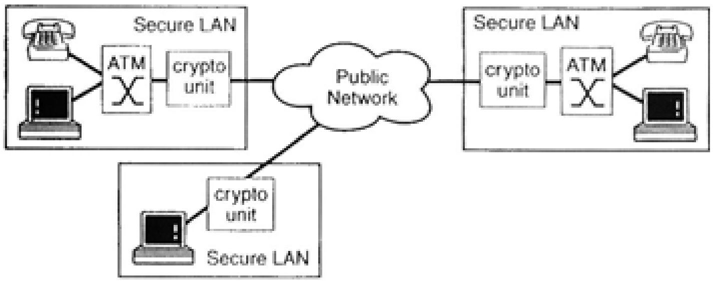


Figure 1: ATM Security Components (Schematic Representation)

In fig 1 shows the Schematic representation of ATM security components in which each VPN (Virtual Private Network) has a switching device to enter into the network (Public Network). Between switching device & network there is a Crypto unit inserted. This crypto unit performs all the encryption & decryption work. Communication has to be between Network to Network or User to Network. For maintaining the security & privacy few functions should be maintained which we will discuss now.

AF-SEC-1 Verification of Identities

The ATM network shall support capabilities to establish and verify the claimed identity of any actor in an ATM network.

In this function the basic authentication should be done. Authentication is done to avoid Masquerade. Few security services should be made available for this purpose.

- User Authentication
- Data Origin Authentication
- Peer Entity Authentication

AF SEC-2 Controlled Access and Authorization

The ATM network shall support capabilities to ensure that actors are prevented from gaining access to information or resources they are not authorized to access.

The access control method decides whether the connection is authenticated or not. If the connection is not authenticated, it will not proceed further otherwise the connection will be initialized. It is very important for the multilevel secure ATM with trusted component.

AF SEC-3 Protection of Confidentiality

The ATM network shall support the capability to keep stored and communicated data confidential.

Protection of confidentiality is used to protect user related information. The confidentiality service provides the protection for the disclosure of exchanged data to the unauthorized access.

AF SEC-4 Protection of Data Integrity

The ATM network shall support granting the integrity of stored and communicated data. Protection of data integrity is used to protect ATM network user related information. The integrity service ensures the correctness of exchanged data, insertion, deletion and modification of the new data.

AF SEC-5 Strong Accountability

The ATM network shall support the capability that an entity cannot deny the responsibility for any of its performed actions as well as their effects.

Strong accountability means Non-repudiation. In this one has to prove that data has actually taken place. It is very important for everyone to be responsible for his work.

AF SEC-6 Activity Logging

The ATM network shall support the capability to retrieve information about security activities stored in the Network Elements with the possibility of tracing this information to individuals or entities.

Activity logging is for controlling security policies. It is necessary to log information about security related events which occurs security relevant operations.

AF SEC-7 Alarm Reporting

The ATM network shall support the capability to generate alarm notifications about certain adjustable and selective security related events.

In alarm reporting, security information is provided which provides information about security relevant events.

AF SEC-8 Audit

The ATM network shall support the capability to analyze and exploit logged data on security relevant events in order to check them on violations of system and network security.

An audit is to test sufficiency of system control.

AF SEC-9 Security Recovery

The ATM network shall support recovery from successful and attempted breaches on security.

A very frequent problem in cell encryption is the loss of cell. If cells are lost decryption will not be possible. Some modes operate on the algorithm of handling lost cells.

AF-SEC-10 Management of Security

The ATM network shall support capabilities to manage the security services derived from the security requirements listed above.

Management of security comprises of important aspects of systems, which includes all activities to establish, maintain & terminate.

Conclusion

This topic of ATM safety and security proves that there is a requirement of strong security specification for ATM network and on the other hand there are many pitfalls and numerous problems.

Authentication, confidentiality and data integrity are the important security framework that fulfills the user needs for secure network. So ATM safety and security provides strong protection of user security and safety and offers the new possibilities to make a network strong.

References

1. The ATM Forum Technical Committee, 'ATM Security Framework 1.0', AF-SEC- 0096.000, February 1998.

2. The ATM Forum Technical Committee, 'ATM Security Specification 1.0', ATM-SEC- 0100.001, February 1999.

3. Donglin Liang, Ohio State University, 'A Survey on ATM Security', August 1997.

4. Daniel Stevenson, Nathan Hillery, and Greg Byrd, 'Secure Communications in ATM Networks', Communication of the ACM, ISSN 0006-0786, Volume 38, Number 2, pages 45-52, February 1995.

5. Mohammad Peyravian, IBM Corporation, and Els Van Herreweghen, IBM Research Laboratory, 'ATM Security Scope and Requirements', ATM Forum/95-0579, June 1995.

