A FORECASTING MODEL FOR INTERNET SECURITY ATTACKS

Alexander D. Korzyk, Sr, Virginia Commonwealth University
J. G. VanDyke & Associates
4738 Cedar Cliff Rd
Chester VA 23831
804.748.8590
Internet address: akorzyk@acm.org

ABSTRACT

All sizes of enterprises and organizations in the Internet World face the plaguing problem of Internet Security. One of the biggest problems of Internet Security is getting personnel to report an attack or a suspected attack. Even large enterprises and organizations which have great resources at their disposal cannot get their personnel to accurately report more than 1% of Internet Security attacks. Unfortunately, most enterprises rely on Internet Advisories to defend themselves from an attack which already occurred in another enterprise. Faced with the dilemma of budgeting to protect against Internet attacks, the CIO needs to have statistics to support their request for large sums to defend against Internet attacks. Without personnel reporting a majority of Internet attacks or suspected attacks, the CIO can only make decisions based on a small amount of data gathered from the Internet attacks that personnel did report. This paper presents a forecasting model for Internet attacks which CIOs can use to quantify their Internet Security budget based on a forecasted number of Internet attacks on their enterprise or organization. With the great increase in the number of businesses making a presence on the Internet and the increase in the number of cyber-customers, the chances of computer security attacks increase daily. How much should an enterprise or organization budget to defend against Internet attacks? This paper researches several time series forecasting models to find the best fit model for forecasting Internet Security attacks.

INTRODUCTION

There are literally millions of enterprises and organizations that already conduct business on the World Wide Web and millions more that will in the future. Many are not sure on how much to spend to defend themselves against Internet Security attacks and many are afraid to conduct business on the Web because of the lack of security in their infrastructure and information systems (Row 97). This paper primarily addresses a major problem faced by all enterprises and organizations which conduct business on the Web and secondarily addresses a significant cost for which those enterprises and organizations not yet on the Web must plan. Now as budget cuts have become commonplace and organizations want to get on the World Wide Web without compromising information security, everyone's information becomes available to everyone else if it is not protected properly. Corporations have not fully embraced Electronic Commerce/Electronic Data Interchange (EC/EDI) (Borg 97). The Federal, state, and local governments and corporations have again been hesitant to implement EC/EDI because of the lack of security technology used on the Internet and World Wide Web (Power 97). The governments and corporations also want to use the Web as the infrastructure on which to run EC/EDI. How much should they budget to defend against Internet Security attacks? What forecasting model should enterprises and organizations use to forecast Internet Security attacks?

The WWW could give enterprises the capability to transfer files and documents for a low monthly unlimited access fee to the WWW. The majority of data transfer in smaller enterprises is currently done by dialing up point to point and paying for each minute of phone line usage or by snail mail (the term for regular postage mail). Larger enterprises use electronic networking to

communicate and collaborate internally and externally. The web would also allow access to the enterprise public database for product information. The enterprise could outsource a web page from companies that specialize in running web servers. This Internet service provider would be responsible for making sure that the web server was secure. The Internet solution provider should use a firewall and other web security products to greatly reduce the chances of an attack from both the outside and inside of the enterprise. The Internet service provider would pay for the expensive software that may be required to secure the web server. An enterprise could, instead, purchase, set up, and maintain their own web server. The enterprise would bear the expense of securing the web server. They would still have to support non-web users as well. This course of action would require a decision between the company hiring an Internet specialist to set up their own web server or train a current Information technology specialist to set up their own web server. The chance of losing data or suffering a financial loss due to an outside attack increases dramatically with the personnel lack of security expertise. This decision is one of many faced by the CIO. This forecasting model would provide the CIO with a basis to use a decision making model for Internet Security.

RESEARCH QUESTION

The primary purpose of this research was to find the best fit forecasting model for Internet Security attacks. The research question concerning enterprises and organizations included the following:

Does a more complex forecasting model improve accuracy?
What is the effect of a management change in policy on the forecast?

RESEARCH METHODS

This research uses business and economic forecasting methodology to develop a forecasting model for Internet Security attacks. Many enterprises cannot develop an accurate budget for conducting business on the Web. This research will enable enterprises to forecast the number of Internet security attacks so they can budget to defend against the attacks

Decomposition

The entire area of Internet Security is complex. In order to decompose Internet Security, this paper looks at the most well known subset, Internet Security attacks. The largest obstacle to building a forecasting model is the lack of data. No Internet Security attack data from commercial enterprises could be found in literature and enterprises which were contacted directly would not release any Internet Security attack data. Internet Security attack data was found in some Government organizations. Data from the Department of Defense is a matter of public record and was available for use to construct this model (Howard, 97).

Eclectic Approach

The entire area of Internet Security is new. Since this research is exploratory in nature an eclectic approach was necessary. The World Wide Web is less than 5 years old yet it contains over 90% of all Internet traffic. The Internet itself has been described as almost 25 years old if one narrowly defines it as including the former Advanced Research Planning Agency Network (ARPANET). Another set of data on computer attacks at the Department of Defense conducted by an internal team showed that attackers were successful 65% of the time with only 4% of the successful attacks detected and 27% of the detected attacks reported by the target (GAO, 1996, p. 19).

Graphical Exploration

A graphical exploration of the data set reveals several important characteristics. The data for Figure 1 comes from the Department of Defense (DISA, 1995). The Government collected the data monthly beginning in January, 1989 to October, 1995. There are two distinct shifts of level, the first occurring at the 37th month and the second from the 55th to the 59th month. The full results of the graphical exploration of the data is at Appendix A.

