VOLUME 26 • NUMBER 42 • OCTOBER 10, 2005 • $4.00 multichannel.com

The Net's NETS

SECOND IN A SERIES

BRAD WALKER

Is Google Cable's Next Nightmare?

NO.

ANALYSIS BY PAUL STRASSMANN

Is Google Cable's Next Nightmare? NO.

SPOTLIGHT

ANALYSIS

BY PAUL STRASSMANN

The Internet Search Giant Isn't About to Set Up the Largest Video-Over-Fiber Network in the World. Not Yet, Anyway.

Is Google a gigantic video-on-demand service in waiting, about to deliver infinite amounts of personally chosen programming across an intercontinental fiber network and with local wireless connections?

You might think so, by reading recent coverage of the Internet search giant's seemingly limitless ambitions:

• "… Google Inc. plans to offer an electronic-payment service which allows consumers to pay for purchases by funding electronic-payment accounts." — *The Wall Street Journal*, 6/20/05

• "… A trail of hidden clues suggests Google is building its own Internet." — *Business 2.0*, 8/25/05

• "… Rumors are circulating that it aims to construct a nationwide fiber-optic network that could support free Wi-Fi access for all." — *Network World*, 8/29/05

• "… Google is reviewing bids for the development of a national DWDM [dense wave-division multiplexing] fiber network … to give Google unprecedented flexibility to push massive amounts of voice, video and data content." — *IP Media Monitor*, 9/19/05

• "… Provision of free and almost unlimited access to the Internet would mean that telephone, cable, broadcast TV, film and music companies would have to come to grips with the fact that their hold on their customers is slipping." — *eWeek*, 9/22/05

• "… Google said it has no plans to provide [wireless Internet access] services outside of the San Francisco area, despite rampant speculation that it aims to do so nationally." — *Wall Street Journal*, 10/03/05

Even *Multichannel News*, in the first part of this series, last week examined how Google has already set up a Web service that allows users, in effect, to instantly create personal cable channels. Its Google Video site (*video.google.com*) lets consumers assemble customized shows on any subject from an online library of videos.

Google also is not alone. Other startups trying to define how video searching will work in the future include Blinkx (*www.blinkxTV.com*), Open Media Network (*omn.org*) and Dave.TV (*www.dave.tv*).

But none of them have the wherewithal of Google, which just raised $4.2 billion in a second public stock offering. So the question of the moment in multichannel competition is: Should cable-system operators be losing sleep over whether Google and its data centers will take over delivery of video programming from their headends?

An examination of the company's public statements about its hardware, software and network infrastructure, academic presentations, its publicly available tools and services, and interviews with business partners, security experts and information technology managers who have worked with the $4.5 billion a year self-professed organizer of the world's digital information yields these conclusions:

• Google does not try to compete with existing media suppliers head-on. It seeks to create services that combine text, images and functionality in new ways, and to attract new customers instead of trying to switch customers from existing services.

• It's most likely to deliver services in which technology can unlock value, such as pulling together a coherent repository of business-related video streams that exist on Web sites or an educational video service that displaces the widespread distribution and use of tapes by schools and colleges.

• Even Google will be challenged to achieve its lofty goal of establishing itself as the "universal switchboard" for finding and playing back the world's burgeoning video content (*Multichannel News*, Oct. 3, page 3).

Should this be causing cable operators nightmares? The emphatic answer is: No.

At least not for five years, anyway.

Here's a look inside the brain and nervous system of Google: what it's capable of; how it's different from other rivals in storing, slicing, dicing and doling out digital information; and what its limits are.

Google's most critical assets are not its algorithms for identifying how many sites are linking to a particular Internet address to find a particular piece of information or its other search formulas.

Instead, its top asset is its network of 200,000 commodity servers, operating in clusters of as many as 32,000 machines each, which allow it to store and process not millions, billions or trillions of bytes of digital records, but quadrillions.

Indeed, the United Kingdom publication *TechWorld* in April 2004 estimated Google handled 5 Petabytes of information on its commodity servers. That has probably doubled by now, meaning Google now likely stores 1,000 times as much information as exists in the print collections of the Library of Congress.

The amount of electronic information in the world doubles every three years, according to the University of California at Berkeley's School of Information Management. Google is racing ahead of that pace.

The Google computing "grid" already offers the world's largest and most advanced high-performance computing civilian capacity, according to a security expert in Washington, D.C., who has consulted with Google on installing the servers.

The 200,000 commodity servers have been produced in Taiwan to custom specifications

See SPOTLIGHT, page 20

SPOTLIGHT

Continued from page 18

at rock-bottom prices, this expert noted. These servers are each managed by a stripped down version of the Red Hat Linux operating system. They are housed in physically distributed clusters in the United States, Europe and the Pacific Rim.