Intelligent Malware Detection System

Sandeep B. Damodhare[1], Prof. V. S. Gulhane[2]

[1]ME Student, Dept of IT, SIPNA's College of Engineering &
Technology, Amravati (MS) INDIA

[2]Associate Professor, Dept of CSE, SIPNA's College of Engineering
& Technology, Amravati (MS) INDIA

ABSTRACT

Malicious programs spy on users' behavior and compromise their privacy. Unfortunately, existing techniques for detecting malware and analyzing unknown code samples are insufficient and have significant shortcomings. We observe that malicious information access and processing behavior is the fundamental trait of numerous malware categories breaching users' privacy (including key loggers, password thieves, network sniffers, stealth backdoors, spyware and root kits), which separates these malicious applications from benign software. Commercial anti-virus software is unable to provide protection against newly launched ("zero-day") malware. In this dissertation work, we propose a novel malware detection technique which is based on the analysis of byte-level file content. The proposed dissertation work will demonstrate the implementation of system for detection of various types of malware.

1. Introduction

Malicious software (i.e., Malware) creeps into users' computers, collecting users' private information, wrecking havoc on the Internet and causing millions of dollars in damage. Malware detection and analysis is a challenging task, and current malware analysis and detection techniques often fall short and fail to detect many new, unknown malware samples. Current malware detection methods in general fall into two categories: signature- based detection and heuristics based detection. The former cannot detect new malware or new variants. The latter are often based on some heuristics such as the monitoring of modifications to the registry and the insertion of hooks into certain library or system interfaces. Since these heuristics are not based on the fundamental characteristics of malware, they can incur high false positive and false negative rates. For example, many benign software access and modify registry entries. Hence, just because an application creates hooks in the registry does not mean that it is malicious (i.e., the application could be a useful system utility). Furthermore, to evade detection, malware may attempt to hook library or system call interfaces that the detector does not monitor. Even worse, since many rootkits hide in the kernel, most such heuristics-based detectors cannot detect them as they do not necessarily modify any visible registry entries or library or system call interfaces.

Malware is software designed to infiltrate or damage a computer system without the owner's informed consent (e.g., viruses, backdoors, spyware, trojans, and worms) [1]. Numerous attacks made by the malware pose a major security threat to computer users. Hence, malware detection is one of the computer security topics that are of great interest. Currently, the most important line of defense against malware is antivirus programs, such as Norton, MacAfee, and Kingsoft's Antivirus. These widely used malware detection software tools use signature-based method to recognize threats. Signature is a short string of bytes, which is unique for each known malware so that future examples of it, can be correctly classified with a small error rate. However, this classic signature-based method always fails to detect variants of known malware or previously unknown malware, because the malware writers always adopt techniques like obfuscation to bypass these signatures [2]. In order to remain effective, it is of paramount importance for the antivirus companies to be able to quickly analyze variants of known malware and previously unknown malware samples. Unfortunately, the number of file samples that need to be analyzed on a daily basis isconstantly increasing [3]. According to the virus analysts at Kingsoft Antivirus Laboratory, the “gray list” that is needed to be analyzed per day usually contain more than 70000 file samples. Clearly, there is a need for an automatic, efficient, and robust tool to classify the “gray list.”

Literature Review/related Work:

Recently, many post processing techniques, including rule pruning, rule ranking, and rule selection have been developed for associative classification to reduce the size of the classifier and make the classification process more effective and accurate [5], [6], [14]. It is interesting to know how these post processing techniques would help the associative classifiers for malware detection. In this paper, we systematically evaluate the effects of the post processing techniques in malware detection and propose an effective way, i.e., CIDCPF, to detect the malware from the “gray list.”

1) Rule Pruning: In order to reduce the size of the classifier and make the classification process more effective and accurate, the removal of the redundant or misleading rules is indispensable. There are five popular rule pruning approaches which mainly focus on preventing these redundant or misleading rules from taking any part in the prediction process of test data objects.

A) χ2 (chi-square) test : The test is always carried out on each generated rule to find out whether the rule's antecedent is positively correlated with the rule's consequent. It is adopted by classification based on multiple association rules (CMAR) algorithm in its rule discovery step.

B)Redundant rule pruning: This rule pruning method discards specific rules with fewer confidence values than general rules. Several algorithms, such as CMAR [1], [2], and [15], adopt this approach for rule pruning.

C) Database coverage: This pruning approach tests the generated rules against the training dataset, and only keeps the rules, which cover at least one training data object not considered by a higher ranked rule for later classification. This method is created by the classification based on associations (CBA) , and used by CMAR [15], and multiclass classification based on association rule (MCAR) .

D) Pessimistic error estimation: The method works by comparing the estimated error of a new rule. If the expected error of the new rule is lower than that of the original rule, then the original rule will be replaced by the new rule. CBA have used it to effectively reduce the number of the generated rules.

E)Lazy pruning: This method discards the rules, which incorrectly classify training objects and keeps all others. It has been used in [5] for rule pruning.

2) Rule Ranking: Rule ranking plays an important role in the classification process, since most of the associative classification algorithms, such as CBA, CMAR [4], [5], multiclass, multilabel associative classification (MMAC) , and MCAR, utilize rule ranking procedures as the basis for selecting the classifier. Particularly, CBA and CMAR use database coverage pruning approach to build the classifiers, where the pruning evaluates rules according to the rule ranking list. Hence, the highest order rules are tested in advance, and then, inserted into the classifier for predicting test data objects.