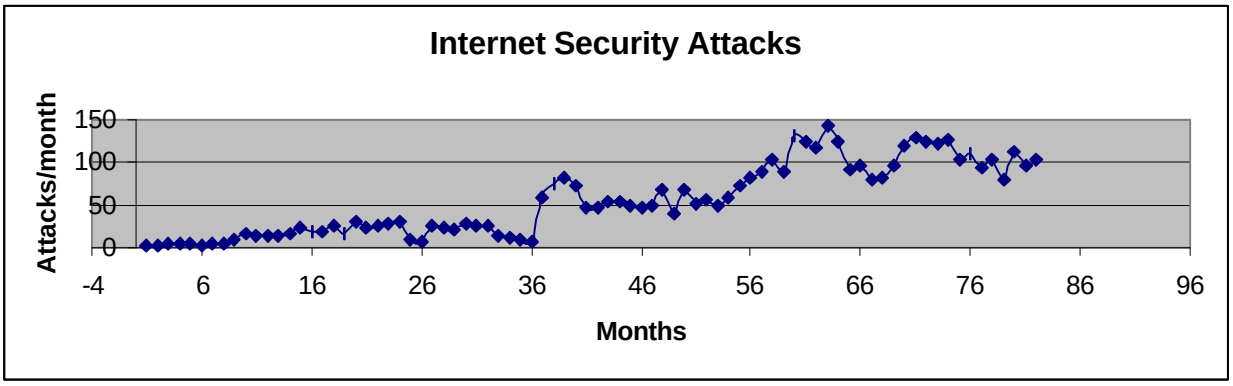


Figure 1--Internet Security Attacks Graph

Trend

Figure 2 shows the same graph as Figure 1 with a trend line based on a twelve month moving average. The trend can be seen to be primarily upward at a slow rate to the 24th month, a flattening out and very slight dip in the 32nd to 36th month, a sharp trend upward from the 37th to the 48th month, another leveling off from the 49th to 55th month, another sharp trend upward from the 56th to the 65th month, then a slight upward trend to the 75th month followed by another flattening out to the end of the data.

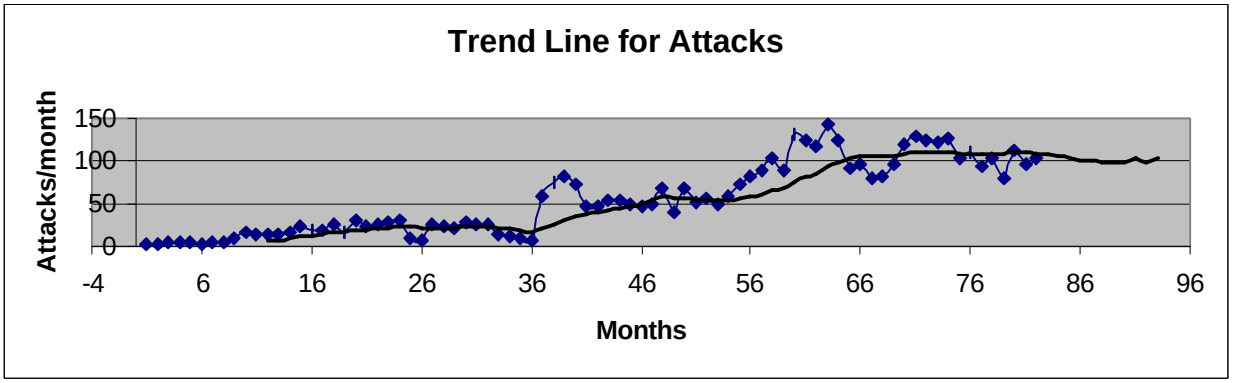


Figure 2—Trend Line for Attacks Graph

Seasonality

The graph of Figure 3 shows the value for each year by month. Three distinct levels appear in this graphical view. The first three years show the first level, the fourth and fifth years show the second level, and the sixth and seventh years show a third level. The first level has little seasonality and only the first year shows an upward trend. The second level starts to show some seasonality but a downward trend cycle for the fourth year and an upward trend cycle for the fifth year. The third level definitely shows seasonality with 150% or more attacks occurring per month in the winter months than in the summer months.

