Cyveillance®
Minding your business on the Net

SIZING THE INTERNET

A White Paper

by

Brian H. Murray
Director, e-Business Strategy Center
Cyveillance

Alvin Moore
Senior Engineer
Cyveillance

July 10, 2000

EXECUTIVE SUMMARY

There are now **more than two billion unique, publicly accessible pages** on the Internet.

Cyveillance recently conducted a study using its proprietary NetSapien™ Technology and identified a number of significant findings. First, there are now more than two billion unique, publicly accessible pages on the Internet. Second, the Internet is continuing to grow at an accelerating rate, despite its already tremendous size and historically high growth rate. If the current growth trends for the Internet continue, the publicly accessible pages on the Internet will double in size to four billion pages by early 2001.

Cyveillance's NetSapien Technology made this study possible. This technology is Cyveillance's proprietary software that the company employs in the delivery of its e-Business Intelligence services. The NetSapien Technology was used to collect data to create a sizeable sample from which Cyveillance constructed a highly effective, dynamic model of the Internet on a continuum. Cyveillance's model of the Internet is unique, in that it goes beyond providing a "snap-shot" by monitoring the Internet's actual growth and acceleration rates on an ongoing basis.

TECHNOLOGY

The Cyveillance study was facilitated through the company's patent-pending NetSapien Technology, which was recently awarded a place in history as part of the Smithsonian Institution's Permanent Research Collection.[1] NetSapien Technology allows Cyveillance to deliver "Extra-Site" e-Business Intelligence by mining and analyzing the Internet to provide key insights about how the Net is impacting the revenue, market share and brand and customer loyalty of Cyveillance clients.

The technique used in the study leveraged the capabilities within NetSapien Technology to reference pages and then identify content and links. The continuous stream of data collected served as the basis for creating the dynamic model of the Internet.

APPROACH

The model leverages the high volume of information collected via NetSapien Technology on an ongoing basis. To estimate the size of the Internet, the model analyzes specific data associated with the links present on the pages reviewed. To illustrate the power and scope of the technology employed for this ongoing study, over a four-month period alone more than 350 million links were considered. By continuously referencing pages and examining the links, the model is able to track the frequency with which unique URLs are encountered, both for the first time and each time thereafter. With this information, it becomes possible to develop an extremely large sample from which Cyveillance is able to effectively construct a dynamic model of the Internet on a continuum. The size of the sample and methodology employed significantly improves the accuracy of the model.

[1] Visit www.cyveillance.com for more information regarding NetSapien Technology.

The Internet's **accelerating growth provides new opportunities** to companies.

Other studies have estimated the size of the Internet based on a static measurement that captures information for a specific point in time or averages between two points in time. By estimating the size on a continuum, the Cyveillance model can more accurately track the rate at which new pages and sites are emerging. In addition to the growth rate, the model is effectively able to estimate the rate of acceleration (or deceleration) on an ongoing basis.

The model also accounts for URLs that are no longer present and is therefore able to track the rate at which pages are being removed. Another important feature incorporated into the model is its ability to account for the more popular sites by factoring the average frequency that the link appears into the extrapolation. Because this factor is incorporated into the model, the outcome is not skewed by the presence of more common links.
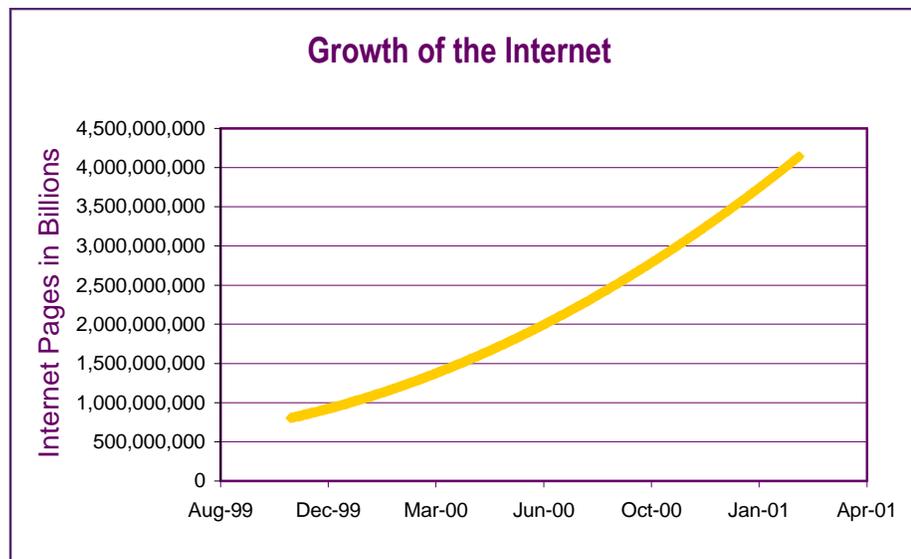
STUDY RESULTS

In addition to collecting additional e-Business Intelligence of interest to Cyveillance clients, the model allows Cyveillance to monitor a host of different Internet measurements associated with the size and growth of the Internet. A few of the more significant results of the study are discussed below. The statistics presented herein are valid as of July 10, 2000.

Size and Growth of the Internet:
- Number of unique pages on Internet: 2.1 billion
- Unique pages added per day: 7.3 million

Note that for the purpose of the study, "number of unique pages" refers to HTTP URLs that are linked to a publicly available URL. More specifically, they are URLs of up to 800 characters in length sent from servers for version 1.1 of the HTTP standard only. The study results do not include FTP sites or secure sites (HTTPS).

Growth of the Internet

Internet Pages in Billions

4,500,000,000	
4,000,000,000	
3,500,000,000	
3,000,000,000	
2,500,000,000	
2,000,000,000	
1,500,000,000	
1,000,000,000	
500,000,000	
0	

Aug-99 Dec-99 Mar-00 Jun-00 Oct-00 Jan-01 Apr-01

The model further allows Cyveillance to make predictions regarding future growth of the Internet. Based on data collected over an eight-month period

of time, the following graph illustrates the number of pages on the Internet versus time and projects growth forward through early 2001.

Some of the study's more interesting observations related to the size of the Internet include the following:
- The data collected fit a second order polynomial and show very little divergence.
- Internet growth is accelerating, indicating that the Internet has not yet reached its highest growth period.
- Modeling the Internet on a continuum allows Cyveillance to project future growth with a higher degree of accuracy compared to other methods. If growth continues its current trend, the Internet will reach three billion pages by late October and four billion pages by February 2001.

Characteristics*
- Average size of pages: 10,060 bytes
- Average number of internal links per page: 23 (median** = 4)
- Average number of external links per page: 5.6 (median** = 1)
- Average number of images on a page: 14.38 (median** = 1)
- Percentage of US vs. international pages: 84.7% /15.37%

Also of interest is the fact that of the more than 350 million links considered, about 10.43 percent generated broken link error messages and less than one percent timed out. Each link was tried three times with three hours between each attempt and a wait time of 10 minutes before timing out.
- *   Note that in order to avoid pages designed to continuously download, only pages fewer than 200 KB were considered in the study.
- **   Note that "median" signifies the middle number inside a sorted sequence of numbers.

CONCLUSION

Based on the analyses to date, in excess of two billion unique, publicly accessible pages exist today on the Internet. The Internet continues to grow at an increasing rate and is on target to double again by early 2001. Additionally, growth continues to accelerate, indicating that the Internet has still not reached its most rapid phase of expansion. By applying NetSapien Technology's ability to mine the Net, reference pages and identify links, Cyveillance is able to model the Internet on a continuum, thereby extracting valuable information, including the size and growth of the Internet.

ABOUT CYVEILLANCE

Cyveillance (www.cyveillance.com), the "Extra-Site" e-Business Intelligence pioneer, helps Global 2000 corporations identify key Internet opportunities and risks. Unique to Cyveillance are its e-Business Strategy Center and its proprietary NetSapien™ Technology, an industrial-strength software that acts like a human but moves at the speed of the Net to extract "Extra-Site"

e-Business Intelligence from across the Internet. "Extra-Site" is a term used by Cyveillance to describe its ability to mine and analyze the Internet's billions of pages, versus other companies that track Web traffic and user data from within a specific corporate Web site. Cyveillance provides clients with competitive and marketing intelligence, as well as brand, partner and supplier management insights. Digital South magazine named Cyveillance among the "Top 50 Companies to Watch in 1999 and 2000," and the company's NetSapien Technology has been included in the Smithsonian Institution's "Permanent Research Collection of Technological Innovations for the New Millennium."

e - B U S I N E S S   I N T E L L I G E N C E   S E R V I C E S

Strategic Solutions
While some vendors in the e-Business Intelligence space today are focused on addressing one or two key issues, only Cyveillance takes a holistic approach—providing a complete solution to fuel executive decision-making and help companies run at maximum efficiency. Our Strategic Solutions address five key areas:
• Competitive Intelligence;
• Brand Management;
• Marketing Intelligence;
• Partner Management; and
• Supplier Management.

Because the Internet impacts every part of your organization—including your bottom line—having a complete solution is a must. By tapping into Cyveillance's comprehensive "Extra-Site" e-Business Intelligence capabilities, your e-Business can begin firing on all cylinders.

And only Cyveillance offers NetSapien™ Technology, the most powerful business search and analysis tool available today. Our e-Business Intelligence Services are built upon this unique, patent-pending technology and coupled with the expertise of our e-Business Strategy Center. Cyveillance solutions provide clients with strategic, competitive insights unavailable from any other company in the market place.

Intellectual Property (IP) Protection Solutions
If you're not sure who's capitalizing on the goodwill of your name or leveraging your valuable digital assets, Cyveillance's Intellectual Property Protection Solutions will identify and prioritize sites across the Internet that are using your IP for their own financial gain. We have a track record and a technology designed to help you monitor and protect your IP.

Regulatory Solutions
The Internet has become the virtual street corner of the new millennium, and many government entities and other organizations are now on the lookout for ways to proactively protect consumers. Cyveillance's Regulatory Solutions are designed to help government entities and other organizations seeking to proactively understand, manage and control cyber crime.

Value-Added Consulting

To complement our standard offerings, Cyveillance provides Value-Added Consulting designed to help you leverage and manage the intelligence you're already receiving from Cyveillance, give you the foundation you need to better understand the market and the competition, transform your business, cross the chasm or address the many other key corporate challenges you face today.

About the Authors

Alvin Moore is a senior engineer at Cyveillance. He has more than 14 years experience in the software industry as a software developer and as a technical consultant specializing in the Windows 32 architecture. Before joining Cyveillance, Mr. Moore worked with Microsoft Corporation, Computer Associates and Legent Corporation, as well as his own consulting firm, MicroSolutions, Inc. In addition to his Masters in Mathematics from University of California at Berkeley, he earned a Bachelor of Arts in Mathematics and a Bachelor of Science in Computer Engineering from Howard University.

Brian Murray is the director of the Cyveillance e-Business Strategy Center. He has 10 years of management and technical consulting experience and is an expert in the area of e-Business Strategy and e-Business Intelligence. Mr. Murray joined Cyveillance from PricewaterhouseCoopers where he was a manager in the Global Risk Management Solutions Practice. He earned his Masters from the Darden Graduate School of Business Administration at the University of Virginia, a Masters of Engineering from The Johns Hopkins University and a Bachelor of Science from Syracuse University.

For additional information, contact Cyveillance today at 1.888.243.0097 or www.cyveillance.com.