For rule ranking, there are five popular ranking mechanisms :

a) confidence support size of antecedent (CSA);
b) size of antecedent confidence support (ACS);
c) weighted relative accuracy (WRA);
d) Laplace accuracy; and

CSA and ACS are belonging to the pure "support-confidence" framework and have been used by CBA and CMAR for rule ranking. WRA, Laplace accuracy, and χ2 measure are used by some associative classification algorithms, such as classification based on predictive association rules (CPAR), to weigh the significance of each generated rule.

3) Rule Selection: After pruning and reordering the generated rules, we can select the subset of the rules from the classifier to predict new file samples. There are three common rule selection approaches :

a) Best first rule: This approach selects the first best rule that satisfies the given data object according to the rule list based on certain rule ranking mechanism to predict the new data object. It is used in CBA for predicting test data objects.

b) All rules: This method collects all rules in the classifier that satisfy the new data object, and then, evaluate this collection to identify its class. CMAR uses weighted χ2 (WCS) testing to predict the class of the new data object.

c) Best k rules: Some associative classification algorithms, like CPAR, select the first best k rules that satisfy the new data object, and then, make predictions using certain averaging process.

Analysis Of Problem:

1) Systematically evaluate the effects of the post processing techniques in malware detection.

2) Propose an effective way CIDCPF, to detect the malware from the "gray list." CIDCPF adapts several different post processing techniques of associative classification, including rule pruning, rule ranking, and rule selection, for building effective associative classifiers.

3) Improve former malware detection system IMDS and update it to IMDS.

4) Perform cases studies on a large collection of executables including 35000 malicious ones and 15000 benign samples, collected by the Antivirus Laboratory of Kingsoft Corporation.

5) Provide a comprehensive experimental study on various antivirus software as well as various data mining techniques for malware detection using our data collection.

Implementation

System Architecture

The system architecture of our malware detection is shown in Fig. 1.Basically, the system first uses the feature extractor to extract the API calls from the collected portable executable (PE) files, converts them to a group of 32-bit global IDs as the features of the training data, and stores these features in the signature database.

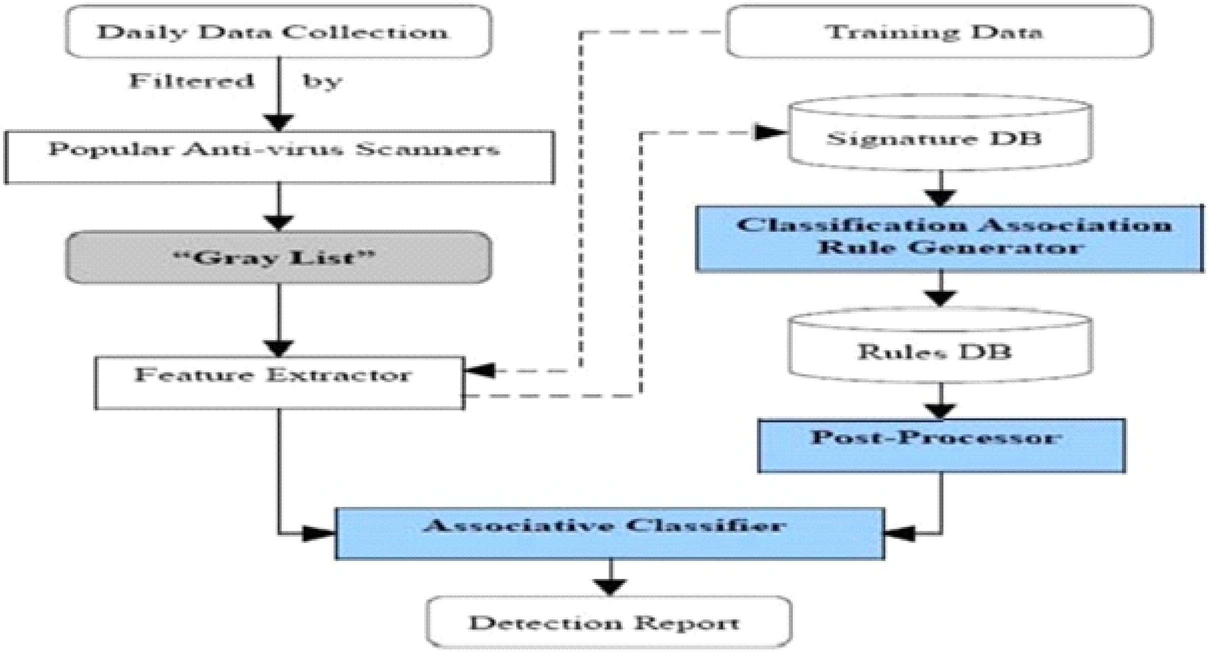


Fig. 1. Flow of malware detection

After data transformation, it then generates the classification association rules from the training Signature database. In the third step, it adapts hybrid post processing techniques of associative classification, including rule pruning, rule ranking, and rule selection to reduce the generated rules. Finally, it builds the classifier using the rules filtered by the postprocessor to detect malware from the "gray list." We will describe the details of each step in the following sections.

Classification Association Rule Generation

Associative classification, as a new classification approach integrating association rule mining and classification, becomes one of the significant tools for knowledge discovery and data mining. It can be effectively used in malware detection , since frequent item sets are typically of statistical significance and classifiers based on frequent pattern analysis are generally effective to test datasets. In this section, we briefly discuss the generation of rules for classification.