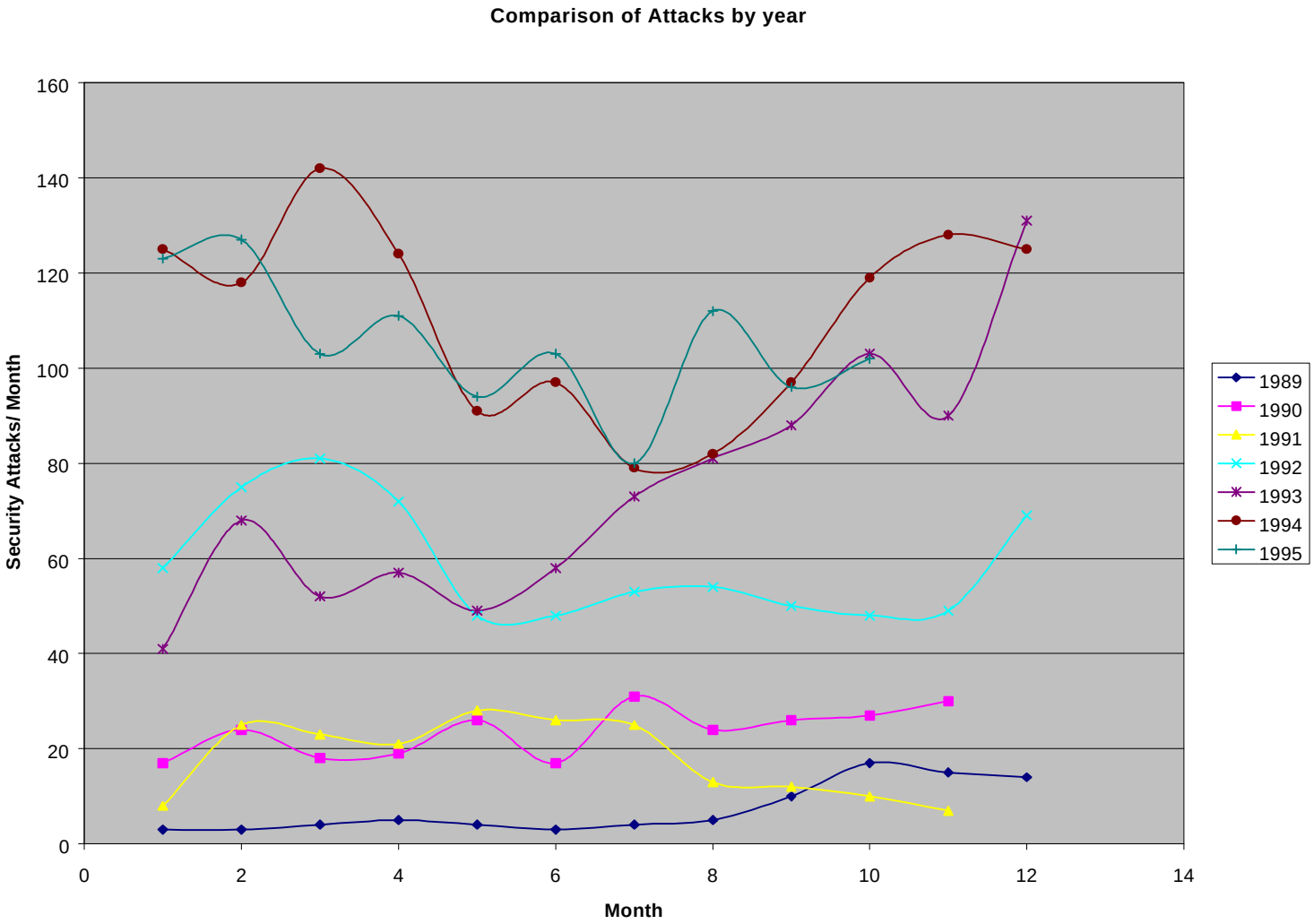


Figure 3—Seasonality graph

Trend-cycle

Figure 4 displays each month of data by year. During the seven years an upward trend is evident overall with cycles occurring every two years. The 3rd, 5th, and 7th years show declines while the 2nd, 4th, and 6th years show increases. The overall trend is upward, approximately doubling every two years.

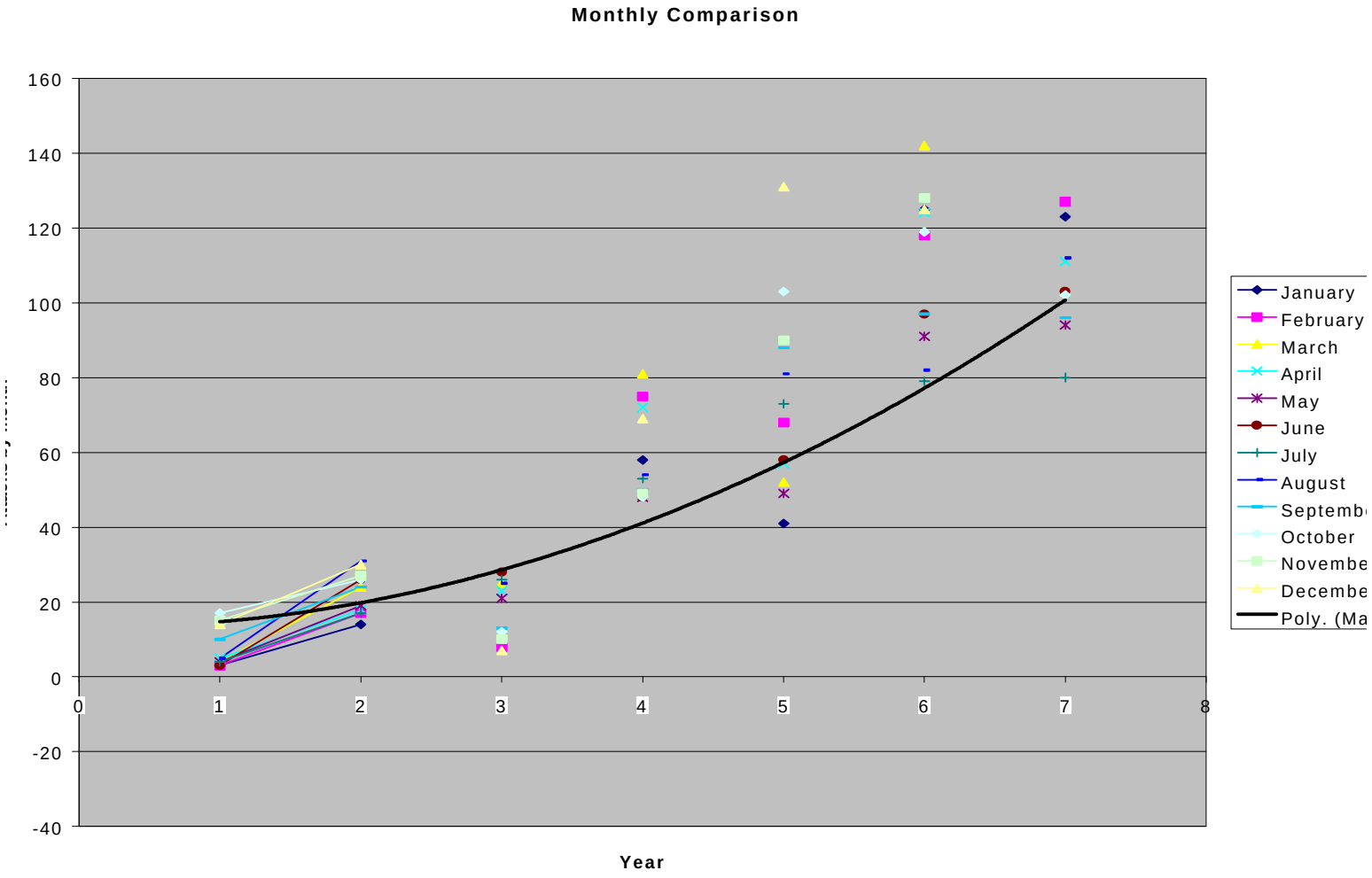


Figure 4—Monthly Trend-Cycle graph

Year-over-year Differences

The trend line in Figure 5 shows an upward trend with cycles similar to Figure 4. The peaks are approximately 24 months apart. If the cycle continues then another sharp upward cycle should begin at the end of the trend line in Figure 5.

