P R O J E C T P O R T F O L I O

covering:

Google world domination

To: English C1 seminar course, Darlene Kilian

By: Thomas Kühnel, Thomas Winkler

Date: June 8, 2010

Abstract

Google is one of the most important and most powerful companies of our century. Started 12 years ago in a garage by two students, it now has total assets of about $40 billion. They provide their users with lots of different services that make life easier and in most cases they don't even want money for this. But there are always downsides too. Google has extremely huge amounts of information they get every day from users that use their services. Now what will happen, if Google stops following it's slogan “Don't be evil.” or their services just stop working or others use Google's services for evil? There are lots of bad things that could happen to your data like Google decides to sell them or some hackers get access to them. And if the services just don't work you don't have access to all your important calendars, emails and documents, so this is also something you don't want to happen.

There are several ways to prevent these worst case scenarios. First of all: Don't trust Google too much! Always think about if you really have to give some information away or if you really need all of the services. And if you really need them, try to find some alternatives so there isn't one big company that knows everything, but many small ones that know a tiny bit.

Table of contents

I Introduction

1 History of Google

Google first started 1996 under the name BackRub and was a project by the two Stanford University students Larry Page and Sergey Brin. They didn't like the search engines of their time because they were good at indexing the websites but the way they sorted the search results was pretty bad. Brin and Page wanted to do this different. Their idea was to rate the importance of a site by using data of how many other pages link to them. With this concept they developed the so called PageRank algorithm.

1998 they renamed the search engine from BackRub to Google, which is actually a misspelling of the word googol(the number 10^{100}). Google got pretty fast popular and 1998 Andy Bechtolsheim, co-founder of Sun Microsystems invested $100,000 in the not yet existing company Google Inc. Later they started the real company and rented the garage of a friend as office for it.

Later this year the US magazine “PC Magazine” listed Google as search engine of choice in its Top 100 Web Sites list. In the beginning of 1999 they rented some offices for the company in Mountain View. They rented more and more space there and finally bought the property in 2006 for $319 million. Google had 11 employees in 1999 now they have about 20,000. The Google search engine currently has a market share of about 85%.

Today Google probably owns the most powerful computer network in the world. They own about 1 million servers all over the globe. To make this infrastructure secure against for example attacks, Google keeps all information about size and position of their data centers secret.

2 Google Services

In 2000 Google started one of it's main ways to earn money, Google Adwords. This is a service where you can buy advertising space, so your ad is shown if the user searches for something related to what you are selling. Companies have to pay Google for every time a user clicks on one of their advertisements.

Another pretty important service of Google is the Google Image Search. This service launched in 2001 and lets the user search for images not only by name, but also by size, format, color or other things.

In the year 2002 they started Google News. This is not a service like other news sites, with articles written specifically for this site, it just collects news from lots of news websites groups them by topic and sorts by importance. With this the user can on the one hand get as much information as possible on a subject like sports or international politics, he's interested in and on the other hand he can see all the different viewpoints the various news sources have on some topic.

On the 1st of April 2004 Google released a new e-mail service called Gmail to the public. What made this service different from other is that they gave 1 gigabyte of space for e-mails to the user for free. This was more than 100 times the space other services provided. With this the developers at Google wanted to make sure that nobody has to delete any old e-mail. To make it possible to find important mails Google implemented a search in the web interface which was also much better than the solutions of it's competitors. They chose this specific date because Google is famous for doing April Fool's jokes and so they chose this date in order that there will be lots of discussion in the news and other places whether Gmail is real or just a joke and with that they got lots of free marketing for it.

Another service that launched in 2004 is Google Code Search which is a search engine for source code that is available to the public. With this developers can find source codes for problems they have, so they don't have to invent the wheel another time.

2005 was the year of Google Maps. This provides maps and satellite photos of the whole earth to everyone for free. It's also possible to plan routes, see traffic jams and lots of other things.

In the year 2006 Google bought the online video service Youtube. They paid $1.65 Billion for this which is the biggest sum so far they paid to acquire another company. Google already had it's own online video service Google Video, but this wasn't really popular in comparison with Youtube. Now both services run parallel with their own interface and content.