A. Data Collection and Transformation

We obtain 50000 Windows PE files of which 15000 are recognized as benign executables and the remaining 35000 are malicious executables. PE is designed as a common file format for all flavors of Windows operating system, and PE malicious executables are in the majority of the malware rising in recent years. All the file samples are provided by theAntivirus Laboratory of Kingsoft Corporation, and the malicious executables mainly consist of backdoors, spyware, trojans, and worms. Based on the system architecture of our previous malware detection system IMDS , we extract the API calls as the features of the file samples and store them in the signature database.

B. Classification Association Rule Generation

For malware detection in this paper, the first goal is to find out how a set of API calls supports the specific class objectives: class1 = malicious, and class2 = benign.

1) Support and confidence: Given a dataset DB, let I = {I1, . . . , Im} be an itemset and I → class(os, oc) be an association rule whose consequent is a class objective. The support and confidence of the rule.

2) Where the function count (I U {class}) returns the number of records in the dataset DB where I U {class} holds.

3) Frequent item set: Given mos as a user-specified minimum support. I is a frequent item set/pattern in DB if os ≥ mos.

4) Classification association rule: Given moc as a user specified confidence.

Let I = {I1, . . . , Im} be a frequent item set. I → class(os, oc) is a classification association rule if oc ≥ moc. Apriori and FP-Growth algorithms can be extended to associative classification. For rule generation, we use the OOA_Fast_FP-Growth algorithm proposed in to derive the complete set of the rules with certain support and confidence thresholds, since it is much faster than Apriori for mining frequent item sets. The number of the rules is also correlated to the number of the file samples.

Postprocessing Techniques Of Associative Classification For Malware Detection System

The goal of our malware detection system is to build classifier using the generated rules to classify the new file samples more effectively and accurately, so the postprocessing of associative classification is very important for improving the system's ACY and efficiency. The postprocessing techniques includes rule pruning, rule ranking, and rule selection.

A. Rule Pruning Approaches

Accompanied with the ability of mining the complete set of the rules, associative classification also has a major drawback that the number of generated rules can be really large and the removal of the redundant or misleading rules is indispensable. Besides the five common rule pruning approaches introduced:

1) $\chi 2$ (chi-square) testing tomeasure the significance of the rule itself;

2) redundant rule pruning to discard the specific rules with fewer confidence values;

3) database coverage to just keep the rules covering at least one training data object not considered by a higher ranked rule;

4) pessimistic error estimation to test the estimated error of a new rule; and

5) lazy pruning to discard the rules incorrectly classifying the training objects, we here propose another rule pruning method before building the classifier, named “insignificant rules pruning.”

Since many generated rules are redundant or minor variations of others and their existence may simply be due to chance rather than true correlation these insignificant rules should be removed.

For example, given the rule: “R1: Job = yes → Loan = approved (supp=35%, conf=80%),” the following rule: “R2: Job=yes, Oversea_asset ≥ 500k→Loan = approved (supp= 32%, conf=81%)” becomes insignificant because it gives little extra information.

We use χ2 measure , which is based on the comparison of observed frequencies with the corresponding expected frequencies, to test whether the rule is significant w.r.t. to its ancestors. Given two rules generated from the training set T consisting of n data objects

R1: A→ r-class (supp = s1 , conf = c1)

R2: AB →r-class (supp = s2 , conf = c2)

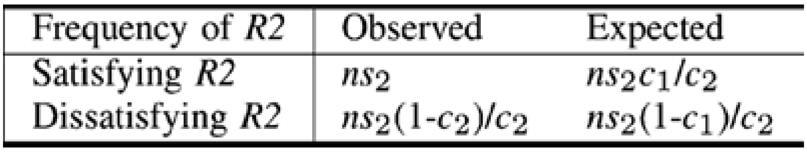
where A, B are the frequent itemsets (A ∩ B = φ) of the generated rules and r-class is the class label of T. If these two rules have the same class label, then we call R1 the ancestor of R2 (or R2 the descendant of R1) [38]. If c1 ≥ c2 , namely the confidence of R1 is not greater than its ancestor R2, then R2 is insignificant and can be pruned.

If c1 < c2 , we set up the hypothesis H0 that the two patterns A and B are independent. We then compute the observed and expected frequencies of R2 as shown in Table I. We later use χ2 measure to test the significance of the deviation from the expected values. Let f0 be an an observed frequency and f be an expected frequency.

The χ2 value is defined as: $\chi 2 = \_(f0 - f)/f.$

Table I

Observed And Expected Frequencies Of R2



Frequency of $R2$	Observed	Expected
Satisfying $R2$	ns_2	ns_2c_1/c_2
Dissatisfying $R2$	$ns_2(1\text{-}c_2)/c_2$	$ns_2(1\text{-}c_1)/c_2$

Given a certain threshold value (e.g., 3.84 at the 95% significance level), if the χ2 measure is above the threshold value, then we reject the hypothesis and keep R2, otherwise, we will accept the assumption and discard R2 From the aforementioned six rule pruning approaches, we empirically study four of them for malware detection: the redundant rule pruning and lazy pruning approaches are not used.

b) Rule Ranking Mechanisms

Within the associative classification framework, regardless of which particular methodology is used to generate the rules, a classifier is usually represented as an order list of the generated rules based on some rule ranking mechanisms. Many associative classification algorithms [4], [5], [15], utilize rule ranking procedures as the basis for selecting the classifier during pruning and later for predicting new data objects. As we discussed in Section II, there are five common ranking mechanisms: CSA, ACS, WRA, Laplace accuracy, and χ2 measure. Here, we give a more detailed introduction.