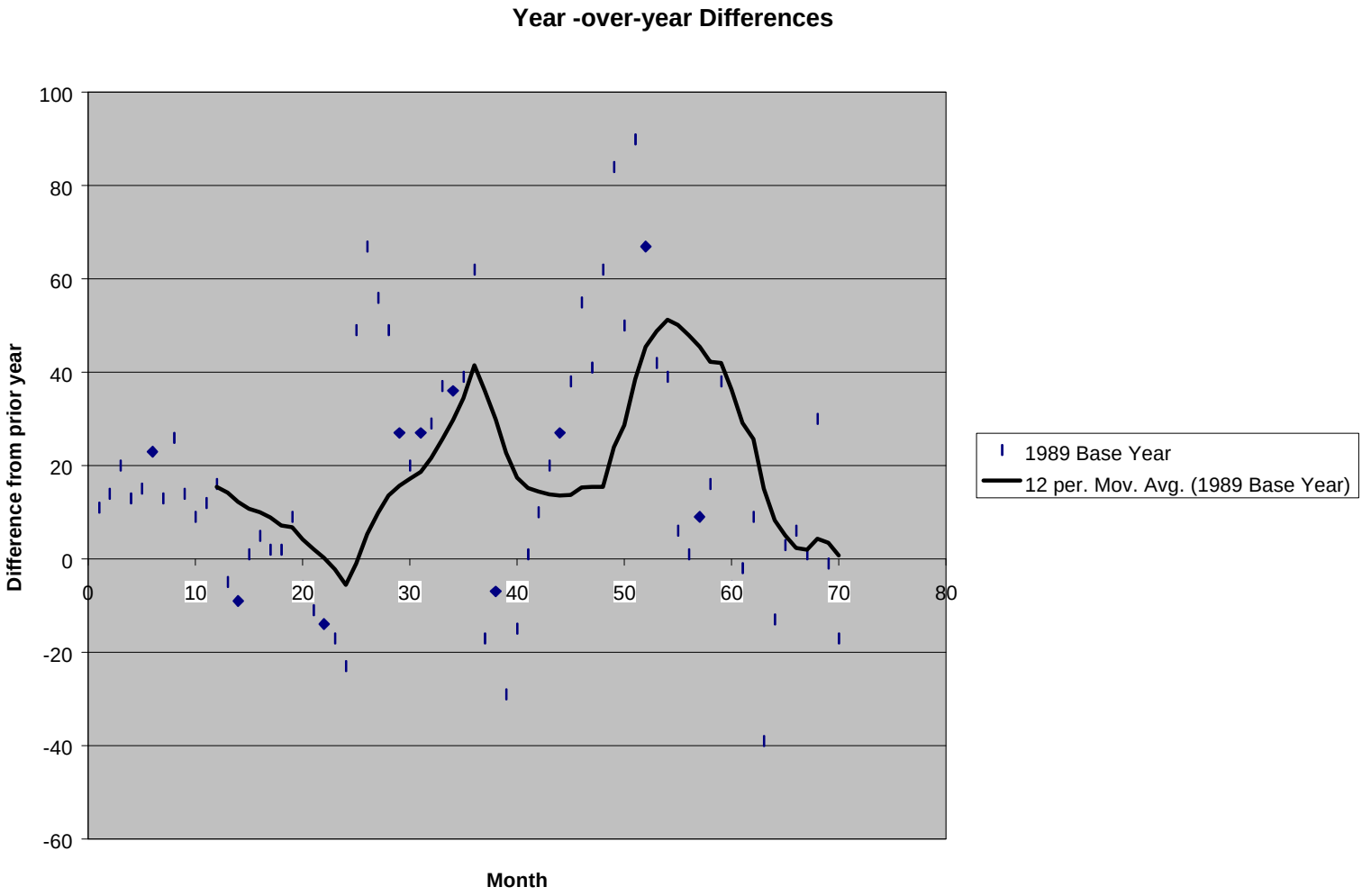


Figure 5—Year-over-year Differences Graph

Evaluation Criteria

This paper examined the decision facing more and more businesses both old and new around the world each year. The number of all companies conducting financial business on the Internet has doubled in one year. Since 1995, the number of companies buying goods and materials on the Internet has increased from 6 to 14%; the number of companies selling goods and materials on the Internet has increased from 5 to 9% (Wilder and Kolbasuk McGee 97).

The decision of how much to budget to defend against Internet Security attacks needs objective evaluation criteria. The paper examines several time series forecasting models. The analysis compares accuracy of the models by calculating the mean square error for each model.

The second criteria is the ability of the model to adjust to shifts caused by policy change from management.

An Ernst and Young survey of all size enterprises conducted in 1996 (Violino 96), revealed that companies attacked had a probability of .05 for a financial loss of over $1,000,000 per attack; companies attacked had a probability of .25 for a financial loss of over $250,000 per attack; and the remaining companies had a probability of .7 for an undetermined amount of financial loss. The Ernst and Young survey further discovered that financial losses come from several causes listed in Table 1 (Violino 96).

Security Problem resulting in financial losses sited above.	Probability of companies with loss from this security problem (independent).
Industrial espionage	.09
Attacks from outside the company	.23
Natural disasters	.29
Attacks from inside the company	.41
Downtime from non-disasters	.6
Accidental errors	.72
Computer viruses	.75
Unknown sources	.2

Table 1--Probability of Financial Losses

This paper focused on two of the security problems caused by the Internet dealing with attacks. The Attacks from outside the company has a probability of .23 and the attacks from inside the company has a probability of .41. Another source sited the estimates for inside attacks as high as .80 back in 1991 (Bresnahan 97). Government agencies and organizations experience only .54 probability of an inside attack costing the government about $72,000 for each security incident (Power 97). For purposes of this research a virus was considered as part of an attack. The chances of contracting a virus is much greater on the Internet because of the high volume of users and traffic. The outside attacker is more likely to be committed with malicious intent (Bresnahan 97). The inside attack also could be committed with malicious intent but the majority of the inside security incidences are accidental versus intentional (Bernstein 97). Outsourcing to an irreputable Internet Provider could raise risks (Caldwell 97).

QUANTITATIVE DATA ANALYSIS

The software tools used to conduct the quantitative data analysis were Microsoft Excel version 7.0, Data Analysis, and Solver add-in for Excel 7.0.

Forecasting Models

In order to evaluate the different forecasting models the author compared the accuracy of each model and the ability to adjust to shifts in a matrix. The best-fit forecasting model is based on the these results. This paper will begin with the simplest method evaluated and progress to the most complex method.

Simple Forecasting Models

The naïve model simply uses the last actual as the forecasted value. From a visual inspection of Figure 2 Trend Line for Attacks Graph it is not likely that the flattened line will continue as forecasted. Figure 1 shows two shift of level approximately two years apart. As management becomes more aware of the cost of Internet Security attacks, increased emphasis on reporting the attacks will occur at certain times which could be a explanation for the dramatic shifts in level.

The Simple Exponential Smoothing model did as poorly as the naïve model (Exhibit 1). During graphical analysis, seasonality was discovered in the last two years of the data (See Figure 3). After deseasonalizing the data, the SES model still performed poorly with a high MSE and extremely low forecast. The Holt model did very well with the lowest MSE and a relatively high forecast. The Holt-dampened model did nearly as poorly as the Naïve and SES models with a low forecast, but a low MSE. The Holt-Winters model did not do very well with a high MSE and a relatively low forecast. The Holt level adjusted exponential smoothing model did the best. Holt-LAES had a high forecast with a low MSE. Finally, the regression model did moderately well for the forecast but poorly with a high MSE. Most of the MSE values resulted an optimum value after using the Solver add-in for Excel 7.0 and solving for minimum MSE by varying the adjusted values of w, v, u, and fe.