In 2008 Google published the Internet browser Google Chrome. It's designed to be very lightweight and fast and so an interesting option for people who don't need any fancy features other browsers provide or for slow computers like netbooks.

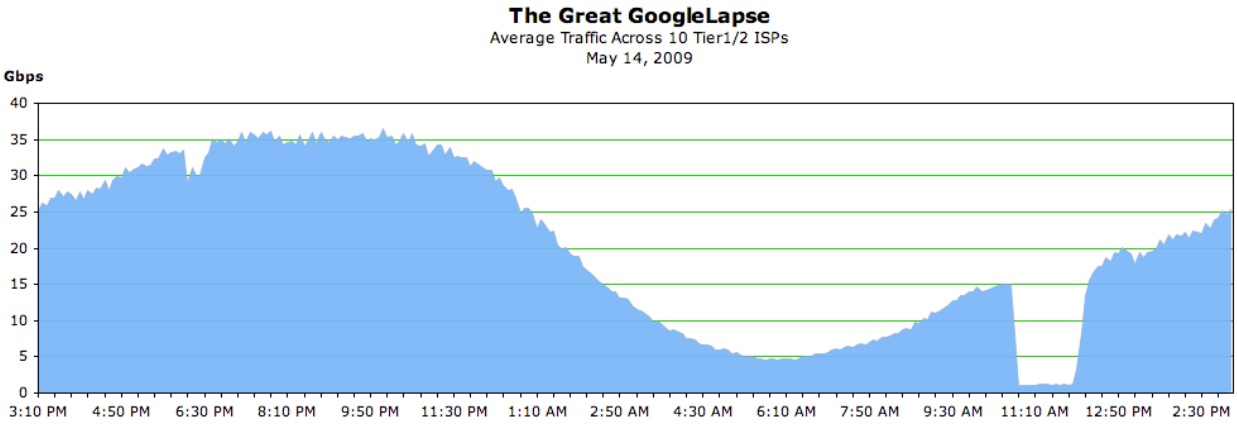
The newest thing that Google bought this year is the video codec VP-8. After they bought it, they made it's source code open source, so everyone can use and improve it. The goal of this step is to get an uniform format to use for audio and video on websites, so people don't have to use proprietary software like Flash anymore.

This list was just a very small part of what Google offers to it's users. At the moment there are about 150 different Google services or products.

II Dependency on Google Services

1 Impact of an outage

So what happens, if Google doesn't work temporarily or even for a longer time period. There is a good example from 2009 to illustrate this problem. On May 14th was a problem at one of Google's data centers which caused 14% of its users to not be able to use Google's services. The following graph shows the Internet traffic of a part of north Americas on May 14th and gives a pretty good impression on the impact of this accident.



Not only is it not possible to use the search engine if Google isn't available, but the other services are also important parts of the everyday life of many people. It's for example pretty bad if you have every important appointment for your job saved in the Google Calendar or use Gmail for e-mail communication in your business. If Google services then don't work it's not only annoying, but can also get pretty costly fast.

Another thing to look at are other web sites that rely on services that Google provides. There are many pages that depend on Google Analytics, a service to provide web site owners detailed statistics about the usage of their sites, which is used by many web pages. If this service isn't available, the websites that use them won't work too in some cases. And there are even important web sites of banks and such that depend of this service.

2 Alternative services

There are several ways to prevent great damages if an outage of a service like Google happens.

The first idea is to just not use these services if there are any alternatives that work even without being online. A calendar is a good example for this. There are enough other ways than the Google online calendar to organize your appointments, like on the cellphone or a traditional calendar made of paper.

The second method is to choose a different provider for different services so if one provider isn't available, there is also just one service not available and not all of them. This is actually pretty ease, because of the huge popularity of Google's services, other companies started to provide similar services after Google did.

A good alternative to the Google Search are so called “meta search engines.” These sites use other search engines and put the results together in one place. This can also produce better results because every search engine has other ways to sort the results. Examples for this type of search engines are MetaCrawler or Excite.