1) CSA: Based on the well-established "support-confidence" framework, CSA first sorts the original rule list based on their confidence values in a descending order. For those rules that share a common confidence value, CSA sorts them in a descending order based on the support values. CSA sorts the rules sharing common values for both confidence and support in an ascending order based on the size of the rule antecedent.

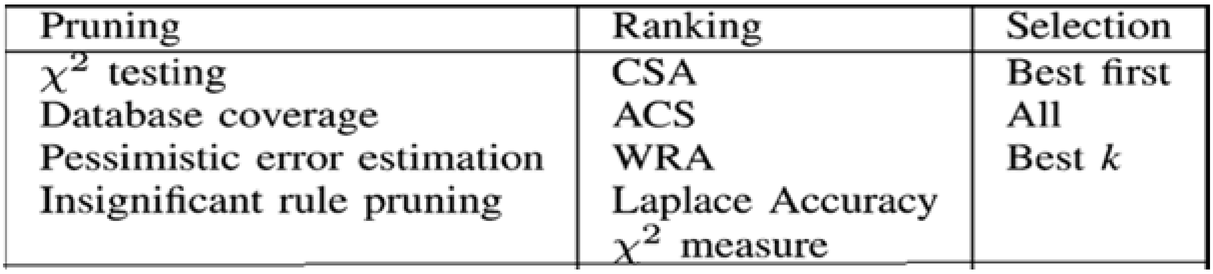
2) ACS: Ensuring that "specific rules have a higher precedence than more general rules" [10], ACS considers the size of the rule antecedent as the most significant factor (using a descending order) followed by the rule confidence and support values, respectively .

3) WRA: WRA assigns an additive weighting score to each rule to determine its expected ACY. The calculation of the value of a rule r is: WRA(r) = supp (r.antecedent)*(conf (r)-supp (r.consequent)) [4]. In the rule reordering stage, the original rule list is sorted based on the assigned WRA value in a descending order.

4) Laplace accuracy: The principle of Laplace accuracy is similar to WRA. The calculation of the Laplace value of a rule r is Laplace (r) =(supp (r.antecedent ∪ r.consequent) + 1)(supp (r.antecedent) + c) where c represents the number of predefined classes.

5) χ2 measure: In associative classification algorithms, if the χ2 measure between two variables (the antecedent and consequent-class of the generated rule) is higher than a certain threshold value, we can conclude that there might be a relation between the rule antecedent and consequent-class, otherwise, it implies that the two variables may be statistically independent. We can order the list of the generated rules in a descending order based on their χ2 values. For the aforementioned five rule ranking mechanisms, we empirically study all of them for building the classifier and later for detecting the new malware.

Table II Adapting Postprocessing Techniques Of Associative Classification For Malware Detection System



Pruning	Ranking	Selection
χ^2 testing	CSA	Best first
Database coverage	ACS	All
Pessimistic error estimation	WRA	Best k
Insignificant rule pruning	Laplace Accuracy	
	χ^2 measure	

c) Rule Selection Methods

After building the classifier by the techniques of rule pruning and rule ranking, we can select the subset of the rules from the classifier to predict the new file samples. As stated in Section II, there are three common rule selection approaches: best first rule, all rules, andbest k rules. For our malware detection system, we will also try all of these methods to predict the new file samples and find the best way for malware detection. The postprocessing techniques, which will be perform in malware detection, can be summarized in Table II.

Proposed Work

In the proposed dissertation work intelligent malware detection system will be implemented. The dissertation work will be carried out as follows.

1. Analysis of available malware detection systems.
2. Evaluation of how these systems complement each other to improve detection rates.
3. Implementation of malware detection system for detection of denial of service and backdoor.
4. Analysis of malware detection results.

Empirical Study Of Postprocessing Techniques For Malware Detection

A. Experiment Setup

We randomly select 17828 executables from our data collection, including 9721 benign executables, 2255 backdoors, 2245 spyware, 3200 Trojans, and 2021 worms in the training dataset. The rest 17828 executables are used for testing purpose of which 7700 are benign files and 2021 are malicious ones. After filtering some of the worthless API calls, we finally extract 5102 API calls from the training dataset. By using the OOA_Fast_FP-Growth algorithm [35], [36], we generate 31 rules with the minimum support and confidence as 0.18 and 0.5, respectively for the benign class, while 8424 rules are derived with the minimum support and confidence as 0.25 and 0.7, respectively for the malicious class.