RESULTS

The results of the data analysis with the Internet Security attack forecasting model clearly indicate that the optimal solution for the enterprise for forecasting Internet Security attacks is to use Level adjusted exponential smoothing Holt's method. It is the only model which adjusts to policy shifts and has a low MSE. As noted in Table 2, a predicted policy change was predicted for month 76 which was consistent with policy shifts which occurred approx. every two years. Further forecasting analysis may be conducted by the user of the model by varying the period of policy shift.

Model Name	Jan 98 Forecast	MSE	Shifts	w	V	u	fe	Exhibit Number
Naïve	102.00	NA	No	NA	NA	NA	NA	Fig. 2
Simple Exp. Smooth	101.22	188.82	No	.8036	NA	NA	NA	1
SES Deseasonalized	112.23	206.54	No	NA	NA	NA	NA	2
Holt	164.67	75.89	Yes	.8479	0	NA	NA	3
Holt-dampened	112.14	85.02	Yes	1	0	NA	0	4
Holt-Winters	140.26	195.72	Yes	.673	0	.2775	NA	5
Holt-Level Adj Exp.S	238.33*	92.40	Yes	.9226	0	NA	NA	6
Regression	161.93	248.14	No	NA	NA	NA	NA	7
*Policy Shift in month 76 predicted judgementally								

Table 2—Results of Different Forecasting Models

RECOMMENDATIONS FOR FURTHER RESEARCH

Further research using empirical data to validate the Internet Security attack forecasting model is needed. As enterprises conduct business on the Internet, researchers should conduct surveys specifically asking for the number of Internet Security attacks per month by type of attack and if possible the amount of financial loss associated with the type of attack.

REFERENCES

Bernstein, David S., Infosecurity News industry survey, Infosecurity News, Vol. 8, No. 3, May 1997, pp. 20-27.

Borg, Kim, Web Readies Wares for Online "Shopholics" But security concerns keep them turned off, Computer Technology Review, Vol. 17, No. 3, March 1997, p. 1, 6-8.

Bresnahan, Jennifer, To Catch a Thief, CIO Magazine, March 1, 1997, pp. 68-72.

Caldwell, Bruce, Violino, Bob, and Kolbasuk McGee, Marianne, Hidden Partners, Hidden Dangers, Information Week, January 20, 1997, pp. 38-52.

Howard, John, An Analysis of Security Incidents On The Internet, 1989-1995, URL:http//www.cert.org/research/JHThesis/Start.html, July 1, 1997.

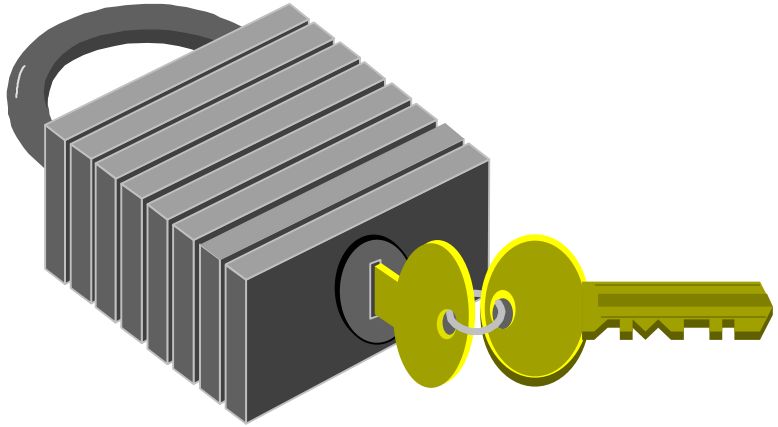
Power, Kevin, FBI finds hackers can't resist a government agency, Government Computing News, April 14, 1997, p. 60.

Row, Heath, The electric handshake, CIO Magazine, January 1, 1997, pp. 48-63.

United States General Accounting Office, INFORMATION SECURITY, Computer Attacks at Department of Defense Pose Increasing Risks, Report to Congressional Requesters, May 1996.

Violino, Bob, The Security Facade, Information Week, October 21, 1996, pp. 36-48.

Wilder, Clinton and Kolbasuk McGee, Marianne, GE The Net Pays Off, January 27, 1997, pp. 14-16.



A Forecasting Model For Internet Security Attacks

by

Alexander D. Korzyk, Sr.

J. G. Van Dyke & Associates

Virginia Commonwealth University

Agenda

- CIO Top 10 Challenges
- CIO Top 10 Critical Techologies
- Research Questions
- Graphical Exploration
- Trend
- Seasonality
- Trend-cycle
- Year-over-year Differences
- Evaluation Criteria
- Results
- Recommendations For Further Research

CIO Top 10 Challenges

Number By Rank	Challenge	Percent
1	**Implementing IT capital planning and investment management**	**76**
2	Measuring IT contribution to mission performance	56
3	**Formulating or implementing an agency IT architecture**	**52**
4	Aligning IT and organizational mission goals	41
5	Championing BPR as a precursor to IT decisions	37
6	Building effective relationships with agency senior executives	35
7	Gaining a seat at the senior management table	32
8	Engaging senior executives on IT strategic directions	30
9	Providing effective IT infrastructure and related services	27
10	Ensuring Year 2000 operations	25

Survey results from AFFIRM October 1996

CIO Top 10 Critical Technologies

Number By Rank	Critical Technology	Percent
1	Internet/Intranet/Web	73
2	**Security Technology**	**68**
3	Electronic Commerce/Electronic Data Interchange	57
4	Distributed Computing	47
5	Data Warehousing	42
6	Client/Server Computing	41
7	Workflow	35
8	Executive Information Systems/DS S	28
9	Groupware	22
10	Relational Databases	21

Survey results from AFFIRM June 1996

Research Questions:

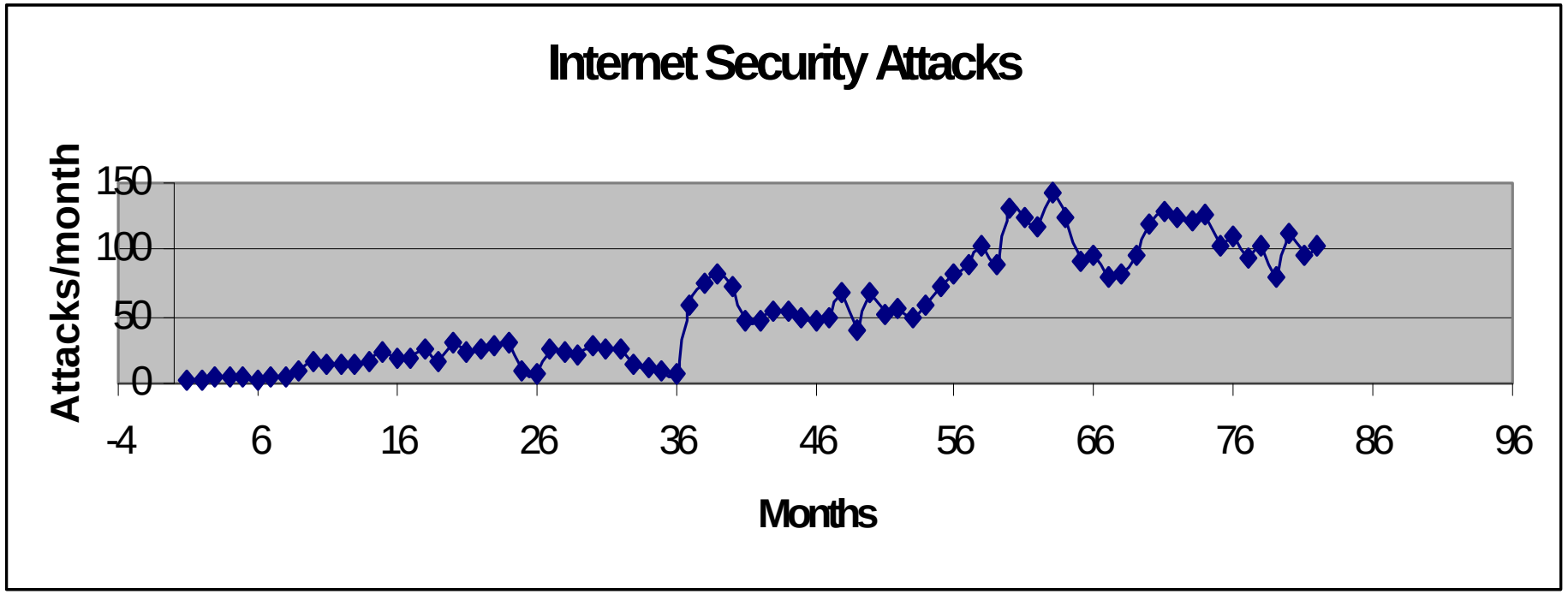
Does a more complex forecasting model improve accuracy?

What is the effect of a management change in policy on the forecast?