A alternative to organize appointments online is HipCal. This service offers pretty much the same as the Google Calendar, just the interface of the site is a bit different.

III Risks of misuse

1 By users

1 Google Earth guiding criminals

As Google Earth provides high-resolution images of many countries taken by satellites, it can potentially be misused for tracking down interesting spots for committing crimes. For example, in June 2009 it was used to find and steal valuable fish in private yards protected by fences, which could not have easily been located if the criminals hadn't had access to bird's eye photos[1]. It was also used to help more inclined felonies, like the terrorist attack in Mumbai, New Delphi on September 26 of 2008[2].

2 Google Street View – WiFi scanning

Now Google has started to scan for wireless local area networks (WLAN) while taking photos for Google StreetView. This data, if published, could very well be used to commit Internet crimes: The criminals look for an open or WEP encrypted (which is considered insecure) network in their proximity, go there with their laptops and log into the open network, using plain, anonymous browser profiles and random MAC addresses (those are normally unique hardware addresses but can be changed by desire in most network interface controllers). Then, if they run away in time before anyone notices anything, which is very likely to succeed if only acting for some minutes,

1 http://www.maximumpc.com/article/news/latest_way_misuse_google_earth_stealing_expensive_fish
2 http://ibnlive.in.com/news/tech-misuse-2611-terrorists-used-google-earth/84374-11.html

everything illegal can be done without facing consequences, as there is *no* possibility to identify the criminals afterwards. Likely the WLAN holder will be liable for any damage caused. This is really a problem of clueless WLAN administrators choosing weak or no encryption at all, but that service would greatly aid criminals in finding such spots.

2 By spammers

1 Exploitation of trust in Google's domains

Some spammers have found a trick to circumvent spam filters by translating a web page by Google which *redirects* to another domain where the spam is actually hosted. Spam filters mainly work by adhering to *whitelists* consisting of trusted Internet addresses and/or *blacklists* containing known-to-be-bad addresses. What's on the blacklist is blocked while what's on the whitelist is let through to the user. What to do with unknown addresses is a decision between security and comfort: While it's safer to block everything unknown as it could possibly be bad, this often keeps out content the user actually wants to see, therefore requiring more interaction to show what has been preventively blocked unnecessarily.

Anyway, that decision doesn't matter in this case as Google's domains are almost always contained in the whitelist because they offer so much good content and are therefore never blocked. Exactly that's what's abused by the spammers: They don't send you direct links to their sites, but filter their pages through Google Translate and include the translated result. This effectively bypasses spam filters deciding by address.

There are ways to fix this problem, for instance by using a content-based spam filter, where spam is detected by certain words or images, or by finer adjustment of whitelist entries, but most standard filters are likely to be affected.

2 Manipulation of search results – *doorway pages*

To have some sort of free advertisement, spammers like to fill an intermediate page with garbage merely consisting of phrases concerning their products which are likely to be entered into a search engine. Having more relevant words on the page means more hits in Google search and consequently more visitors. Such a doorway page normally instantly redirects to another site where advertisements of the products may be found.

Google tries to variate it's search engine's ranking algorithm from time to time to make it harder to build and manage such pages. Nevertheless, even well-known companies like Automobile.de and BMW have used this unfair technique of boosting their search results[3]. Even though that is now already a quite old method, it's still being actively used. Not that it would be dangerous, but it exploits the fact that so many people use Google's search engine to find whatever needed on the Internet.

3 By hackers

1 Searching for common security vulnerabilities – *Google Hacking*

Google's search engine can easily be misused to find insecure web hosts. Imagine the following scenario: A hacker searches for ready-to-use code examples intended for use by web developers, such as up- and download or text highlighting scripts or whatever comes to mind. Lots of those can

3 http://www.golem.de/0602/43155.html

be found at so-called tutorial pages about web technologies like PHP or JavaScript. The hacker skims the examples, specifically looking for security holes. When something potentially exploitable has been found, the corresponding code is noted. Then, depending on whether it's a server or client, compiler or interpreter language, a search string is built based on the code and/or expected output. Alternatively, the search string might consist of common error messages (like *"Warning: filesize(): Stat failed for ..."* which indicates that a file access failed, due to missing rights or the file simply not existing) or version strings of affected software.