To systematically evaluate the effects of postprocessing techniques for malware detection, we conduct the following three sets of experimental studies using our collected data obtained from the Antivirus Laboratory of Kingsoft Corporation. The first set of study is to compare the ACY and efficiency of the three different associative classifier building algorithms: CBA, CMAR, and CPAR [37], when used for malware detection system. Since none of the three algorithms adopt the insignificant rule pruning approach, in the second set of study, we prune the insignificant rules before building the classifier. From these two sets of studies, we will choose the best rule pruning and rule selection methods for malware detection. In third set of experiments, we will compare the five rule ranking mechanisms and find the best ranking method for malware detection. Please note in all the experiments, rule mining, selection, and ranking are performed only within training data. From the three set of experiments, we will propose an effective classifier building method and incorporate it to our improved malware detection system IMDS.

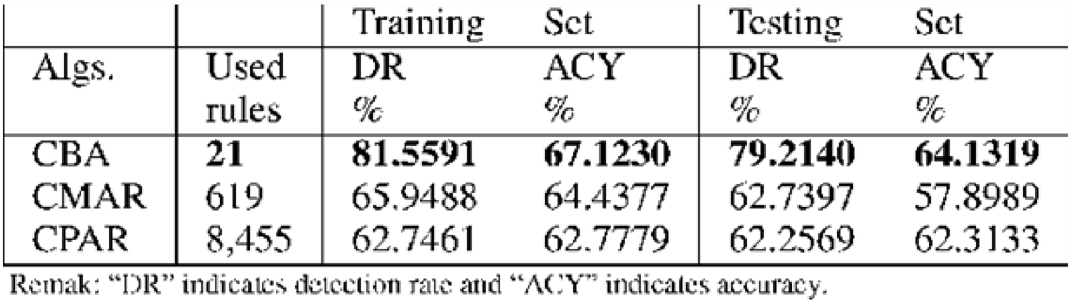
B. Comparisons of CBA, CMAR, and CPAR for Malware Detection

Since the algorithms of CBA, CMAR, and CPAR have been successfully used in associative classification and represent different kinds of postprocessing techniques for building the classifiers, in the first set of experiments, we use them for malware detection and compare their ACY and efficiency. The postprocessing techniques adopted by these three algorithms are listed in Table III.

The datasets described in Section VI-A are used for training and testing. In this paper, we use DR and ACY defined as follows to evaluate each classifier building method.

1) True positive (TP): The number of executables correctly classified as malicious code.
2) True negative (TN): The number of executables correctly classified as benign executables.
3) False positive (FP): The number of executables mistakenly classified as malicious executables.
4) False negtive (FN): The number of executables mistakenly classified as benign executables.
5) DR: TP/(TP + FN).
6) ACY: TP + TN/(TP + TN + FP + FN).

Table Iv Results Of Cba, Cmar, And Cpar Classifier Building Method Used In Malware Detection System



		Training	Set	Testing	Set
Algs.	Used rules	DR %	ACY %	DR %	ACY %
CBA	**21**	**81.5591**	**67.1230**	**79.2140**	**64.1319**
CMAR	619	65.9488	64.4377	62.7397	57.8989
CPAR	8,455	62.7461	62.7779	62.2569	62.3133

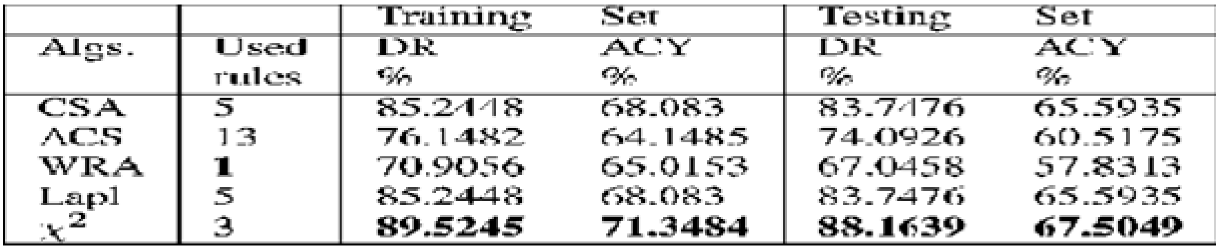
Remak: “DR” indicates detection rate and “ACY” indicates accuracy.

The experimental results shown in Table IV indicate that CBA classifier building method performs better than the other two for malware detection.

Comparisons of Different Rule Ranking Mechanisms

In this section, we compare the five rule ranking mechanisms and find the best one for malware detection. Results in Table V illustrate that χ2 measure rule ranking mechanism performs best.

Table V Results By Using Different Rule Ranking Mechanisms In Malware Detection System



		Training	Set	Testing	Set
Algs.	Used rules	DR %	ACY %	DR %	ACY %
CSA	5	85.2448	68.083	83.7476	65.5935
ACS	13	76.1482	64.1485	74.0926	60.5175
WRA	1	70.9056	65.0153	67.0458	57.8313
Lapl	5	85.2448	68.083	83.7476	65.5935
χ^2	3	**89.5245**	**71.3484**	**88.1639**	**67.5049**