A Forecasting Model for Internet Security Attacks

J. G. Van Dyke & Associates

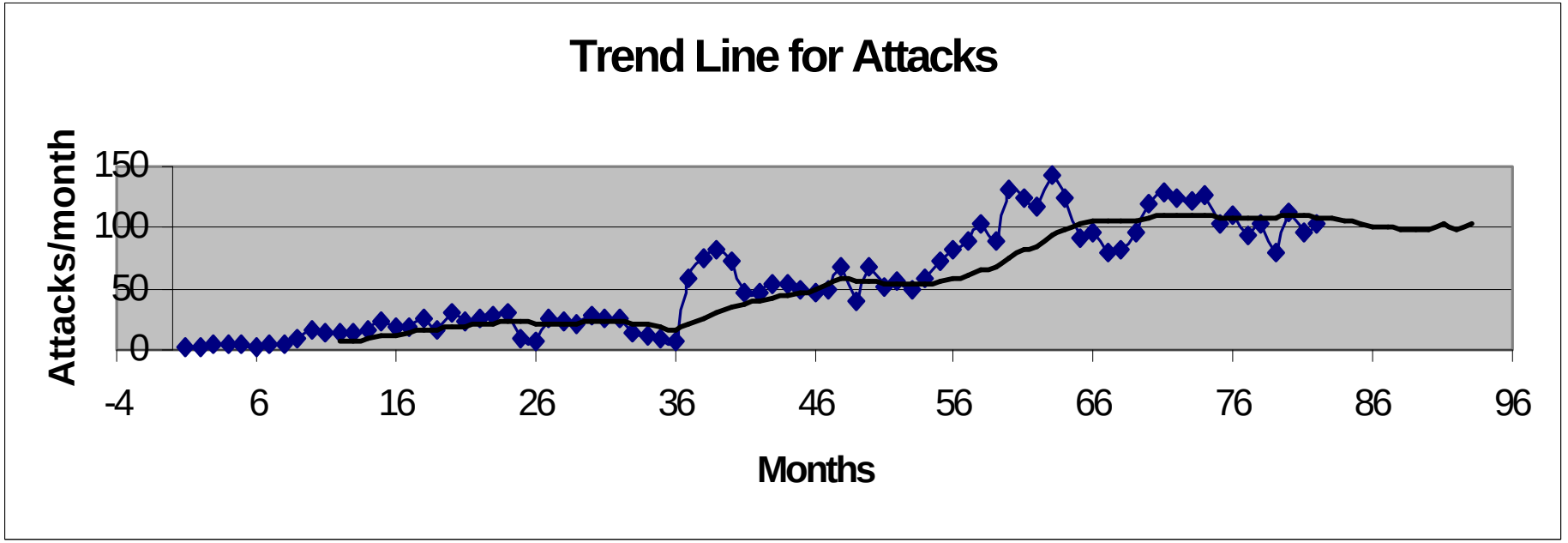
Virginia Commonwealth University



A Forecasting Model for Internet Security Attacks

J. G. Van Dyke & Associates

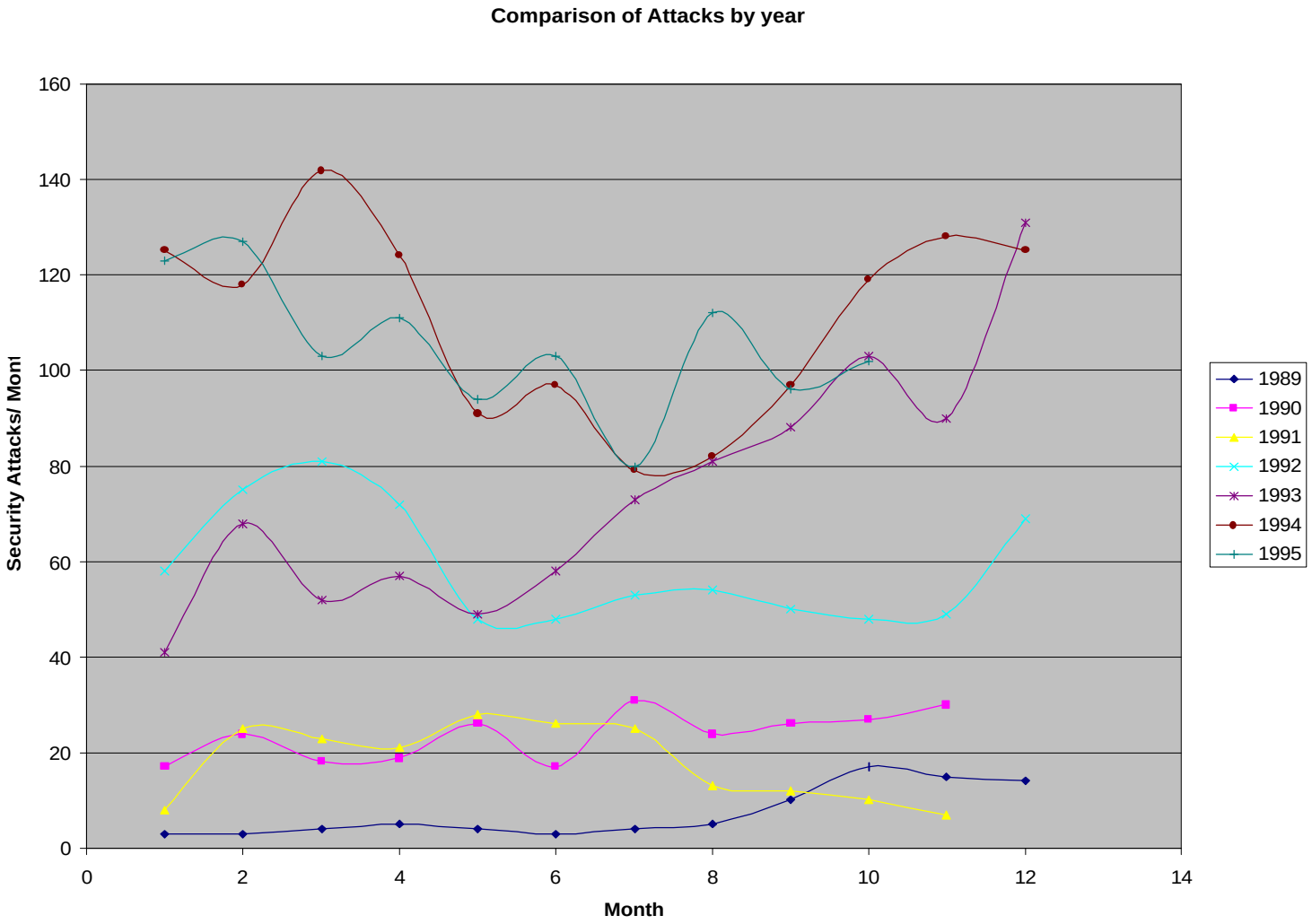
Virginia Commonwealth University



A Forecasting Model for Internet Security Attacks

J. G. Van Dyke & Associates

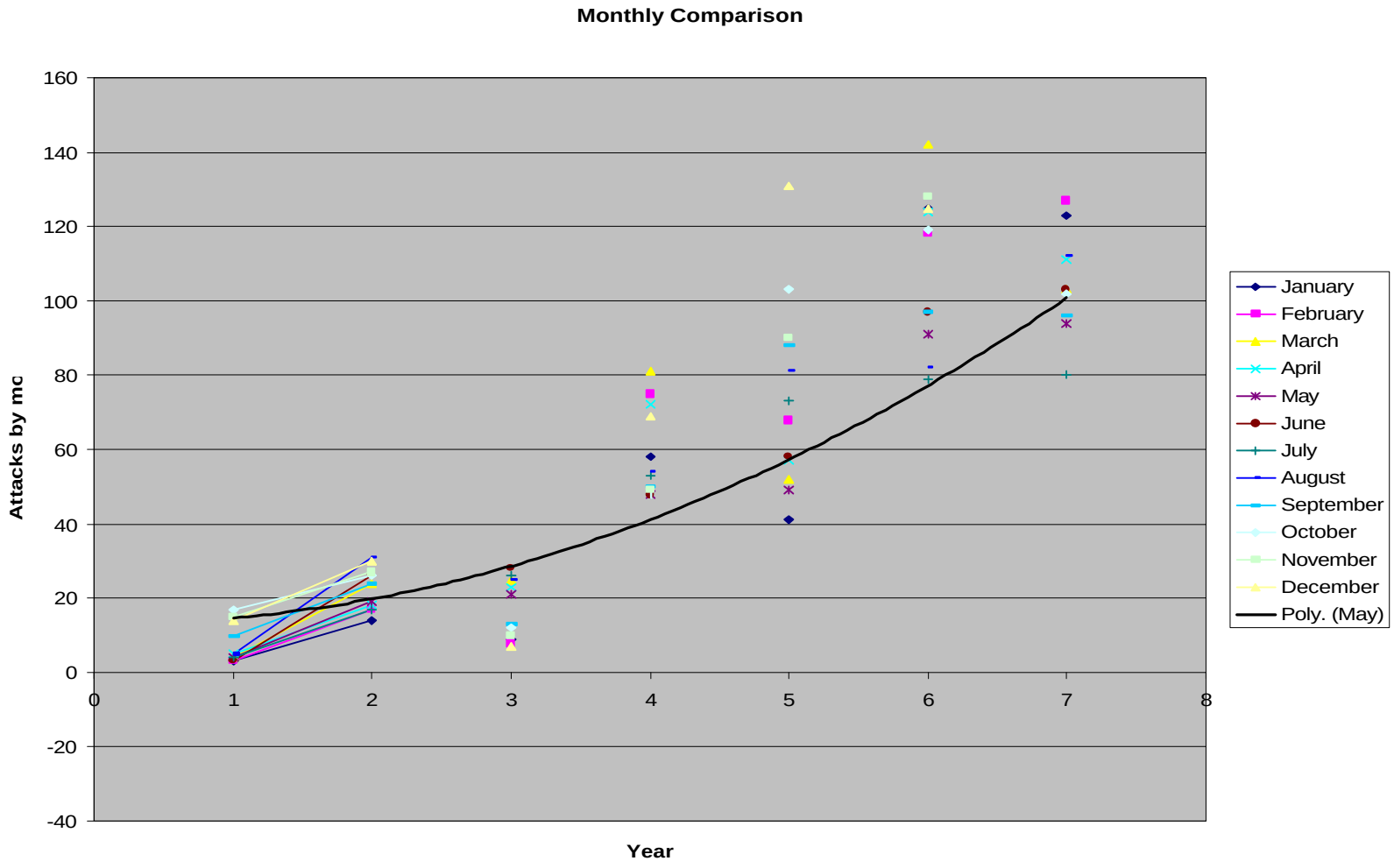
Virginia Commonwealth University



A Forecasting Model for Internet Security Attacks

J. G. Van Dyke & Associates

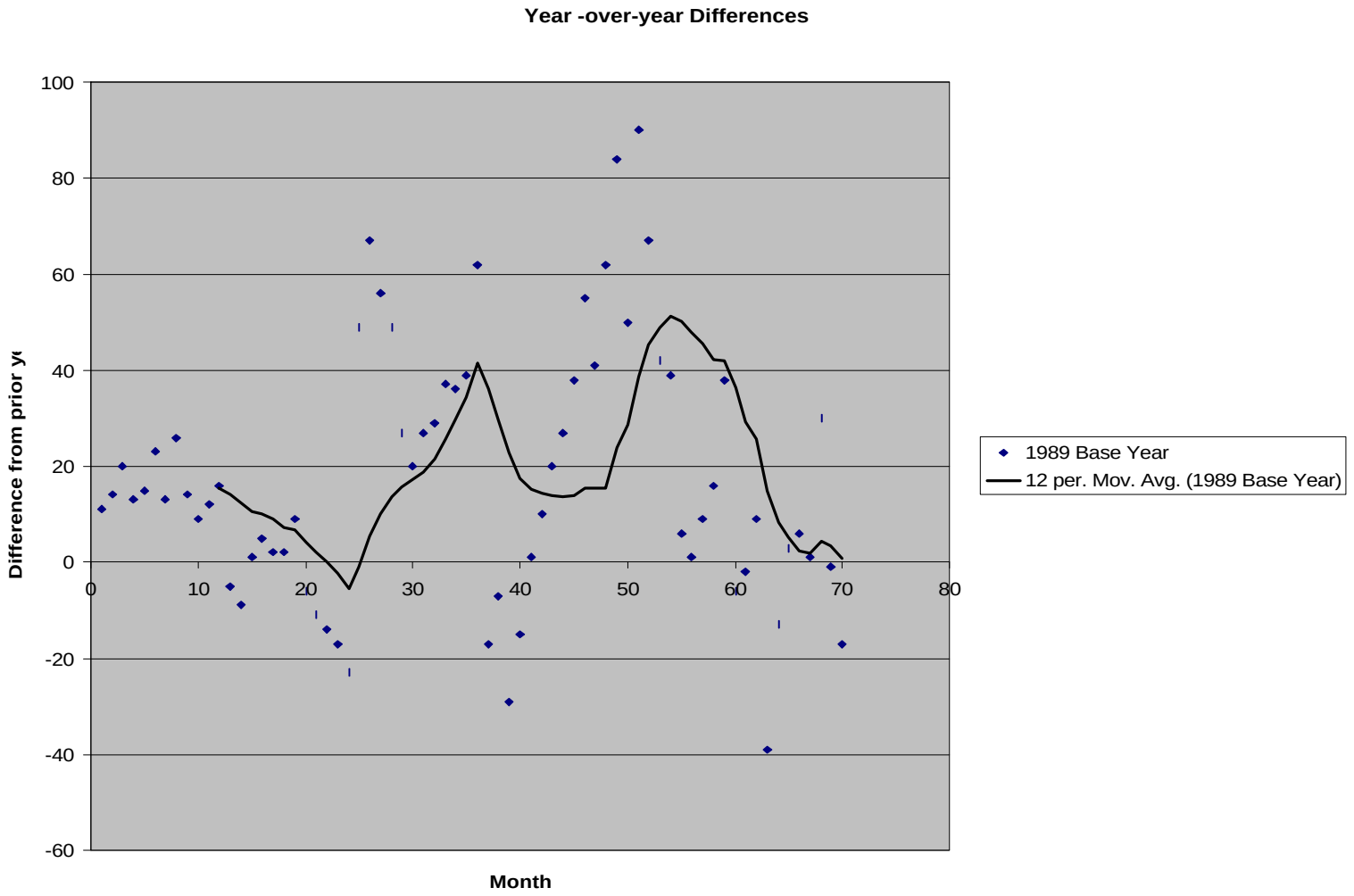
Virginia Commonwealth University



A Forecasting Model for Internet Security Attacks

J. G. Van Dyke & Associates

Virginia Commonwealth University



A Forecasting Model for Internet Security Attacks

Security Problem resulting in financial losses sited above.	Probability of companies with loss from this security problem (independent).
Industrial espionage	.09
Attacks from outside the company	.23
Natural disasters	.29
Attacks from inside the company	.41