The Net's NETS

SECOND IN A SERIES

150M QUERIES DAILY

Each of these clusters can handle more than 150 million queries a day, according to Google fellow Urs Holzle. Maintenance costs are reduced through factory refurbishing of failed servers. Operating staff is kept to the barest minimum because of the use of simple, consistent boxes of hardware and software that can be easily monitored — and replaced as needed, according to Dr. Jeff Dean, another Google fellow, in a presentation at the University of Washington in October 2004.

The Google servers are cheap because they use commodity processors and disk drives. All data files and file indexes are replicated at least three times to protect against failure and to move content as close as possible to where a service must be delivered. Software manages the redundancy with little human intervention. And the commodity box and software management allow capacity to be added, on a just-in-time basis, like in a car factory — only bring in the additional hardware at the moment it's needed.

Google's data centers also are more energy efficient than any comparable installation, according to an April 2003 paper published by three Google authors in *IEEE Micro*, a journal of the Institute of Electrical and Electronics Engineers. Total automation of network controls, standardization, modular configurations, energy efficiencies and the simplicity of the applications architecture make the Google computing grid as expandable as any that has been designed to date.

The result: Computing performance that is hard to match in a world where millions of users may be requesting services at the same time. In two quick tests, identical inquiries found 25,800 good responses for Google, 12,100 for Yahoo! and 4,320 for MSN. This capability is delivered in over 40 languages in subsecond response time, regardless of location, and even while Google might be coordinating as many as 1,000 separate servers to complete a single search.

The key to such performance is the practice of "dynamic indexing" of more than 8 billion pages of text and 1 billion graphic images found on the Web. That indexing is possible because a large part of the computing load — including "crawling" sites to see what they contain — can take place during off-peak periods.

DATA, INDEXES & SHARDS

In the Google network, "data" (text and images) are separated from the "indexes" that keep track of where the data is physically located. Both the data as well as the indexes are broken into pieces called "shards," which are farmed out to separate clusters. This increases reliability. Since the "shards" are placed near where the most activity takes place, reaction speed is improved as well.

This leads, in general, to the expectation that Google could well have the capacity to deliver video streams to millions of customers at the same time — if true, that would be a clear threat to cable operators.

But consider this: The equivalent of a sheet of paper's worth of text, encoded in the HyperText Markup Language, requires only 16,000 digital bytes to store on a disk. Sure, that's four times as many bytes as is needed to store the plain text of that page of information.

Compare that to video, though. Even if it is just a "switchboard," Google still will be trying to store lots of one-hour shows beyond *Desperate Housewives* (if its owners allow it). A single one-hour show will consume 10 million bytes of storage, even when compressed.

That's in standard-definition format, not High Definition Television. Adding higher resolution and the capability for an interactive search on all the actors, features and elements of a movie or TV program will probably double — or more — that count, based on the experience we have with storage-requirements expansion. Then, there are multiple languages for the audio track and "meta" files that will index all the content. It's going to make organizing all the world's books look like a children's tea party.

My estimate is that it would take almost 2 billion hours of video storage per year to capture what operators in 1,000 markets, each offering 200 channels, would need to store online as their capacity.

At the receiving end — the "last mile" of the video delivery chain — the strains for capacity would saturate all channels, as millions of subscribers would expect to receive their personal video streams without deterioration in quality during the peak hours.

Google could innovate and overcome such limitations. The most successful one today is Bit Torrent, a popular protocol used for swapping huge (video) files today among users (*www.bittorrent.com/introduction.html*).

It has a unique characteristic: the more people that use it, the less overwhelming does traffic between any two points get.

So it is conceivable, if Google did establish or gain control of an international fiber grid of its own, that in a few years it could deliver video efficiently.

That's why you should take, with a truckload of salt, the news that Google is hiring Vinton G. Cerf, the so-called "Father of the Internet," as merely an evangelist. Cerf, who played a pivotal role in the development of the original Internet communications protocols, is just the kind of luminary who, behind the scenes, can and is likely to advise a team of network-design stars how to pull off tasks that those protocols were not originally designed to handle.

But you can discount the significance of Google contracting for large amounts of "dark fiber" capacity.

Right now, just with text and still images, Google's demand for cheap and high-bandwidth circuits can be explained easily by the need of those clusters of 32,000 servers to communicate with each other, as Google's volume of searches nearly triples in a year.

I also discount the development of the Google Secure Access service that allows users to make safe connections over the air to the Internet, using wireless fidelity, or Wi-Fi, cards. This and its application to provide wireless access service in San Francisco support the idea that Google wants to be a national or international supplier of wireless access service.

But Google is not in the business of setting up Wi-Fi antennae everywhere. It's much more likely to let someone else do that job, and then simply connect some such wireless access service to its fiber network, through partnership, merger or acquisition, when the need or desire to do so arrives. It could wait for the higher-capacity WiMax access services, which will ship video over the air, to get going.

It's also not likely that Google will be able to provide a TiVo-like search-and-record service on its clusters anytime soon, either. The demands of fulfilling millions of consumers' simultaneous requests at this point is technically beyond the reach of any network.

Most practically, Google's ad-based model may not be able to extract enough money from advertisers to cover the fees it could well have to pay to intellectual property owners for the replay of their content. To get viewers to pay for movies or videos, Google would also have to develop a mechanism for billing customers — and for collecting from them.

Sure, users now register for its Gmail service (*gmail .google.com*), but a complete, efficient global billing system could take years to develop and implement.

That is just another reason why I do not expect much of an encroachment on cable operators' existing revenues.

GOOGLE IS DIFFERENT

But make no mistake. Google is different from Yahoo, America Online or Microsoft Corp.'s MSN. As a business devoted to finding, storing and retransmitting media in all its forms to all comers, it's ferociously devoting itself to continuous upgrading and innovation.

Google engineers and staff are encouraged to devote 20% of their time to exploring new opportunities and initiating innovative Google services. A freely accessible cluster of computers is accessible to all employees for experimentation. Some of the results can be found at Google Labs (*labs.google.com*), which lists 16 experimental applications such as Site-Flavored Google Search (*www.google.com/services/siteflavored.html*), which lets a Web site operator easily tailor Google's search results to a particular "flavor" of content.

There's also the practical Froogle Mobile US (*labs.google.com/frooglewml.html*), which shows a a cell phone user whether that person might be about to pay too much in a store for a pair of shoes or other merchandise.

Google also fosters experimentation by providing easy mechanics, known

Indexing Google

The Internet's "digital switchboard," by the numbers:

Servers:	200,000
Largest Cluster:	32,000
A Cluster Handles:	150 million queries, daily
Data Centers:	12-plus
Information Indexed, Dynamically:	8 billion pages of text, 1 billion images
Total Information Stored:	10 quadrillion numbers and letters
Equivalent Of:	1,000 Library of Congress print collections

SOURCES: Google, industry estimates, *Multichannel News* research

SPOTLIGHT

as Application Programming Interfaces, to allow developers to build new services on top of its network. Google Tools (*www.google.com/intl/en/options/*) are readily available for general use and without any restrictions, in contrast with the controlled approach used by most vendors. Google even has openly released its counter-hacking procedures, which have convinced programmers that Google's network is sufficiently technically secure to give away instructions on how to connect to its network services (*code.google.com/apis.html*).

The result: an ongoing rat-a-tat release of new applications, from facsimile copies of L.L. Bean, Land's End, Gump's and other catalogs (*catalogs.google.com*) to automated answers to complex inquiries to satellite maps to blog searching to book scanning. I estimate that Google publishes a major new application every month.

What this says is: Cable operators should not be losing sleep because Google is likely to encroach on their existing franchise. They should start thinking about the ways in which Google will innovate with the delivery of services that cable will not be offering at all.

Instead of attacking the cable, TV, movie or publishing business head-on, Google's tactic will likely be to fill gaps where the Internet can deliver a service more economically, at a higher standard of quality.

For instance, business Web pages now include a huge collection of streaming media. These valuable videos remain mostly inaccessible because they are not catalogued. This is a vacuum Google is likely to fill.

Another likely target: Giving schools and universities an alternative to mailing videotapes. Why ship boxed collections when downloads or streaming video will do?

In effect, what Google is likely to offer soon are specialized video services for audiences in narrowly defined markets, where electronic distribution has not be economically feasible, before; or where, if it innovates, it can make it feasible. Placing all educational videos on the Google net is such an example.

The idea, though, that somehow Google could displace cable TV services for consumers on a mass scale in hundreds of millions of homes in this country and others is a figment, at least for the next five years.

Cable operators can sleep peacefully — for a while, any way — while the capacity for the distribution of video over the Internet is limited.

Your bedrooms and houses are safe, today. You won't lose the customers you serve well. But you could well lose customers who will seek out Google for services you cannot satisfy. ■

Turf Watch

Video search engines:

Blinkx: www.blinkx.tv

Brightcove: www.brightcove.com

Columbia University's Webseek: Persia.ee.columbia.edu:8008

DaveTV: www.dave.tv

Google: video.google.com

iFilm: www.ifilm.com

Open Media Network: www.omn.org

Yahoo: video.search.yahoo.com

Next up:

Apple Computer: The progenitor of the iTunes music software is expected this week to announce it plans to deliver a version of its popular iPod digital player that will also play back videos from the Web. Which could lead to a video version of iTunes, as well.

www.appleinsider.com/article.php?id=1304

Paul Strassmann is distinguished professor of information sciences at George Mason University and an adviser on information technology to the U.S. Defense Department. He was the chief information officer and vice president of strategic planning at Xerox and has served as the CIO of Kraft and the National Aeronautics and Space Administration. He is based in New Canaan, Conn.