Comparisons of Intelligent Malware Detection System (IMDS) With Other Systems

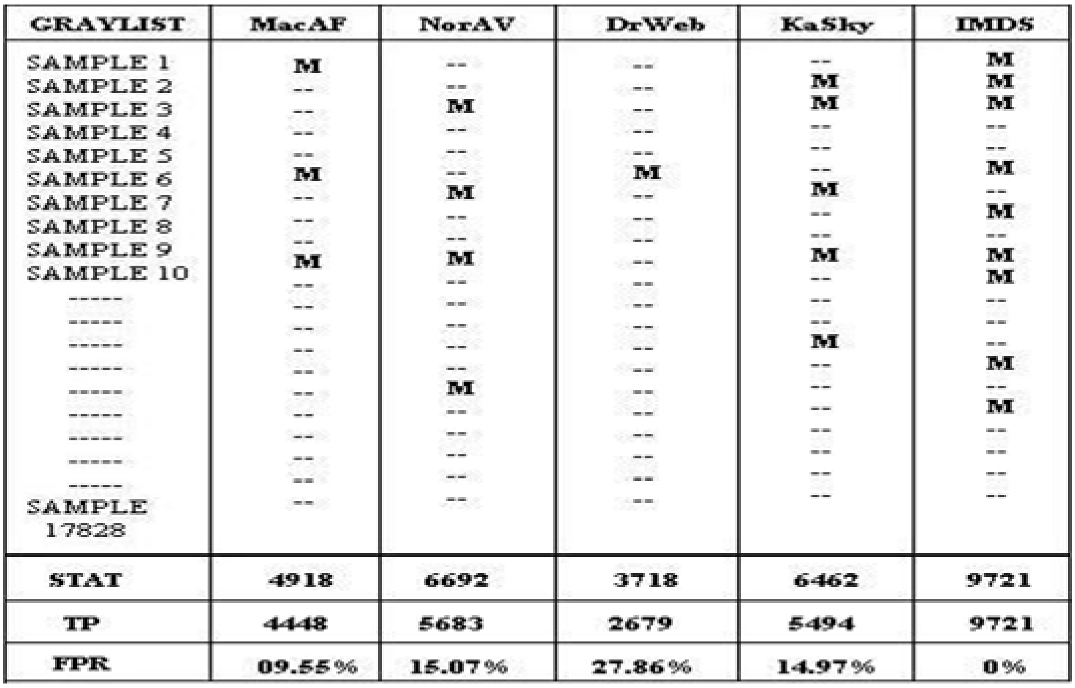
In this section, we conduct two sets of experiments to compare our IMDS system with other malware detection system: 1) the first set of experiments is to examine the abilities of detecting the malware from the "gray list" of our IMDS system, in comparison with some of the popular software tools, such as McAfee Virus Scan, Norton AntiVirus, Dr.Web, and Kaspersky AntiVirus. We use fair versions of the base of signature on the same day (17 May, 2012) for testing. The efficiency by using different scanners have also been examined. 2) In the second set of experiments, resting on the analysis of API calls, we compare our malware detection system IMDS with other classification based methods and Decision tree.

Comparisons of Detection Results From the Gray List

Since the goal of our improved malware detection system is to help our virus analysts picking up as many malware samples as possible from the "gray list," which consists of millions of executables, we perform the experiments based on the "gray list" in this section. We randomly select 17828 file samples from the "gray list." Table VII shows the detection results of different antivirus scanners. Of the 17828 samples from the "gray list," 4572 are detected as malware by the five scanners in total. These detected results of the scanners should be reviewed by our virus analysts, since they have false positive rate (FPR). The FPR of each scanner results from the disability of recognizing the benign software adoptingobfuscation technique in clients, such as the instant message (IM) software "QQ". Our virus analysts perform analysis on the detected files and find that 3015 of them are correctly detected. We then calculate the DR and ACY for each scanner. The statistical results are shown in Table V. From Tables VI and VII , we can see that our IMDS system outperform other antivirus software for malware detection from the "gray list."

Table VI

Detection Result From The "GRAYLIST":-



GRAYLIST	MacAF	NorAV	DrWeb	KaSky	IMDS
SAMPLE 1	M	--	--	--	M
SAMPLE 2	--	--	--	M	M
SAMPLE 3	--	M	--	M	M
SAMPLE 4	--	--	--	--	--
SAMPLE 5	--	--	--	--	--
SAMPLE 6	M	--	M	--	M
SAMPLE 7	--	M	--	M	--
SAMPLE 8	--	--	--	--	M
SAMPLE 9	--	--	--	--	--
SAMPLE 10	M	M	--	M	M
-----	--	--	--	--	M
-----	--	--	--	--	--
-----	--	--	--	--	--
-----	--	--	--	M	--
-----	--	--	--	--	M
-----	--	M	--	--	--
-----	--	--	--	--	M
-----	--	--	--	--	--
-----	--	--	--	--	--
	--	--	--	--	--
SAMPLE 17828	--	--	--	--	--
STAT	4918	6692	3718	6462	9721
TP	4448	5683	2679	5494	9721
FPR	09.55%	15.07%	27.86%	14.97%	0%

IMDS => Intelligent Malware Detection System

M => Indicates the file in the "GRAYLIST" is detected as MALWARE.

STAT => Total Numbers of file detected as malware.