Downtime from non-disasters	.6
Accidental errors	.72
Computer viruses	.75
Unknown sources	.2

Probability of Financial Losses (Violino, 96)

- EVALUATION CRITERIA
  - LOWEST MSE
  - ABILITY OF MODEL TO ADJUST TO SHIFTS BY MANAGEMENT POLICY CHANGE

A Forecasting Model for Internet Security Attacks

Model Name	Jan 98 Forecast	MSE	Shifts	w	V	u	fe
Naïve	102.00	NA	No	NA	NA	NA	NA
Simple Exp. Smooth	101.22	188.82	No	.8036	NA	NA	NA
SES Deseasonalized	112.23	206.54	No	NA	NA	NA	NA
Holt	164.67	75.89	Yes	.8479	0	NA	NA
Holt-dampened	112.14	85.02	Yes	1	0	NA	0
Holt-Winters	140.26	195.72	Yes	.673	0	.2775	NA
Holt-Level Adj Exp.S	238.33*	92.40	Yes	.9226	0	NA	NA
Regression	161.93	248.14	No	NA	NA	NA	NA
*Policy Shift in month 76 predicted judgementally							

Results of Different Forecasting Models

Results

Holt-Level Adjusting
Exponential Smoothing

A Forecasting Model for Internet Security Attacks

Recommendations for Further Research

- Gather empirical data from enterprises
- Determine if combination of models may be better than one single model