By entering such a search string into Google it's easy to find websites where these insecure scripts are used and then to exploit the problems. Possibilities range from gathering information which is not intended by the web master to be seen by anyone out there (think about a hacker getting access to a user database containing e-mail addresses, passwords and so on) to attacking the server. So Google is really great at aiding hackers who just want to attack random websites – a phenomena grown up in the last years, called *Google Hacking*.

2 Getting access to user data

Since Google collects so much data about individuals and is so widely used by companies, the fear of a leak of user data to unauthorized third parties grows. Hackers could blackmail Google and/or it's users if they succeeded in retrieving enough relevant data. It's of very high priority to Google to have a secure network. However, considering the complexity and constant development of the services, vulnerabilities can at any time be implemented and will often not be noticed instantly, if at all.

As a well-known example, in January of 2010, some Chinese achieved to get access to some Gmail accounts. They could have copied messages out of them and/or used them to trick/spam the account holders' friends.

4 By Google itself

1 Selling information

At least in Germany, Google has sold personal information like users' names, birth dates, telephone numbers, email addresses and occupations to whoever has payed for them[4]. After all, it is a company aiming at maximizing it's profit. As long as a user doesn't explicitly forbid selling data gathered by using the service, it's perfectly legal to do so according to German laws, and thus in fact done.

2 Co-operation with advertisers

A lot of data Google collects about it's users is used to adapt advertisements to customers' interests. Not only search phrases, but also email messages and data from other services are used. The advertising enterprise named *Doubleclick,* bought by Google in 2007[5], now profits from this pool of information. Also it uses *tracking cookies* (small chunks of data stored in the browser) to uniquely identify and observe what Doubleclick-featured pages the user visits. So the data by advertisers is no longer collected *per site*, but globally connected *all over the web*.

This seems officially not to be regarded as a misuse of user data, given also that everything is processed by computers automatically and no one has insight to the raw data collected. But it is a deep invasion in personality and some people feel being observed and that for good reasons.

4 http://translationmusings.com/2008/09/07/does-google-sell-its-users-personal-data/
5 http://en.wikipedia.org/wiki/DoubleClick#Acquisition_by_Google.2C_Inc.

3 Exploitation of users' dependency

Many people all over the world rely on Google's services, both for personal and for commercial uses. It's amazing to see what happens when Google doesn't work just for some hours; it's unimaginable what would happen if it not worked for days! So what if Google suddenly saw a reason to demand a high fee for using it's services? What if it for whatever unlikely reason took down some of it's services?

Most users don't think about such cases. Of course Google wouldn't want to stop it's services, as they give profit by usage. But anyway, it's always best to think of the worst-case scenario: What *could* happen? There are alternative services, but could they stand up to the sudden flood of "Googlers" suddenly utilizing them?

Fortunately, there has not been any bigger problem yet, so it's unknown what would happen. One can only speculate...

4 If there was a total business failure...

...and Google would have to raise money to clear debts, what would then happen to the collected data? It so often happened already that bankrupt companies sold their users' addresses to spammers. But no failed business has ever had such comprehensive profiles about it's users.

Again, there's nothing more than to speculate about this. And really, to hope this will never ever happen...

IV Conclusion

1 Split up power

It is generally safer to use several small service providers than one big all-in-one solution like Google offers. That is because it's not easily possible for the individual providers to connect the data they have about you with data collected by another provider. Thus they don't know everything but only a little about you.

As an example, if you login to Google Mail, then do a web search in Google, your search strings can be associated to your E-Mail account with ease. However, if you use a different mail provider than search engine, the search engine can't really know who you are (well, apart from the usual data it might collect anyway, like your IP address, provider and browser identifications, and so on).

Also, the amount of collected data splits up to multiple companies, which in turn means that if one for instance will be hacked, only that part of data of you can actually be revealed. And so smaller providers are less likely to be hacked or spammed at all because they appear less attractive to most attackers than very used ones.

2 Think carefully

Try to keep independent of just one big company. Instead, focus on some smaller ones. What would happen if you needed Google's services every day and suddenly it would decide to postulate a lot of money for usage? Likely you would pay with a grim face, because you don't have any other choice. However, if you're independent and for example suddenly your mail provider started to become non-free or increased fees beyond your limits, you could just drop it and use the alternative one you maybe preventively have already registered for.

As always in life, always think about what data you give away. Before entering anything, ask yourself the question: *Would I tell that to a stranger down the road asking me?* If not, then don't enter it into any Internet service either! People tend to be more willing to give sensitive data about themselves to some online form than to a person in real life, because they mistakenly feel it would be more anonymous. But it's actually not at all! While it is most likely that you'd never meet exactly that stranger again or that person just forgets what you told, assume that the Internet never forgets. There might always be a *copy* of your published information and then, after you long deleted it, it might come up again years later, and you'd hate yourself for having been so naive to put it on the net at all.

So it's best to keep this at an absolute working minimum. That means for example that it's safer not to fill in optional fields asking for your phone number, interests, a photo of you, your relationship status, your eye color, or the place where you live (however, the latter one is often required for legal reasons).

When sending secret data over the net, always encrypt it on your workstation with a trusted application (like *GnuPG*) and never send it across the net in an unencrypted form, regardless of the provider you use, as data is passed through a lot of “hands” on the way to the target host and might be copied anywhere.

A Appendix

1 Glossary

Browser profile: a web browser program stores a lot of data about it's use, such as bookmarks, *cookies*, history, website logins, most recently displayed objects. This collection is called the browser's profile and some parts of it (especially the cookies) may be seen by websites visited.

client: system accessing a remote *server*

codec: abbreviation for coder/decoder, computer program able to compress and decompress data

compiler: translates human-readable *source code* into a program executable by a computer

cookie: also referred to as web cookie, browser cookie, HTTP cookie. It's a small text stored in a *browser profile* which can be accessed by websites being visited. This way it is possible for the website to easily identify individual users.

data center: a room or building where a lot of *server*s are situated, having air conditioners, fail-safe power supplies, fire protection and so on

domain: a common network name organizing a collection of network devices

encryption: transforms readable data into non-readable data by an algorithm using a given key, which can be a password, file, fingerprint or similar. The non-readable data can be decrypted only with the correct key or by finding and exploiting a weakness in the algorithm.

indexing: building a catalog containing only those parts of web pages relevant for searches

interpreter: executes human-readable *source code* while reading it

IP address: software network addresses assigned to hosts which are unique at a given moment, so if both time and IP address are recorded, this pair of information can be used to find the user by asking the corresponding Internet provider

JavaScript: a programming language running in web browsers to interact with users

open source: a computer program whose source code is freely available

PHP: short for Hypertext Preprocessor, a programming language running on web servers dynamically generating web pages

provider: a company or organization offering services

server: a system designed to be used by many users (*client*s)

source code: text describing a program, written by a programmer and understood by a *compiler*

spam: undesired advertisements, often sent in huge quantities

WEP: abbreviation for *wired equivalent privacy*, a deprecated, easily breakable encryption for wireless networks

2 Literature

http://en.wikipedia.org/wiki/History_of_Google
http://en.wikipedia.org/wiki/Google_Services
http://www.pcworld.com/article/164946/google_outage_lesson_dont_get_stuck_in_a_cloud.html
http://www.maximumpc.com/article/news/latest_way_misuse_google_earth_stealing_expensive_fish
http://ibnlive.in.com/news/tech-misuse-2611-terrorists-used-google-earth/84374-11.html
http://www.symantec.com/connect/de/blogs/more-spammer-abuse-googles-services
http://www.golem.de/0602/43155.html
http://en.wikipedia.org/wiki/Doorway_page
http://en.wikipedia.org/wiki/Google_hacking
http://www.guardian.co.uk/technology/blog/2010/mar/25/gmail-china-hacking
http://translationmusings.com/2008/09/07/does-google-sell-its-users-personal-data/
http://en.wikipedia.org/wiki/DoubleClick#Acquisition_by_Google.2C_Inc

All checked to be working as of June 14th, 2010.