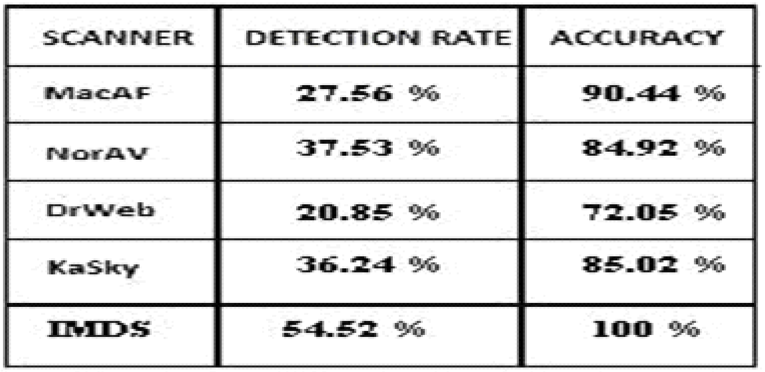
TP => Numbers of correctly detected file.

FPR => Mistakenly classified as MALWARE.(FALSE POSITIVE RATE)

– => Scanner default files

Table VII

Statistical Results Of Dr And Acy



SCANNER	DETECTION RATE	ACCURACY
MacAF	27.56 %	90.44 %
NorAV	37.53 %	84.92 %
DrWeb	20.85 %	72.05 %
KaSky	36.24 %	85.02 %
IMDS	54.52 %	100 %

STATISTICAL RESULTS OF DR AND ACY

Application:

Many instances of malware take advantage of features provided by common applications, such as e-mail clients, Web browsers, and word processors. By default, applications often are configured to favor functionality over security. Accordingly, organizations should consider disabling unneeded features and capabilities from applications, particularly those that are commonly exploited by malware, to limit the possible application attack vectors for malware. Organizations should also consider identifying applications that are typical malware propagation methods (e.g., Web browsers, e-mail clients and servers) and configuring them to filter content and stop other activity that is likely to be malicious. Spam is often used for phishing and spyware delivery (e.g., Web bugs often are contained within spam), and it sometimes contains other types of malware. Using spam filtering software on e-mail servers or clients or on network-based appliances can significantly reduce the amount of spam that reaches users, leading to a corresponding decline in spam-triggered malware incidents.

Conclusion:

We systematically evaluate the effects of the post processing techniques (e.g., rule pruning, rule ranking, and rule selection) of associative classification in malware detection and propose an effective way, i.e., CIDCPF, to detect the malware from the "gray list." We are using post processing techniques of associative Classification in malware detection. Experiments on a large real data collection from Antivirus Laboratory at Kingsoft Corporation demonstrate among the most common and popular associative classification building methods, our CIDCPF method achieves better performance on detection ability andefficiency because of its concise, but effective classifier. In addition, our IMDS system, which adopts CIDCPF method for building classifiers can greatly reduce the number of generated rules and make it easy for our virus analysts to identify the useful ones.

References:

[1] M. Antonie and O. Zaiane, "An associative classifier based on positive and negative rules," in Proc. 9th ACM SIGMOD Workshop Res. Issues Data Mining Knowl. Discovery, 2004, pp. 64–69.

[2] D. Brumley, C. Hartwig, M. G. Kang, Z. Liang, J. Newsome, D. Song, and H. Yin. BitScope: Automatically dissecting malicious binaries. Technical Report CMU-CS-07- 133, School of Computer Science, Carnegie Mellon University, March 2007.

[3] D. Brumley, C. Hartwig, Z. Liang, J. Newsome, D. Song, and H. Yin. Botnet Analysis, chapter Automatically Identifying Trigger-based Behavior in Malware. 2007.

[4] M. Costa, J. Crowcroft, M. Castro, A. Rowstron, L. Zhou, L. Zhang, and P. Barham. Vigilante: End-to-end containment of internet worms. In Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP'05), October 2005.

[5] J. R. Crandall and F. T. Chong. Minos: Control data attack prevention orthogonal to memory model. In Proceedings of the 37th International Symposium on Microarchitecture (MICRO'04), December 2004.

[6] M. Egele, C. Kruegel, E. Kirda, H. Yin, and D. Song. Dynamic Spyware Analysis. In Proceedings of the 2007 Usenix Annual Conference (Usenix'07), June 2007.

[7] P. Ferrie. Attacks on virtual machine emulators. Symantec Security Response, December 2006.

[8] H. Cheng, X. Yan, J. Han, and P. S. Yu, "Direct discriminative pattern mining for effective classification," in Proc. ICDE-2008, pp. 169–178.

[9] H. Cheng, X. Yan, J. Han, and C. Hsu, "Discriminative frequent pattern analysis for effective classification," in Proc. ICDE-2007, pp. 716–725.

[10] M. Christodorescu, S. Jha, and C Kruegel, "Mining specifications of malicious behavior," in Proc. ESEC/FSE-2007, pp. 5–14.

[11] F. Coenen and P. Leng, "An evaluation of approaches to classification rule selection," in Proc. 4th IEEE Int. Conf. Data Mining 2004, pp. 359–362.

[12] X. Jiang and X. Zhu, "vEye: Behavioral footprinting for self-propagating worm detection and profiling," Knowl. Inf. Syst., vol. 18, no. 2, pp. 231– 262, 2009.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

<u>Address of the Editorial Office:</u>

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Notes: