

Advertising on the web and GOMC

Prof. mahmed@ashesi.edu.gh

Ashesi Univ, Accra, Ghana

2/2011

Ads

- How is web different from physical ads?
- Targeted ads
- Search Ads
- Partner ads

Ads for search

A screenshot of a Google search results page for the query "flowers". The search bar at the top contains the word "flowers". Below it, a message says "About 332,000,000 results (0.30 seconds)" and a "Search" button. To the right is a link to "Advanced search". On the left, there's a sidebar with categories: Everything (selected), Images, Videos, News, Books, Places, Blogs, Realtime, and Discussions. Below that is a location section for "Accra" with a "Change location" link. At the bottom, there's a "Any time" section with links for "Latest", "Past 24 hours", "Past 2 days", and "Past week".

The main content area shows search results. The first two results are highlighted with purple ovals and labeled "Ads". The first ad is for "Flowers to USA for \$19.99 - Send Flowers to Your Loved Ones." from www.proflowers.com. The second ad is for "Send Flowers to the USA" from www.fromyouflowers.com. A third result, "Flower - Wikipedia, the free encyclopedia", is also highlighted with a purple oval.

Below these results, there's a section titled "Free Pictures of Flowers and Flower Wallpaper Photos" with a link to www.flowers.vg/. At the very bottom, there's a "Images for flowers" section showing five small thumbnail images.

Side ads

Google search results for "seeds". The search bar shows "seeds". Below it, a sidebar lists various Google services: Everything, Images, Videos, News, Books, Places, Blogs, Realtime, and Discussions. A location setting for Accra is shown. A "Any time" filter is applied, with "Latest" selected. The main search results include:

- Souvenir Seeds Store**
Super Strains From Only \$19 £10 €12 Discrete, Secure Worldwide Delivery
www.theseedsstore.com
- Seed - Wikipedia, the free encyclopedia**
A **seed** is a small embryonic plant enclosed in a covering called the **seed coat**, usually with some stored food. It is the product of the ripened ovule of ...
en.wikipedia.org/w/index.php?title=Seed&oldid=5000000 - Cached - Similar
 - Seed (disambiguation)
 - List of edible seeds
 - Seedling
 - Granivore
 - Seed testing
 - Seedbed
 - Seed saving
 - Seed orchard[More results from wikipedia.org »](#)
- The Seeds - Wikipedia, the free encyclopedia**
The **Seeds** were an American rock band. The group which repertoire spread ...
en.wikipedia.org/w/index.php?title=The_Seeds&oldid=5000000 - Cached - Similar
- Burpee Seeds and Plants - Vegetable, Flower, Home Gardening ...**

Seeds, plants and gardening supplies for home gardens - Large, exclusive selection: Heirloom and organic **seeds** and plants. Vegetable, flower, herb, fruit, ...
www.burpee.com/ - United States - Cached - Similar

A purple circle highlights the advertisement for "Garden Starts Nursery" on the right side of the search results.

seeds

About 113,000,000 results (0.40 seconds)

Advanced search

Ads

Garden Starts Nursery

Selling vegetable, flower **seeds**

Visit us for Garden **seeds**

www.gardenstarts.com

See your ad here »

Google Adwords

Learn about AdWords

How it works	You create your ads You create ads and choose keywords, which are words or phrases related to your business. Get keyword ideas
Why it works	Your ads appear on Google When people search on Google using one of your keywords, your ad may appear next to the search results. Now you're advertising to an audience that's already interested in you.
Costs and payment	You attract customers People can simply click your ad to make a purchase or learn more about you.
For local businesses	Sign up now Next topic »

Keywords are what people search for on Google:



Your ad appears beside relevant search results.



You create your ads

- You create ads and choose keywords, which are words or phrases related to your business. [Get keyword ideas](#)
- **Your ads appear on Google**
- When people search on Google using one of your keywords, your ad may appear next to the search results. Now you're advertising to an audience that's already interested in you.
- **You attract customers**
- People can simply click your ad to make a purchase or learn more about you.

Google AdWords offers you

Targeted reach

- Now you can advertise to people searching on Google. Even if you already appear in Google's search results, AdWords can help you target new audiences on Google and our advertising network.

Greater control

- You can edit your ads and adjust your budget until you get the results you want. You can also display a variety of ad formats and even target your ads to specific languages and geographic locations.

Measurable value

- There's no minimum spending requirement or time commitment. And with the cost-per-click option, you're only charged if people click your ads. This means every dollar of your budget goes toward bringing new prospects to you.

Ad costs

- **Set your budget**
- There's no minimum spending requirement – the amount you pay for AdWords is up to you. You can, for instance, set a daily budget of five dollars and a maximum cost of ten cents for each click on your ad.
- **Avoid guesswork**
- We provide keyword traffic and cost estimates so you can make informed decisions about choosing keywords and maximizing your budget. ([Estimate keyword costs](#))
- **Pay only for results**
- You're charged only if someone clicks your ad, not when your ad is displayed.
- Payment options vary by country and currency. [Learn more](#)

Local and regional targeting!

- Set your ads to appear only to people searching in a particular region. Now it's easy to target online customers within, say, 25 miles of your front door.



Define a custom area to target around your business.

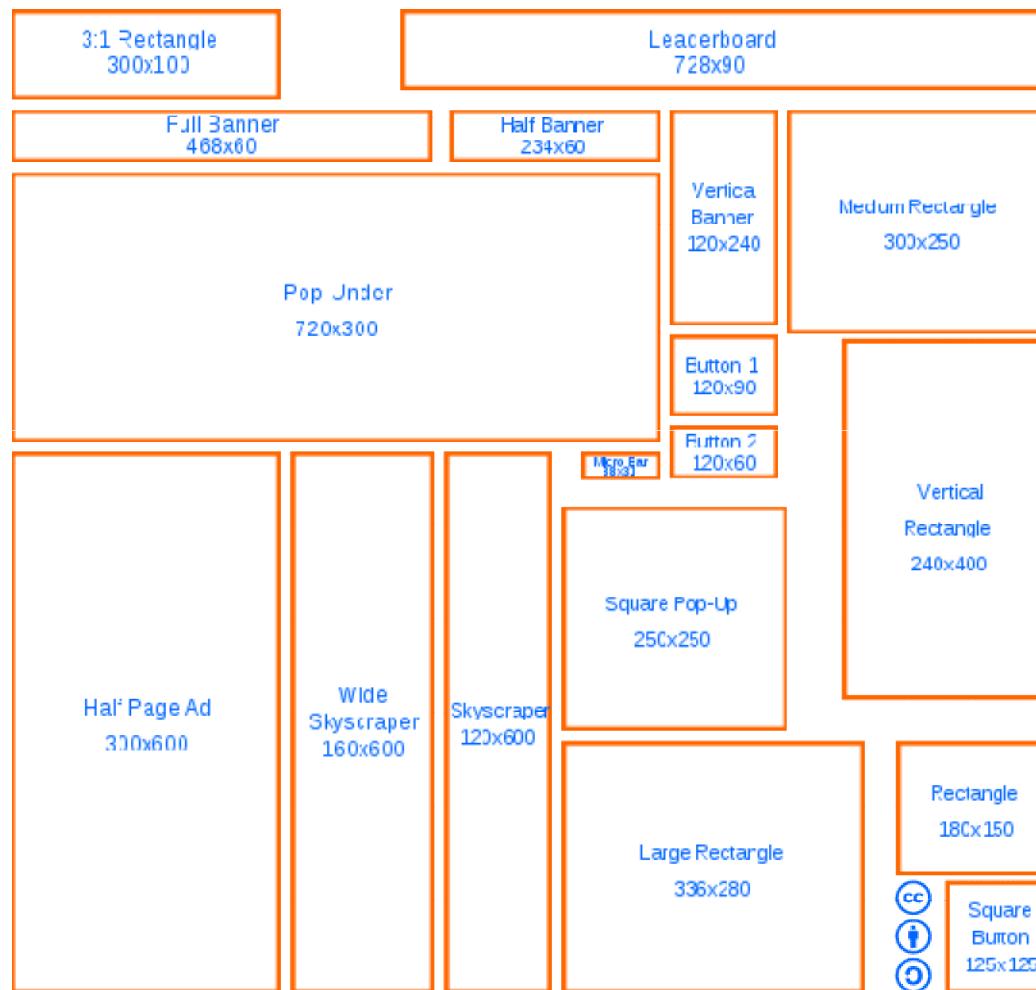


Promote your location with local business ads.

Acronyms

- **CTR - Clickthrough.**
- CPA - Cost per action
- CPC - Cost per **click**
- CPM - Cost per thousand
- CPI - Cost Per impression
- PPC - Pay per **click**
- Short click
- Click fraud.
- http://en.wikipedia.org/wiki/Clickthrough_rate

Banner ads - sizes



Ad Types

- **Interstitial ads** (**interstitial** means "in between") are a way of placing full page messages between the current and destination page.
- Pop up ads
- Flash ads
- Text ads

from http://en.wikipedia.org/wiki/Online_advertising

GOMC - Marketing contest

Timeline for GOMC 2011

- You can run your campaign over any three consecutive weeks between
- the 28th of January and 4th of June, 2010.
- Teams must submit their final report before June 11, 2010.
- Global and regional winners are announced July 2010.

Your Learning Objectives are:

- Given the opportunity, choose to discuss online marketing and media planning.
- Using examples, share the learning experience of group work and business consulting.
- Using examples, explain the following terms: banner advertisement, click-through-rate, conversion, landing page, optimization techniques, ROI and text advertisements.

Your Learning Objectives are:

- Using examples, contrast mass advertising and context-sensitive advertising.
- Using examples, illustrate technical and cultural factors affecting the success of online advertising campaigns.
- Using examples, illustrate the difficulties of developing a web-based marketing campaign that will stand out among the billions of web pages available.

Team

- Nominate a team captain. Your professor will receive instructions from Google and let you know the next steps.
- You'll need a team captain as part of this process. Your team captain will need to have a Google Account.
- See www.google.com/accounts/NewAccount.

Starting GOMC: Find a Business

- Select a business. Your team must agree with your professor on who you will work with.
- You must present the business with a copy of the 'Letter to Businesses' (included with this guide) and have them verbally agree to work with you.

Understand the Business

- Meet with your business and write your Pre-Campaign Strategy.
- To be successful in the Challenge, you will need to understand what the business does and what it hopes to achieve from online marketing.
- Allocate time to meet with them, write your Pre-Campaign Strategy and submit it to your professor and Google before you start your campaign. (Your professor will provide instructions on how to submit your reports to Google).

Setup Adwords

- Set up your AdWords account and begin your campaign.
- Once you receive your US\$200 account access, you should review the section in this guide titled ‘Making the Most of Your Campaigns’ for details on how to structure your account.
- Your campaign must run for three consecutive weeks between January 28 and June 4, 2010.
- Over these three weeks your team will check the results, run reports and optimize your campaign.

Reports

- Write your Post-Campaign Summary. Within no more than three weeks after your campaign ends and no later than June 11, 2010.
- Your team must write and submit the Summary to your professor and Google to be considered for regional and global judging. (Instructions on how to submit your reports to Google will be provided to your professor).
- Remember: Google MUST receive both your Pre-Campaign Strategy and Post-Campaign Summary on time!

Avoid

- Web Hosting
- Web Design Agencies
- Insurance Companies
- Mortgage Agencies
- Debt Consolidation Companies
- Multi-level Marketers -
http://en.wikipedia.org/wiki/Multi-level_marketing
- Distributors
- Affiliate Companies -
http://en.wikipedia.org/wiki/Affiliate_marketing

Landing page

- Finally, please note that the ‘landing page quality score’ of the website can affect your account performance.
- When selecting your business, ensure its website is suitable by reading the **landing page/website guidelines** at
<http://adwords.google.com/support/bin/answer.py?answer=46675&topic=9356>
- For further information on websites that typically have poor landing page quality, please see
<http://adwords.google.com/support/bin/answer.py?answer=66238>

Judging

- The contest has two components, Campaign Statistics that Google will assess and two written reports that academics will assess.

Campaign Statistics

1. Account Structure
2. Optimization Techniques
3. Account Activity and Reporting
4. Performance and Budget
5. Relevance

Account structure

- An Excellent account structure mirrors your client's website structure where possible. Your campaign(s) should be grouped according to product lines/themes or geography, and contain multiple Ad Groups specific to relevant subcategories.
- For example, a campaign for 'accessories' could have separate Ad Groups for 'bags' and 'jewelry'. Each Ad Group would then contain ad texts/variations specific to these subsections, and a targeted and specific keyword list.

B) Optimization techniques

- We will monitor how well you implement the suggested optimization techniques and best practices.
- In particular, we will monitor which of these techniques you implemented (e.g. keyword matching options) and how you optimized the Google network to your advantage, e.g. how effectively you used the content network.

E) Relevance

- Achieving a strong click-through rate is a key measure of how relevant your ads are and we will consider your click-through-rate when judging your account.
- To create relevant and effective advertising often means revisiting your campaigns and tweaking where necessary.

Written reports

- Pre-Campaign Strategy (30 points total, maximum four pages)... Communication and Readability (5 points)
- Client Overview (12 points) that describes your client business and
- Proposed AdWords Strategy (13 points) that helps your team craft and defend your draft AdWords Strategy. Combined, both components should be a maximum of four pages.. submit the Pre-Campaign Strategy to their professor and to Google.

Client profile

1. (2 points), a few sentences including some of the following. Please note that clients may not want to share some information.
2. Name, location
3. Sales and number of employees
4. Goods and services offered
5. Key online marketing personnel
6. Age of the company
7. url, website age, website management
8. Company presence and sales via online and offline channels
9. Other relevant information

Market analysis

1. (4 points, a couple of paragraphs including some of the following)
2. Current and potential customers
3. Current and potential competitors
4. Overview of the industry (key characteristics, competitive/saturated/mature)
5. Projected and historical online spend for the industry
6. Market position/specialties
7. Unique selling points of the goods/services offered
8. Seasonality of their goods/services or seasonality that the company has identified
9. Other relevant market information

Current marketing

1. Current marketing (4 points, a couple of paragraphs including some of the following)
2. Website uses, e.g. sales, customer service
3. Website strengths and weaknesses
4. Website visibility, such as Google PageRank, incoming links, a few keyword search results, online advertising, and offline promotion of the url.
5. If available, summary information from Google Analytics or other third party web tracking software
6. Email campaigns
7. Offline advertising
8. Other online or offline marketing

Strategy

- Conclusion on how the AdWords campaign should align with the client's business (2 points, a few sentences) ..
- Proposed AdWords Strategy (13 points, about two pages including sample AdWords and keywords)

Proposed AdWords Strategy

- Number of Ad Groups and the focus for each Ad Group.
- Keywords and negative keywords
- Text for at least two AdWords versions for some Ad Groups
- Daily and weekly plans for spending their campaign budget
- Network(s) for their AdWords ads

Proposed AdWords Strategy

- Target audience settings
- Ad Serving options
- Keyword Bidding
- Geotargeting
- Goals for impressions, clicks, CPC and CTR
- Proposed success metrics
- Other relevant information

Communication & Readability (5 points)

- The Pre-Campaign Strategy should have a logical flow, be easy to follow, and avoid grammatical mistakes.
- Post-Campaign Summary (70 points total, maximum eight pages).
- Post-Campaign Summary has five components: an Executive Summary (8 points), Industry Component (28 points) and Learning Component (14 points), Communication and Readability (10 points), and relevant use of Tables, Figures and Charts (10 points).

Industry Component (28 points, max five pages)

- This is the team's chance to share the results with their client and expand upon the Executive Summary.
- The ideal approach is to write the Industry Component first and then summarize this content for the Executive Summary.
- As a rule, you would include most if not all of your Charts, Tables and Figures in your Industry.

Campaign overview

1. Review the major campaign goals (strategic goals as well as metrics: CTR, CPC, and Impressions, etc.) set prior to the project and discuss your general strategies for approaching each goal.
2. Operational details (campaign dates, money spent, ad groups used). Review the basic schedule and cost structure you followed, your methods for monitoring the account, etc.

Evolution of Your Campaign Strategy:

- What were the major changes you made during the campaign and what led to these changes?
- How did these changes affect your campaign?

Key Results

Summarize your results based on three weeks of data, such as:

1. Overall performance of the campaign and individual ad groups.
2. Performance of the initial campaign and changes in performance following your optimization efforts.
3. Keyword combinations that were effective and ineffective.
4. Your success stories and quick, but clear references to failures you experienced.

Discussing performance

Refer to metrics such as:

1. Impressions
2. Clicks
3. Click Through Rate
4. Average Cost per Keyword
5. Total Cost of Campaign
6. Other metrics provided by the client, such as conversions

Conclusion

1. Synthesize the Industry Component, tie together the entire package and focus your client's attention on the key project aspects.
2. Take this opportunity to repackage information from the data section to display your practical lessons learned with the client.

Future Recommendations

1. Provide simple actionable and well-justified advice on your client's future online marketing, particularly in relation to AdWords and the website.
2. Learning Component (14 points, maximum two pages).

Reporting

1. Learning objectives and outcomes
2. Group dynamics - what problems ..
3. Client dynamics - what problems ...
4. Future recommendations - what would they do differently in the future to improve their campaign strategy, learning experience, group dynamics and client dynamics?

Future recommendations

What would they do differently in the future to improve their campaign strategy, learning experience, group dynamics and client dynamics?

Letter to Business

1. Teams must deliver the ‘Letter to Businesses’ to prospective clients.
2. Remember: think of yourselves as consultants, and the business as the client. That is, you work for the business and not the other way around.
3. Make sure the business understands everything that will happen and how you will follow up with them once the campaign ends.

Adwords

- AdWords is Google's sponsored link programme. Customers use AdWords to display ads on the pages of search engine results.
- These ads are published next to the organic search results, and, if required by the customer, on websites on the search network and sites associated with Google.

Adwords

- The advertiser only pays each time the user clicks on the ad, and accesses the advertised web page. Advertisers have full control over the targeting of their campaigns (by geography and language), the budget, the keywords that trigger the publication of their ads and the ad text.
- All these adjustments can be made easily and in real time. This section briefs you on what AdWords is, and how to set up and optimize your account.

Search vs Ads

- Organic Results vs Sponsored Links
- There are two kinds of Google search results: natural results (also known as organic) and AdWords results, classified as sponsored links.
- Sponsored links appear in the right-hand column, and sometimes at the top of the page.

Adwords

Google AdWords Advantages

1. Control
2. Segmentation
3. AdWords
4. Relevance
5. Monitoring

Targeting

- It is also possible to direct campaigns at geographic segments and languages.
- For example, an advertiser who offers home repair services in Seville can display ads only for users located in Seville and who carry out Google searches in Spanish related to home repair services.
- The more targeted the geographical and language settings, and keywords lists, the more chance of a profitable campaign.

Relevance

Google AdWords campaigns offer advertisers the possibility of displaying their ads for a user's search term in a highly relevant manner.

The ads are shown to an audience carrying out an active search, whose receptiveness at that time is very high.

CPC

- Costs (Cost Per Click/Budget)
- The CPC is the maximum an advertiser will pay for a click on an ad. In choosing a maximum CPC for an ad group or a keyword, advertisers ensures that they will never pay more than this defined amount each time the ad registers a click.
- CPC is also a key value that affects ad position.

Daily budget

- The daily budget is the amount that the advertiser pays daily for a given campaign.
- For example, if an advertiser has two campaigns in an account, one with a daily budget of €10.00 and the other €20.00, the total daily budget is €30.00.
- The monthly cost of the whole campaign is €900.00.
- According to this model, if this advertiser established a maximum CPC of €0.20 per keyword, and the entire monthly budget of €900.00 was used, the advertised web pages would have received at least 4,500 visits from users who carried out searches related to the advertiser's business.

Keyword Bidding

The Google AdWords system uses dynamic bidding. There are no pre-established keyword prices; rather, the advertisers decide on the maximum price they will pay for their ads to appear on the search results page (from €0.01).

This price is a key factor that determines ad position in relation to the competition. Just as with the daily budget, the maximum CPC bid can be modified frequently and with no cost to the advertiser.

Ad Rank

- The position in which the ad appears on the Google search results page, Ad Rank, depends on two main factors: maximum CPC bid and the quality of the ad and keywords.
- The Google AdWords system compares the performance and relevance of the keywords and the ad text with the searches carried out to establish a Quality Score.
- This score, combined with the ad's maximum CPC bid, determines the ad position.

Ad Quality

- Calculating the CPC and Quality Score
- Ad Ranking = CPC x Quality Score
- This equation rewards advertisers with an organised and relevant campaign.
- Thus, an advertiser with a maximum CPC lower than that of competitors can publish ads in higher positions if the ads are a higher quality.

Quality Score

- The Quality Score awarded to ads depends mainly on the percentage of clicks, also called click-through rate or CTR. The higher the percentage of clicks (CTR) that an ad receives, the higher the Quality Score.
- Therefore, advertisers should ensure they have very well targeted keywords lists and attractive ad texts, in order to achieve good CTR scores. The CTR formula follows:
- $\text{CTR} = \text{Clicks} / \text{Impressions} \times 100$

- CTR = Clicks / Impressions x 100
- Other Quality Score factors include the relevance of the ad and its keywords in relation to the searches carried out, the account history (overall CTR of all ads and keywords) and other indicators such as landing page quality.
- The following example demonstrates how advertiser A, with a lower maximum CPC bid, displays ads higher than advertiser B.

Ad groups

Sec1:19 (19 of 37) | 100% | Find

Google AdWords Account Structure

The following diagram shows the different levels of an AdWords account. This sample account has two campaigns, each with two ad groups. Each campaign has a specific daily budget, start and end dates, Google Network preferences (depending on whether the advertiser wants its ads published exclusively on the search engine and on other sites in the Google Network) as well as location and language targeting.

Each ad group contains one or more ad texts associated with a keyword list and the CPC.

The diagram illustrates the hierarchical structure of a Google AdWords account:

- Account Level:** Contains fields for Unique email address, Password, and Billing information.
- Campaign Level:** Contains fields for Daily budget, Location targeting, Language targeting, Distribution preference, and End dates. There are two separate Campaign sections, one for each campaign.
- Ad group Level:** Contains fields for One set of keywords and placements, One or more ads, and One display URL domain. There are four Ad group sections, arranged in a 2x2 grid, corresponding to the two campaigns and their respective ad groups.

Dos and Don'ts

for account structure and campaign management

Do...

- create multiple Ad Groups per campaign
- group Campaigns by theme, geography or product line
- make it easy to maintain
- continue refining your keywords and ad text

Don'ts

Don't...

- create just one Ad Group and a big list of keywords
- mismatch keywords in one Ad Group
- run dozens and dozens of keywords with a low CPC
- stop checking your campaign statistics

Choosing Keywords

- To select keywords that will deliver an ad, put yourself in the place of users who are going to search. Which terms would they use?
- The best solution is to make a short list of keywords that are neither too general nor too specific.
- Then expand on this list, including possible variations of these words (such as both singular and plural terms, different spellings, synonyms, etc.).
- Keywords need to be tightly themed and relevant to the ads. Include specific keywords that directly relate to the specific theme of your ad group and landing page.
- For optimal ad visibility, include relevant keyword variations, along with singular and plural versions.

Relevance is key

- The keywords should relate closely to the goods or services you are advertising.
- If you are selling roses, use specific keywords such as Buy red roses rather than generic keywords like Flowers.
 - Generic keywords may be searched for more frequently, but the people searching for may not necessarily be interested in what you are offering.
 - You will receive better results with keyword phrases such as 'Red roses to your door' or 'Long-stemmed roses'.
- To determine the Ad Rank, the position your ad appears in, AdWords takes into account how relevant your ad and your website are to each search query (quality factor).
 - If your keywords are not relevant, your ad may appear a long way down the column of sponsored links or not at all.

Keyword tool

- Use the Keyword Tool
- Use the Keyword Tool to find relevant keyword ideas. You can access the tool at
- <https://adwords.google.com/select/KeywordToolExternal?hl=en-GB>. Experiment!

Keyword Match Types

Broad Match is the default match setting.

If your ad group contains the keyword ‘tennis shoes,’ your ad is eligible to appear when a user’s search query contains either or both words ('tennis' and 'shoes') in any order, and possibly with other terms.

Your ads could also show for singular/plural forms, synonyms, etc.

Broad match keyword:

1. tennis shoesAds may show on searches for:
2. tennis
3. shoes
4. buy tennis shoes
5. tennis shoe photos
6. running shoes
7. tennis sneakers

Phrase Match

- If you enter keywords in quotation marks, as in ‘tennis shoes,’ your ad is eligible to appear when a user searches for the phrase tennis shoes, with the words in that order.
- Your ad can also appear for searches with other terms as long as the search includes the exact phrase you specified.

Phrase match keyword

Ads may show on searches for:

1. ‘tennis shoes’ red tennis shoes
2. buy tennis shoes
3. tennis shoes photo

Ads won’t show on searches for:

1. shoes for tennis
2. tennis shoe
3. tennis sneakers

Phrase match is more targeted than broad match, and more flexible than exact match.

Exact Match

- If you surround your keywords in brackets -- such as [tennis shoes] -- your ad is eligible to appear when a user searches for the specific phrase 'tennis shoes,' in this order, and without other terms in the query.

Ads won't show on searches for:

- [tennis shoes] tennis shoes red tennis shoes
- tennis shoe
- buy tennis shoes

- Negative match

- If your keyword is ‘tennis shoes’ and you add the negative keyword ‘-used,’ your ad will not appear for any searches that contain the word ‘used.’

Keywords: Ads may show on searches for:

1. tennis shoes tennis shoes
2. -used buy tennis shoes
3. tennis

Ads won’t show on searches for:

1. used tennis shoes
2. shoe used for tennis

Ad text

Ad Text

A Google AdWords ad comprises a headline, two description lines, a Display URL and a Destination URL that the user will be directed to after clicking on the ad.

Ads can contain (including spaces) a headline of up to 25 characters, an ad text of 70 characters, and 35 characters for the Display URL. This all appears in four lines: one for the headline, two for the ad text (in two lines of 35 characters) and one line for the Display URL.

New text ad	
Headline	Get visits from Google
Description line 1	Advertise in AdWords
Description line 2	Your campaign active in few minutes
Display URL	adwords.google.com
Destination URL	<input type="text"/> http:// adwords.google.com

Ad preview

[Get visits from Google](#)
Advertise in AdWords
Your campaign active in few minutes
[adwords.google.com](#)

Ad text

A Google AdWords ad comprises a headline, two description lines, a Display URL and a Destination URL that the user will be directed to after clicking on the ad.

Ads can contain (including spaces) a headline of up to 25 characters, an ad text of 70 characters, and 35 characters for the Display URL.

This all appears in four lines: one for the headline, two for the ad text (in two lines of 35 characters) and one line for the Display URL.

How to Write Attractive Ad Text

Ads must be direct and targeted. It is important to include practical information for potential customers, while at the same time bring their attention to features that differentiate the business from its competitors.

If advertisers search Google with the chosen keywords in their listing, they can check the level of competition for those keywords and write ads that take into account the texts used by other companies.

A campaign's success or failure depends largely on the quality of its ads.

Headline

Try to include the advertised good or service in the headline.

A common mistake is to include the business name, which usually already appears in the Display URL, in the headline.

Users rarely search for the business name except for extremely well known brands.

If the user's search term is in the ad text, the term appears in bold when the ad is published, giving it high search result exposure.

This effect is maximised in the headline, which already appears in a larger font.

Description Lines

Lines 2 and 3 must describe the good or service.

It is very important for this text to be clear and informative.

Attract the consumer's attention by including a price, an offer, a call to action, any information that gives you an advantage over the competition.

URL

- The Display URL field must include the website's domain. This URL is visible to the user in the ad.
- The Destination URL, which is not visible in the ad, links to the web page that users will be taken to when they click the AdWord.
- The URL can be the same as the Display URL, but this field can also take the user directly to a specific web page that contains explicit offers, contacts, registration, etc.

Tips on writing effective ad texts

Distinguish yourself from the competition.

Tell users what they can expect from your website and your company with keywords such as excellent service, good value, professional, competent, and quick.

Incorporate a call to action in your ad.

It's better to say Buy flowers rather than Flowers for sale.

If you are selling in a specific town, give the name of the town!

Tips on writing effective ad texts

Use the most important keywords in the ad text.

If search terms appear in the ad, they will appear in bold. This bold text increases the ad's chance of being noticed and subsequent clicks.

Experiment with ad variations.

Let your audience decide the best ad text.

Create multiple versions of your ad and then investigate which ones receive the most clicks.

Remember: Costs incur only when someone clicks on your ad.

Good vs Bad Ad

Good ad text

Masters In Marketing

Approved courses, work placement.

10% discount if you register online!

www.mastersacademy.net

Bad ad text

Academy.net

Academic courses

Tel: +34 902 34 34 34.

www.mastersacademy.net

Content Network and Placements

Google's content network includes millions of high-quality websites, news pages and blogs that partner with Google to display targeted AdWords ads.

Advertisers that display their ads on the content network, increase their advertising reach to target potential customers who visit these sites.

AdWords advertisers can manually select sites on the content network or let Google's targeting technology display their ads on the most relevant pages for advertisers goods and services.

Create Adwords account

The next step is to set the time zone and currency. For the Challenge, please choose US dollars (US\$) as your preferred currency and DO NOT enter any billing information.

Write down the CID number.

Creating your First Campaign

1. Click on the Create your first campaign button.
2. Choose a name for your campaign.
3. Choose what geographical locations you want your ads displayed.
4. Select which languages you want your customers to read.
5. **PAUSE** every campaign you created.
Failure to pause your campaigns will cost your budget and hinder your campaigns.

CPC too low

If your CPC bid limit is too low, your budget may not be used up in the first few days.

If this happens, increase your CPC bid limit after a few days or add more keywords.

Experiment

If your budget is fully used up, you can either increase your budget in order to get more website visitors, or you can optimize your keywords, i.e. pause more expensive keywords and add new keywords.

You can change your CPC bid at any time. The best thing is to get started and begin experimenting.

Invite team members

- Enter the email address and name of the person you want to invite.
- Enter professor's email.
- Professor will verify your accounts.
- Enter the email addresses of your team mates.

Companies

- Beaconbooksgh.com NGO Beacon Books
- www.axisocl.com AXIS
- www.ashesi.edu.gh NGO Ashesi
- (please search online) COFA
- sci-fiwebtech.com. Sic-Fiwebtech
- thevillageofhope.com/ NGO Village of Hope
- (please search online) Leading ladies network
- www.womenstrust.org NGO Women's trust
- Ghanablogging NGO Ghanablogging.com
- GrassrootsAfrica (verify the website) NGO GrassrootsAfrica.org.gh

Homework

Each group – submit writeup of $\frac{1}{2}$ page on the website you want to advertise, good writeups will get first choice.

From Monday lab:

- submit printout of two coupons you created.\
- Submit printout of your website/landing page you created in xampp.
- Submit text of 5 good and 5 bad (spam) ads on the internet (we will use it later).

Bad ad examples

 study guide copper hno3 sulphat

About 2,430 results (0.15 seconds) [Advanced search](#)

rything	<u>Chemistry Formula List Part 1</u> <u>(Examville.com)</u>	Ads
ges	22 Sep 2010 ... Only All chloride salts are not decomposable by heat certain sulphate salts are decomposed ... Example: Zinc sulphate , Copper (II) sulphate , Iron (III) sulphate NH ₄ Cl ... It is soluble in - diluted HNO ₃ White precipitate is 2- formed.	<u>Chemicals</u> Order premium quality products from the specialty chemicals company <u>www.lanxess.in</u>
eos	<u>Examville.com Organic Chemistry Study Review Guide - Benzene ...</u> <u>www.docstoc.com/.../Chemistry-Formula-List-Part-1-(Examvillecom)</u> - Cached	See your ad here »
vs		
pping		
iks		
CBS		
3S	<u>Nomenclature Worksheet # 4</u> 24 Sep 2009 ... CuNO ₃ copper (I) nitrate 9. francium germanate Fr ₂ GeO ₃ 29 ...	
albme		

Untargeted ad (which keyword was used?)

Google

html tag name value reference Search

About 165,000,000 results (0.23 seconds) Advanced search

Everything

- Images
- Videos
- News
- Shopping
- Books
- Places
- Blogs
- Realtime
- Discussions

The web

[Links in HTML documents](#)

An anchor **name** is the **value** of either the **name** or **id** attribute when used in the ... that **id** and **name** must be the same when both appear in an element's start **tag**: ... in the **HTML DTD**, the **name** attribute may contain character **references**. ...

Introduction to links and anchors - The A element - Document relationships: the ...

www.w3.org/TR/html401/struct/links.html - Cached - Similar

[HTML input tag](#)

In **HTML**, the <input> **tag** has no end **tag**. In **XHTML**, the <input> **tag** must be properly closed, like this <input />. ... **name**, **name**, Specifies a **name** for an input element, STF ... **value**, **value**, Specifies the **value** of an input element, STF ... Tutorials, references, and examples are constantly reviewed to

Ads

Name

1000s of Baby Names by Gender, Origin, Meaning, Free Registration.
babynamewinder.babycenter.in

[See your ad here »](#)

Auctions

What is an auction

- An **auction** is a process of buying and selling goods or services by offering them up for bid, taking bids, and then selling the item to the highest bidder. In economic theory, an auction may refer to any mechanism or set of trading rules for exchange.
- The word "auction" is derived from the Latin *augēre*, which means "to increase" or "augment"

Traditional Auction



- There are several variations on the basic auction form, including
- time limits, minimum or
- maximum limits on bid prices, and special rules for determining the winning bidder(s) and sale price(s).
- Participants in an auction may or may not know the identities or actions of other participants.
- Depending on the auction, bidders may participate in person or remotely through a variety of means, including telephone and the internet.
- The seller usually pays a commission to the auctioneer or auction company based on a percentage of the final sale price.

Ancient auction, Candle auction

- Romans
- British 1600s
- In some parts of England during the seventeenth and eighteenth centuries auction by candle was used for the sale of goods and leaseholds. This auction began by lighting a candle after which bids were offered in ascending order until the candle spluttered out. The high bid at the time the candle extinguished itself won the auction

Sotheby and Christie

- Sotheby's, now the world's second-largest auction house,[11] held its first auction in 1744. Christie's, now the world's largest auction house,[11] was established around 1766.

Fish auction in Hawaii



English Auction

- **English auction**, also known as an *open ascending price auction*. This type of auction is arguably the most common form of auction in use today.
- Participants bid openly against one another, with each subsequent bid higher than the previous bid.
- An auctioneer may announce prices, bidders may call out their bids themselves (or have a proxy call out a bid on their behalf), or bids may be submitted electronically with the highest current bid publicly displayed. At least two bidders are required.

English ..

- The auction ends when no participant is willing to bid further, at which point the highest bidder pays their bid.
- Alternatively, if the seller has set a minimum sale price in advance (the 'reserve' price) and the final bid does not reach that price the item remains unsold.
- Sometimes the auctioneer sets a minimum amount by which the next bid must exceed the current highest bid.
- The most significant distinguishing factor of this auction type is that the current highest bid is always available to potential bidders.
- The English auction is commonly used for selling goods, most prominently antiques and artwork.

Dutch auction

- **Dutch auction** also known as an *open descending price auction*.
- In the traditional Dutch auction the auctioneer begins with a high asking price which is lowered until some participant is willing to accept the auctioneer's price.
- The winning participant pays the last announced price.
- The Dutch auction is named for its best known example, the Dutch tulip auctions.

Dutch

- "Dutch auction" is also sometimes used to describe online auctions where several identical goods are sold simultaneously to an equal number of high bidders.
- In addition to cut flower sales in the Netherlands, Dutch auctions have also been used for perishable commodities such as fish and tobacco.
- In practice, however, the Dutch auction is not widely used

Sealed first-price auction

- also known as a *first-price sealed-bid auction* (FPSB). In this type of auction all bidders simultaneously submit sealed bids so that no bidder knows the bid of any other participant. The highest bidder pays the price they submitted. This type of auction is distinct from the English auction, in that bidders can only submit one bid each. Furthermore, as bidders cannot see the bids of other participants they cannot adjust their own bids accordingly. Sealed first-price auctions are commonly used in tendering, particularly for government contracts and auctions for mining leases

Vickrey auction

- also known as a *sealed-bid second-price auction*. This is identical to the sealed first-price auction except that the winning bidder pays the second highest bid rather than their own. This is very similar to the [proxy bidding](#) system used by [eBay](#), where the winner pays the second highest bid plus a bidding increment (e.g., 10%). Although extremely important in auction theory, in practice Vickrey auctions are rarely used

Multi-unit auctions

- Multi-unit auctions sell more than one identical item at the same time, rather than having separate auctions for each. This type can be further classified as a uniform price auction or a discriminatory price auction.

All-pay auction (politics)

- is an auction in which all bidders must pay their bids regardless of whether they win. The highest bidder wins the item. All-pay auctions are primarily of academic interest, and may be used to model lobbying/bribery (bids are political contributions) or competitions such as a running race

Buy Now

- **Buyout auction** is an auction with a set price (the 'buyout' price) that any bidder can accept at any time during the auction, thereby immediately ending the auction and winning the item. If no bidder chooses to utilize the buyout option before the end of bidding the highest bidder wins and pays their bid.

Buy now

- Buyout options can be either *temporary* or *permanent*. In a temporary buyout auction the option to buy out the auction is no longer available after the first bid is placed.
- In a permanent buyout auction the buyout option remains available throughout the entire auction until the close of bidding.
- The buyout price can either remain the same throughout the entire auction, or vary throughout according to preset rules or simply at the whim of the seller

Combinatorial auction

- is any auction for the simultaneous sale of more than one item where bidders can place bids on an "all-or-nothing" basis on "packages" rather than just individual items.
- That is, a bidder can specify that he or she will pay for items A and B, but only if he or she gets *both*.
- In combinatorial auctions determining the winning bidder can be a complex process where even the bidder with the highest individual bid is not guaranteed to win.

Combinatorial auction e.g.

- For example, in an auction with four items (W, X, Y and Z), if Bidder A offers \$50 for items W & Y, Bidder B offers \$30 for items W & X, Bidder C offers \$5 for items X & Z and Bidder D offers \$30 for items Y & Z, the winners will be Bidders B & D while Bidder A misses out because the *combined* bids of Bidders B & D is higher (\$60) than for Bidders A and C (\$55).

No-reserve auction (bargain),

- NR also known as an *absolute auction*, is an auction in which the item for sale will be sold regardless of price.
- From the seller's perspective, advertising an auction as having no reserve price can be desirable because it potentially attracts a greater number of bidders due to the possibility of a bargain

Seminar

- 15 minutes/person
- About 15 slides
- Any ecommerce topic
 - Not discussed yet
 - From the text book
 - Research topic
 - Business topic
 - Ghanian law
 - Ghanian e-business
 - Sign up for the presentation

Seminar

- Upload presentation to courseware.
- All presentations can appear in exams.
- Put your ideas/questions in google moderator (see link on courseware).
- See machine learning books for download from courseware.

What else

- Adwords final presentation
- Final exam
- B2B XML Api
- E-voting
- E-Cash, payments
- Ghanian context
- Shopping carts

NR

- If more bidders attend the auction a higher price might ultimately be achieved because of heightened competition from bidders.
- This contrasts with a *reserve auction*, where the item for sale may not be sold if the final bid is not high enough to satisfy the seller.
- In practice, an auction advertised as "absolute" or "no-reserve" may nonetheless still not sell to the highest bidder on the day,
- for example, if the seller withdraws the item from the auction or extends the auction period indefinitely, although these practices may be restricted by law in some jurisdictions or under the terms of sale available from the auctioneer.

Reserve auction

- is an auction where the item for sale may not be sold if the final bid is not high enough to satisfy the seller - that is, the seller *reserves* the right to accept or reject the highest bid.
- In these cases a set 'reserve' price known to the auctioneer, but not necessarily to the bidders, may have been set in advance below which the item may not be sold.

Reserve auction

- The reserve price may be *fixed* or *discretionary* - in the latter case, the decision to accept a bid is deferred to the auctioneer, who may accept a bid that is marginally below it.
- A reserve auction is safer for the seller than a no-reserve auction as they are not required to accept a low bid, but this could potentially result in a lower final price than might otherwise be the case if this means that less interest is generated in the sale

Reverse auction

- is a type of auction in which the role of the buyer and seller are reversed, with the primary objective to drive purchase prices downward.
- While ordinary auctions provide suppliers the opportunity to find the best price among interested buyers, reverse auctions give buyers a chance to find the lowest-price supplier.
- During the reverse auction, suppliers may submit multiple offers, usually as a response to competing suppliers' offers, bidding down the price of a good or service to the lowest price they are willing to offer.
- By revealing the competing bids in real time to each participating supplier, reverse auctions promote “information transparency”.
- This, coupled with the dynamic bidding process, improves the chances of reaching the fair market value of the purchase

Silent auction

- is a variant of an English auction where bids are written on a sheet of paper. At the predetermined end of the auction the highest listed bidder wins the item.
- This auction is often used in [charity](#) events, with many items auctioned simultaneously with a common finish time.
- The auction is "silent" in that there is no auctioneer, the bidders writing their bids on a bidding sheet often left on a table near the item.
- Other variations of this type of auction may include sealed bids. The highest bidder pays the price he or she submitted

Top-Up Auction

- is a variation on the all-pay auction, primarily used for charity events.
- Bidders must pay the difference between their bid and the next lowest bid, whether they win or not.
- Only the winning bidder does not have to pay the "top-up" fee, but does have to pay for the item.

Walrasian auction

- *Walrasian tâtonnement* is an auction in which the auctioneer takes bids from both buyers and sellers in a market of multiple goods.
- The auctioneer progressively either raises or drops the current proposed price depending on the bids of both buyers and sellers, the auction concluding when supply and demand exactly balance.
- As a high price tends to dampen demand while a low price tends to increase demand, in theory there is a particular price point somewhere in the middle where supply and demand will match.

Auctions can differ in #users

- In a *supply* (or *reverse*) auction, m sellers offer a good that a buyer requests
- In a *demand* auction, n buyers bid for a good being sold
- In a double auction n buyers bid to buy goods from m sellers
- Prices are *bid* (or *offered*) by buyers and *asked* by sellers. Auctions may also differ by the procedure for bidding (or asking, as the case may be):
 - In an *open* auction participants may repeatedly bid and are aware of each other's previous bids.
 - In a *closed* auction buyers and/or sellers submit sealed bids

Pricing

- Auctions may differ as to the price at which the item is sold, whether the first (best) price, the second price, the first *unique* price or some other. Auctions may set a reservation price which is the least/maximum acceptable price for which a good may be sold/bought.
- Without modification, *auction* generally refers to an open, demand auction, with or without a reservation price (or *reserve*), with the item sold to the highest bidder.

Uses

- The antique business, where besides being an opportunity for trade they also serve as social occasions and entertainment
- In the sale of collectibles such as stamps, coins, classic cars, fine art and luxury real estate
- The wine auction business, where serious collectors can gain access to rare bottles and mature vintages, not typically available through retail channels

Uses

- In the sale of all types of real property including residential and commercial real estate, farms, vacant lots and land.
- For the sale of consumer second-hand goods of all kinds, particularly farm (equipment) and house clearances and online auctions.
- Sale of industrial machinery, both surplus or through insolvency.

Uses

- In commodities auctions, like the fish wholesale auctions
- In livestock auctions where sheep, cattle, pigs and other livestock are sold. Sometimes very large numbers of stock are auctioned, such as the regular sales of 50,000 or more sheep during a day in New South Wales.
- In wool auctions where international agents purchase lots of wool
- Thoroughbred horses, where yearling horses and other bloodstock are auctioned.

Uses

- In legal contexts where forced auctions occur, as when one's farm or house is sold at auction on the courthouse steps.
- Travel tickets. One example is SJ AB in Sweden auctioning surplus at Tradera (Swedish eBay).

Large uses

- Sales of businesses
- Spectrum auctions, in which companies purchase licenses to use portions of the electromagnetic spectrum for communications (e.g., mobile phone networks)
- Private electronic markets using combinatorial auction techniques to continuously sell commodities (coal, iron ore, grain, water...) to a pre-qualified group of buyers (based on price and non-price factors)
- Timber auctions, in which companies purchase licenses to log on government land
- Timber allocation auctions, in which companies purchase timber directly from the government Forest Auctions

- Electricity auctions, in which large-scale generators and consumers of electricity bid on generating contracts
- Environmental auctions, in which companies bid for licenses to avoid being required to decrease their environmental impact. These include auctions in emissions trading schemes.

- Debt auctions, in which governments sell debt instruments, such as bonds, to investors. The auction is usually sealed and the uniform price paid by the investors is typically the best non-winning bid. In most cases, investors can also place so called *non-competitive bids*, which indicates an interest to purchase the debt instrument at the resulting price, whatever it may be
- Auto auctions, in which car dealers purchase used vehicles to retail to the public.

Bidding

- Bid shading is placing a bid which is below the bidder's actual value for the item. Such a strategy risks losing the auction, but has the possibility of winning at a low price. Bid shading can also be a strategy to avoid the Winner's curse.

Chandelier bidding

- A practice, especially by high-end art auctioneers, of raising false bids at crucial times in the bidding process in order to create the appearance of greater demand or to extend bidding momentum for a work on offer. To call out these nonexistent bids, auctioneers might fix their gaze at a point in the auction room that is difficult for the audience to pin down.
- In the United Kingdom, this practice is legal on Property Auctions up to but not including the reserve price, and is also known as "Off The Wall" bidding

Collusion

- Whenever bidders at an auction are aware of the identity of the other bidders there is a risk that they will form a "ring" and thus manipulate the auction result, a practice known as collusion.
- By agreeing to bid only against outsiders, never against members of the "ring", competition becomes weaker, which may dramatically affect the final price level.
- After the end of the official auction an unofficial auction will take place among the "ring" members.
- The difference in price between the two auctions will then be split among the members.

Dummy bid

- In an English auction a dummy bid is a bid made by a dummy bidder acting in collusion with the auctioneer or vendor, designed to deceive genuine bidders into paying more.
- In a First price auction a dummy bid is an unfavourable bid designed so as not to become the winning bid.
- (The bidder does not want to win this auction, but he wants to make sure that he will be invited to the next auction).

Suggested opening bid (SOB)

- There will usually be some kind of (rough) estimate as to what the object will fetch. In an ascending open auction it is considered important that there should be at least a 50 percent increase in the bids from start to finish.
- To accomplish this the auctioneer must start the auction by announcing a Suggested Opening Bid, SOB, that is low enough to be immediately accepted by one of the bidders. Once there is an Opening Bid there will quickly be several other higher bids submitted.
- Experienced auctioneers will often select an SOB that is about 45 percent of the (lowest) estimate. Thus there is a certain margin of safety to ensure that there will indeed be a lively auction with many bids submitted. Several observations indicate, that the lower the SOB, the higher the final winning bid will be.
- This is due to the increase in number of bidders attracted by the low SOB. When 50 bidders compete with each other the winning bid will be about twice as high as when only two bidders compete. Sometimes with [English auction](#) there will be more than 50 bidders.

Bid distribution

- A Chi-square distribution shows many low bids but few high bids. Bids "show up together"; without several low bids there will not be any high bids.

What price to begin at?

- Another approach to choosing a SOB: The auctioneer may achieve good success by asking the expected final sales price for the item, as this method suggests to the buyer the amount of the item's particular value.
- For instance, say an auctioneer is about to sell a \$1,000 car at a sale.
- Instead of asking \$100, hoping to entice wide interest (for who wouldn't want a \$1,000 car for \$100?), the auctioneer may still suggest the opening bid of \$1,000; although the first bidder may finally begin bidding at a mere \$100, the final bid may more likely approach \$1,000.

Multiplying large numbers In php

A

B

multiply

A=1111111111111111111111111111

B=2222222222222222222222222222

A+B =2.3333333333E+24

big Sum =23333333333333333333333333

A*B =2.46913580247E+47

big Mult=246913580246913580246913308641975308641975308642

Show results

```
1.  <?php  
2.  $A = $HTTP_GET_VARS["A"];  
3.  $B = $HTTP_GET_VARS["B"];  
4.  show('A',$A);  
5.  show('B',$B);  
6.  show('A+B ', $A + $B);  
7.  show('big Sum ', sum($A,$B) );  
8.  show('A*B ', $A * $B);  
9.  show('big Mult ', mult($A,$B) );  
10. ?>
```

Mult two big numbers

```
1. // mult("12","34") = 408
2. function mult($A,$B) {
3.     $M = ""; $b = 0;
4.     while( last_digit($B,$b) ) {
5.         $M = sum( mult_one_digit($A, $b) , $M.'0');
6.     }
7.     return $M;
8. }
```

Sum two big numbers

```
1. // sum("11", "22") = 33.
2. function sum($A, $B) {
3.     $a = 0; $b = 0; $carry = 0; $S = "";
4.     while( strlen($A)>0 || strlen($B)>0) {
5.         last_digit($A,$a);
6.         last_digit($B,$b);
7.         $t = $a + $b + $carry;
8.         $carry = (int) ($t / 10);
9.         $t = $t % 10;
10.        $S = $t . $S;
11.    }
12.    if( $carry > 0)
13.        $S = $carry . $S;
14.    return $S;
15. }
```

Multiply one digit at a time

```
1. // mult_one_digit("12",2) = 24
2. function mult_one_digit($A,$b) {
3.     $S = "";
4.     while($b-- > 0)
5.         $S = sum($S,$A);
6.     return $S;
7. }
```

Get least significant digit

```
1. // last_digit("1234",...) .. returns true
2. // and (A="123",a=4).
3. function last_digit(&$A, &$a) {
4.     $a = 0;
5.     if (strlen($A)<1)
6.         return false;
7.     $a = (int) substr($A,-1);
8.     $A = substr($A,0,-1);
9.     return true;
10. }
```

Shopping e-carts

[Forrester Research](#) estimates that the United States online retail industry will be worth \$279 billion in 2015. From wikipedia **Online shopping**

History

- 1990 [Tim Berners-Lee](#) created the first www
- 1994 - online banking and online pizza shop [Pizza Hut](#).
- 1994, [Netscape](#) introduced SSL encryption of data transferred online, which has become essential for secure online shopping.
- In 1994 the German company [Intershop](#) introduced its first online shopping system.
- 1995 [Amazon](#) launched its online shopping site,
- 1996 [eBay](#) appeared.

E-Saleman?

- In a conventional retail store, clerks are generally available to answer questions. Some online stores have real-time chat features, but most rely on e-mail or phone calls to handle customer questions.



Why eshops?

- Online stores are usually available 24 hours a day
- Choices
- Convenience
- Pricing
- Shipping to remote areas.

Aggregation

- High-volume websites, such as [Yahoo!](#), [Amazon.com](#) and [eBay](#), offer hosting services for online stores to all size retailers.
- These stores are presented within an integrated navigation framework. Collections of online stores are sometimes known as [virtual shopping malls](#) or [online marketplaces](#).

Comparison shopping

- One advantage of shopping online is being able to quickly seek out deals for items or services with many different vendors (though some local search engines do exist to help consumers locate products for sale in nearby stores).
- Search engines, online price comparison services and discovery shopping engines can be used to look up sellers of a particular product or service.

Why aggregation

- Many companies that don't have internal resources or expertise (such as do-it-yourselfers) work with a web development firm to handle all or some of the facets of the online shopping set-up, including integration of the e-commerce platform and hosting.
- Full service digital companies can design, develop and set-up ecommerce sites so that they're up and running with existing merchant accounts or new ones."

Product Reviews

- Reviews on electronics (57%) such as DVD players, cell phones or PlayStations and so on, reviews on cars (45%), and reviews on software (37%) play an important role and have influence on consumers who tend to make purchases and buy online.
- from wikipedia on Nielsen survey.

E-shops

- Sticking with known stores, or attempting to find independent consumer reviews of their experiences; also ensuring that there is comprehensive contact information on the website before using the service, and noting if the retailer has enrolled in industry oversight programs such as trust mark or trust seal.

Peer recommendations

- In addition to online reviews, peer recommendations on the online shopping pages or social media play a key role for online shoppers while researching future purchases of electronics, cars and travel or concert bookings.
- On the other hand - 40% of online shoppers indicate that they would not even buy electronics without consulting online reviews first.

Good e-shops

- Ensuring that the retailer has an acceptable **privacy policy posted**. For example note if the retailer does not explicitly state that it will not share private information with others without consent.
- Ensuring that the vendor address is protected with SSL (see above) when entering credit card information. If it does the address on the credit card information entry screen will start with "**HTTPS**".
- Using **strong passwords**, without personal information. Another option is a "pass phrase," which might be something along the lines: "I shop 4 good a buy!!" These are difficult to hack, and provides a variety of upper, lower, and special characters and could be site specific and easy to remember.

Evaluating eshops

- Before buying from a new company, evaluate the website by considering issues such as: the professionalism and user-friendliness of the site; whether or not the **company lists a telephone number** and/or street address along with **e-contact information**;
- whether a fair and reasonable **refund** and return policy is clearly stated; and whether there are hidden price inflators, such as excessive shipping and handling charges.

Shopping cart systems

- Simple systems allow the offline administration of products and categories.
- The shop is then generated as HTML files and graphics that can be uploaded to a webspace.
- These systems do not use an online database.

Shopping cart

Item	Unit Price	Quantity	Total	
 Naruto Volume 2	\$10	<input type="text" value="1"/>	\$10	Delete
 Hunter X Hunter Volume 1	\$10	<input type="text" value="2"/>	\$20	Delete
	Sub-total		\$30	
	Shipping		\$5	
	Total		\$35	Update Cart

[<< Continue Shopping](#)

[Proceed To Checkout >>](#)

You may have seen it on another shopping cart solution that to remove an item the customer must set the quantity to zero then click the 'Update Cart' button. That is the **wrong way** to do it because it makes a very simple action difficult.

Outsourced shopping carts

- Google checkout
- Paypal API
- Amazon stores



Secure Checkout

Order Details - Example.com, (800) 555-1234 1576 Random Road, Boston, MA 01234

Qty	Item	Price
1	Cake pan - 10-in. Nonstick Angel Food Cake Pan by Flying M...	\$15.99
	Shipping & handling 2nd Day (est. \$1.49)	\$1.49
	Tax:	-
	Subtotal:	\$17.48

Create a **Google** Account to complete this purchase

Current email:

Choose a password:

[Password strength:](#)

Re-enter password:

Credit card number:

Expiration date: / CVC: [What's this?](#)

Name on card:

Billing address line 1:

Billing address line 2:
(optional)

Phone number:

e.g. 650-555-1212. Required for account verification.

I agree to the [Terms of Service](#).

[Agree and continue »](#)

You can still make changes to
your order on the next page.

Or sign in
If you already have a **Google** Account

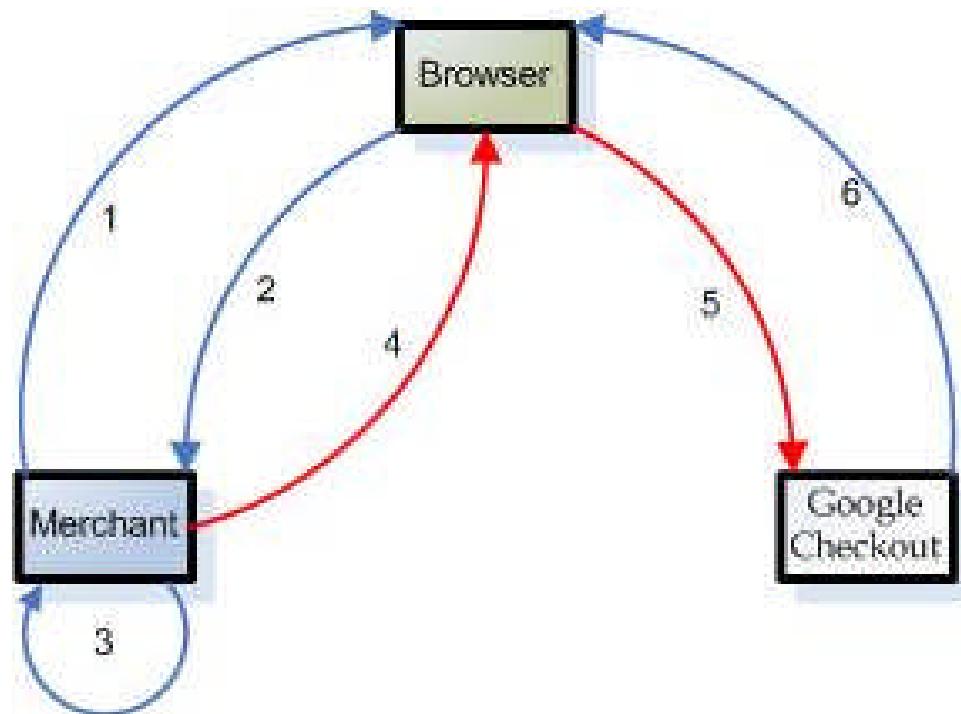
Email:

Password:

[Sign in and continue](#)

[Forgot your password?](#)

3rd party cart interaction



Integrated shopping cart

- A high end solution can be bought or rented as a standalone program or as an addition to an [enterprise resource planning](#) program.
- It is usually installed on the company's own webserver and may integrate into the existing [supply chain](#) so that ordering, payment, delivery, accounting and warehousing can be automated to a large extent.
 - Database integration.
 - Payment integration
 - Delivery tracking

Online shopping payments

- Billing to mobile phones and landlines
- Cash on delivery (C.O.D., offered by very few online stores)
- Check
- Debit card
- Direct debit in some countries
- Electronic money of various types
- Gift cards
- Postal money order
- Wire transfer/delivery on payment

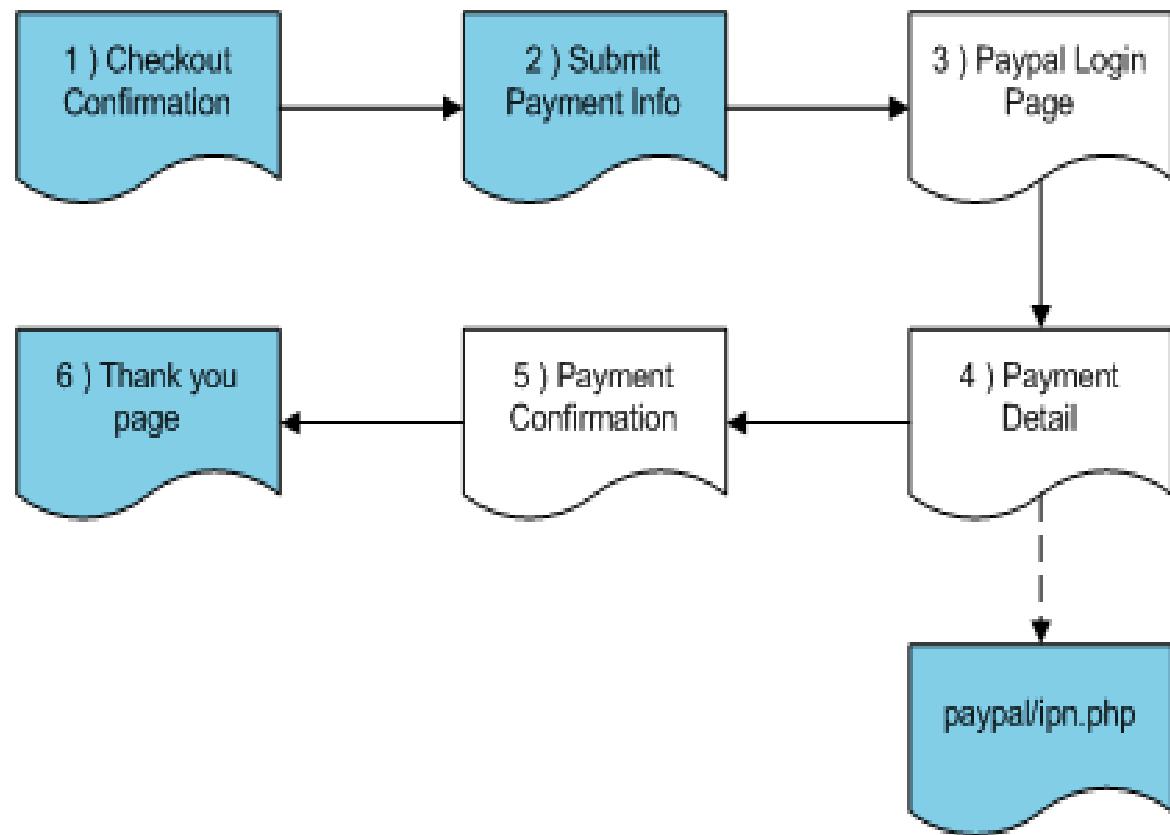
Payment screen

Payment Information	
<input type="checkbox"/> Same as shipping information	
First Name	<input type="text"/>
Last Name	<input type="text"/>
Address1	<input type="text"/>
Address2	<input type="text"/>
Phone Number	<input type="text"/>
Province / State	<input type="text"/>
City	<input type="text"/>
Postal / Zip Code	<input type="text"/>
Payment Method	
<input checked="" type="radio"/> Paypal <input type="radio"/> Cash on Delivery	

Confirm checkout

Ordered Item			
Item	Unit Price	Total	
2 x Naruto Volume 2	\$10	\$20	
1 x Hunter X Hunter Volume 1	\$10	\$10	
		Sub-total	\$30
		Shipping	\$5
		Total	\$35

Shipping Information	
First Name	Cutty
Last Name	Flam
Address1	Old Warehouse
Address2	Under The Bridge
Phone Number	777-FRANKY



PCI Payment Card Industry Data Security Standard

- The PCI security standards are a blanket of regulations set in place to safeguard payment account data security.
- The council that develops and monitors these regulations is composed of the leading providers in the payment industry: American Express, Discover Financial Services, JCB International, MasterCard Worldwide and Visa Inc. Inc. International.
- Essentially, they define the best practices for storing, transmitting, and handling of sensitive information over the internet

Creating a eshop – mysql db

- CREATE TABLE books (
- id int(6) unsigned NOT NULL auto_increment,
- title varchar(100) NOT NULL default "",
- author varchar(100) NOT NULL default "",
- price decimal(3,2) NOT NULL default '0.00',
- PRIMARY KEY (id)
-) TYPE=MyISAM;
- INSERT INTO books VALUES (1, 'Where God Went Wrong', 'Oolon Colluphid', '24.99');
- INSERT INTO books VALUES (2, 'Some More of God\'s Greatest Mistakes', 'Oolon Colluphid', '17.99');
- INSERT INTO books VALUES (3, 'Who Is This God Person Anyway?', 'Oolon Colluphid', '14.99');

PHP Sessions

- Server side user data.
- To ‘switch it on’, simply add `session_start();` at the top of your code;
- you can then store values (or arrays) in session variables:

Cookies and sessions

- The default behaviour of PHP sessions is to store the session ID (a long string of numbers and letters that identifies you to the server) as a cookie on your computer;
- however, even if you have cookies disabled this functionality will still work – PHP will instead append the session ID to each link on the site (in the form ‘mypage.php?PHPSESSID=’) so that it can continue to accurately track visitors from page to page.

Cart is a session variable

- // List of product ids in the cart
- `$_SESSION['cart'] = "1,1,3,1,2";`
- E.g. a cart containing “1,1,3,1,2” has four items;
- three of product #1, and
- one each of products #2 and #3.

Cart php

```
1. function writeShoppingCart() {  
2.     $cart = $_SESSION['cart'];  
3.     if (!$cart) {  
4.         return '<p>You have no items in your shopping  
cart</p>';  
5.     } else { // Parse the cart session variable  
6.         $items = explode(',',$cart);  
7.         $s = (count($items) > 1) ? 's':";  
8.         return '<p>You have <a href="cart.php">' .  
9.             count($items). ' item' . $s.  
10.            ' in your shopping cart</a></p>';  
11.    }  
12.}
```

Product links

- Add to cart
1. \$cart = \$_SESSION['cart'];
 2. if (\$cart) {
 3. \$cart .= ','.\$_GET['id']; // Append new item
 4. } else {
 5. \$cart = \$_GET['id']; // new cart with 1 item
 6. }
 7. \$_SESSION['cart'] = \$cart; // save it

Cart contents as an array

- // array consists of product_id=>quantity pairs
 1. \$cart = \$_SESSION['cart'];
 2. if (\$cart) {
 3. \$items = explode(',',\$cart);
 4. \$contents = array();
 5. foreach (\$items as \$item) {
 6. \$contents[\$item] = (isset(\$contents[\$item])) ?
 7. \$contents[\$item] + 1 : 1;
 8. }

Total bill, priced using the database

- \$total = 0;
- \$output[] = '<table>';
- foreach (\$contents as \$id=>\$qty) {
- \$sql = 'SELECT * FROM books WHERE id = '.\$id;
- \$result = \$db->query(\$sql);
- \$row = \$result->fetch();
- extract(\$row);

Print the bill as a table

- \$output[] = '<tr>';
- \$output[] = '<td>Remove</td>';
- \$output[] = '<td>'.\$title.' by '.\$author.'</td>';
- \$output[] = '<td>£'.\$price.'</td>';
- \$output[] = '<td><input type="text" name="qty'.\$id.'" value="'.\$qty.'" size="3" maxlength="3" /></td>';

Total bill is

- \$output[] = '<td>£'.(\$price * \$qty).'</td>';
- \$total += \$price * \$qty;
- \$output[] = '</tr>';
- }
- \$output[] = '</table>';
- \$output[] = '<p>Grand total:£'.\$total.'</p>';

Total

Your Shopping Cart - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Your Shopping Cart

Qty	Product	Price Each	Remove?
3	FIFA Soccer 2002	\$36.50	Remove
2	SSK Tricky	\$45.50	Remove
4	Tony Hawk 3	\$23.95	Remove

[<< Keep Shopping](#) Total: \$296.30

Done Local intranet



Remove item from cart?

- Remove
- \$cart = \$_SESSION['cart'];
- \$action = \$_GET['action'];
- switch (\$action) {

Case 'add'

- case 'add': // ADD
- if (\$cart) {
- \$cart .= ','.\$_GET['id'];
- } else {
- \$cart = \$_GET['id'];
- }
- break;

Case 'delete'

- case 'delete':
- if (\$cart) {
- \$items = explode(',',\$cart);
- \$newcart = '';
- foreach (\$items as \$item) {
- if (\$_GET['id'] != \$item) { // SKIP deleted item
- if (\$newcart != "") {
- \$newcart .= ','.\$item;

Refill the cart, except the deleted item

- } else {
- \$newcart = \$item; } } }
- \$cart = \$newcart; }
- break; }
- \$_SESSION['cart'] = \$cart;

Cart as a form

- \$output[] = '<form
action="cart.php?action=update" method="post"
id="cart">';
- \$output[] = '<table>';
- foreach (\$contents as \$id=>\$qty) {
- \$sql = 'SELECT * FROM books WHERE id =
' . \$id;
- \$result = \$db->query(\$sql);
- \$row = \$result->fetch();
- extract(\$row);
- \$output[] = '<tr>';

Form 2

- \$output[] = '<td>Remove</td>';
- \$output[] = '<td>'.\$title.' by '.\$author.'</td>';
- \$output[] = '<td>£'.\$price.'</td>';
- \$output[] = '<td><input type="text" name="qty'.\$id.'" value="'.\$qty.'" size="3" maxlength="3" /></td>';
- \$output[] = '<td>£' .(\$price * \$qty). '</td>';
- \$total += \$price * \$qty;

Form 3

- \$output[] = '</tr>';
- }
- \$output[] = '</table>';
- \$output[] = '<p>Grand total:
£'.\$total.'</p>';
- \$output[] = '<div><button
type="submit">Update
cart</button></div>';
- \$output[] = '</form>';

Case 'update'

- case 'update':
- if (\$cart) {
- \$newcart = ";
- foreach (\$_POST as \$key=>\$value) {
- if (stristr(\$key,'qty')) { // Change Quantity?
- \$id = str_replace('qty','', \$key);

Updating

- \$items = (\$newcart != "") ?
explode(',',\$newcart) : explode(',',\$cart);
- \$newcart = ":";
- foreach (\$items as \$item) {
- if (\$id != \$item) {

Updated

- if (\$newcart != "") {
- \$newcart .= ','.\$item;
- } else {
- \$newcart = \$item; } } }
- for (\$i=1;\$i<=\$value;\$i++) {
- if (\$newcart != "") {
- \$newcart .= ','.\$id;
- } else {
- \$newcart = \$id; } } } }
- \$cart = \$newcart;
- break;

Checkout place order

The screenshot shows a Google Checkout interface for placing an order. The page title is "Review and place order". At the top right is the "Google Checkout" logo. Below it, the "Order Details" section shows the address: "Google, Inc., 555 555-5555, 1600 Amphitheatre Pky, Mountain View, CA 94043 US" with a "Change order" link and a "1" button. The main table lists items with columns for "Qty", "Item", and "Price". The table includes two items: "Dry Food Pack AA1453" and "MegaSound 2GB MP3 Player". A "Shipping & Handling" row shows "SuperShip (\$10.00)" with a dropdown menu and a "\$10.00" price. A "Tax (NY)" row shows a value of "\$15.74". The total is listed as "Total: \$238.74". Numbered callouts (1 through 8) point to various elements: 1 points to the "Change order" link; 2 points to the quantity of the first item; 3 points to the item name; 4 points to the item description; 5 points to the price of the second item; 6 points to the shipping method; 7 points to the coupon input field; and 8 points to the tax amount.

Qty	Item	Price
2	Dry Food Pack AA1453 - A pack of highly nutritious dried food for emergency	\$35.00
1	MegaSound 2GB MP3 Player - Portable MP3 player - stores 500 songs	\$178.00
Shipping & Handling SuperShip (\$10.00)		\$10.00
Tax (NY)		\$15.74
Total: \$238.74		

Thanks for ordering

lmno@gmail.com | [My Account](#) | [Help](#) | [Sign out](#)

 Google Checkout

✓ **Dave, thanks for your order!**

- A copy of your receipt has been emailed to you and saved in your Google Checkout [purchase history](#).
- You can check your receipt at any time for [up-to-date order status](#).

[Return to Example.com »](#)

Evoting



In 1869 Thomas Edison received US patent 90,646 for an “electronic voting device.” He tried to sell his invention to the Massachusetts legislative bodies, unsuccessfully. A century later, we are once again attempting to apply electronic wizardry to expedite the democratic process.

Evoting from wikipedia

- **Electronic voting** (also known as **e-voting**) is a term encompassing several different types of [voting](#), embracing both electronic means of casting a vote and electronic means of counting votes.
- Electronic voting technology can include [punched cards](#), [optical scan voting systems](#) and specialized voting kiosks (including self-contained [direct-recording electronic voting systems](#), or DRE). It can also involve transmission of [ballots](#) and votes via telephones, private [computer networks](#), or the [Internet](#).

Diebold voting machine (RIP)



E voting by Mark Ryan

- **Electronic voting promises**
 - *convenient* way of recording and tallying votes
 - *security* against fraud and manipulation
 - *transparency* for voters and candidates

Uses

- It could be used in a variety of kinds of elections
 - small committees or on-line communities
 - student elections, trade unions, local government
 - full national elections
- Cost savings in ballot paper, people.

Govt

- *Governments worldwide are investing in e-voting*
- *Electronic systems potentially allow large scale undetectable fraud*
 - In contrast, fraud in manual systems limited by requirement to generate or dispose of paper, which is quite hard to do undetectably in presence of TV cameras.
- *There are protocols which are capable of guaranteeing strong properties*
 - But few companies are marketing them, and few countries are interested in them for their government elections

Requirements 1

- **1. Accuracy:** the declared outcome corresponds properly to the votes cast.
- **2. Eligibility:** only eligible voters can vote, and only once.
- **3. Fairness:** no early results; i.e. no voter can be influenced by votes already made.

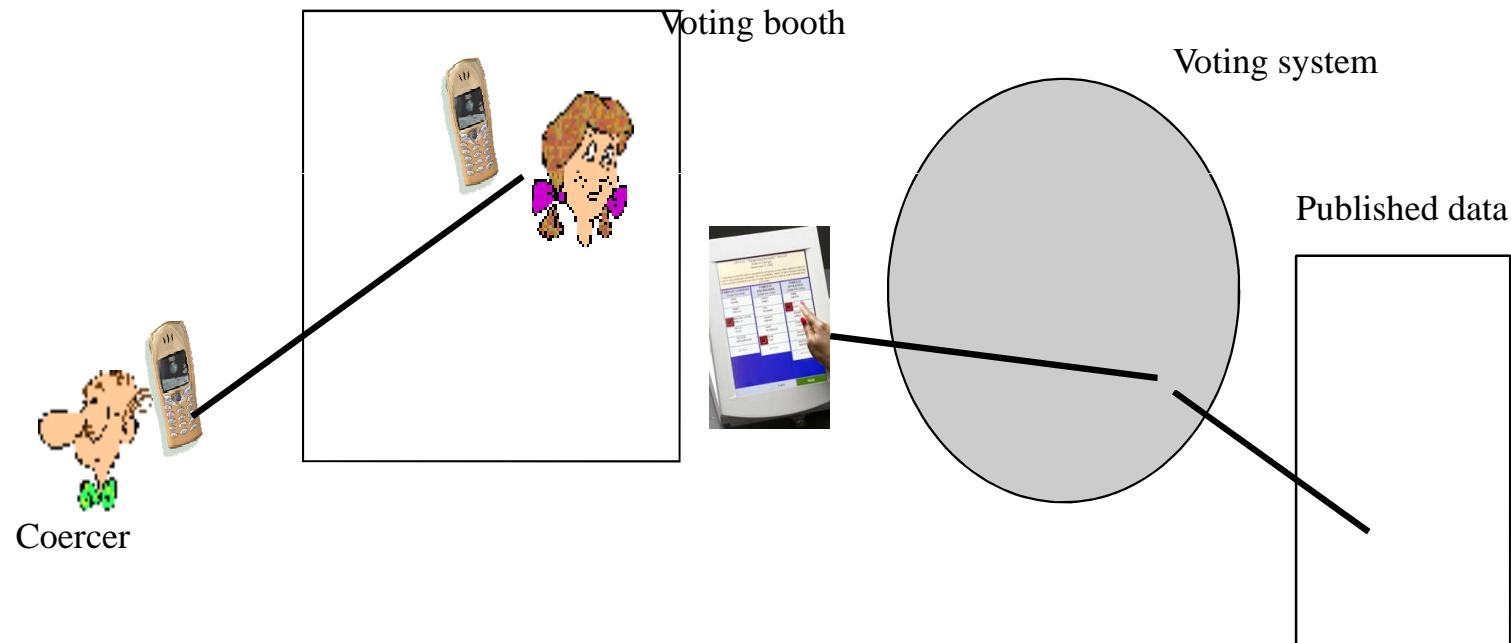
Properties 2 -Transparency

- **4. Individual verifiability:** a voter can verify that her vote was counted.
- **5. Universal verifiability:** a voter can verify that the published result is the tally of the votes cast.

Properties 3

- **6. Privacy**
 - no-one can find out how Alice voted.
- **7. Receipt-freeness**
 - Alice doesn't get a receipt (or any other by-product of the voting process); thus:
 - Alice cannot prove afterwards to a coercer how she voted
 - Thus, *receipt-freeness* is like *privacy*, but even with Alice's cooperation
- **8. Coercion-resistance**
 - Alice cannot prove how she voted, even if interaction with the coercer is allowed during the voting process
 - Even stronger than *receipt-freeness*.

Coercion resistance and the voting booth



Coercion-resistance

- Alice interacts with the coercer (e.g. by mobile phone) during the election.
- The coercer can participate in Alice's vote:
 - She can tell him messages she receives during the process (although he might not believe her)
 - He can instruct her on what messages to send back (although she might not obey).
- He might have independent means of verifying her reports and her actions



Some “non-functional” requirements

- **9. Robustness:** Voters cannot disrupt the election.
Faulty behaviour tolerated.
- **10. Vote-and-go:** Voters participate just once in an election.

Requirements of evoting

- 1. Completeness: All valid votes are counted correctly.
- 2. Soundness: The dishonest voter cannot disrupt the voting.
- 3. Privacy: All votes must be secret.
- 4. Un-reusability: No voter can vote twice.
- 5. Eligibility: No one who isn't allowed to vote can vote.
- 6. Fairness: Nothing must affect the voting.
- 7. Verifiability: No one can falsify the result of the voting.
- From Mark Hershberg Mit Phd thesis

additional properties

- 8. Receipt-Freeness: The voter does not need to keep any record of his vote.
- 9. Non-Duplication: No one can duplicate anyone else's vote.
- 10. Public Participation: Everyone knows who did, and did not, vote.
- 11. Private Error Correction: A voter can prove his vote was miscounted without revealing how he voted.
- (Properties 9, 10, and 11 are taken from Schneier 1996.)

Voting machines



Self-Adjudicating Protocols

- The most basic protocols require no external parties. Security is created by multiple layers of encryption and/or signing.
- Anonymity is generated by repeated reordering of the votes during various steps of the algorithm.
- All voters need to have public keys which are assumed to have been distributed before the election begins.

Basic evoting protocol

- The most basic protocols require no external parties. Security is created by multiple layers of encryption and/or signing. Anonymity is generated by repeated reordering of the votes during various steps of the algorithm. Michael Merritt designed one of the earliest schemes.

Basic evoting

- Each voter performs the following steps, as summarized by Schneier.
- 1. He attaches a random string, R, to his vote, V.
- 2. Then he encrypts his vote with public keys of Voters 1 through N, in that order.
- 3. Again, he repeats step two, but this time includes a random string within each layer of encryption.
- At this point the votes look like:
- $\text{EN}(\text{RN}, \text{EN-1}(\dots(\text{R}_2, \text{E}_1(\text{R}_1, \text{EN}(\text{EN-1}(\dots(\text{E}_1(V, R))\dots))))\dots))$
- where R_i is random string of voter i , and E_j is the encryption of the parenthesized expression using Voter j 's public key.

Step 4

4. All votes are passed from voter to voter, starting with voter N and ending with Voter 1. Each voter decrypts the message and strips off the random string, making certain it is the one he had used. The voter then scrambles the votes and sends them onto the next voter (with Voter 1 sending the votes on to Voter N).

- 5. Again each Voter from N down to 1 decrypts his layer, but then signs the message and sends it on. Voter i checks the validity of the signature of voter $i+1$ and if it is valid decrypts, signs, and passes the message onward.
- 6. All voters confirm the signature of Voter 1 and check the list of votes for their initial random string to insure their vote was counted.

No Stuffing

- The number of votes is constant throughout the process and so ballot stuffing or dropping is easily detected. Votes cannot be replaced by a malicious party. An attempt to do so in the second round of decryptions (step 5) will be discovered as the signed object will not be correct. The signatures at this stage make it easy to trace back and find the malicious party.

Central Vote Repository

- Excessive computation can be avoided by creating a Central Vote Repository (CVR).
- This system requires far less computational work. Again the voters are presumed to have a public/private key pair $\{k,d\}$.
- 1. The CVR asks each voter whether or not he will participate in the upcoming election.
- 2. A list of all participants is made public.
- 3. Each voter receives an ID number using an All-Or-Nothing-Disclosure-of- Secrets (ANDOS) protocol.

CVR 2

- 4. Each voter anonymously sends the CVR his ID number, I , along with the encryption of his vote, V , paired with his ID number.
- 5. The CVR publishes all encrypted votes $E_k(I, V)$.
- 6. After step 5 is complete, each voter anonymously sends $\{I, d\}$ to the CVR.
- 7. All votes are decrypted and their values published alongside them.

Multi server election MSE 1

- In the spirit of checks and balances, the next improvement to election schemes comes from using two centers, instead of one. Now, instead of a single CVR, there are two entities: a Validation Agency (VA) and a Tabulation Facility (TF).
- A valid vote must pass through both bodies to be counted. The first recognizes the voter's right to vote, without seeing the actual ballot, and gives the voter some token confirming this authorization.

MSE 2

- The second party is then anonymously passed the validation token and the vote. This type of scheme, of course, assumes that the two groups are set up so as not to collude with one another.
- 1. Each voter, after providing his identity, asks the VA for an authorization number.
- 2. The VA randomly generates authorization numbers and distributes them.
- 3. The list of all such authorization numbers is given to the TF.

Blind Signatures

- Our voting system's need for signatures has an additional constraint in that it must be a blind signature.
- Bob doesn't trust Alice.
- Bob trusts Trent, who trusts Alice.
- Bob is willing to accept a message from Alice only if Trent signs it.
- Alice doesn't want Trent to see it.
- She can use a blind signature.

Blind signature by Trent

- In the real world, Alice can seal her message in an envelope filled with carbon paper.
- Trent can then sign the outside of the envelope and his signature will get transferred to Alice's message, without Trent ever actually seeing it.
- Alice can then remove the message from the envelope, and give the signed message to Bob,
- Bob can verify Trent's signature.

Blind Commitment

- A commitment is a way in which one party can commit to an object (e.g. string of bits, message, contract) without anyone else seeing what that object really is.
- However any attempt by the first party to change the object can be detected.
- A good analogy would be for Alice to lock away a message in a safe requiring two keys, which she would then split between her and Bob. Alice cannot open the safe to

Disrupting elections with DDOS

The system lies dormant until the attacker decides that it is time to strike. At that point, the attacker sends a signal to the master, using a publicly available tool, indicating a target to attack.

The master conveys this information to all of the daemons, who simultaneously flood the target with more Internet traffic than it can handle.

The effect is that the target machine is completely disabled.

- From “Security Considerations for Remote Electronic Voting over the Internet” – by Avi Rubin

Zombies - Backorifice 2000

- Backorifice 2000 (BO2K) is packaged and distributed as a legitimate network administration toolkit. In fact, it is very useful as a tool for enhancing security. It is freely available, fully open source, extensible, and stealth (defined below).
- The package is available at <http://www.bo2k.com/>. BO2K contains a remote control server that when installed on a machine, enables a remote administrator (or attacker) to view and control every aspect of that machine, as though the person were actually sitting at the console.

Hacks

- If an attacker can add these two lines to the preferences file on somebody's machine, he can control every aspect of the web experience of that user.
- In c:\program_files\netscape\prefs.js
- user_pref("network.proxy.http",
"www.badguy.com");
- user_pref("network.proxy.http_port",
1799);

Hacks 2

- User's who open attachments and download software from the network are not the only ones putting their votes at risk. AOL, for instance, is in a position to control a large fraction of the total votes, because all of their users run AOL's proprietary software.
- There are dozens of software vendors whose products run on many peoples' home machines.
- For example, there are millions of personal computers running Microsoft office, Adobe Acrobat, RealPlayer, WinZip, Solitaire, and the list goes on.

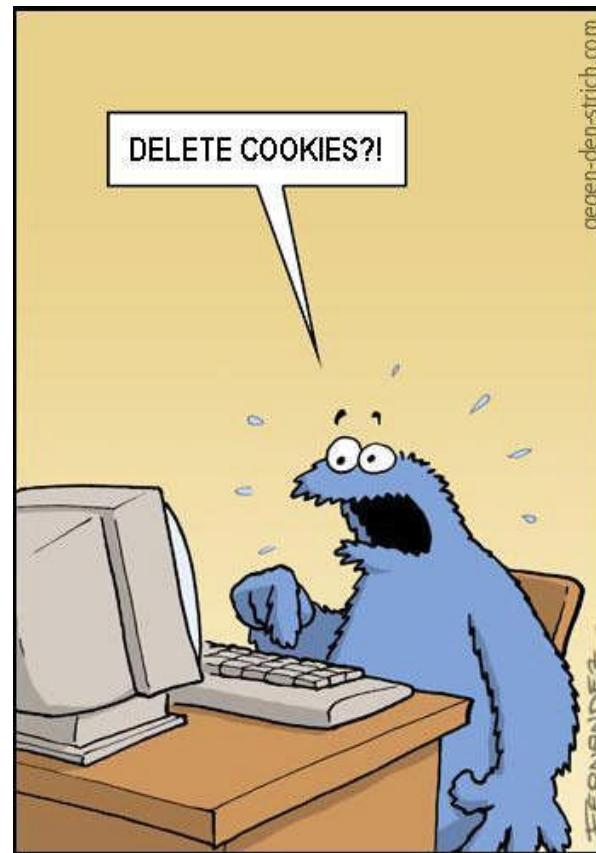
Hacks 3

- These vendors are in a position to modify any configuration file and install any malicious code on their customers' machines, as are the computer manufacturers and the computer vendors.
- Even if the company is not interested in subverting an election,
- all it takes is one rogue programmer who works for any of these companies.

Hacks 4

- Most of the software packages require an installation procedure where the system registry is modified, libraries are installed, and the computer must reboot. During any stage of that process, the installation program has complete control of all of the software on that machine.

Users and security



Social engineering

- *Social Engineering* is the term used to describe attacks that involve fooling people into compromising their security.
- Talking with election officials, one discovers that one of the issues that they grapple with is the inability of many people to follow simple directions.
- It is surprising to learn that, for example, when instructed to circle a candidate's name, people will often underline it.

User education

- While computers would seem to offer the opportunity to provide an interface that is tightly controlled and thus less subject to error, this is counter to the typical experience most users have with computers.
- For non-Computer Scientists, computers are often intimidating and unfamiliar.
- User interfaces are often poor and create confusion, rather than simplifying processes.

Who is SSL?

- In fact, most users would probably not distinguish between a page from an SSL connection to the legitimate server and a non-SSL page from a malicious server that had the exact same look as the real page.

Spoofing emails

- There are several ways that an attacker could spoof the legitimate voting site.
- One way would be to send an e-mail message to a user telling that user to click on a link, which would then bring up the fake voting site. The adversary could then collect the user's credentials and in a sense, steal the vote.

Spoofing

- An attacker could also set up a connection to the legitimate server and feed the user a fake web page, and act as a man in the middle, transferring information between the user and the web server, with all of the traffic under the attacker's control.
- This is probably enough to change a user's vote, regardless of how the application is implemented.

Attacking DNS

- Attack is possible by targeting the Internet's Domain Name Service (DNS). The DNS is used to maintain a mapping from IP addresses, which computers use to reference each other.

DNS poisoning

- The DNS is known to be vulnerable to attacks, such as cache poisoning, which change the information available to hosts about the IP addresses of computers.
- The reason that this is serious is that a DNS cache poisoning attack, along with many other known attacks against DNS, could be used to direct a user to the wrong web server when the user types in the name of the election server in the browser.
- Thus, a user could follow the instructions for voting, and yet receive a page that looked exactly like what it is supposed to look like, but actually is entirely controlled by the adversary.

Cyber cafe

- A malicious librarian or cyber café operator could set up public computers that appear to accept votes, but actually do nothing with the votes.

Smart cards

- tamper-resistant devices, such as smart cards. Cryptographic keys can be generated and stored on these devices, and they can perform computations, such that proper credentials can be exchanged between a client and a voting server.
- Expensive.
- Malicious code installed on the computer could misuse the smart card.

Voting from the cell phone

- cell phone with no general-purpose processor, equipped with a smart card, offer more promise of solving the technical security problems.
- Digital divide.

More security

- Trusted computing.
- RSA keys in the CPU.

Web searching

- Google
- Yahoo
- Bing
- Amazon

Crawl

- Get every page on the internet?
- Find list of pages to download
- Remove duplicate pages
 - www.amazon.com/watch/seiko
 - www.amazon.co.uk/watch/seiko
 - amazon.com/watch/seiko
 - amazon.com/seiko/watch
- Crawling private pages
- robots.txt
- Sitemaps.xml

How often to crawl?

- Amazon.com/products/cameras
- Cnn.com

Index

- **Indexing**
 - Parse each webpage
 - Remove stop words
 - Delete html templated.
 - Index words on page – Inverted index
- See <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>,
<http://www.dcs.gla.ac.uk/Keith/Preface.html>,
http://en.wikipedia.org/wiki/Inverted_index,

Text Processing

- Lexical analysis & tokenization
 - Split text into words, downcase letters, filter out punctuation marks, digits, hyphens
- Stopword elimination
 - Better retrieval accuracy, more compact index
 - Ex: “to be or not to be”

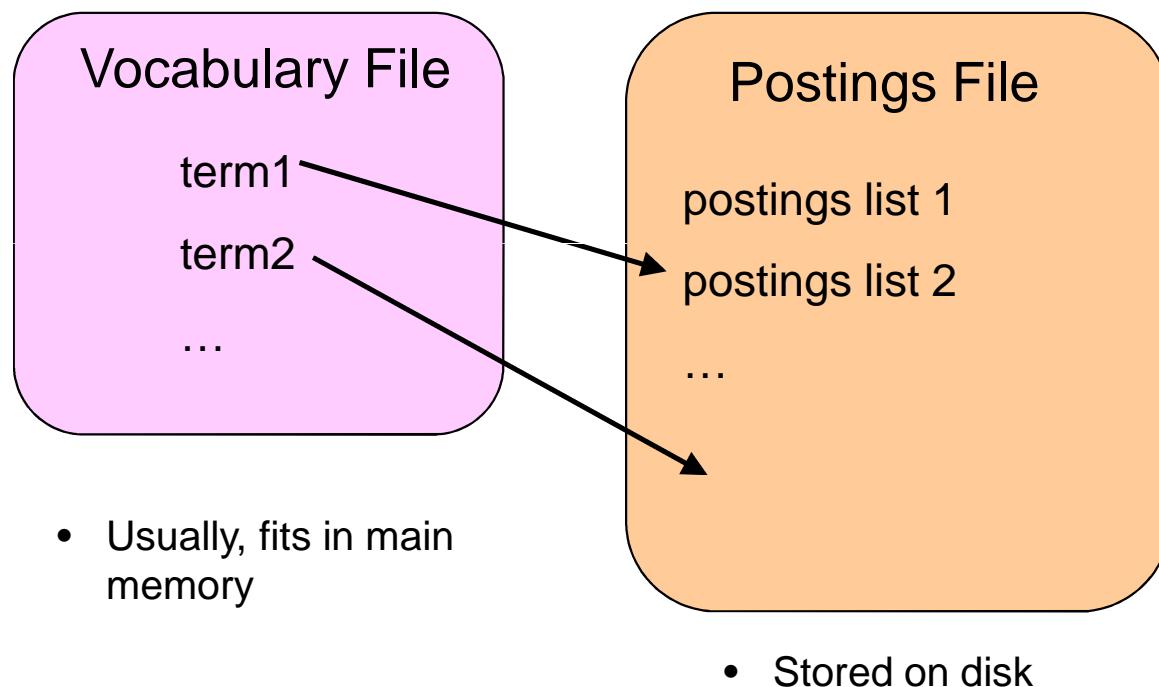
Text analysis

- Stemming
 - Ex: “computer”, “computing”,
 - “computation” → compute
- Bigram detection
 - New York city
- Identifying names, cities, businesses
 - George Washington was the president.
 - Drive on the George Washington bridge in NY
 - George’s Washington® machines.

Indexing

- Index terms selection
 - Keywords vs. full text
- SIP – statically improbable phrases.
 - E.g. Non linear analysis.
- Duplicate page (plagiarism) analysis.
 - RSS copies, wiki copies.
- Detect spam pages

Inverted Index Structure



Web server

- Scaling to millions of users.
- Scaling to many countries.
- Scaling to serve billions of web pages from one server?

Logging

- Webserver records what data was requested, e.g.
- apache/logs/access.log
- 192.68.12.21 - - [17/Feb/2011:15:49:06 +0000] "GET /manual/bind.html HTTP/1.1" 200 10404
- IPaddress - - [Date Time] "Cmd URL" code size
- Apache/logs/error.log
- [Thu Feb 24 18:44:43 2011] [error] [client 127.0.0.1] PHP Warning:
file_get_contents(lang.tmp) [<a href... failed to open stream: No such file or directory
in xampp\\index.php on line 2

Log analysis

- Collect all logs
- See what is popular where
 - Google trends
- Audit
- Accounting, billing
- Legal

Log analysis

- Privacy
 - Ipadress
 - Queries
 - Messages

Logging for Advertising

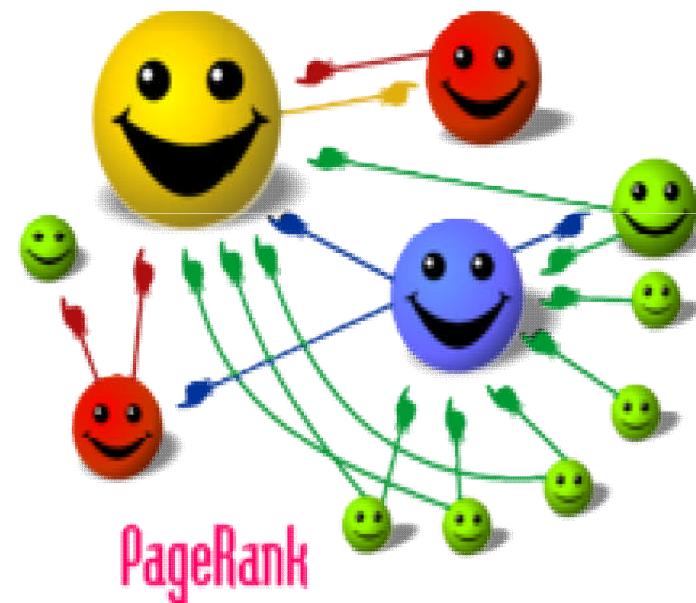
- [Query -> click] correlation
- [Ad lists -> clicks] analysis
- [clicks / impressions] ctr computation
- Billing advertisers for clicks
- Click fraud

Logging for fraud detection

- Short clicks
- Ip addresses
- Correlation with ipaddress, browser?
- Botnet and malware detection

Rank each page

- <http://en.wikipedia.org/wiki/PageRank>,



Search

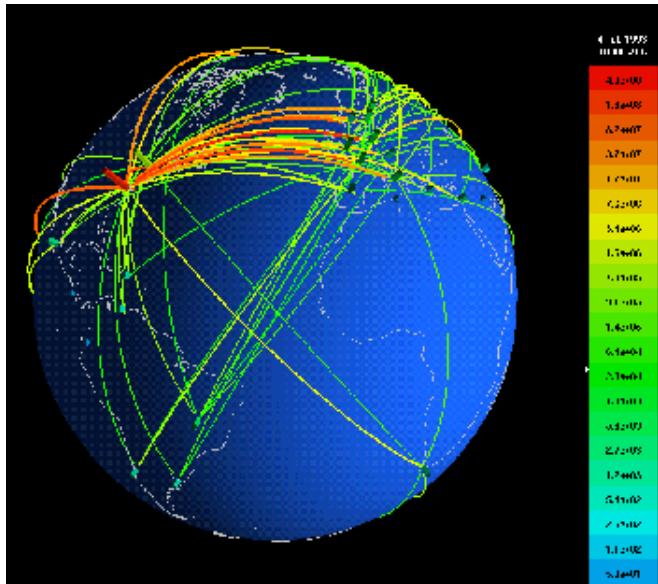
- Slides from
- <http://www.ee.technion.ac.il/courses/049011>

Examples of large data sets: Credit card transactions



- 47.5 billion transactions in 2005 worldwide
- 115 Terabytes of data transmitted to VisaNet data processing center in 2004

Examples of large data sets: Internet traffic



Traffic in a typical router:

- 42 kB/second
- 3.5 Gigabytes/day
- 1.3 Terabytes/year

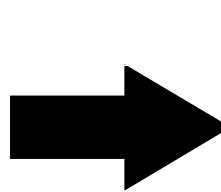
Examples of large data sets: The World-Wide Web



- 25 billion pages indexed
- 10kB/Page
- 250 Terabytes of indexed text data
- “Deep web” is supposedly 100 times as large

Reasons for the emergence of large data sets: Better technology

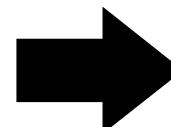
- Storage & disks
 - Cheaper
 - More volume
 - Physically smaller
 - More efficient



Large data sets are
affordable

Reasons for the emergence of large data sets: Better networking

- High speed Internet
- Cellular phones
- Wireless LAN

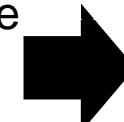


More data **consumers**
More data **producers**

Reasons for the emergence of large data sets:

Better IT

- More processes are automatic
 - E-commerce and V-commerce
 - Online and telephone banking
 - Online and telephone customer service
 - E-learning
 - Chats, news, blogs
 - Online journals
 - Digital libraries
- More enterprises are computerized
 - Companies
 - Banks
 - Governmental institutions
 - Universities



More data is
available in
digital form

World's yearly
production of data:
5 billion Gigabytes

Reasons for the emergence of large data sets:

Growing needs

- Science
 - Astronomy
 - Earth and environmental studies
 - Meteorology
 - Genetics
- Business
 - Billing
 - Mining customer data
- Intelligence
 - Emails
 - Web sites
 - Phone calls
- Search
 - Web pages
 - Images
 - Audio & Video



More **incentive** to construct
large data sets

Characteristics of large data sets

- Huge
- Distributed
 - Dispersed over many servers
- Dynamic
 - Items add/deleted/modified continuously
- Heterogeneous
 - Many agents access/update data
- Noisy
 - Inherent
 - Unintentional
 - Malicious
- Unstructured / semi-structured
 - No database schema

New challenges

Restricted access

- Large data sets are kept on magnetic and optical storage devices



- Access to data is **sequential**
- Random access is **costly**

New challenges

Stringent efficiency requirements

- Traditionally, “efficient” algorithms
 - Run in (small) polynomial time.
 - Use linear space.
- For large data sets, efficient algorithms
 - Must run in linear or even sub-linear time.
 - Must use up to poly-logarithmic space.

New challenges

Search the data

- Traditionally, input data is:
 - Either small and thus easily searchable
 - Moderately large, but organized in database tables.
- In large data sets, input data is:
 - Immense
 - Disorganized, unstructured, non-standardized

Hard to find what you want

New challenges

Mine the data

- Association rules
 - “Beers and diapers”
- Patterns
- Clusters
- Statistical data
- Graph structure

New challenges

Clean the data

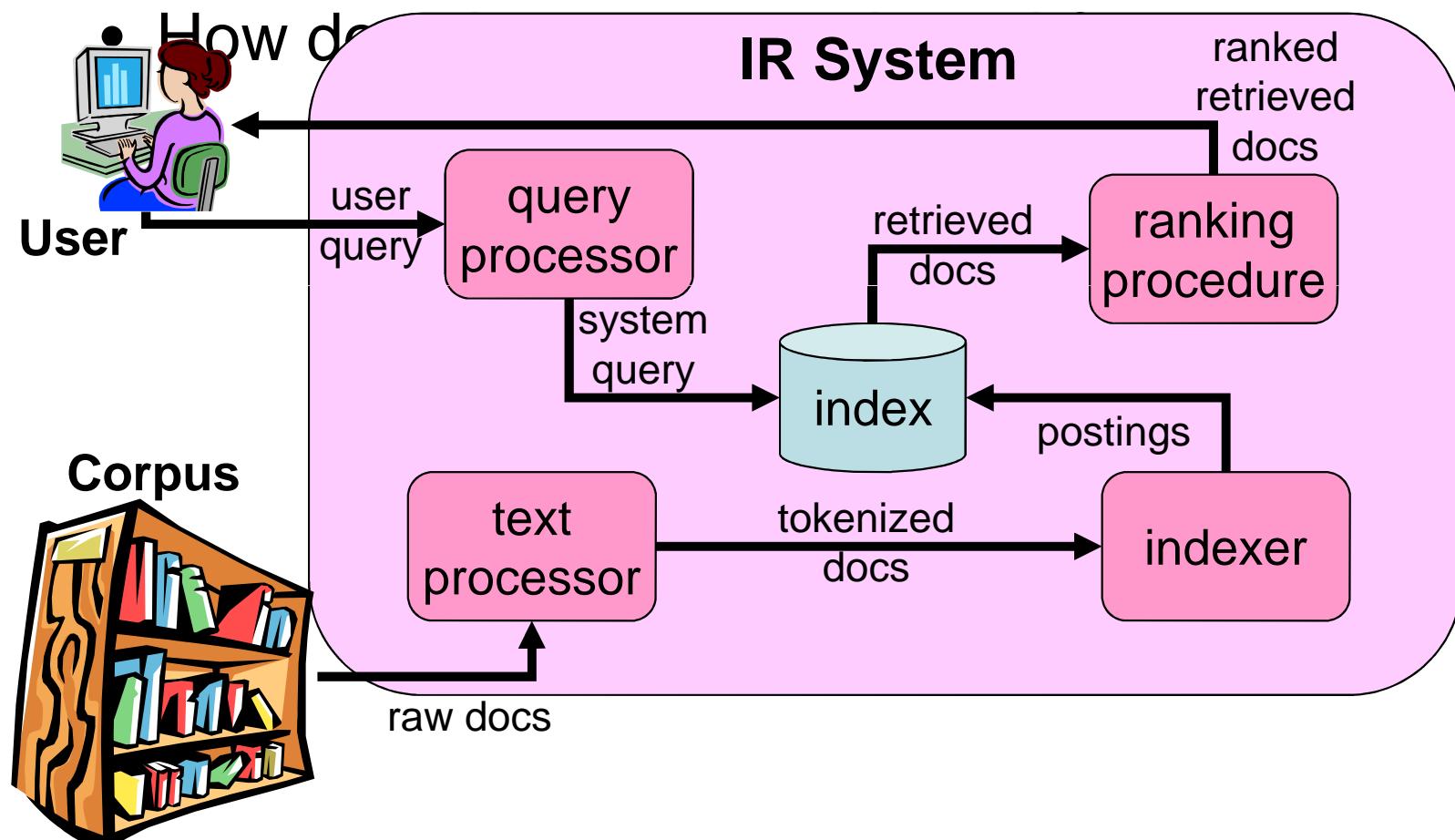
- Noise in data distorts
 - Computation results
 - Search results
 - Mining results
- Need automatic methods for “cleaning” the data
 - Spam filters
 - Duplicate elimination
 - Quality evaluation

Information Retrieval vs. Data Retrieval

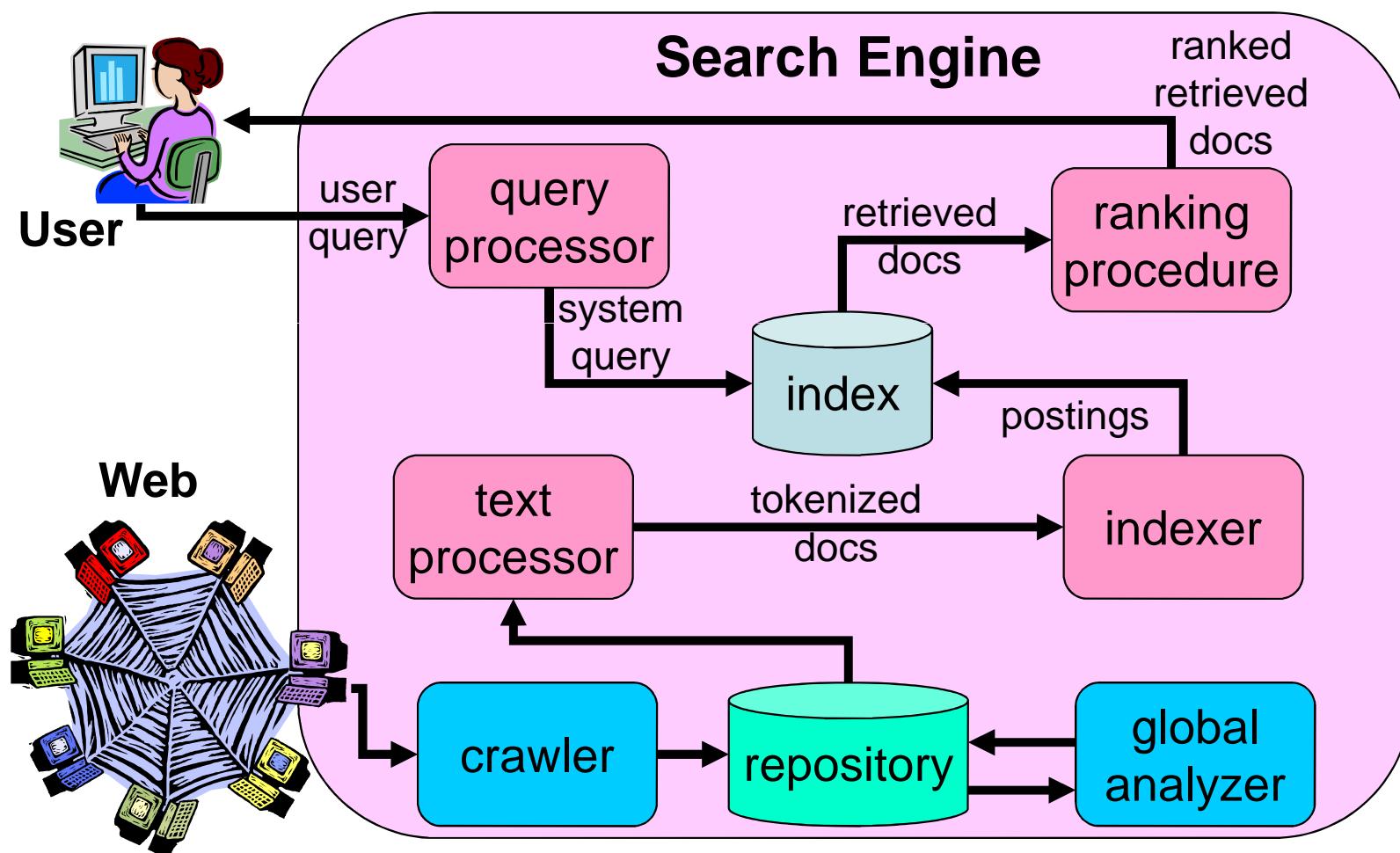
	Information Retrieval	Data Retrieval
Data	Free text, unstructured	Database tables, structured
Queries	Keywords, Natural language	SQL, Relational algebras
Results	Approximate matches	Exact matches
Results	Ordered by relevance	Unordered
Accessibility	Non-expert humans	Knowledgeable users or automatic processes

Search

from <http://webee.technion.ac.il/courses/049011/lectures/>



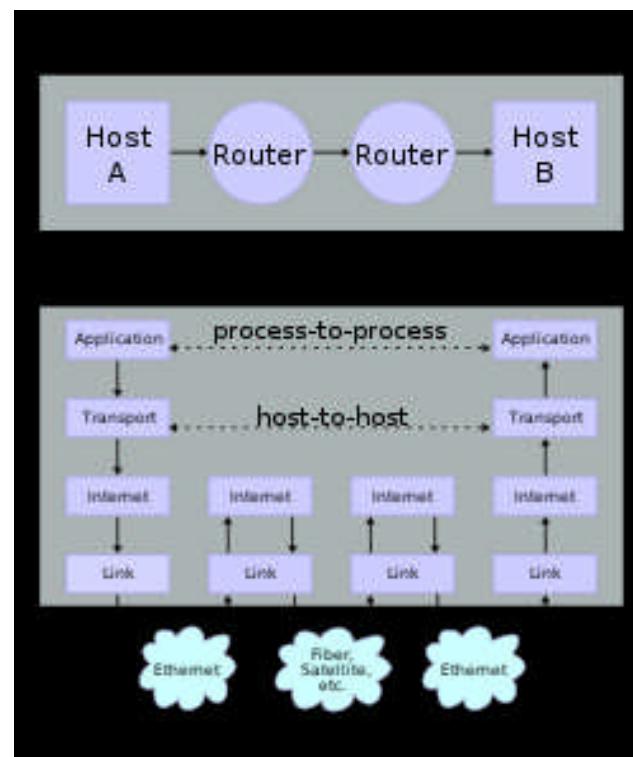
Search Engines



Securing the Internet for E-commerce

Notes from the web.

Internet



From: wikipedia internet protocol

IP. Internet is unreliable

- Data sent in IP packets
- IP Packets can get:
 - delayed,
 - lost,
 - routed wrongly,
 - duplicated,
 - corrupted,
 - stolen.

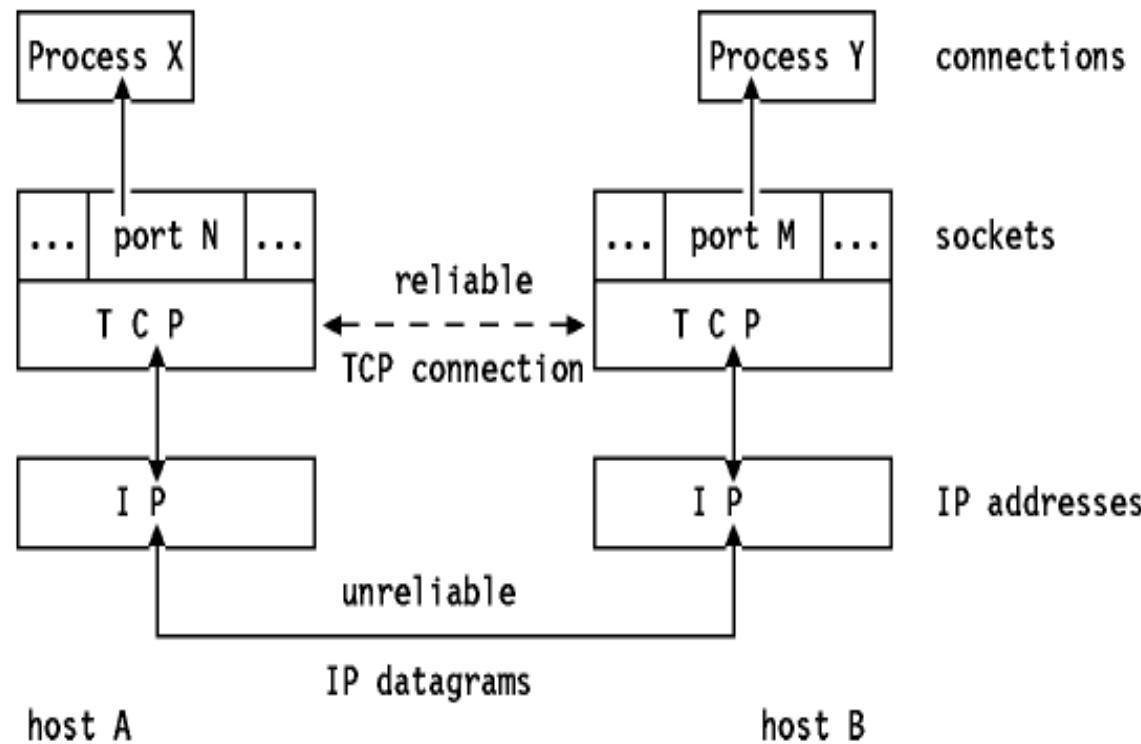
Layers of protocols

Internet protocol suite	
5. Application layer	DHCP • DNS • FTP • HTTP • IMAP4 • IRC • MIME • POP3 • SIP • SMTP • SNMP • SSH • TELNET • TLS/SSL • BGP • RPC • RTP • RTCP • SDP • SOAP • BitTorrent • ...
4. Transport layer	TCP • UDP • DCCP • SCTP • ...
3. Network layer	IP (IPv4 • IPv6) • ARP • IPSec • ICMP • IGMP • RSVP • IGP • RARP • ...
2. Data link layer	ATM • Bluetooth (PAN-Profile) • DTM • Ethernet • FDDI • Frame Relay • GPRS • Modems • PPP • Wi-Fi • ...
1. Physical layer	Bluetooth RF • Ethernet physical layer • ISDN • Modems • RS232 • SONET/SDH • USB • Wi-Fi • Power line communication • ...

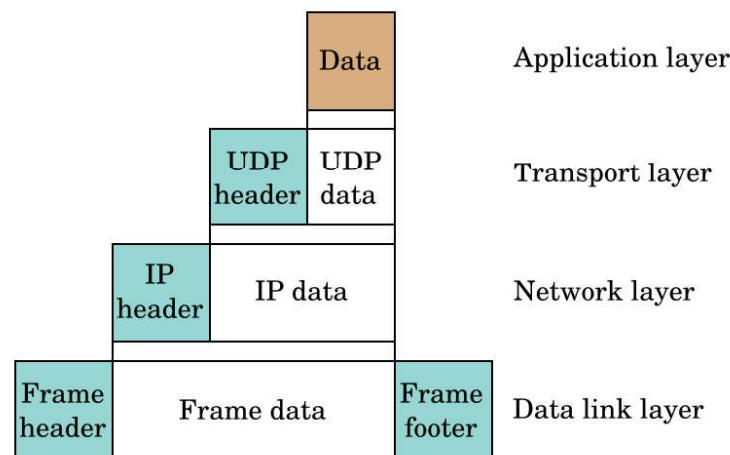
TCP over IP

- TCP adds reliability to IP
 - Small packets appear as a continuous stream
 - Reliability (takes care of packet loss, sequencing)
 - Flow control (congestion, window sizing)
 - Multiplexing (multiple data streams)
 - Duplex connection

Layers: Process -> tcp -> ip

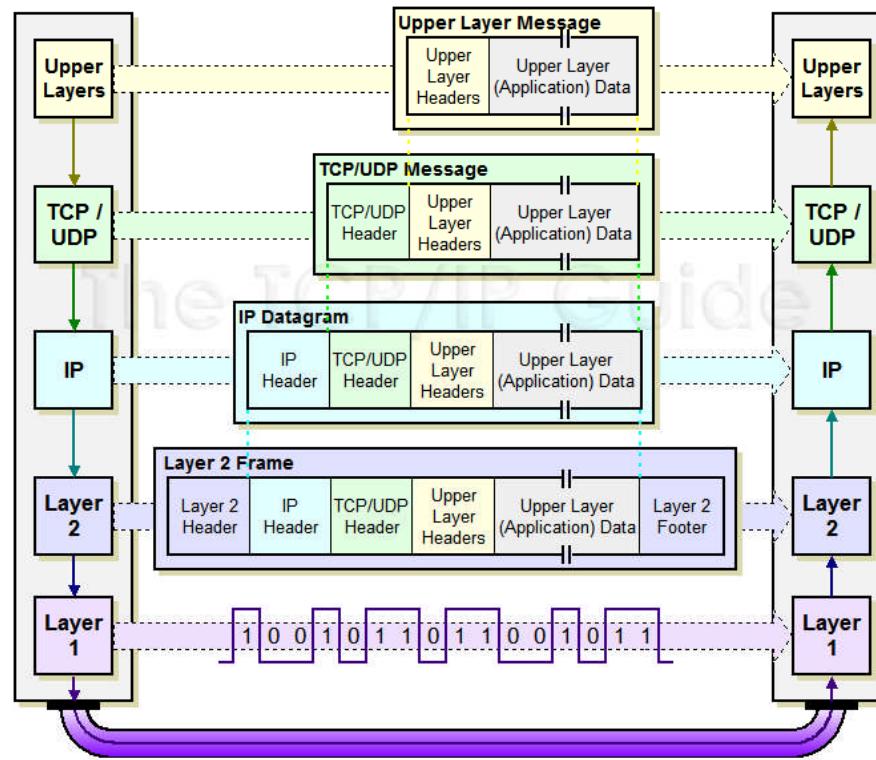


Packet encapsulation

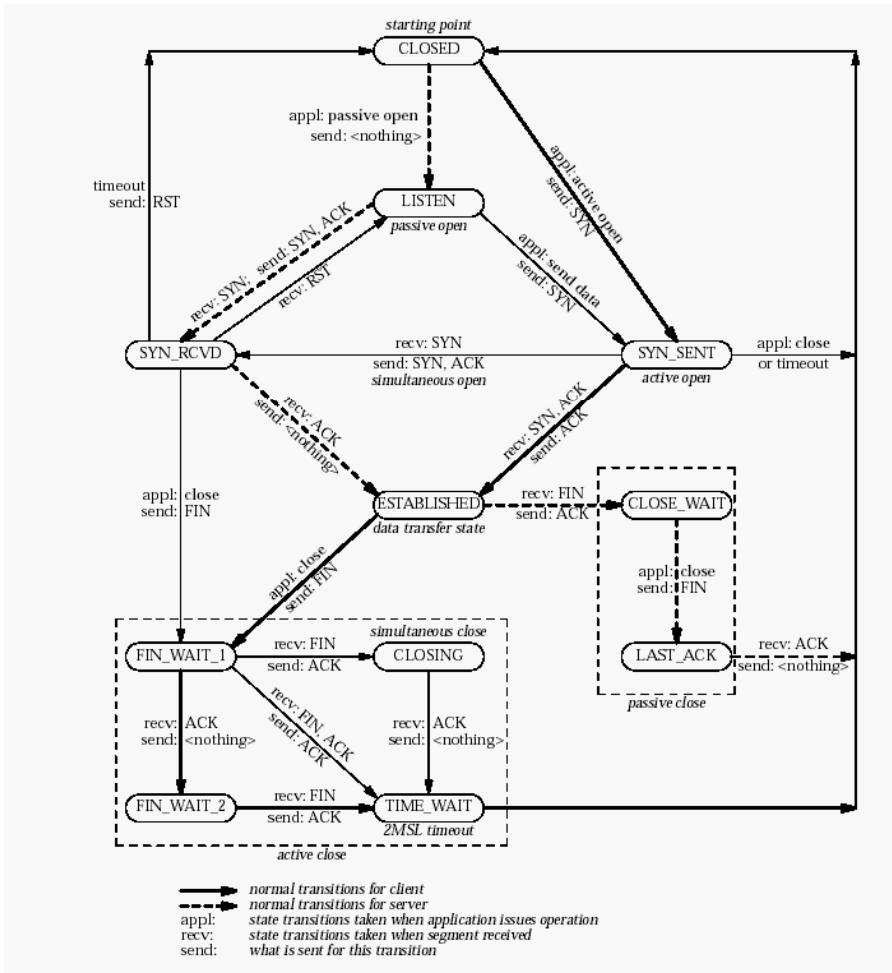


Layers of encapsulation

(from tcip guide)



TCP/IP protocol state machine



Sniffers

- Wireshark for windows
- Tcpdump on linux and unix.
- Network card modes:
 - Normal - only frames destined for the NIC's MAC address, and broadcasts, are passed up.
 - Promiscuous mode - all packets are passed up, if it's not destined for your MAC address, some wifi card don't support it
 - Monitor mode - for raw viewing of 802.11 frames

Use of network sniffers

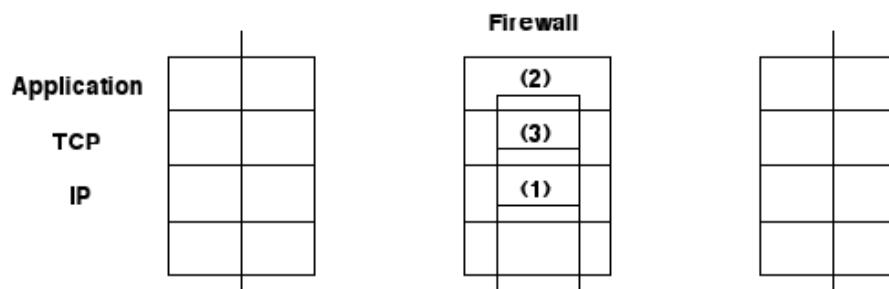
- Find network problems
- Find insecure protocols to deprecate:
 - Telnet, FTP, HTTP
- Find bad devices, like hacked routers
- Look for bad packets

Security Issues

- Arp poisoning
- DNS poisoning
- Session hijacking
- MITM (Man In The Middle) attack
- Phishing
- Malware

Firewall

1. Packet filter
 - o Just filters at the level of TCP/IP
 - o least "intelligent"
2. Application-level filter
 - o Filters at the level of applications
 - o most "intelligent"
3. Circuit-level gateway
 - o between 1 and 2



Limitations of firewalls

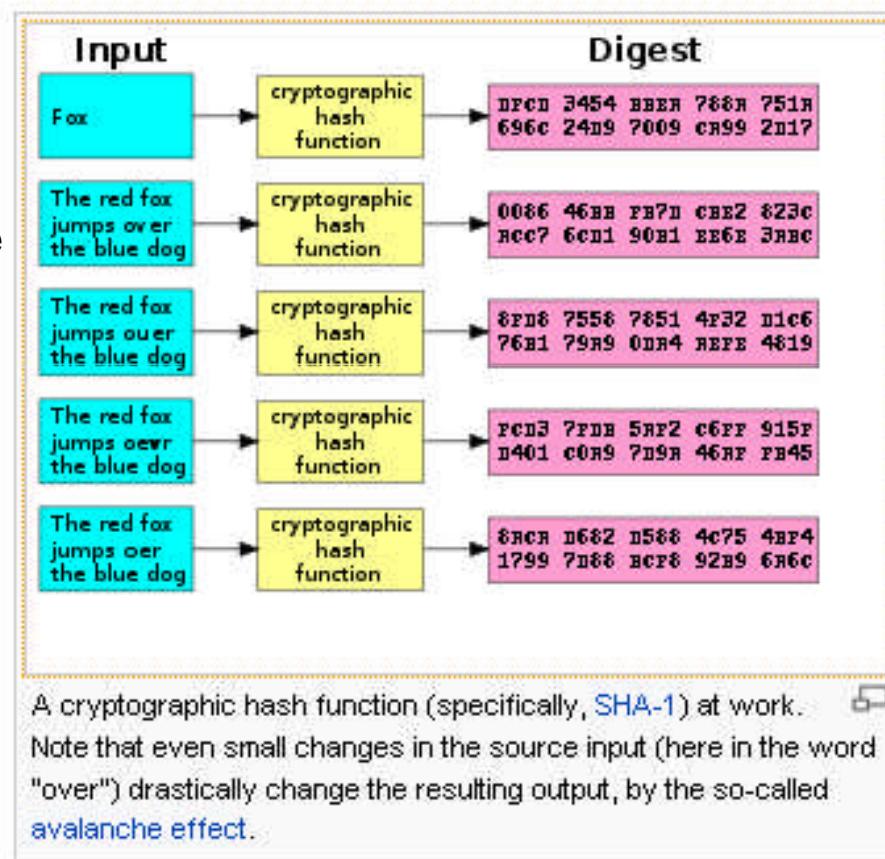
- The firewall cannot protect against attacks that bypass the firewall, like modems.
- People carrying CD, USB RAM, IPODs.
- Encrypted traffic, cannot be filtered.
- No protection against internal threats, virus-infected programs, impossible to scan all the incoming data.

Protecting data with Encryption

- We need to lock the data
- Make sure it is not tampered
- Make sure it available only to the recipients.

Cryptographic hash functions

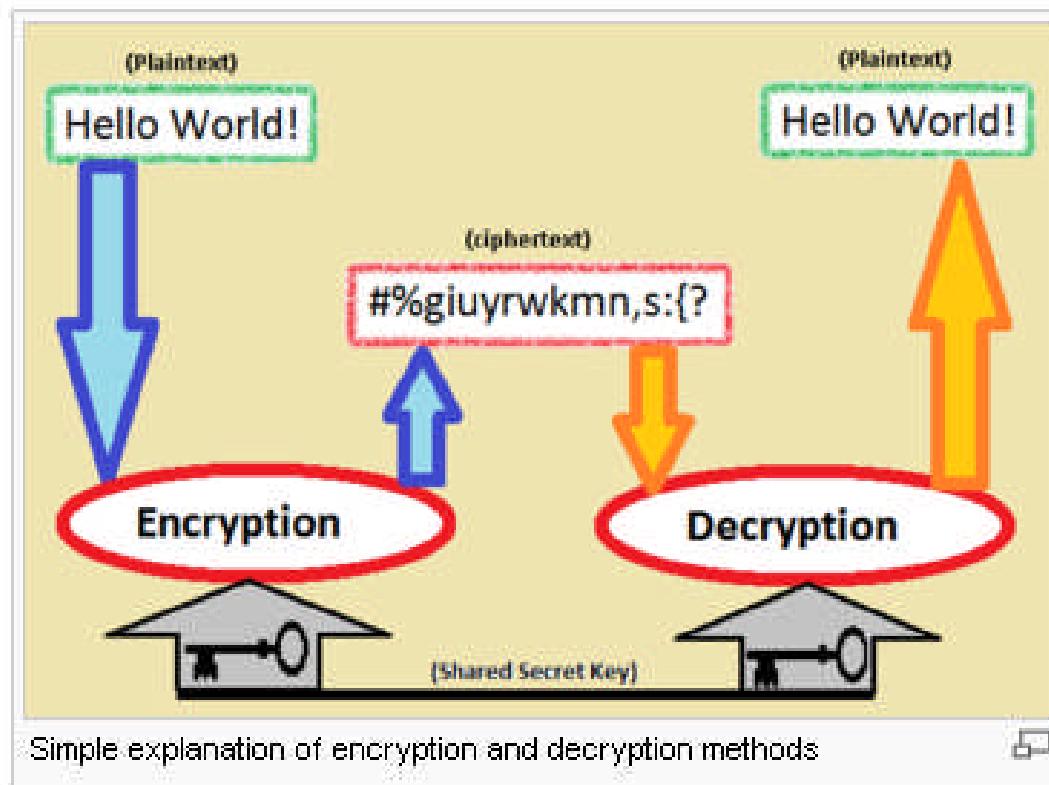
- One way functions - given output, very hard to compute input.
- Hard to find another input that gives same output.
- Hard to find 2 different inputs with same output.



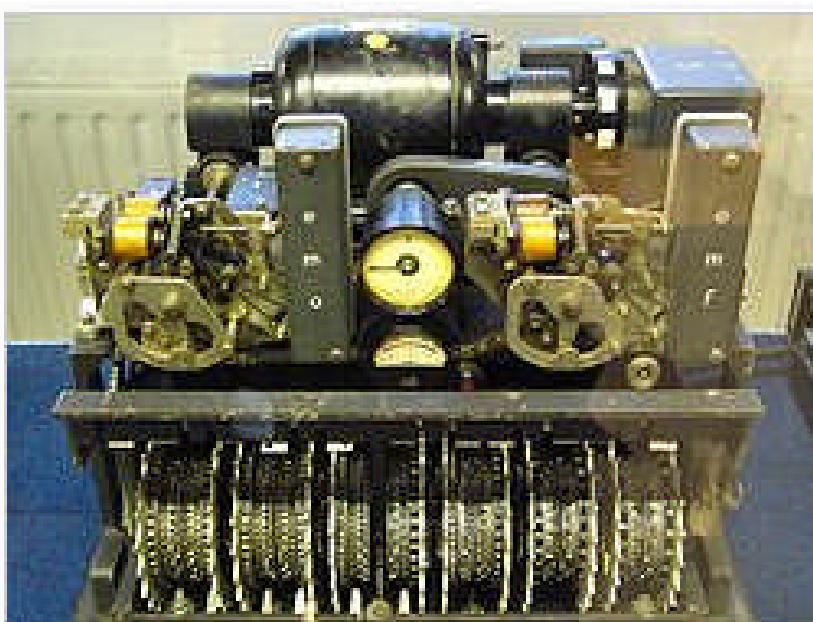
Examples of hashing functions

- MD5 128 bits
- SHA 160
- SHA2 256,512
- Used to generate signatures.

Encryption



Encryption before computers



German Lorenz cipher machine, used in World War II to encrypt very-high-level general staff messages

Symmetric key encryption

- Symmetric key encryption uses one key, called secret key - for both encryption and decryption. Users exchanging data must keep this key secret. Message encrypted with a secret key can be decrypted only with the **same** secret key.
- Examples: [DES](#) - 64 bits, [3DES](#) - 192 bits, [AES](#) - 256 bits, IDEA - 128 bits, Blowfish, Serpent

Symmetric vs Public key encryption



A) Secret key (symmetric) cryptography. SKC uses a single key for both encryption and decryption.



B) Public key (asymmetric) cryptography. PKC uses two keys, one for encryption and the other for decryption.

Public key Cryptography

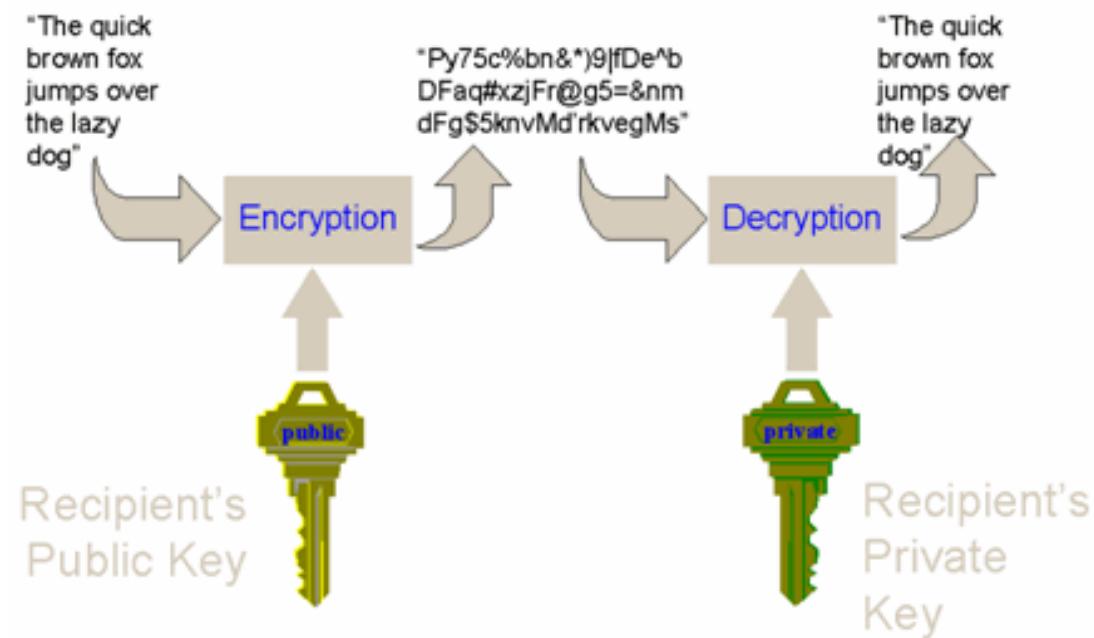
http://en.wikipedia.org/wiki/Public-key_cryptography

- The two main branches of public key cryptography are:
- Public key encryption: a message encrypted with a recipient's public key cannot be decrypted by anyone except a possessor of the matching private key—presumably, this will be the owner of that key and the person associated with the public key used. This is used for [confidentiality](#).
- [Digital signatures](#): a message signed with a sender's private key can be verified by anyone who has access to the sender's public key, thereby proving that the sender had access to the private key (and therefore is likely to be the person associated with the public key used), and the part of the message that has not been tampered with. On the question of [authenticity](#), see also [message digest](#).

Properties of Encryption

- No leakage of information in transit.
- Same message looks different if sent twice.
- Size of message does not indicate anything.
- Time stamped to avoid relay attacks.
- Cannot be cracked by brute force trials.

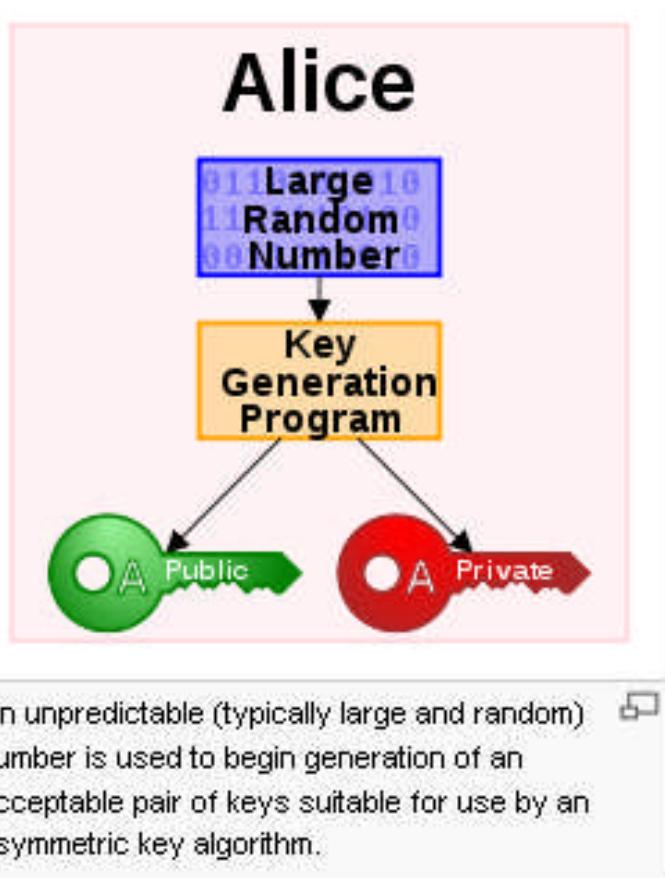
Public key cryptography



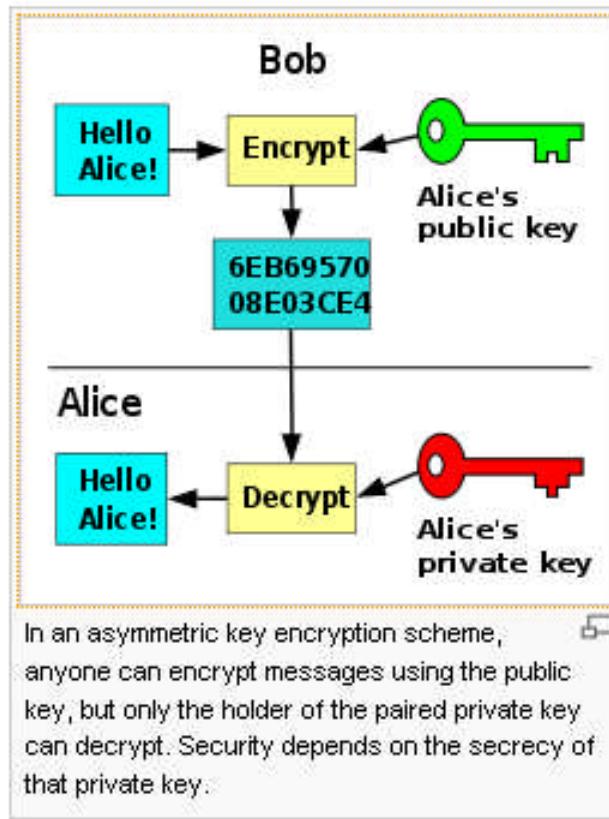
Mailbox analogy

- An analogy to public-key encryption is that of a locked [mailbox](#) with a mail slot. The mail slot is exposed and accessible to the public; its location (the street address) is in essence the public key. Anyone knowing the street address can go to the door and drop a written message through the slot; however, only the person who possesses the key can open the mailbox and read the message.
- An analogy for digital signatures is the sealing of an envelope with a personal [wax seal](#). The message can be opened by anyone, but the presence of the seal authenticates the sender.
- A central problem for use of public-key cryptography is confidence (ideally proof) that a public key is correct, belongs to the person or entity claimed (i.e., is 'authentic'), and has not been tampered with or replaced by a malicious third party. The usual approach to this problem is to use a [public-key infrastructure](#) (PKI), in which one or more third parties, known as [certificate authorities](#), certify ownership of key pairs. Another approach, used by [PGP](#), is the "[web of trust](#)" method to ensure authenticity of key pairs.

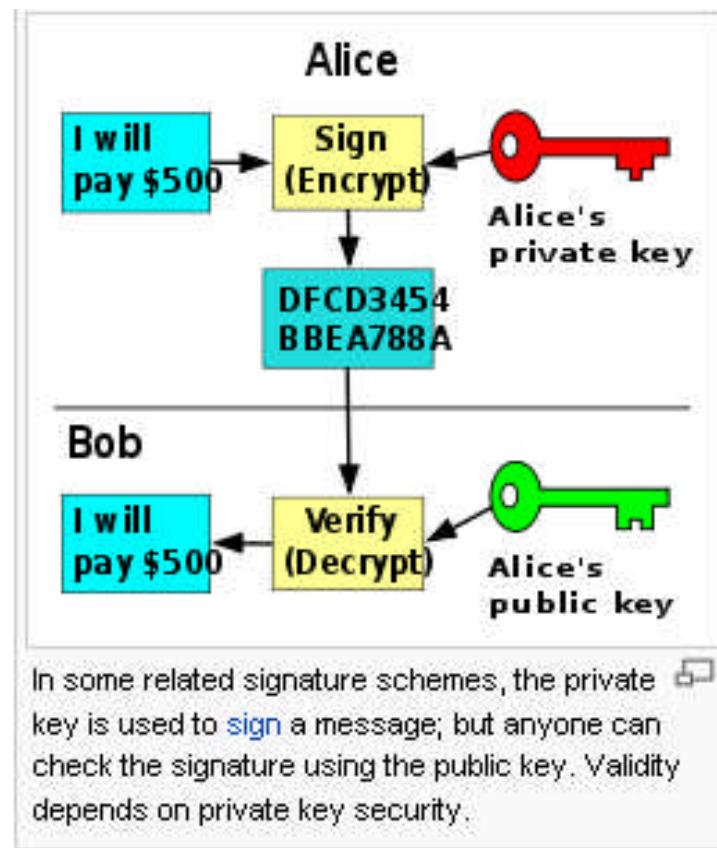
Alice makes a public key lock



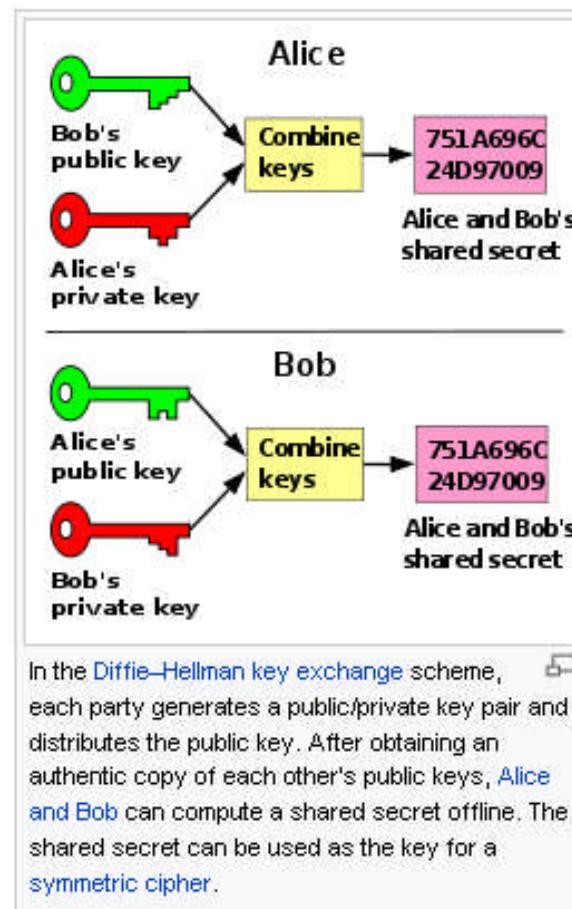
Any can lock, only Alice can unlock



Signing: Only Alice can lock anyone can unlock



Diffie Hellman key exchange, each user have their own locks



RSA 1 (Rivest Shamir Adelman)

RSA is usually used to encrypt a private **key** and then send that with along with a message encrypted *by* the private **key**. It uses a variable **key** length, and a variable block size that is not greater than the **key** length.

- Choose two large primes, p and q, kept secret.
- Compute $n = p * q$, assume factoring n is hard.
- Compute $\phi(n) = (p - 1)(q - 1)$

Modular math

- $5 \bmod 3 ?$
- $3 * 3 \bmod 5 =$
- $3 * 3 * 3 \bmod 5 =$
- $(3 * 3 \bmod 5) * 3 \bmod 5 =$
- $(3^4) \bmod 5 =$
 $= ((3^2) \bmod 5)^2 \bmod 5)$
 $= (9 \bmod 5)^2 \bmod 5$

Factor demo

- Install cygwin on windows,
- \$ factor 1111111111111111
- 1111111111111111: 3 31 37 41 271 2906161
- factor 928440727
928440727: 928440727
- factor 928440729
 - 928440729: 3 3 337 443 691

RSA 2

- Pick prime e , such that $\gcd(e, \phi(n)) = 1$.
- Determine $d = e^{-1} \pmod{\phi(n)}$
- Publish *public-key* (e, n)
- Hide *private-key* (d, n)
- Encrypt plaintext= m using public-key to:
Ciphertext = $c = m^e \pmod{n}$.
- Decrypt c using private key: $m = c^d \pmod{n}$.

RSA Example

- Choose two large prime, $p = 61$ and $q = 53$.
- $n = p * q = 61 * 53 = 3233$
- $\phi(p * q) = (61 - 1)*(53 - 1) = 3120.$
- Let $e = 17$, $\gcd(e, n)=1$.
- Compute $d=2753$ ($17*d=1 \bmod \phi(3233)$)
Hard without knowing factors of n .
- **public key** is $(n = 3233, e = 17)$,
 - encryption function is $m^{17}(\bmod 3233)$.
- **private key** is $(n = 3233, d = 2753)$
 - decryption function is $c^{2753}(\bmod 3233)$.

RSA example continued

- Encrypt 'A' = 65, using (e=17,n=3233)
 - $c = 65^{17} \pmod{3233} = 2790$.
- Decrypt 2790, using (d=2753,n=3233)
 - $m = 2790^{2753} \pmod{3233} = 65$, to get 'A'.

Primes and factoring

- > OpenSSL> prime 17 (is 17 a prime?)
 - 11 is prime (Hex 11 = 16 + 1 = 17 Dec).
- > gpg --gen-prime 1 30 (30 bit prime).
 - 3756E197 (which is 928440727 DEC).

Primes and factoring

- > OpenSSL> prime 17 (is 17 a prime?)
 - 11 is prime (Hex 11 = 16 + 1 = 17 Dec).
- > gpg --gen-prime 1 30 (30 bit prime).
 - 3756E197 (which is 928440727 DEC).
- factor 928440727

928440727: 928440727
- factor 928440729
 - 928440729: 3 3 337 443 691

Openssl to generate public key

Generate a new public/private keypair:

- \$ openssl genrsa -out key.pem
- Generating RSA private key, 512 bit long modulus
- ..+++++.....++++++
- e is 65537 (0x10001)

Extract the modulus,e,primes from your key:

- \$ openssl rsa -in key.pem -noout -text
- publicExponent: 65537 (0x10001)
- Modulus=....long-string-of-hex-digits...

Openssl to generate public key

Generate a new public/private keypair:

- \$ openssl genrsa -out key.pem
- Generating RSA private key, 512 bit long modulus
- ..+++++.....++++++
- e is 65537 (0x10001)

Extract the modulus,e,primes from your key:

- \$ openssl rsa -in key.pem -noout -text
- publicExponent: 65537 (0x10001)
- Modulus=....long-string-of-hex-digits...

Which layer to secure

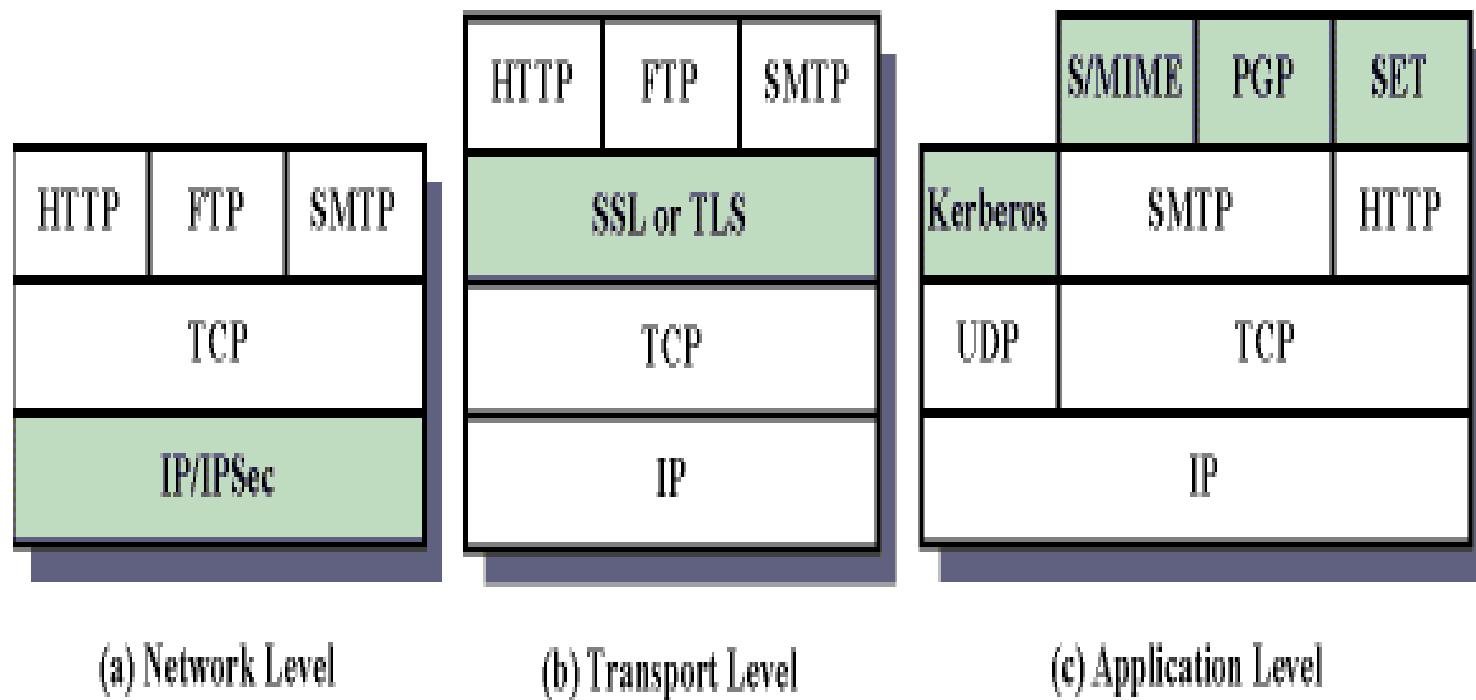


Figure 14.1 Relative Location of Security Facilities in the TCP/IP Protocol Stack

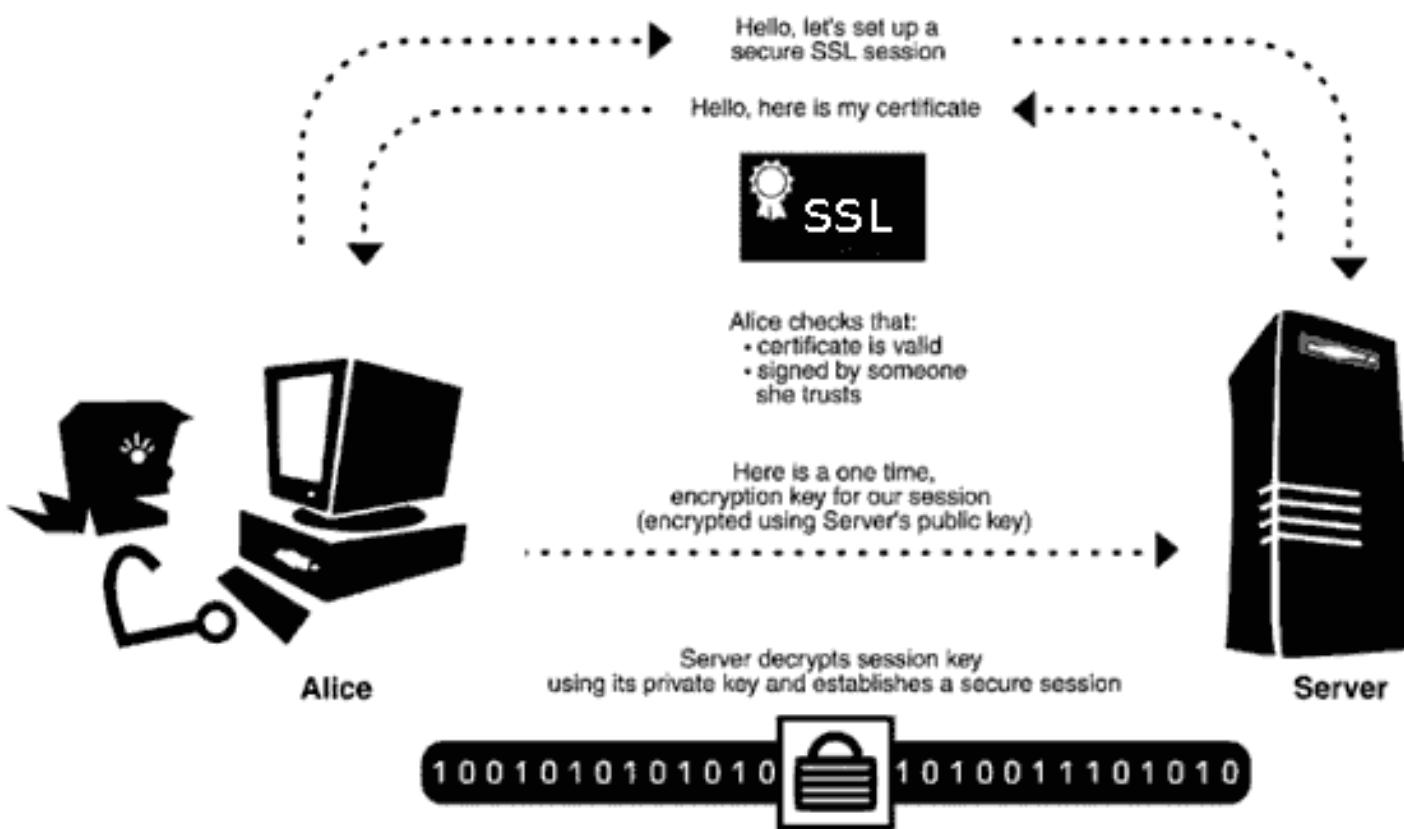
SSL and TLS

- SSL stands for Secure Sockets Layer and TLS stands for Transport Layer Security.

Netscape released SSL version 2 in 1994 and SSL version 3 in 1995. TLS is an IETF standard designed to standardize SSL as an Internet protocol. It is just a modification of SSL 3.

- The TLS acronym is the result of arguments between Microsoft and Netscape over the naming of the protocol because each company proposed its own name. These days, SSL and TLS seem to be used interchangeably.

SSL diagram



HTTPS security guarantees

from eff.org

- **Server authentication** allows the browser and the user to have some confidence that they are talking to the true application server. Without this guarantee, there can be no guarantee of confidentiality or integrity.
- **Data confidentiality** means that eavesdroppers cannot understand the communications between the user's browser and the web server, because the data is encrypted.
- **Data integrity** means that a network attacker cannot damage or alter the content of the communications between the user's browser and the web server, because they are validated with a cryptographic [message authentication code](#).

Using SSL

- You specify that you want to connect to a server using SSL by replacing http with https in the protocol component of a URI. The default port for HTTP over SSL is 443.

SSL protocol

- **Phase 1. Establish Security Capabilities**
Phase 2. Server authentication and key exchange
Phase 3. Client authentication and key exchange
Phase 4. Finish

SSL Protocol Summary

- The process to establish an SSL connection is the following: The user uses his browser to connect to the remote server.
- The handshake phase starts, and the browser and server exchange keys and certificate information.
- The browser checks the validity of the server certificate, including that it has not expired, that it has been issued by a trusted CA, and so on.
- Optionally, the server can require the client to present a valid certificate as well.
- Server and client use each other's public key to securely agree on a symmetric key.
- The handshake phase concludes and transmission continues using symmetric cryptography.

SSL data into records

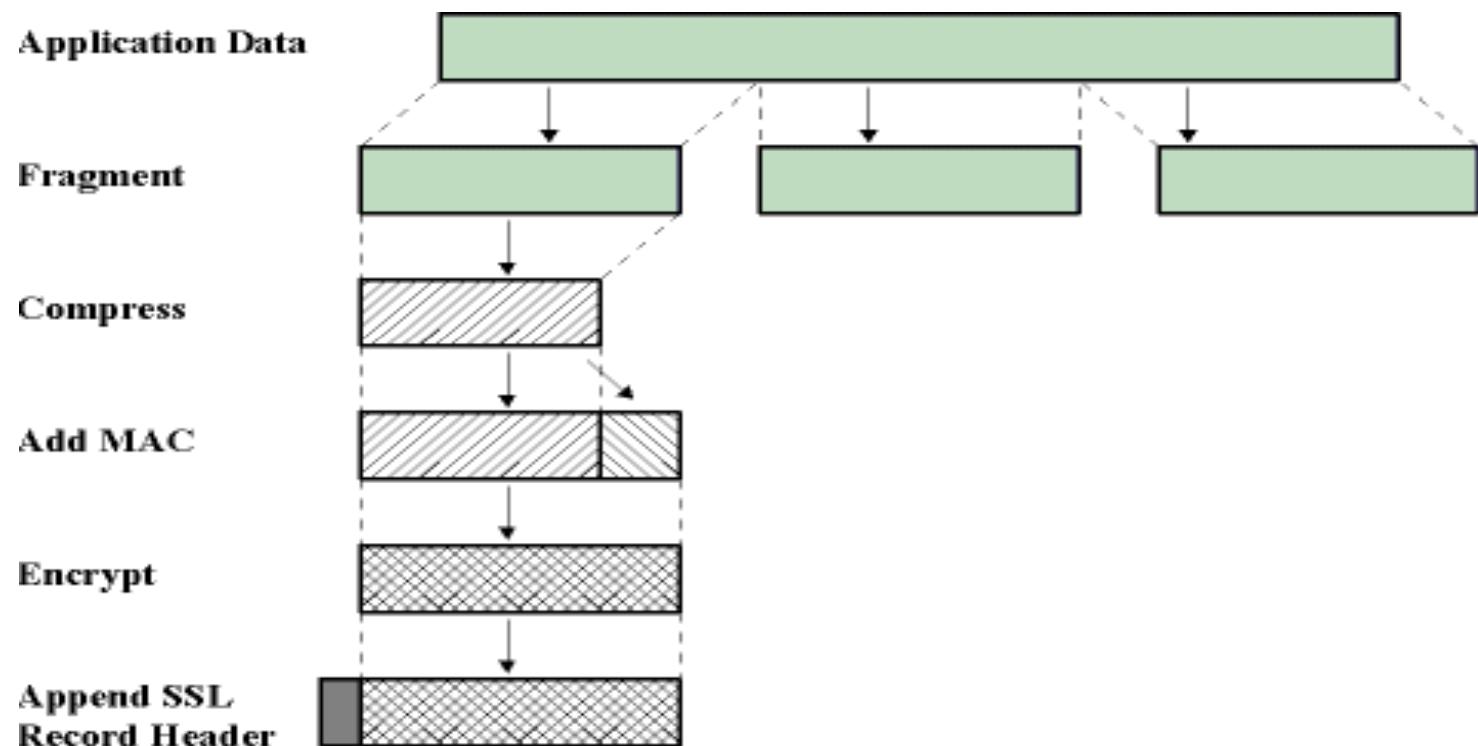


Figure 14.3 SSL Record Protocol Operation

SSL Key exchange

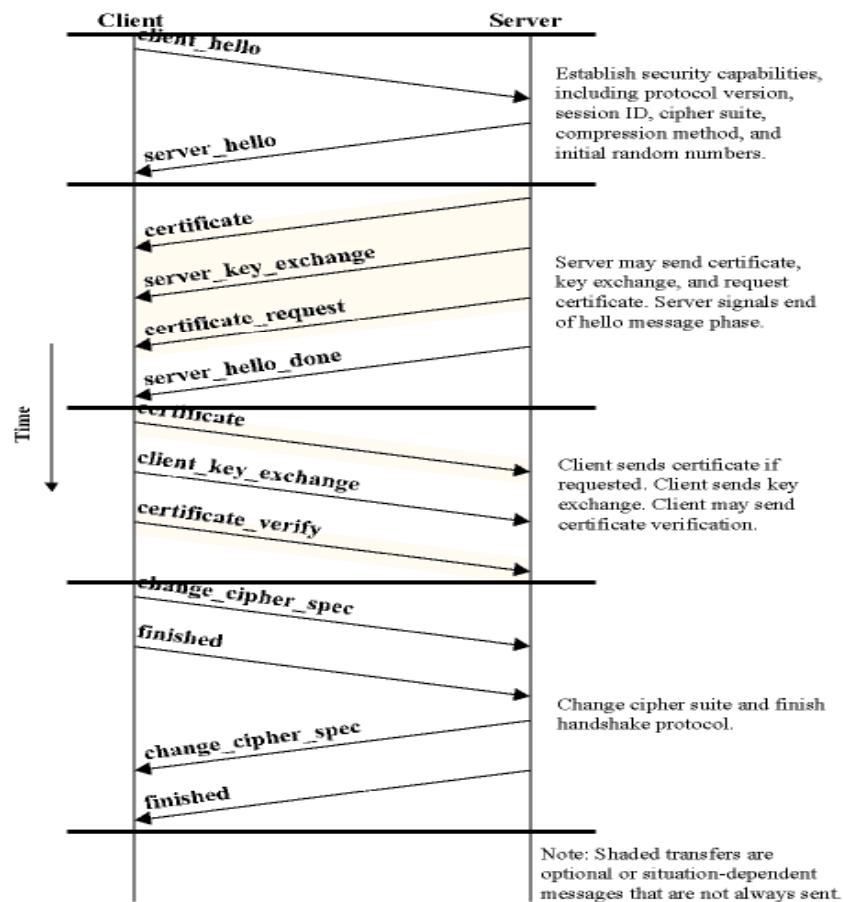
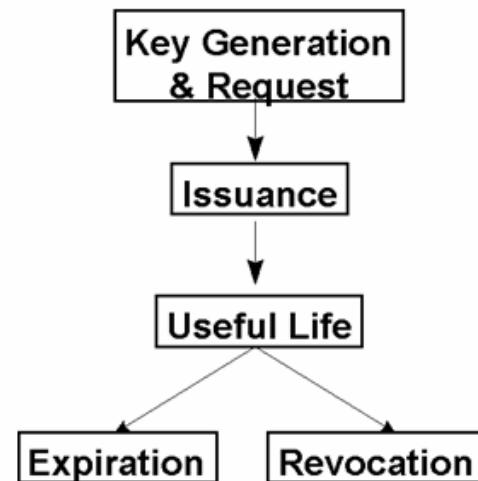


Figure 14.6 Handshake Protocol Action

Certificate



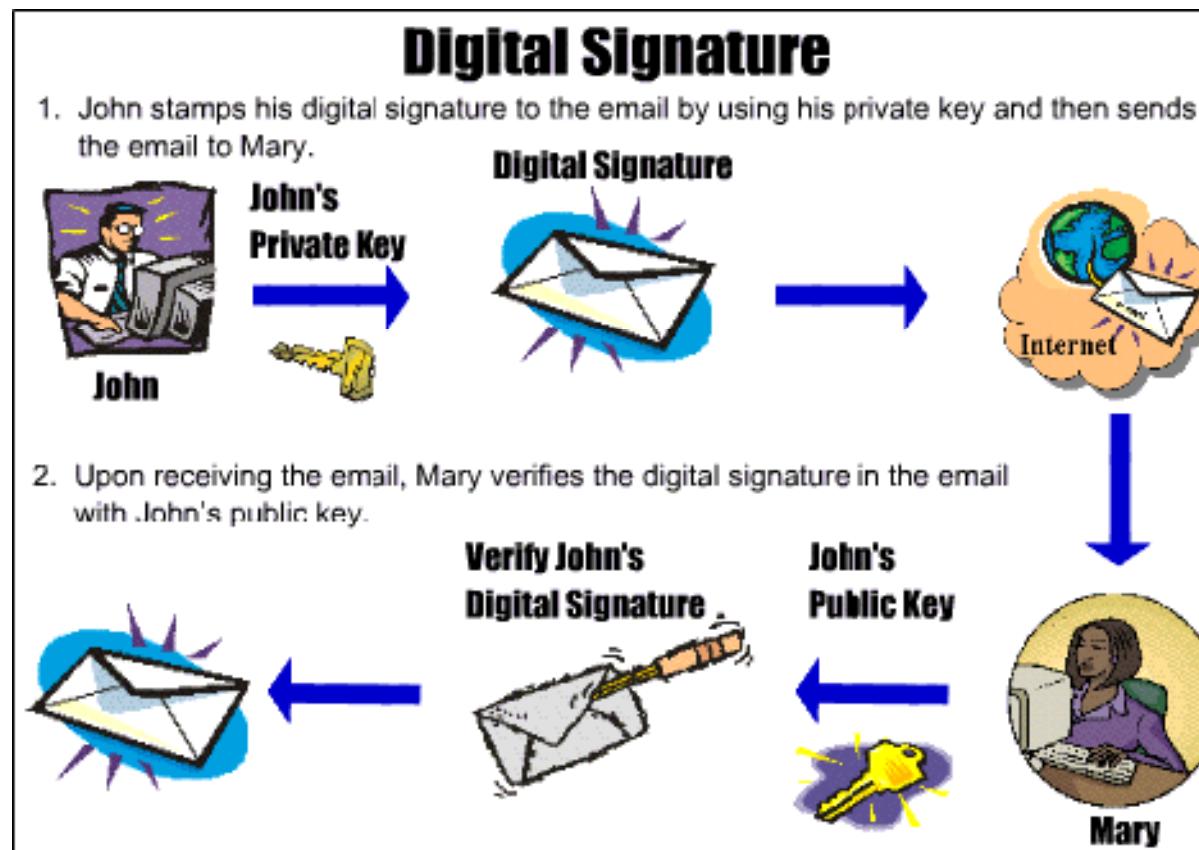
Life of a public key



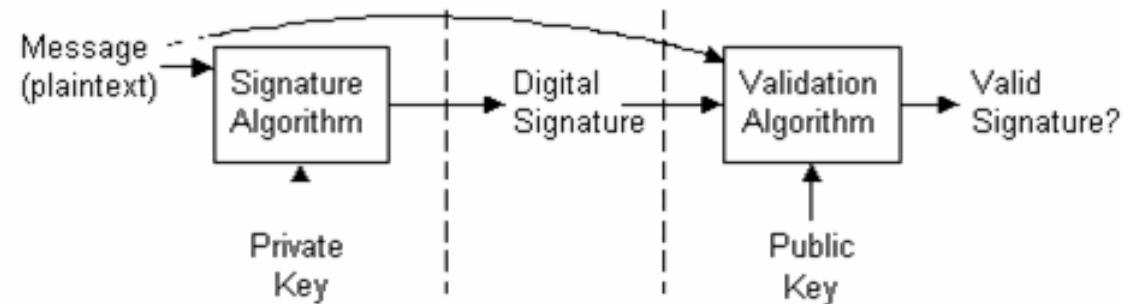
CA certification authority

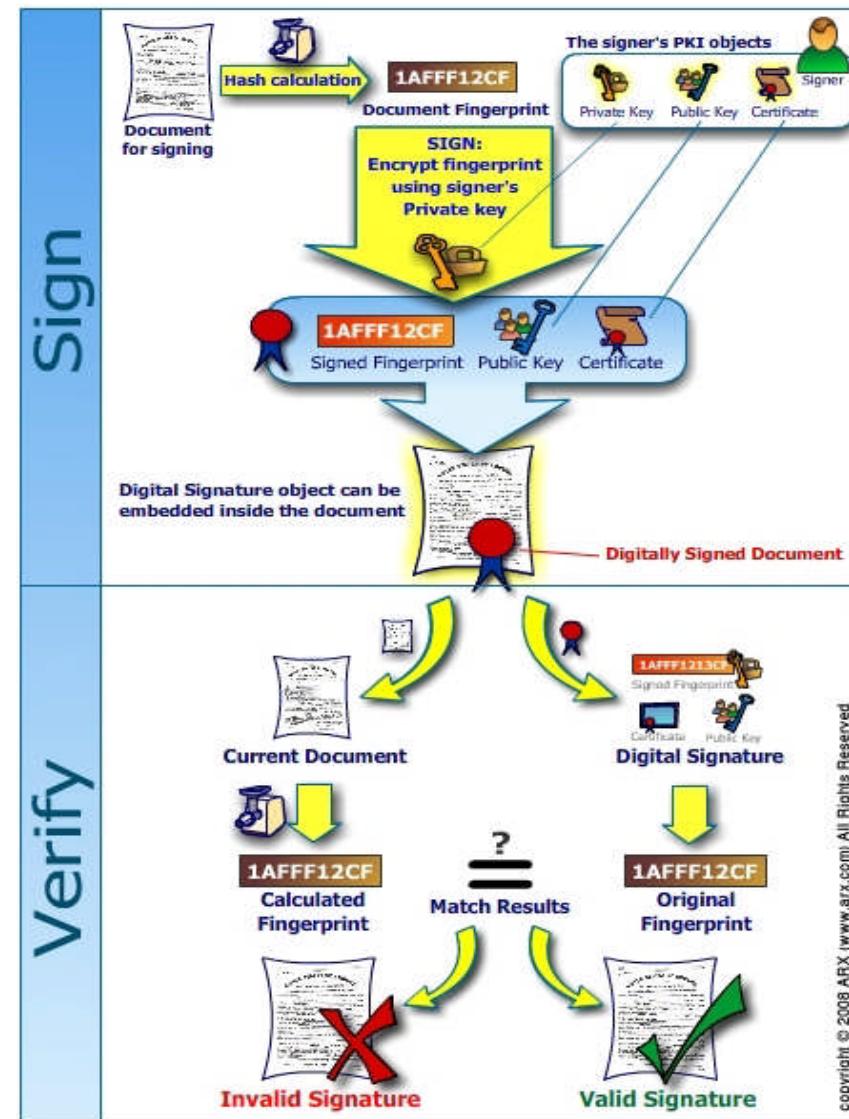
- CA is a trusted server that generates certificates of the form $\{\text{name}, \text{public key}\}_{\text{CA}}$ where CA is the certification authority's signature (private) **key**. All hosts are preconfigured with the certification authority's **public key**, therefore any host can check the signature on these certificates. Note that a CA is more attractive than a KDC because a CA it doesn't need to be on-line. Certificates can be stored anyplace and forwarded anywhere as they are needed.

Digital Signatures

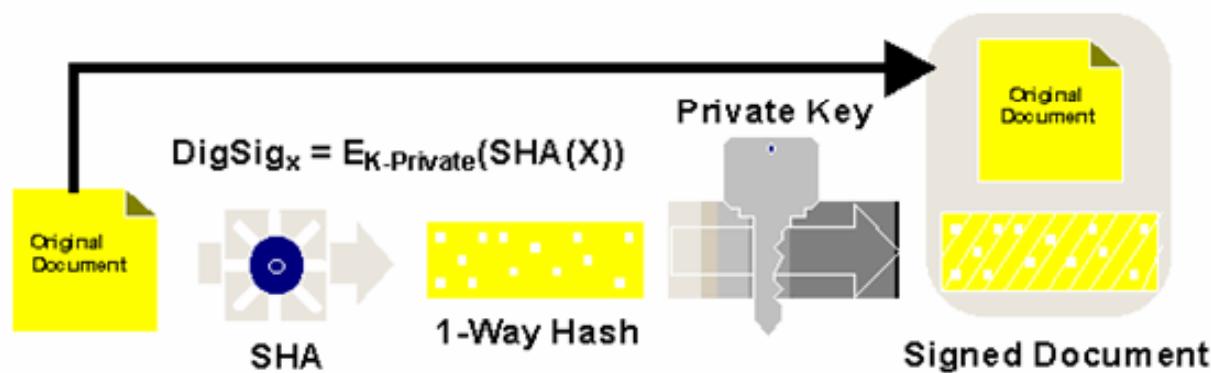


Digital Signatures





Signing using SHA and RSA



Random numbers

- PRNG are predictable, `/dev/urandom`
 - Use a fixed function and time to generate random numbers.
 - Example: $r = \text{md5}(\text{time} + \text{hostname} + \text{process id})$
- RNG `/dev/random`, uses system entropy
 - Use physical system to generate random number
 - Example: $r = \text{md5}(\text{mouse and keyboard delays})$

How safe are your passwords

- WORD, SIZE, BITS, CRACK-TIME
 - 1 Single word of any language, 8 char, 24-bits, Seconds
 - 2 Random [a-z] 8 char, 37-bits, Minutes
 - 3 Random [a-z] 16 char, 75-bits, Decades

Password length and strength

Completely random printable string

SIZE, BITS, CRACK-TIME

- 6 char, 40-bits, Minutes
- 8 char, 52-bits, Hours
- 12 char, 78-bits, Decades
- 15 char, 97-bits, Centuries
- 20 char, 130-bits, Un-crackable

How safe are your passwords

- WORD, SIZE, BITS, CRACK-TIME
- Single word of any language, 8 char, 24-bits, Seconds
- Random [a-z] 8 char, 37-bits, Minutes
- Random [a-z] 16 char, 75-bits, Decades

GPG (Gnu privacy guard)

PGP (pretty good privacy)

- GPG
- C:\> gpg --version
- C:\> gpg –gen-key
 - » gpg: key 43F2B829 marked as ultimately trusted
 - » public and secret key created and signed.
- ~/.gnupg
- gpg --export --armor > Public-key.asc
- gpg --import file.asc
- gpg --sign-key RedHat

GPG usage 2

- Generate a private key:

```
gpg --gen-key
```

- Get your public key as ascii text:

```
gpg --armor --output pubkey.txt --export you@ashesi.edu.gh
```

- Send your keys to a key-server

```
gpg --send-keys youremail --keyserver hkp://subkeys.pgp.net
```

- Import Friend's key

```
gpg --import friend.asc OR gpg --search-keys 'friend@ashesi.edu.gh'  
--keyserver hkp://subkeys.pgp.net
```

See <http://www.gnupg.org/gph/en/manual.html> for more help.

GPG usage 3

- Encrypt message.txt for your friend:

```
gpg --encrypt --recipient friend@ashesi.edu.gh  
message.txt
```

- Reading mail from your friend

```
gpg --decrypt reply.txt
```

- Signing a file

```
gpg --armor --detach-sign my-file.zip
```

- Verify the sign

```
gpg --verify crucial.zip.asc crucial.zip
```

Key server

<http://pgp.mit.edu/>

MIT PGP Public Key Server

Help: [Extracting keys](#) / [Submitting keys](#) / [Email interface](#) / [About this server](#) / [FAQ](#)
Related Info: [Information about PGP](#) / [MIT distribution site for PGP](#)

Extract a key

Search String:

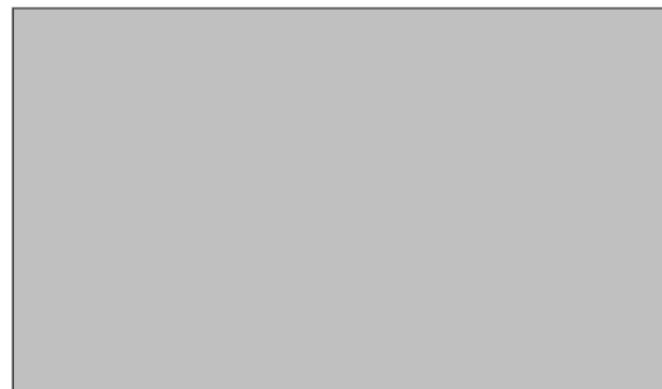
Index: Verbose Index:

Show PGP fingerprints for keys

Only return exact matches

Submit a key

Enter ASCII-armored PGP key here:



Securing the cookie

- Cookie is a temporary permit
- `setcookie("autologin", md5($_SERVER['REMOTE_ADDR']), time() + 3600);`
- Cookie must be **secured**, so https cookie is not sent over http connection.
- CSRF (**Cross-site request forgery**) example
``
from http://en.wikipedia.org/wiki/Cross-site_request_forgery

B2B Communications

Technologies for b2b Comm

- Traditional paper in mail
- Email?
- IMs?
- http
- Xml
- Protobuffer

Checkout

- **Introduction to Google Checkout**
- Google Checkout lets your customers buy items from you quickly and securely using a Google username and password.
- You can use Google Checkout to charge customers' credit cards, track orders through your fulfillment process and receive order payments in your bank account.
- As such, Google Checkout touches each step of the customer's shopping experience, beginning with the customer's search for an item and continuing through the order checkout and fulfillment processes.

HTML or XML

- Merchants can choose between two types of Google Checkout implementation options:
- **HTML APIs** enable merchants to send information to Google Checkout and receive information from Google Checkout using name/value pairs in HTML forms rather than XML.
- Merchants can also use the HTML API to submit name=value pairs via a server-to-server HTTP POST request.

HTML

- HTML implementations are particularly recommended for small merchants who do not want to generate XML.
- Merchants can not digitally sign orders in HTML implementations, so merchants who use this implementation and do not submit server-to-server requests should plan to review orders manually.

XML

- **XML APIs** enable merchants to access all Google Checkout features.
- XML implementations are recommended for merchants who need to be able to digitally sign orders before sending them to Google.
- XML implementations are also recommended for merchants who want to offer coupons or discounts and for merchants who plan to integrate Google Checkout with their internal order processing and billing systems.

XML Syntax

- <TAG ATTRIBUTE=VALUE> TAG-DATA </TAG>
- EXAMPLE of tag,attribute,value,data:
 - <students>
 - <astudent>
 - <name post="President">Gina Smith</name>
 - <phone>555-123-456</phone>
 - <ssn>111-23-4567</ssn>
 - </astudent>
 - </students>

Xml Tag

- A piece of text that describes a unit of data, or element, in XML.
- The tag is distinguishable as markup, as opposed to data, because it is surrounded by angle brackets (< and >).
- For example, the element <name>My Name</name> has the start tag <name>, the end tag </name>, which enclose the data "My Name".
- To treat such markup syntax as data, you use an entity reference or a CDATA section.

Xml Attributes

- A qualifier on an XML tag that provides additional information.
- For example, in the 'slide' tag
`<slide title="Ecom">`,
'title' is an attribute, and
'Ecom' is its value.

Multiple ways of writing data

- ELEMENT vs ATTRIBUTES,
 1. .. LEAF.
 2. Fixed <size>20</size> .. UGLY,

MIXED.

- 3. Unlimited nesting of ELEMENTS:

```
<font>  
  <name>fixed</name>  
  <size>20</size>  
</font>
```

HTML: if not displayed is an attribute

- Many DTD don't use ATTRIBUTES.
- Use ATTRIBUTES to specify interpretation of value <temperature scale="centrigrade">30</temperature>

No fixed

Yes <name> fixed</name>
<size>30</size>

XML CDATA

- Character References: ࡊ (TM symbol, unicode in hex)
- Entity names [<&>"']: < & >
" '
- <![CDATA[Raw strings like "'&"
#include <stdio>]]>
- <? Processing Instructions: ?>
- <!-- Comments -->

XML vs html

XML Case Sensitive

HTML No case

XML Balanced tags.

HTML Can skip tags like
, </p>.

HTML Single tags (open and close together), eg.

XML vs HTML 2

XML Attributes must quoted, eg. <p
width="100">A</p>

HTML Optional quotes, eg. <p width=100>A</p>

XML All attributes are named,
eg. <input type="radio" checked="true">

HTML Attributes can be unnamed,
eg. <input type="radio" checked>

Parsing XML in Java

- import java.io.*;
- import javax.xml.parsers.*;
- import org.w3c.dom.*;
- import org.xml.sax.*;

```
DocumentBuilderFactory factory =
    DocumentBuilderFactory.newInstance();
DocumentBuilder builder =
    factory.newDocumentBuilder();
Document doc =
    builder.parse("file:///mosh/java/data/xml.xml");
```

XML traversal

```
// Prints all the children:  
//   child name=name  
//   child name=size  
NodeList children = root.getChildNodes();  
for(int i=0; i<children.getLength(); i++){  
    Node child = children.item(i);  
    if( child instanceof Element ){  
        Element childElement = (Element) child;  
        System.out.println("child=" + childElement.getTagName());  
    }  
}
```

Checkout buyer

- The Google Checkout order process begins when the customer shops on your site and adds items to a virtual shopping cart.
- In addition to your regular checkout options, your web pages will display a [Google Checkout button](#).
- This button will be contained within a form on your page.

Click BUY button

- When the customer clicks the Google Checkout button, the customer's browser will submit the shopping cart information to Google.
- You can include the customer's order information in an HTML form on your page or submit the order information using a server-to-server HTTP POST request.
- Google Checkout will then display the **Sign In/Sign Up** page, which shows the items in the customer's order and the shipping options available for the order.
- The **Sign In/Sign Up** page, which is shown below, also allows the customer to create a new Google Account or log in to an existing account.



Secure Checkout

Order Details - Example.com, (800) 555-1234 1576 Random Road, Boston, MA 01234

Qty Item

1 Cake pan - 10-in. Nonstick Angel Food Cake Pan by Flying M...

Price

\$15.99

Shipping & handling 2nd Day (est. \$1.49)

\$1.49

Tax: -

Subtotal: \$17.48

Create a **Google** Account to complete this purchase

Current email:

Choose a password:

[Password strength.](#)

Re-enter password:



Credit card number:

Expiration date:

 mn / yy

CVC: [What's this?](#)

Name on card:

Billing address line 1:

Billing address line 2:
(optional)

City

State

Zip

Phone number:

e.g. 850-555-1212. Required for account verification.

I agree to the [Terms of Service](#).

[Agree and continue »](#)

You can still make changes to

Or sign in

If you already have a **Google** Account

Email:

Password:

[Sign in and continue](#)

[Forgot your password?](#)

BUY with Cookie

- If Google detects a cookie indicating that the customer already has a Google Account and the account already contains credit card information,
- the right side of this page will display the email address for that account.

Place order

lmino@gmail.com | [My Account](#) | [Help](#) | [Sign out](#)

Google Checkout™

[Review and place order](#)

Order Details Example.com (800) 555-1234 1576 Random Road, Boston, MA 01234

Qty	Item	Price
1	Cake pan - 10-in. Nonstick Angel Food Cake Pan by Flying M...	\$15.99
	Shipping & handling (CA) Media Mail (\$1.49)	\$1.49
	Tax (CA) :	\$1.31
	Total:	\$18.79

Keep my email address confidential
Google will forward all email from Example.com to lmino@gmail.com. [Learn more](#)

I want to receive promotional email from Example.com.

Ship to: **Dave Bowman - [Change](#)**
1600 Amphitheatre Pkwy
Mountain View, CA 94043

Pay with: **Visa xxx-1234 - [Change](#)**

[« Edit order](#) [Place your order now \\$18.79](#)

Billing Information & Privacy
Your credit card will be charged by Google and your billing information will be kept private. "GOOGLE * Example.com" will appear by the charge on your credit card statement. To process your order, Example.com will have access to your shipping information, email address, and phone number. Your phone number may be used in relation to your order, but will never be used to deliver unsolicited commercial messages to you. [Learn more](#)

Example.com Return Policy
Example.com will accept returns on all products in NEW condition within 15 days of the delivery date, minus original shipping and handling fees, return shipping, and a 15% restocking fee. All exchanged and returned merchandise must be in original factory condition, including all packaging materials, inserts and manuals, warranty cards (not filled-out) and all accessories. Please do not tape or write anything on the item or the package. We will charge you for replacement of damaged, altered missing, written-on or taped-on contents or cartons. We reserve the right to refuse any such returns. [View entire return policy »](#)

©2006 Google [Terms of Use](#) - [Privacy Policy](#) - [Google Home](#)

Order data

- The target URL for the Change order link is the value of the `<edit-cart-url>` tag.
- The **Qty** column displays the value of the `<quantity>` tag for each item in the order.
- The bold text in the **Item** column displays the value of the `<item-name>` tag for each item in the order.
- The text following the name of each item is the `<item-description>` for that item.
- The **Price** column displays the value of the `<unit-price>` tag for each item in the order as well as the appropriate currency symbol for that price.
- The pulldown menu in the **Shipping & Handling** row displays a list of the shipping methods that you offer, which are contained in the `<shipping-methods>` tag in your API request. Each item in the pulldown menu displays the name and price of the shipping method.
- word **Tax** identifies the state of the shipping address selected by the buyer.

Order data

- The Enter coupon or gift card ... link displays if either the **<accept-gift-certificates>** or **<accept-merchant-coupons>** tag has a value of **true** in the Checkout API request or you have defined coupons using the [Google Checkout Coupon Creator](#).
- The **Tax** row displays the total calculated tax for the order based on the information included in the **<tax-tables>** that you send to Google with each order.

Ordered

lmno@gmail.com | [My Account](#) | [Help](#) | [Sign out](#)



✓ **Dave, thanks for your order!**

- A copy of your receipt has been emailed to you and saved in your Google Checkout [purchase history](#).
- You can check your receipt at any time for [up-to-date order status](#).

[Return to Example.com »](#)

Test the production account

- **curl** -d '<hello xmlns=
["http://checkout.google.com/schema/2"](http://checkout.google.com/schema/2)'/
https://**P_MERCHANT_ID**:**P_MERCHANT_KEY**
@[checkout.google.com/api/checkout/v2/request/](https://checkout.google.com/api/checkout/v2/request/Merchant/)
[Merchant/](https://checkout.google.com/api/checkout/v2/request/Merchant/)**P_MERCHANT_ID**
- <?xml version="1.0" encoding="UTF-8"?> <bye
xmlns="<http://checkout.google.com/schema/2>"
serial-number="c567262a-dd13-4084-b8d3-
6ccfbabc69d03" />

Submitting data

- <form method="POST"
action="https://checkout.google.com/api/checkout/v2/checkout/Merchant/MERCHANT_ID">
- The [Google Checkout Sample Code](#) page provides ASP, Java, .NET, PHP and Perl code samples to help you with your integration.

XML shopping cart

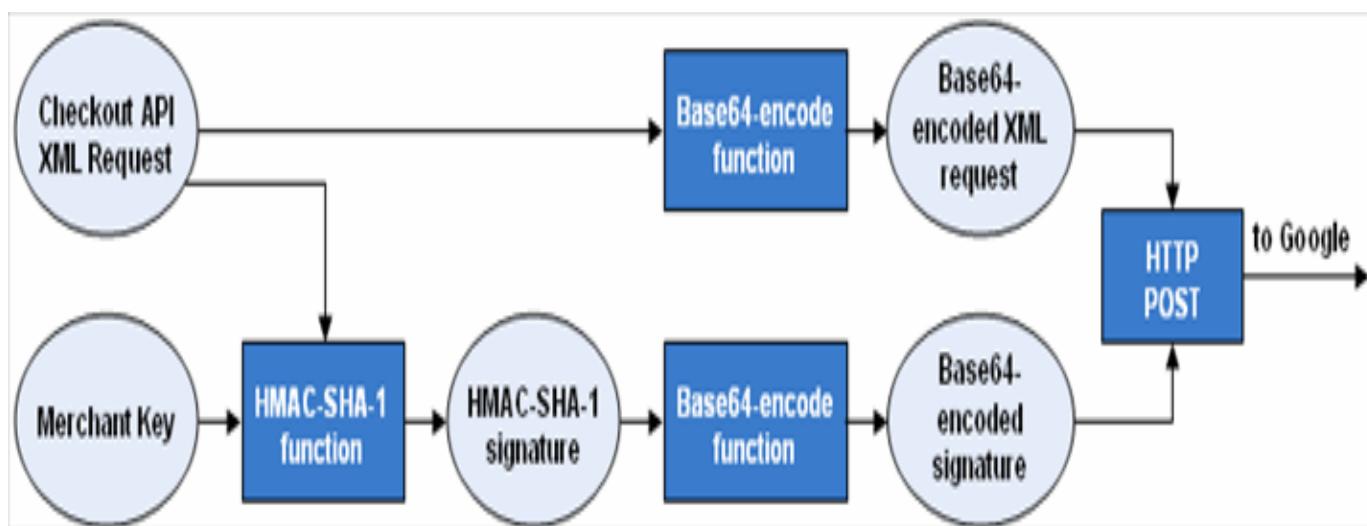
- <?xml version="1.0" encoding="UTF-8"?>
- <checkout-shopping-cart xmlns="http://checkout.google.com/schema/2">
- <shopping-cart>
- <items>
- <item>
- <item-name>HelloWorld 2GB MP3 Player</item-name>
- <item-description>HelloWorld, the simple MP3 player</item-description>
- <unit-price currency="USD">159.99</unit-price>
- <quantity>1</quantity>
- </item>
- </items>
- </shopping-cart>

Continued ..

- <checkout-flow-support>
- <merchant-checkout-flow-support>
- <shipping-methods>
- <flat-rate-shipping name="SuperShip Ground">
- <price currency="USD">9.99</price>
- </flat-rate-shipping>
- </shipping-methods>
- </merchant-checkout-flow-support>
- </checkout-flow-support>
- </checkout-shopping-cart>

Securing the shopping cart

- Create an HMAC-SHA-1 signature, a cryptographically secure value that enables Google to verify that the your Checkout API request was not altered before Google received it.
- To create the signature, call the appropriate function in your development environment, passing the Checkout API XML structure and your **Merchant Key** as parameters to the function.



Base 64 encode the data

- Base64-encode the Checkout API XML request. Most development environments provide a function for base64-encoding.
- Using the same function as in the previous step, base64-encode the HMAC-SHA-1 signature that you created in step i.
- Include the base64-encoded Checkout API request and the base64-encoded HMAC-SHA-1 signature in the form on your page that displays a Google Checkout button.
- The HTML example below shows a sample [Google Checkout button](#) form.
- The form contains two hidden input fields. The first field, which is assigned the name **cart** contains the base64-encoded Checkout API XML request. The second field, which is assigned the name **signature**, contains the base64-encoded HMAC-SHA-1 signature.

Signed XML

- <form method="POST"
- action="https://sandbox.google.com/checkout/api/checkout/v2/checkout/Merchant/1234567890">
- <input type="hidden" name="cart"
- value="PD94bWwgdmVyc2lvbj0iMS4wIj8+CjxjaGVja291dC1zaG9wcGlu">
- <input type="hidden" name="signature"
- value="kdjsf590GFDGK23l2kgit259fjSDKET0592jalkfwe3539Gjekwu">

- <input type="image" name="Google Checkout" alt="Fast checkout through Google"
-
- src="http://sandbox.google.com/checkout/buttons/checkout.gif?merchant_id=1234567890
-
- &w=180&h=46&style=white&variant=text&loc=en_US" height="46" width="180">
- </form>

Optional

- Set up a web service to perform custom calculations for tax, shipping, coupons and gift certificates after the customer has reviewed the order in Google Checkout.

XML best practices

- All XML messages between you and Google Checkout must use UTF-8 (Unicode) encoding. Specify UTF-8 encoding .
- Time/date values use the [ISO 8601 standard](#), which specifies time as an offset from UTC.
- All money elements must have a **currency** attribute.

XML validation

- **Validating XML Messages**
- curl -d '@**FILENAME**'
[https://MERCHANT_ID:MERCHANT_KEY
@checkout.google.com/api/checkout/v2/re
quest/Merchant/MERCHANT_ID/diagnose](https://MERCHANT_ID:MERCHANT_KEY@checkout.google.com/api/checkout/v2/request/Merchant/MERCHANT_ID/diagnose)

Schema

- <?xml version="1.0" encoding="UTF-8"?>
- <diagnosis
 xmlns="http://checkout.google.com/schema/2"
- serial-number="49ba18e3-016b-4c52-a697-
 159a3lk38bf9">
- <input-xml>
- <charge-order google-order-
 number="552406916759246">
- <amount currency="USD">5.51</amount>
- </charge-order>
- </input-xml>
- </diagnosis>

XML API

- A Checkout API request sends order information from your website to Google, thereby enabling your customer to complete an order using Google Checkout. The following XML shows a simple Checkout API request ..

API request

- <?xml version="1.0" encoding="UTF-8"?>
- <checkout-shopping-cart xmlns="http://checkout.google.com/schema/2">
- <shopping-cart>
- <items>
- <item>
- <item-name>HelloWorld 2GB MP3 Player</item-name>
- <item-description>HelloWorld, the simple MP3 player</item-description>
- <unit-price currency="USD">159.99</unit-price>
- <quantity>1</quantity>
- </item>
- </items>
- </shopping-cart>

Continued

- <checkout-flow-support>
- <merchant-checkout-flow-support>
- <shipping-methods>
- <flat-rate-shipping name="SuperShip Ground">
- <price currency="USD">9.99</price>
- </flat-rate-shipping>
- </shipping-methods>
- </merchant-checkout-flow-support>
- </checkout-flow-support>
- </checkout-shopping-cart>

XML Api data

- A Checkout API request contains two main sections:
- The `<shopping-cart>` element contains the list of items in the customer's shopping cart. You can specify the following information for each item in the cart:
 - The name of the item.
 - A description of the item.
 - The per-unit cost of the item.
 - The number of units of the item that the buyer is ordering.
 - A value, such as a SKU, that you use to uniquely identify the item.
 - An alternate tax table that should be used to calculate tax for the item. If you do not specify an alternate tax table for an item, Google Checkout will use your default tax table to calculate tax for the item.

XML Api data 2

- A flag indicating that the item is a digital good and additional information that explains how the customer will be able to access the digital content after completing the order. If you sell digital goods, please see the [Digital Delivery](#) feature document for more information about formatting Checkout API requests to sell digital goods.
- Additional, [proprietary information](#) that you would like associated with the item in the order. This information will not be displayed to the buyer but will be included in subsequent API exchanges that include information about the order.
- The [`<checkout-flow-support>`](#) element contains additional information that is needed during the checkout process, such as your shipping options and tax tables. Please see the following sections of this document for additional information about checkout-flow-support features:
- [Shipping and Digital Delivery](#)
[Taxes](#)
[Coupons, Gift Certificates and Price Adjustments](#)

Other APIs...

- Taxes
- Shipping and Handling

Coupons

- You can specify a coupon code to identify the offer.
- You can specify a discount value as an absolute amount (in dollars or British pounds). U.S. merchants can also specify a discount as a percentage of the order total. For U.S. merchants, the discount will be provided before taxes or shipping charges are added to the order total. For U.K. merchants, the discount will be applied after taxes and shipping charges are added to the order total.
- You can specify that the coupon may only be used if the order total exceeds a specified amount.
- You can limit the number of times that each customer can use the coupon.
- You can make the coupon available to all customers or only to customers who have never purchased from you through Google Checkout.

Callbacks with SSL

- Google Checkout uses SSL to send [merchant-calculation-callback](#) requests.
- As such, your web service must use HTTPS and you must have a valid SSL certificate.
- Your production service for handling merchant calculation callbacks must use port 443, which is the default port for HTTPS.

Testing and debugging

- Error reporting to merchant
- Debugging problems
- Auditing
- Reporting all transactions
- Fraud detection

Using Windows effectively

Windows Disk consist of drives:

C:\ D:\ E:\

And each disk has directories, e.g.

c:\windows

c:\Documents and Settings\...\Desktop

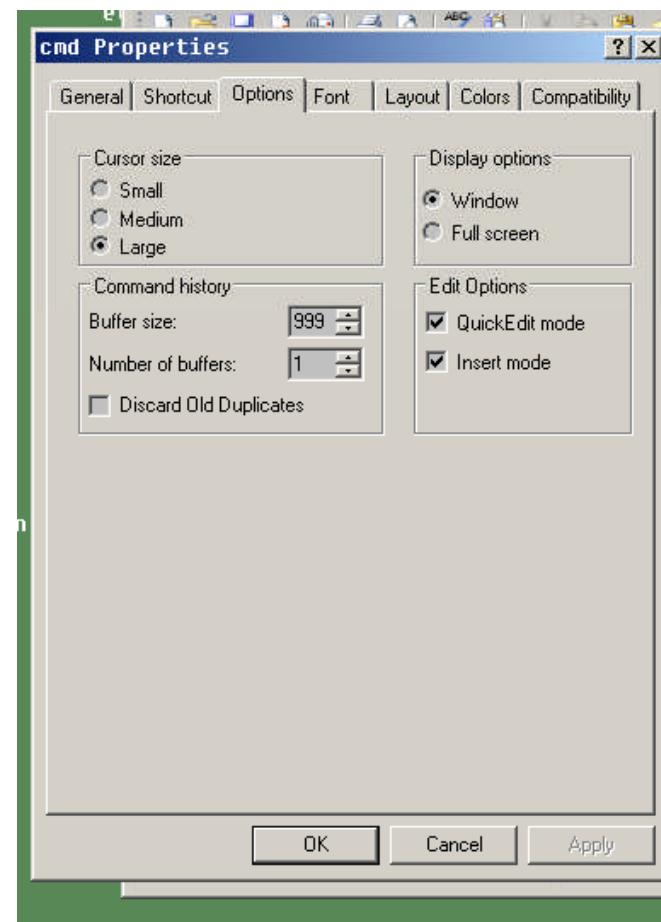
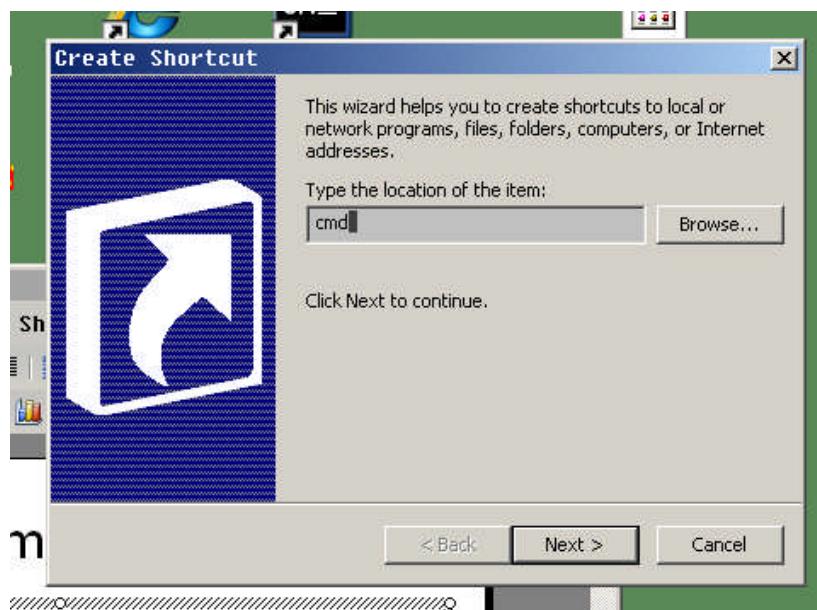
And each directory contains files,

Files have a name and extension, eg.

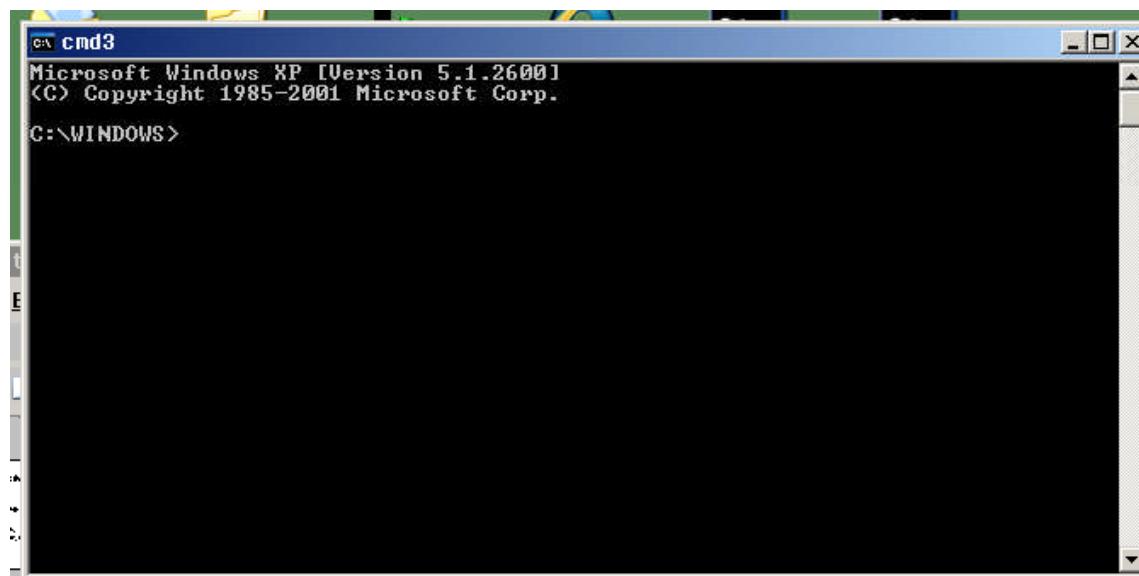
C:\WINDOWS\system32\drivers\etc\hosts

Command line

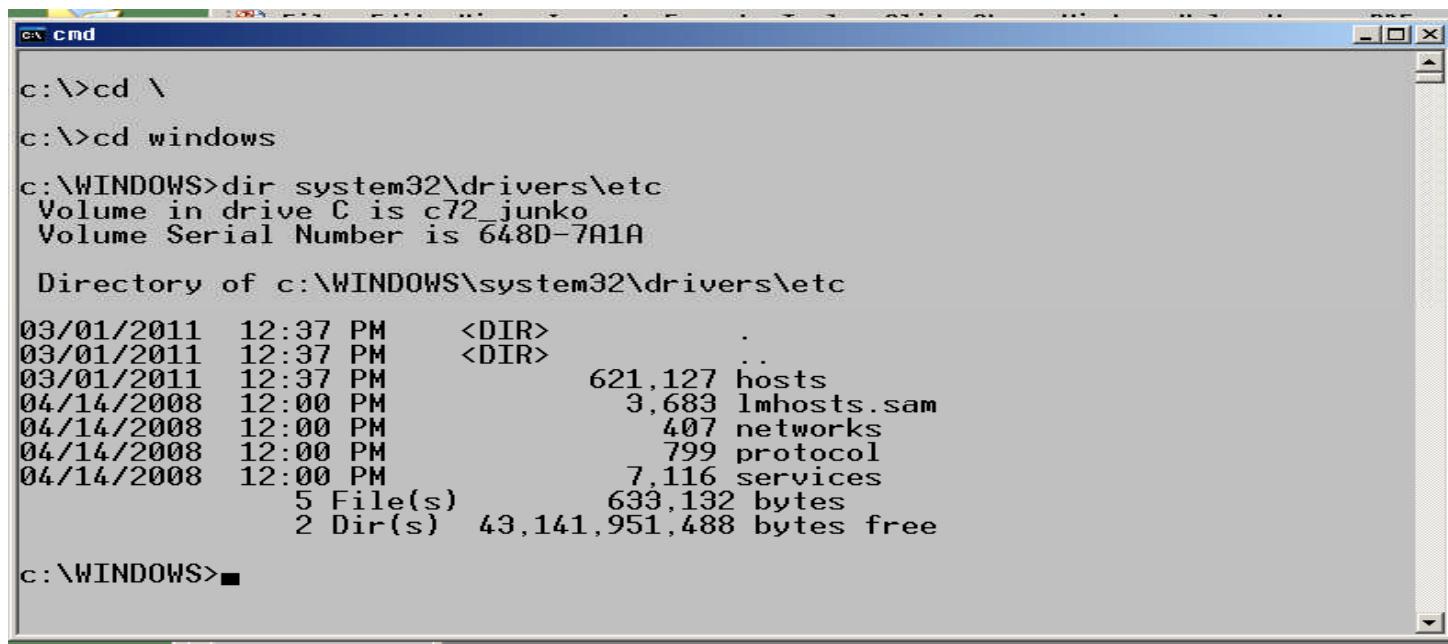
- Right click -> new -> shortcut -> cmd -> cmd
- Right click -> properties -> quick-edit



- Run cmd



cd and dir



```
c:\>cd \
c:\>cd windows
c:\WINDOWS>dir system32\drivers\etc
Volume in drive C is c72_junko
Volume Serial Number is 648D-7A1A

Directory of c:\WINDOWS\system32\drivers\etc

03/01/2011  12:37 PM    <DIR>
03/01/2011  12:37 PM    <DIR>
03/01/2011  12:37 PM            621,127 hosts
04/14/2008  12:00 PM            3,683 lmhosts.sam
04/14/2008  12:00 PM            407 networks
04/14/2008  12:00 PM            799 protocol
04/14/2008  12:00 PM            7,116 services
              5 File(s)   633,132 bytes
              2 Dir(s)  43,141,951,488 bytes free

c:\WINDOWS>
```

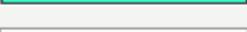
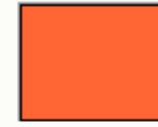
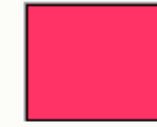
Primes

```
C:\mosh\mtext>gpg --gen-prime 1  
160
```

++++

```
F8EB025BF1A501F6CEAA8A535B3  
48DE368951407
```

ENTER OR ROLL OVER COLOR VALUE, OR CLICK ON THE COLOR SWATCH BELOW

Current Color				
	 51.255.204 #33FFCC	 51.204.255 #33CCFF	 51.102.255 #3366FF	 102.51.255 #6633FF
 R: 51 G: 255 B: 204 <input type="button" value="Set RGB"/>	 51.255.102 #33FF66	 0.245.184 #00F588	 0.184.138 #00888A	 204.51.255 #CC33FF
 <input type="button" value="# 33FFCC"/>	 102.255.51 #66FF33	 184.0.46 #B8002E	 245.0.61 #F5003D	 255.51.204 #FF33CC
<input type="button" value="Set HEX"/> <input type="button" value="Lighten Scheme"/> <input type="button" value="Darken Scheme"/>	 204.255.51 #CCFF33	 255.204.51 #FFCC33	 255.102.51 #FF6633	 255.51.102 #FF3366

Converting Hex

- bash\$ echo "ibase=16 ; 8C6F " | bc
- c:\>bc -ql bc 1.06
- ibase=16
- 8C6F
- 35951
- obase=2
- 8C6F
- 1000110001101111

bc

- c:\> bc -q
- 3^4 % 5
- 1
- 2^64
- 18446744073709551616

Encryption

Internet is unreliable

- Packets can get:
 - corrupted,
 - stolen.
- Sniffers - wireshark for windows

Security Issues

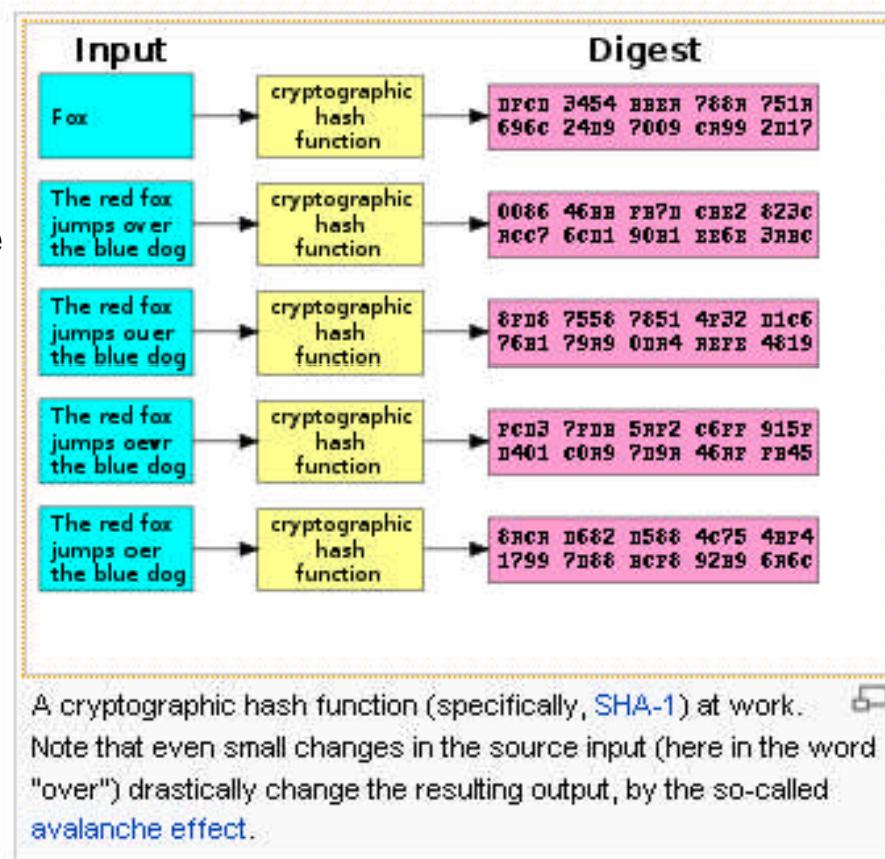
- DNS poisoning
- MITM (Man In The Middle) attack
- Phishing
- Malware

Protecting data with Encryption

- We need to lock the data
- Make sure it is not tampered
- Make sure it available only to the recipients.

Cryptographic hash functions

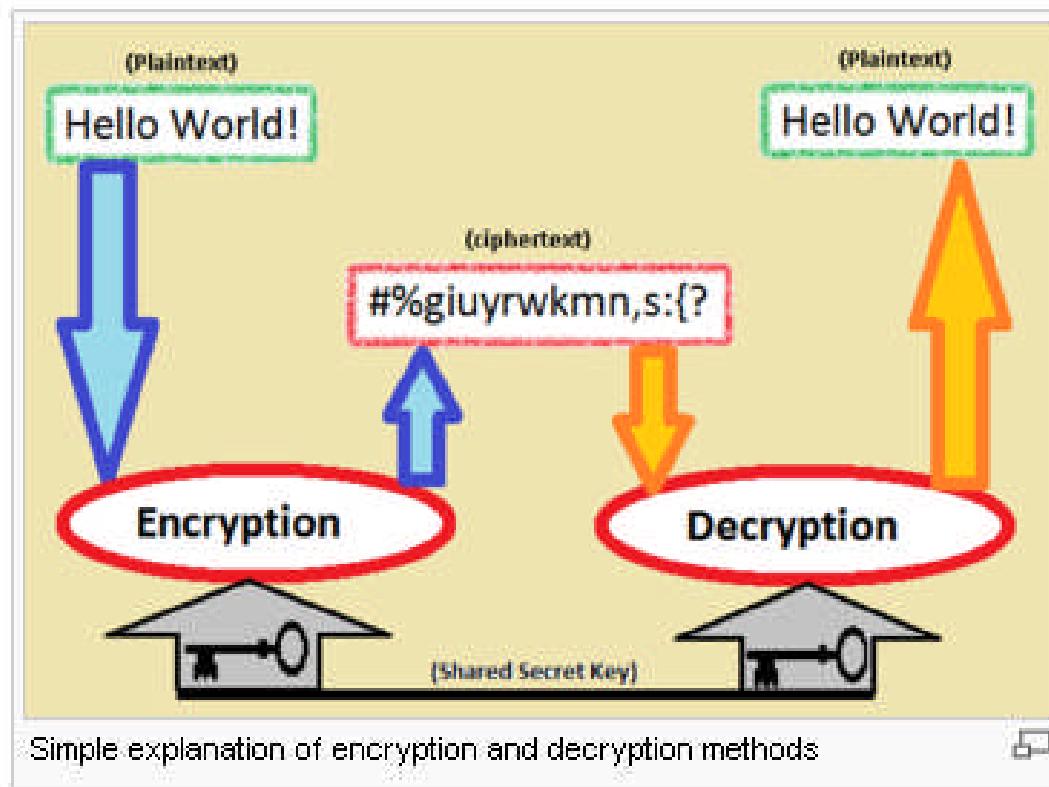
- One way functions - given output, very hard to compute input.
- Hard to find another input that gives same output.
- Hard to find 2 different inputs with same output.



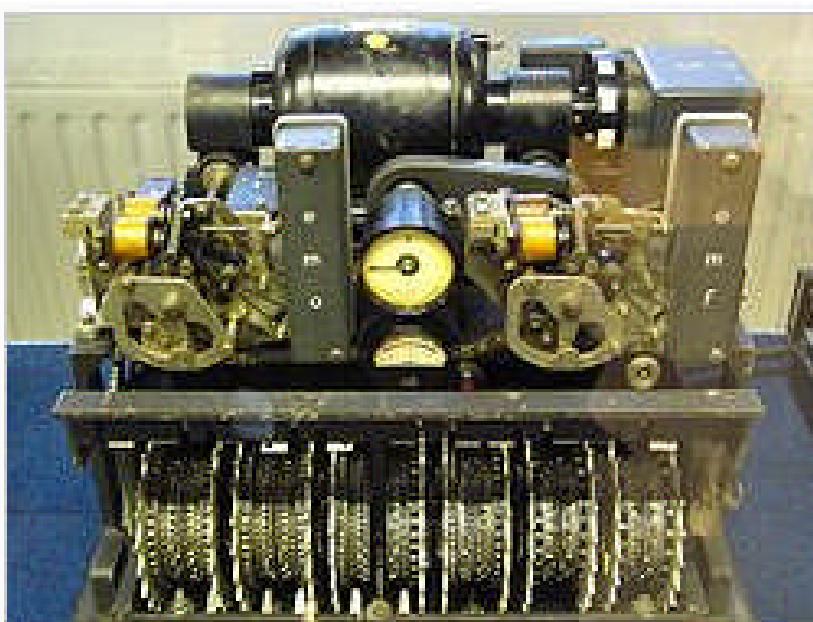
Examples of hashing functions

- MD5 128 bits
- SHA 160
- SHA2 256,512
- Used to generate signatures.

Encryption



Encryption before computers



German Lorenz cipher machine, used in World War II to encrypt very-high-level general staff messages

Symmetric key encryption

- Symmetric key encryption uses one key, called secret key - for both encryption and decryption. Users exchanging data must keep this key secret. Message encrypted with a secret key can be decrypted only with the **same** secret key.
- Examples: [DES](#) - 64 bits, [3DES](#) - 192 bits, [AES](#) - 256 bits, IDEA - 128 bits, Blowfish, Serpent

Symmetric vs Public key encryption



A) Secret key (symmetric) cryptography. SKC uses a single key for both encryption and decryption.



B) Public key (asymmetric) cryptography. PKC uses two keys, one for encryption and the other for decryption.

Public key Cryptography

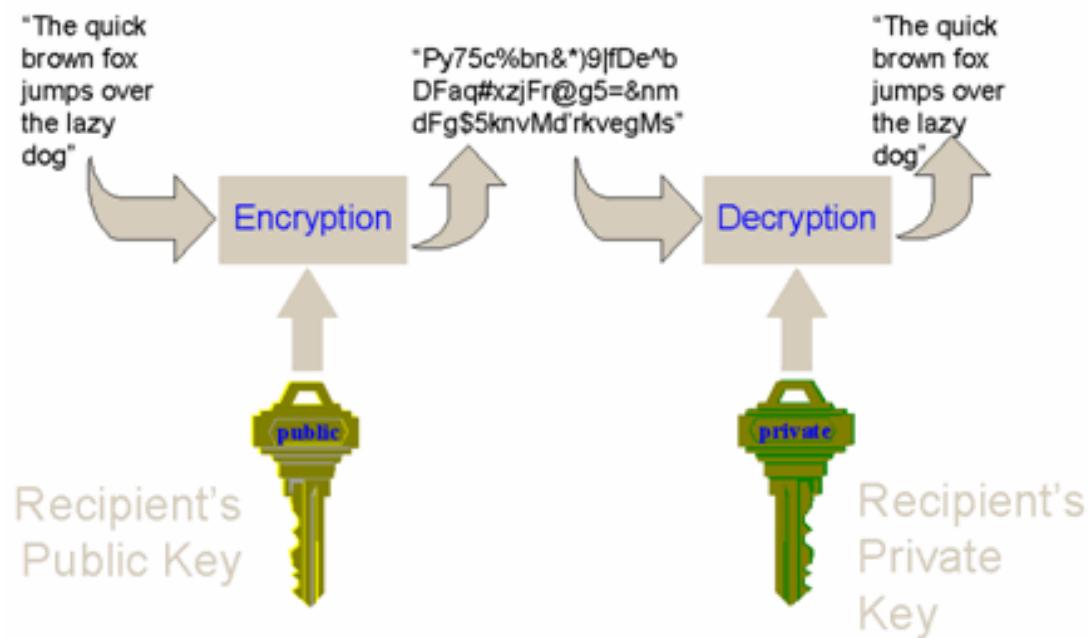
http://en.wikipedia.org/wiki/Public-key_cryptography

- The two main branches of public key cryptography are:
- Public key encryption: a message encrypted with a recipient's public key cannot be decrypted by anyone except a possessor of the matching private key—presumably, this will be the owner of that key and the person associated with the public key used. This is used for [confidentiality](#).
- [Digital signatures](#): a message signed with a sender's private key can be verified by anyone who has access to the sender's public key, thereby proving that the sender had access to the private key (and therefore is likely to be the person associated with the public key used), and the part of the message that has not been tampered with. On the question of [authenticity](#), see also [message digest](#).

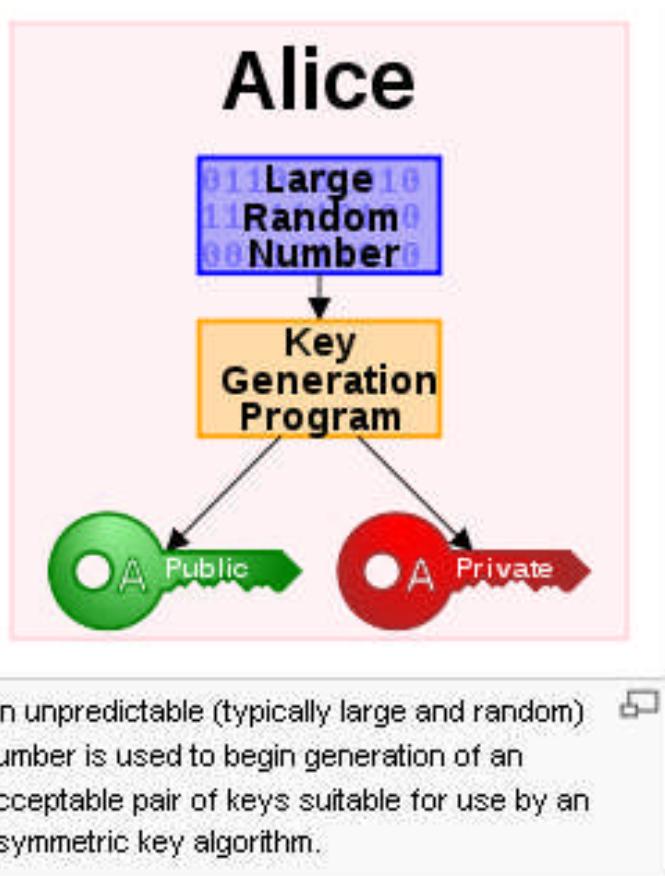
Properties of Encryption

- No leakage of information in transit.
- Same message looks different if sent twice.
- Size of message does not indicate anything.
- Time stamped to avoid relay attacks.
- Cannot be cracked by brute force trials.

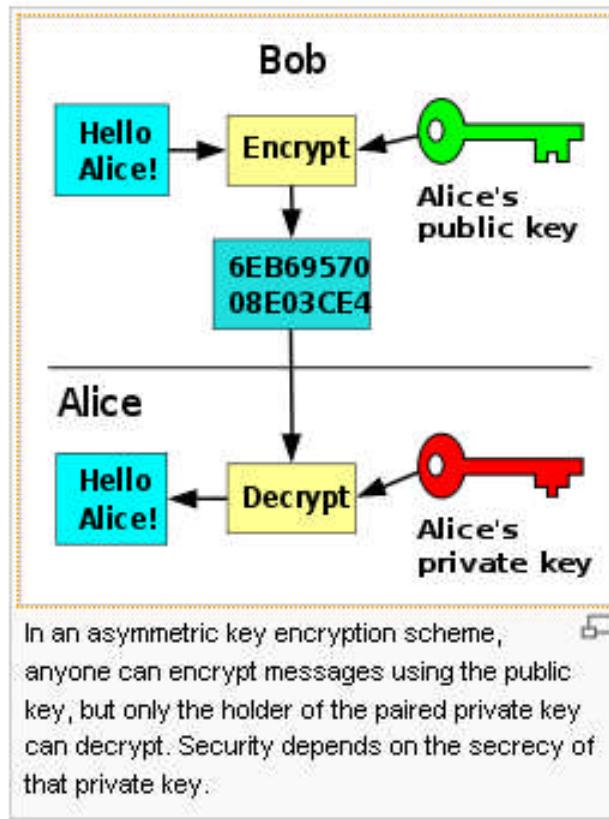
Public key cryptography



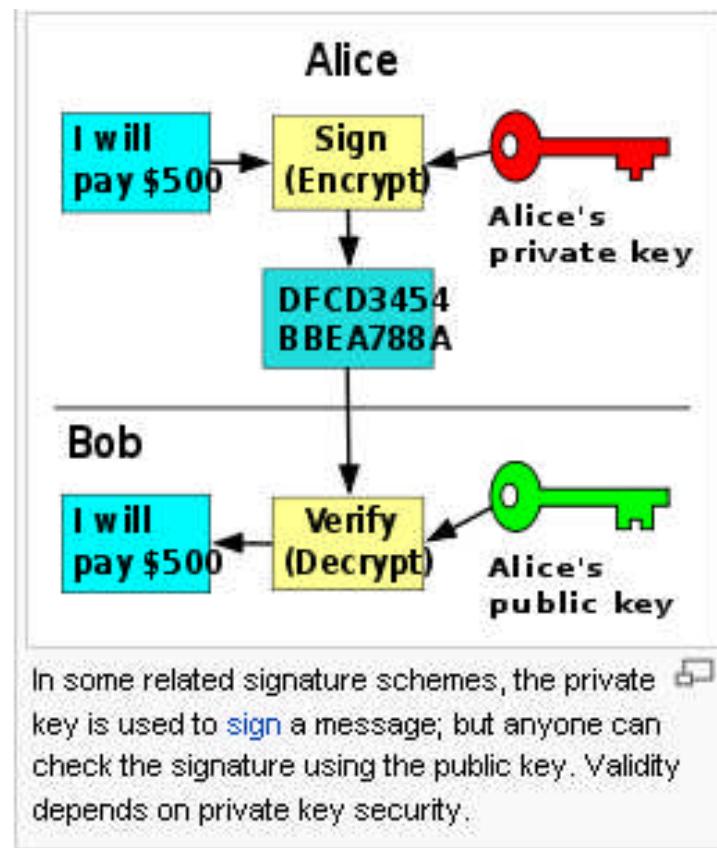
Alice makes a public key lock



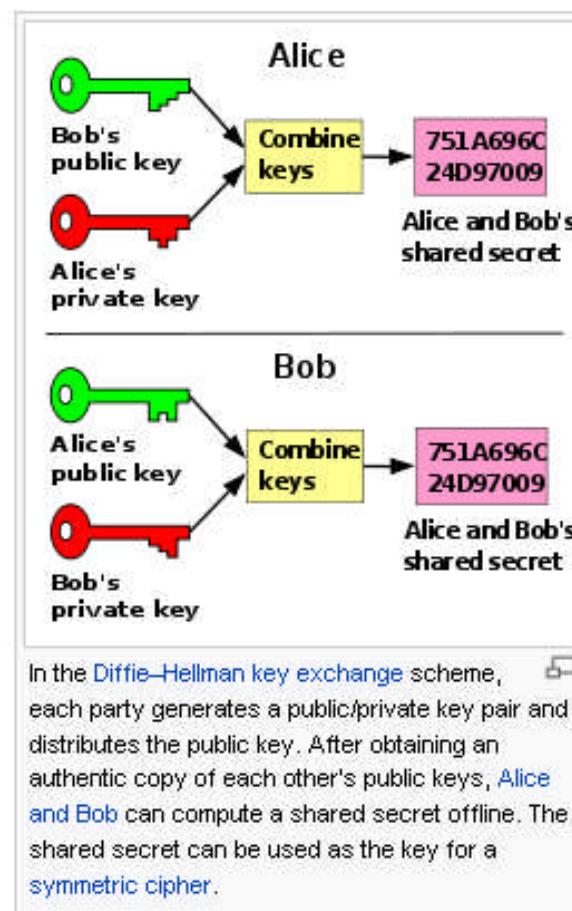
Any can lock, only Alice can unlock



Signing: Only Alice can lock anyone can unlock



Diffie Hellman key exchange, each user have their own locks



Primes and factoring

- > OpenSSL> prime 17 (is 17 a prime?)
 - 11 is prime (Hex 11 = 16 + 1 = 17 Dec).
- > gpg --gen-prime 1 30 (30 bit prime).
 - 3756E197 (which is 928440727 DEC).

Openssl to generate public key

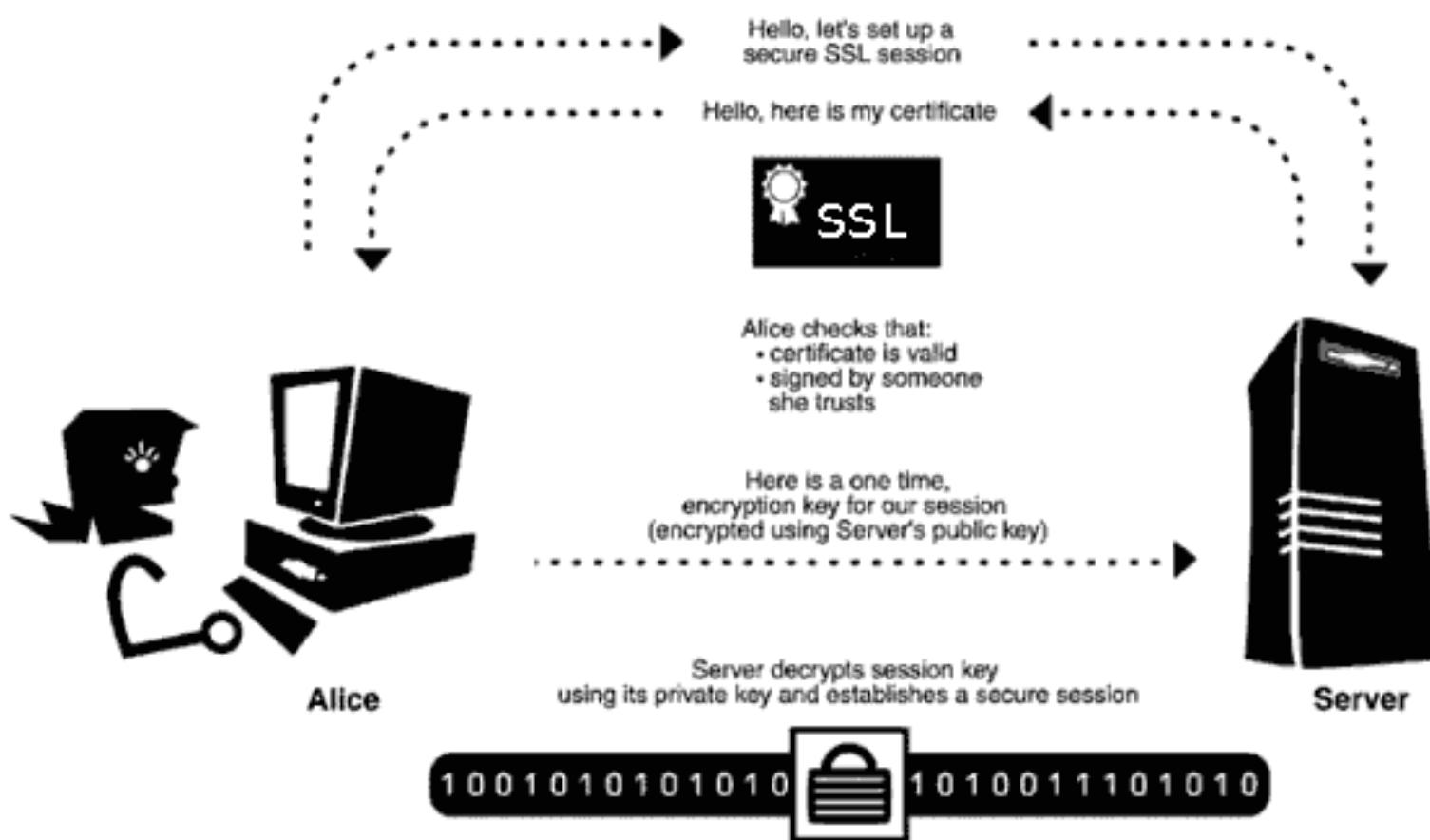
Generate a new public/private keypair:

- \$ openssl genrsa -out key.pem
- Generating RSA private key, 512 bit long modulus
- ..+++++.....++++++
- e is 65537 (0x10001)

Extract the modulus,e,primes from your key:

- \$ openssl rsa -in key.pem -noout -text
- publicExponent: 65537 (0x10001)
- Modulus=....long-string-of-hex-digits...

SSL – Secure socket layer



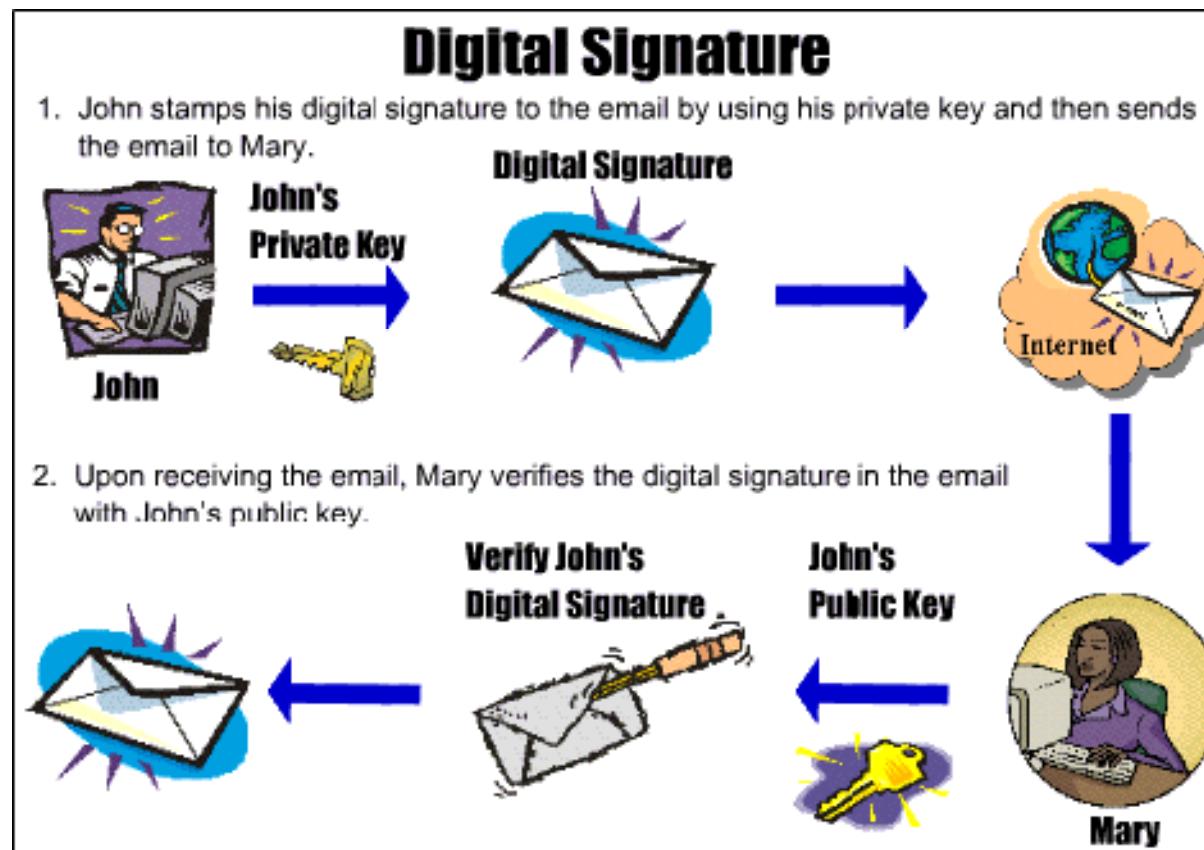
SSL Protocol Summary

- The process to establish an SSL connection is the following: The user uses his browser to connect to the remote server.
- The handshake phase starts, and the browser and server exchange keys and certificate information.
- The browser checks the validity of the server certificate, including that it has not expired, that it has been issued by a trusted CA, and so on.
- Optionally, the server can require the client to present a valid certificate as well.
- Server and client use each other's public key to securely agree on a symmetric key.
- The handshake phase concludes and transmission continues using symmetric cryptography.

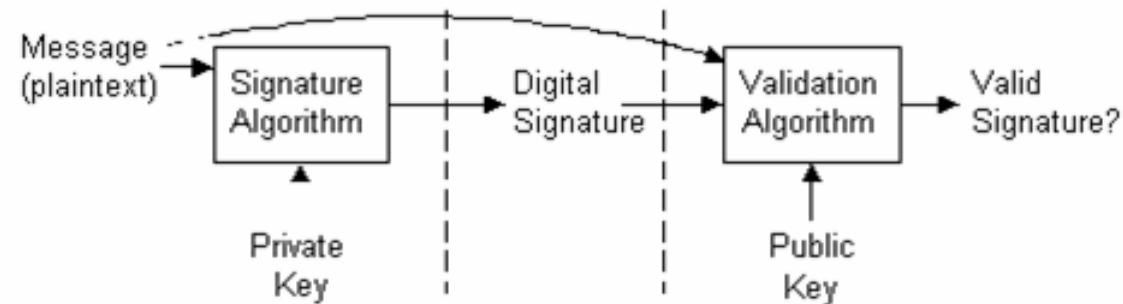
Certificate

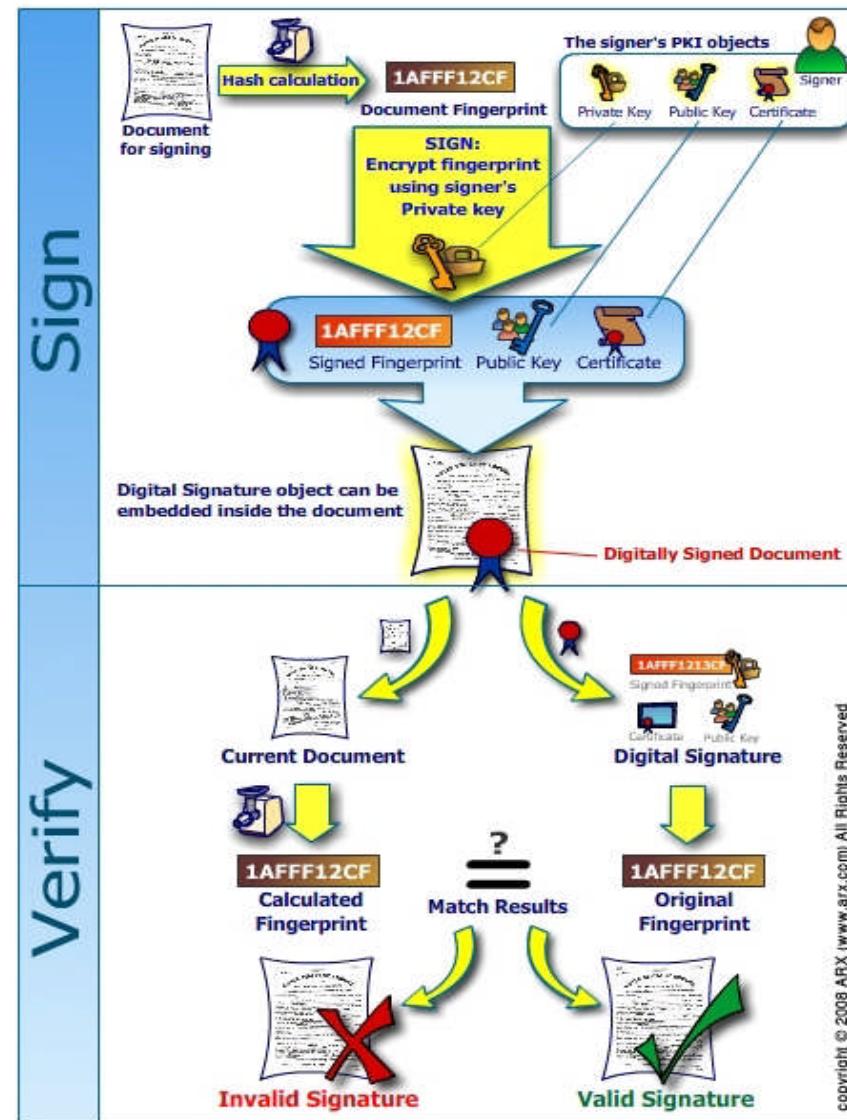


Digital Signatures

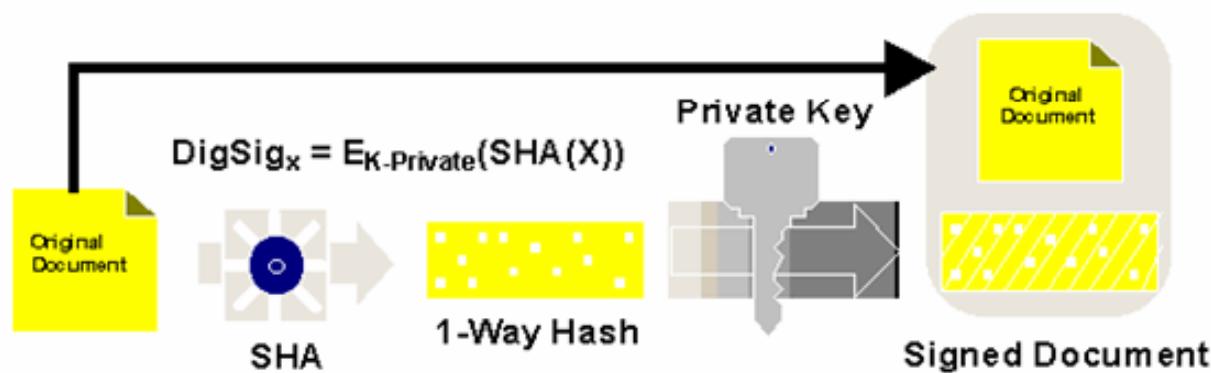


Digital Signatures





Signing using SHA and RSA



Password length and strength

Completely random printable string

SIZE, BITS, CRACK-TIME

- 6 char, 40-bits, Minutes
- 8 char, 52-bits, Hours
- 12 char, 78-bits, Decades
- 15 char, 97-bits, Centuries
- 20 char, 130-bits, Un-crackable

GPG (Gnu privacy guard)

PGP (pretty good privacy)

- GPG
- C:\> gpg --version
- C:\> gpg -gen-key
 - » gpg: key 43F2B829 marked as ultimately trusted
 - » public and secret key created and signed.
- ~/.gnupg
- gpg --export --armor > Public-key.asc
- gpg --import file.asc
- gpg --sign-key RedHat

GPG usage 2

- Generate a private key:

```
gpg --gen-key
```

- Get your public key as ascii text:

```
gpg --armor --output pubkey.txt --export you@ashesi.edu.gh
```

- Send your keys to a key-server

```
gpg --send-keys youremail --keyserver hkp://subkeys.pgp.net
```

- Import Friend's key

```
gpg --import friend.asc OR gpg --search-keys 'friend@ashesi.edu.gh'  
--keyserver hkp://subkeys.pgp.net
```

See <http://www.gnupg.org/gph/en/manual.html> for more help.

GPG usage 3

- Encrypt message.txt for your friend:

```
gpg --encrypt --recipient friend@ashesi.edu.gh  
message.txt
```

- Reading mail from your friend

```
gpg --decrypt reply.txt
```

- Signing a file

```
gpg --armor --detach-sign my-file.zip
```

- Verify the sign

```
gpg --verify crucial.zip.asc crucial.zip
```

Key server

<http://pgp.mit.edu/>

MIT PGP Public Key Server

Help: [Extracting keys](#) / [Submitting keys](#) / [Email interface](#) / [About this server](#) / [FAQ](#)
Related Info: [Information about PGP](#) / [MIT distribution site for PGP](#)

Extract a key

Search String:

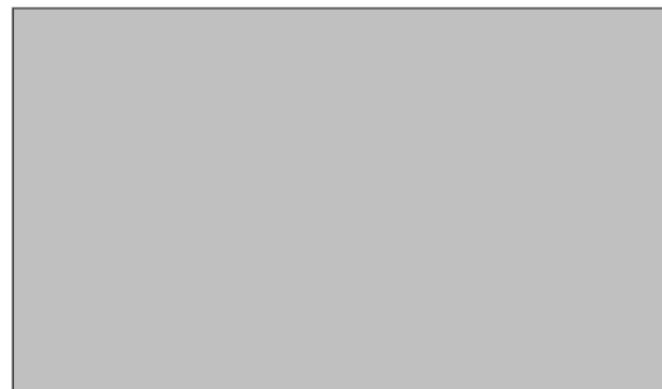
Index: Verbose Index:

Show PGP fingerprints for keys

Only return exact matches

Submit a key

Enter ASCII-armored PGP key here:



Interviewing

Resume

- Short sentences to the point.
- Don't write long paragraphs about your project and internal project names.
- Spell check
- Grammar check
- Dates when appropriate.
 - Internship with NGO, 5/2010-7/2010.
- Plain txt, doc and pdf file (no ms-word or ppt).
- Contact email(s), phone numbers, local address.

Resume

- Action oriented, start sentence with action.
 - I arranged a water tanker for the school.
- Result oriented, highlight what you solved
 - **Solved** water problem by arranging tanker.
- Quantitative, quantify your work
 - Wrote **1000** lines of C++.
 - Saved **2000\$** by reusing old hard drives.

Resume

- Be **honest**, do not write something you cannot explain when asked to elaborate. It is not ok to say you forgot (within 3 years of college).
- E.g. Wrote a big project in C++.
 - Q. What version of compiler did you use, why?
 - Q. What do you think of STL?
- E.g. **Expert** in Java
 - Q. How would you debug a deadlock in your code?

Resume

- Make it interesting, add few projects you did outside of class work, e.g. visited orphanage, wrote a blog on Ghanian birds.
- References – ask references in advance, be prepared.

Not for resume

- Race
- Gender
- Citizenship, unless for local jobs
- Compensation
- Family history
- Transcripts
- Photo?

Internet Reputation

- Google – search your name + ashesi on google, to make your webpage comes up first.
- Create your
 - google.com/profiles
 - linkedin.com
 - blogger.com
 - sites.google.com
- Check other sites that refer to your name.
- Facebook.com – make sure your wall is clean.

Do your homework on the company

- Company size, products, years, employees, offices.
- Earnings, stock performance
- Check what others say about them and the products, use google.com

Dressing



Ashesi brand

- *US President Barack Obama wears his Ivy League badge on his shirt sleeves and business suits. And it has been no different during his ongoing India trip.* The conservatism of the ensemble was in keeping with the tradition of his Harvard University background. The president studied law there. Harvard men are known for their smart power suits that combine slim silhouettes, straight lines and slim fits, says its website.
- Create an Ashesi Tie/Pin/Blazer.

Going for an interview

- Sleep well before the interview.
- Do not try to study too much on last day.
- Deep breath to reduce stress – meditate and be calm.
- Dress well and formally, not flashy.

Do's

- Discuss technical issues related to job
- Discuss global issues related to development
- Ask questions about the job, work culture.
- Company future.
- Your future in the company
- Interesting Technology, Finance questions.
- Be courteous and listen carefully
- Ask about who you will work with, the position and duties required of you.
- Compensation

Don't

- Discuss politics,
- Religion, Race, Gender, Jokes.
- Personal opinions.
- Previous compensation.
- Over-eat during the interview lunch

Technical questions

- **Think loudly** - solve one step at a time.
- Write on board.
- Understand the question.
- Ask and clarify
- List your assumptions
- List alternatives.
- Don't panic if you didn't solve one problem.

1. What are they looking for?

- Communication skills?
- Friendly to work with.
- Problem finder and specifier, not just solver.
- Self motivated.
- Interested in work.
- Interesting to work with.

2. What are they looking for?

- Technically competent.
- Wider vision, not just technically sound.
- Well read technically.
- Drive solutions to completion.

You are also interviewing them

- Do you like the company, so take the first job – ask questions, do they listen to you – if not, maybe it is not your type of company.
- Make sure you like their **questions**, food, style, dressing
 - if not, it is sign their culture is not right for you,
 - Maybe they will bore you soon.

Sample questions

- Q. How would you multiple two large 100 digit numbers?
- Q. Is the internet safe for children?
- Q. How many cars are there in Accra?
- Discussion

Preparation

- <http://slashdot.org> – programmers talk site.
- news.google.com – current events.
- Wikipedia.org - read, contribute.
- Unix programming Environment book.
- PHP and MySQL book.
- Contribute to open source projects.

Cloud computing

Using millions of computers in Service Oriented Arch

Dr.Mohsin Ahmed

mosh.ahmed@gmail.com

25-Aug-2011

Scale

- 64 bit machines
 - Lots of RAM – 10s of GB
 - Lots of HD – few TB
 - No video, audio, monitors, serial ports
 - Gps NIC, USB for debugging
- Linux or BSD operating system
- Racks of 100 computers
- 100,000 machines in a data center

Hardware Issues

- Power / rack = $300W * 100 = 30KW$
- Power / dc = $300W * 100,000 = 30MW$
- Heat dissipation
- Power supply
- High speed network fibre
 - $100,000 * 1Mbs/machine = 100 Gps$
- Physical security for data.

Hardware failures

- Reliability – 0.01% failure = 10/dc/day.
- Detect RAM,DISK,POWER,Mother board errors/failures.
- Replace components without down time.

Dealing with Failures

- So if machine m1234 fails, another free machine will be named m1234, and all programs from old to new machine. Users will not even realize the machine has changed!

Fault tolerant distributed DNS

- DNS maps names to machines (ip address) in DC. Machines may have names like “m1291919”
- Transparently removes bad machines from use.
- Install software on new machines
- Replace failed machines with new ones
- Inform operator of failures

Security

- Ssh with public and private keys on all machines
- Two factor authentication for users
- Proxy servers at firewalls
- Log all user access.
- Log all jobs, for security audits.
- Accounting – CPU is expensive, 10\$/hour, with 20K machines = 20,000\$/hour charge.

Monitoring

- All machines must report back in realtime to master monitors, which compute statistics like CPU usage, users, bandwidth, power, temperature, failure rates.
- Error rate > threshold => send email, page operator, call operator, dynamically throttle traffic, instead of falling over under load. (Domino effect).

wikipedia/Leader_election

- In distributed computing, **leader election** is the process of designating a single process as the organizer of some task **distributed** among several computers (nodes).
- Before the task is begun, all network nodes are unaware which node will serve as the "**leader**," or coordinator, of the task.
- After a **leader election** algorithm has been run, however, each node throughout the network recognizes a particular, unique node as the task **leader**.

The Chubby Lock Service for Loosely-Coupled Distributed Systems

- **Chubby** lock service, which is intended to provide coarse-grained locking as well as reliable (though low-volume) storage for a loosely-coupled distributed system.
- **Chubby** provides an interface much like a distributed file system with advisory locks, but the design emphasis is on availability and reliability, as opposed to high performance.
- Many instances of the service have been used for over a year, with several of them each handling a few tens of thousands of clients concurrently.

Details

- Proto buffers
- GFS
- Bigtable
- Mapreduce

Protocol buffers

What Are Protocol Buffers?

From <http://code.google.com/apis/protocolbuffers/>

- Protocol buffers are Google's language-neutral, platform-neutral, extensible mechanism for serializing structured data – **think XML, but smaller, faster, and simpler.**
- You define how you want your data to be structured once, then you can use special generated source code to easily write and read your structured data to and from a variety of data streams and using a variety of languages – Java, C++, or Python.

Protocol buffers

- IPC – Interprocess communication
- RPC – remote procedure calls
- Data passing across network in protocol-buffers (open source library for C++, Python, Java from google).
- PB are compressed structs/objects
- PB 100x more efficient than XML, JSON.

Protocol buffers

- PB data can be exchanged between servers in different languages.
- Versioning of data, so old clients can keep operating with newer servers.
- Secure RPC - like SSL.
- Servers and clients have certificates
- Data is encrypted with public key.

Why Proto Buffers?

<http://code.google.com/apis/protocolbuffers/docs/overview.html>

- New fields could be easily introduced, and intermediate servers that didn't need to inspect the data could simply parse it and pass through the data without needing to know about all the fields.
- Formats were more self-describing, and could be dealt with from a variety of languages (C++, Java, etc.)
- Automatically-generated serialization and deserialization code avoided the need for hand parsing.

PB in RPC

- In addition to being used for short-lived RPC (Remote Procedure Call) requests, people started to use protocol buffers as a handy self-describing format for storing data persistently (for example, in Bigtable).
- Server RPC interfaces started to be declared as part of protocol files, with the protocol compiler generating stub classes that users could override with actual implementations of the server's interface.

polyline.pb example

- message Point {
- required int32 x = 1;
- required int32 y = 2;
- optional string label = 3;
- }
- message Line {
- required Point start = 1;
- required Point end = 2;
- optional string label = 3;
- }
- message Polyline {
- repeated Point point = 1;
- optional string label = 2;
- }
- // From wikipedia

Using polyline.pb.cc

- This subsequently compiled with protoc.
- A C++ program can then use it like so:
 - #include "polyline.pb.h"
 - Line* createNewLine(const std::string& name) {
 - Line* line = new Line;
 - line->mutable_start()->set_x(10);
 - line->mutable_start()->set_y(20);
 - line->mutable_end()->set_x(30);
 - line->mutable_end()->set_y(40);
 - line->set_label(name);
 - return line;
 - }

GFS

Distributed file system

wiki/Google_File_System

- **Google File System (GFS or GoogleFS)** is a proprietary distributed file system developed by Google Inc. for its own use.^[1] It is designed to provide efficient, reliable access to data using large clusters of commodity hardware.
- As opposed to many file systems, **GFS** is not implemented in the kernel of an operating system, but is instead provided as a userspace library.

GFS Masters and Chunkservers

- one **Master** node and a large number of
- **Chunkservers** store the data files, with each individual file broken up into fixed size chunks (hence the name) of about **64 megabytes**, similar to clusters or sectors in regular file systems.
- Each chunk is assigned a unique **64-bit label**, and logical mappings of files to constituent chunks are maintained.
- Each chunk is **replicated** several times throughout the network, with the minimum being three, but even more for files that have high demand or need more redundancy.

GFS Master

- The Master server doesn't usually store the actual chunks, but rather all the [metadata](#) associated with the chunks, such as the tables mapping the 64-bit labels to chunk locations and the files they make up, the locations of the copies of the chunks, what processes are reading or writing to a particular chunk, or taking a "snapshot" of the chunk pursuant to replicating it (usually at the instigation of the Master server, when, due to node failures, the number of copies of a chunk has fallen beneath the set number).
- All this metadata is kept current by the Master server periodically receiving updates from each chunk server ("Heart-beat messages").

Chunk servers

- Keep adding chunk servers and more masters to increase capacity.
- See <http://labs.google.com/papers/gfs.html>

Big table

The excel sheet of the web

Billions of rows and columns
stored on thousands of machines

wikipedia.org/wiki/BigTable

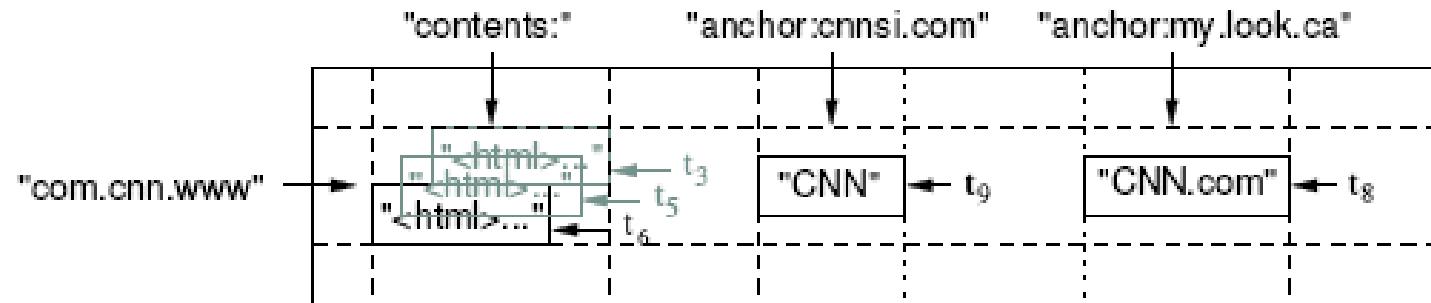
- **BigTable** is a compressed, high performance, and proprietary database system built on Google File System (GFS), Chubby Lock Service, SSTable

BT

- **BigTable** maps two arbitrary string values (row key and column key) and timestamp (hence three dimensional mapping) into associated arbitrary byte array.
- It is not a relational database and can be better defined as a sparse, distributed multi-dimensional sorted map.
- **BigTable** is designed to scale into the [petabyte](#) range across "hundreds or thousands of machines, and to make it easy to add more machines [to] the system and automatically start taking advantage of those resources without any reconfiguration"

BT Cell (row, 3 columns)

from <http://labs.google.com/papers/bigtable.html>

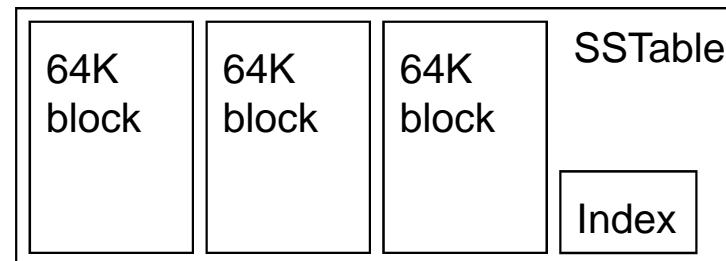


BT

- <Row, Column, Timestamp> triple for key - lookup, insert, and delete API
- Arbitrary “columns” on a row-by-row basis
 - Column family:qualifier. Family is heavyweight, qualifier lightweight
 - Column-oriented physical store- rows are sparse!
- Does not support a relational model
 - No table-wide integrity constraints
 - No multirow transactions

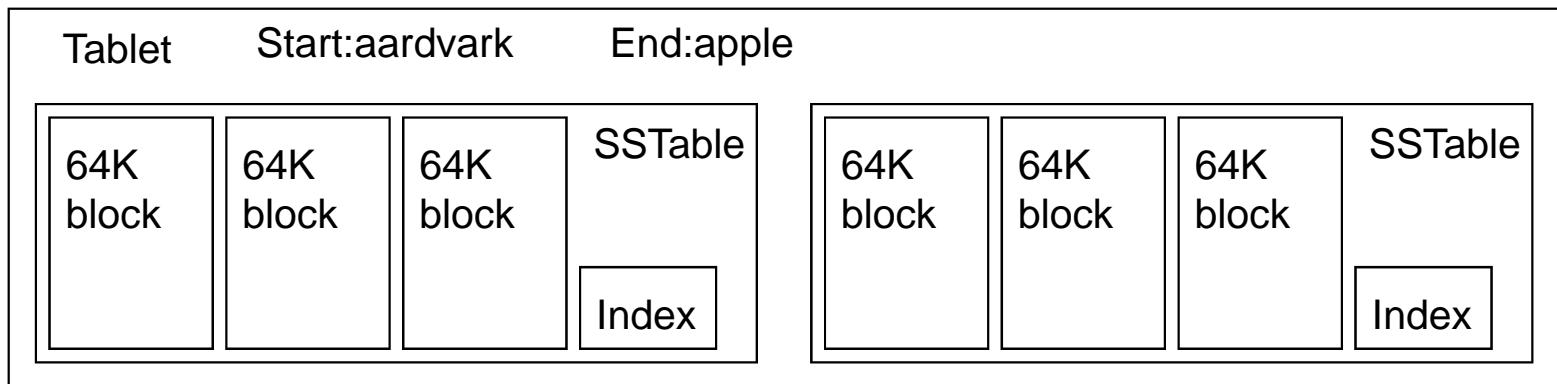
SSTable

- Hash map stored on disk
- Immutable, sorted file of key-value pairs
- Chunks of data plus an index
 - Index is of block ranges, not values



Tablet

- Contains some range of rows of the table
- Built out of multiple **SSTables**



Uses

- Store the whole internet in a BT
- Store huge hash map in a BT
- Input to mapreduce
- Output from mapreduce
- Access is very fast
- Cache server

MR

MAP REDUCE

Map Reduce

- Parallel processing of independent data
 - Trillions of web docs
 - Trillions of logs records
- For
 - Billing ads
 - Detecting fraud
 - Searching
 - Machine learning (spelling, voice, spam)
 - Indexing web documents

Map Reduce

- Mapreduce patented by google
- Hadoop from Apache – Yahoo
 - Hadoop is not fault tolerant, java, HDFS, Amazon S3.
- Programming – 2 base classes
 - Map
 - Reduce
- <http://en.wikipedia.org/wiki/MapReduce>
- <http://code.google.com/edu/submissions/mapreduce-minilecture/listing.html>

Sequential MR in python

- C:\> Python
- >>> print 'primes=',filter(None,map(lambda y:y*reduce(lambda x,y:x*y!=0,map(lambda x,y=y%x,range(2,int(pow(y,0.5)+1))),1),range(2,1000)))
- primes=[2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53,

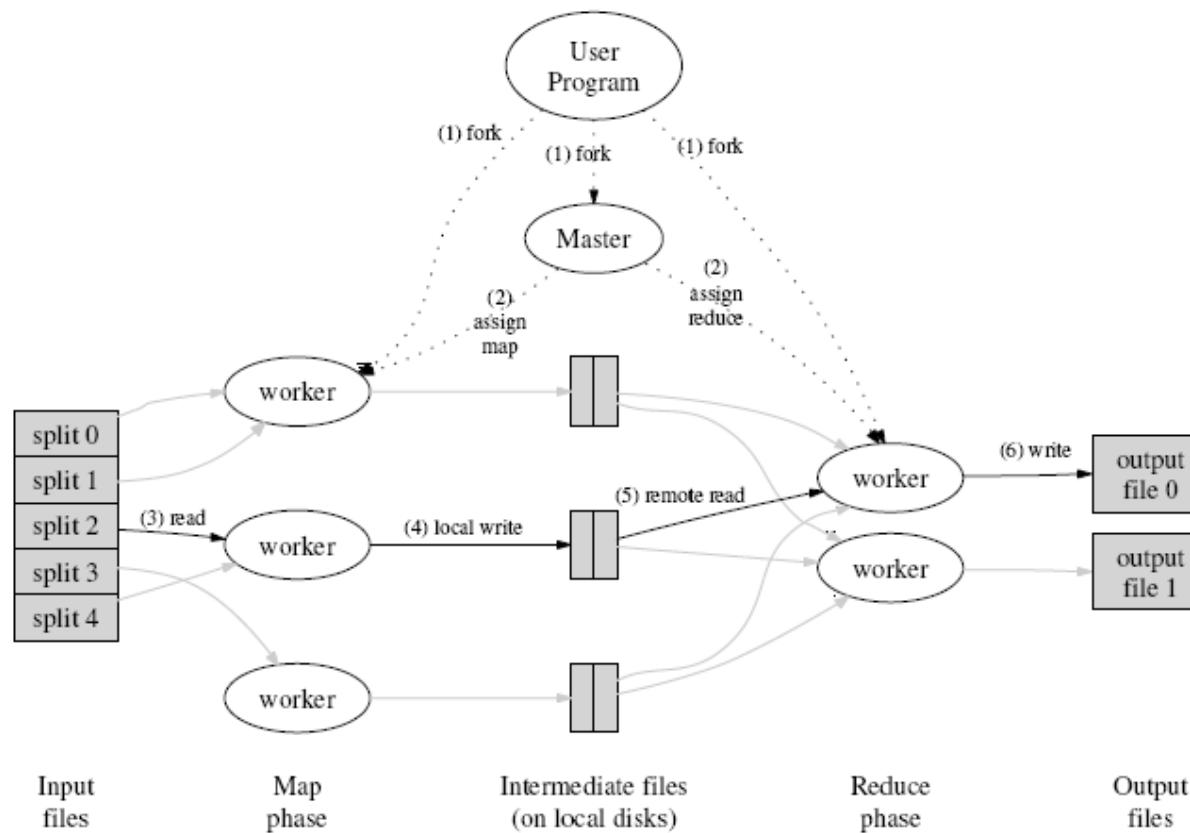
Parallel Wordcount as Map Reduce

- // Word count
- map(String key, String doc):
- foreach word in doc:
- sendToReduce(word, "1");

- reduce(String word, Iterator counts):
- wordcount = 0;
- for c in counts:
- wordcount += c;
- Output(word, wordcount);

MR dataflow

<http://code.google.com/edu/parallel/mapreduce-tutorial.html>



MR 1

- The MapReduce library in the user program first shards the input files into M pieces of typically 16 megabytes to 64 megabytes (MB) per piece.
- It then starts up many copies of the program on a cluster of machines.
- One of the copies of the program is special: the **master**.
- The rest are **workers** that are assigned work by the master.
- There are M **map** tasks and R **reduce** tasks to assign.
- The master picks idle workers and assigns each one a **map** task or a **reduce** task.

MR 2

- A worker who is assigned a **map** task reads the contents of the corresponding input shard.
- It parses key/value pairs out of the input data and passes each pair to the user-defined **Map** function.
- The intermediate key/value pairs produced by the **Map** function are buffered in memory.
- Periodically, the buffered pairs are written to local disk, partitioned into R regions by the partitioning function.
- The locations of these buffered pairs on the local disk are passed back to the master, who is responsible for forwarding these locations to the **reduce** workers.

MR 3

- When a **reduce** worker is notified by the master about these locations, it uses remote procedure calls to read the buffered data from the local disks of the **map** workers.
- When a **reduce** worker has read all intermediate data, it sorts it by the intermediate keys so that all occurrences of the same key are grouped together.
- If the amount of intermediate data is too large to fit in memory, an external sort is used.

MR 4

- The **reduce** worker iterates over the sorted intermediate data and for each unique intermediate key encountered, it passes the key and the corresponding set of intermediate values to the user's **Reduce** function. The output of the **Reduce** function is appended to a final output file for this **reduce** partition.

MR 5

- When all **map** tasks and **reduce** tasks have been completed, the master wakes up the user program. At this point, the MapReduce call in the user program returns back to the user code.
- After successful completion, the output of the MapReduce execution is available in the R output files.

MR uses 1 - **Distributed Grep**

- The **map** function emits a line if it matches a given pattern.
- The **reduce** function is an identity function that just copies the supplied intermediate data to the output.

MR uses 2- Count of URL Access Frequency

- : The **map** function processes logs of web page requests and outputs <URL, 1>.
- The **reduce** function adds together all values for the same URL and emits a <URL, total count> pair.

MR uses 3 - Reverse Web-Link Graph

- The **map** function outputs <target, source> pairs for each link to a target URL found in a page named "source".
- The **reduce** function concatenates the list of all source URLs associated with a given target URL and emits the pair: <target, list(source)>.

MR uses 4 - Term-Vector per Host

- A term vector summarizes the most important words that occur in a document or a set of documents as a list of <word, frequency> pairs.
- The **map** function emits a <hostname, term vector> pair for each input document (where the hostname is extracted from the URL of the document).
- The **reduce** function is passed all per-document term vectors for a given host.
- It adds these term vectors together, throwing away infrequent terms, and then emits a final <hostname, term vector> pair.

MR uses 5 - Inverted Index

- The **map** function parses each document, and emits a sequence of <word, document ID> pairs.
- The **reduce** function accepts all pairs for a given word, sorts the corresponding document IDs and emits a <word, list(document ID)> pair.
- The set of all output pairs forms a simple inverted index.
- It is easy to augment this computation to keep track of word positions.

Social Network

Dr. Mohsin Ahmed
moshahmed@gmail.com

Has social life improved?

- You have friends across the world you never met.
- You can keep in touch with relatives you never wrote letters to before.
- You can keep in touch with school friends, you never send email to.
- You can write your opinions on every matter.

What has improved

- You find out new things from new friends.
- You can find what your friends think of the latest movie.
- You can keep up to date on homeworks, ask questions in real time, interact with students anywhere in the world.

Social change

- Power of the democracy – organized protest with like minded people for freedom (examples).
- Tweet and Blog about pressing problems, inform others what is happening in realtime (examples).
- Government wants to know too. (examples).

Less time for outdoor activities



What has got worse

- You have less time for social events, because you are too busy updating your status.
- You can let the world know trivia
- Attention deficit disorder
- You talk strange LOL WFH

Collaborative information revolution

- You can listen to MIT, UCB class lectures.
 - On video.google.com, youtube.com
- Find information on wikipedia

Wikipedia

- So many scanned books – too much information
- Less is perfect
- 90% people make one word/line corrections.
- Open source, downloadable
- If you are not using wikipedia, you are not part of the information revolution.

Knol - Google's wikipedia

- Not enough free contributors
- Quickly outdated
- Critical mass needed
- Not free, mix of © and CC data.
- Private algorithms, moderation
- Not downloadable
- Commercial pages.
- Impossible for machine to detect wrong information – need users to correct it.
- Locked down articles for commercial reason.
- Author centric – each author controls the document.

Wikipedia vs Britannica

	Wikipedia	Britannica
Cost	Free, download	Online, payment
License	CC	©
Authors	Worldwide	Britain/US, western viewpoint.
Authority	Democratic	Corp Ed..
	less errors than BT.	Few errors
	Huge	Large
	Instant	Depends on editors

Faceless contributors

- Collaborate on wikipedia,
 - no authorship.
- GPL software
 - Anyone can add,
 - no one can remove.
- wiki books.
- Examples:
 - Linux, Python, GCC, G++
 - Gutenberg books and audio books.

Ideal social network

- Better communications.
- Keep your contacts up to date
- Lets you manage your data
 - Download your data
 - Delete your data and data about you.
- Maintain your personality
- Helps you lead a better social life.
- Find friends and information

Different kinds of social networks

- Email, Chat
- MySpace
- Blogger
- Buzz, +1
- Twitter
- Newsgroups
- Orkut
- Facebook

Private 1-1 Communication

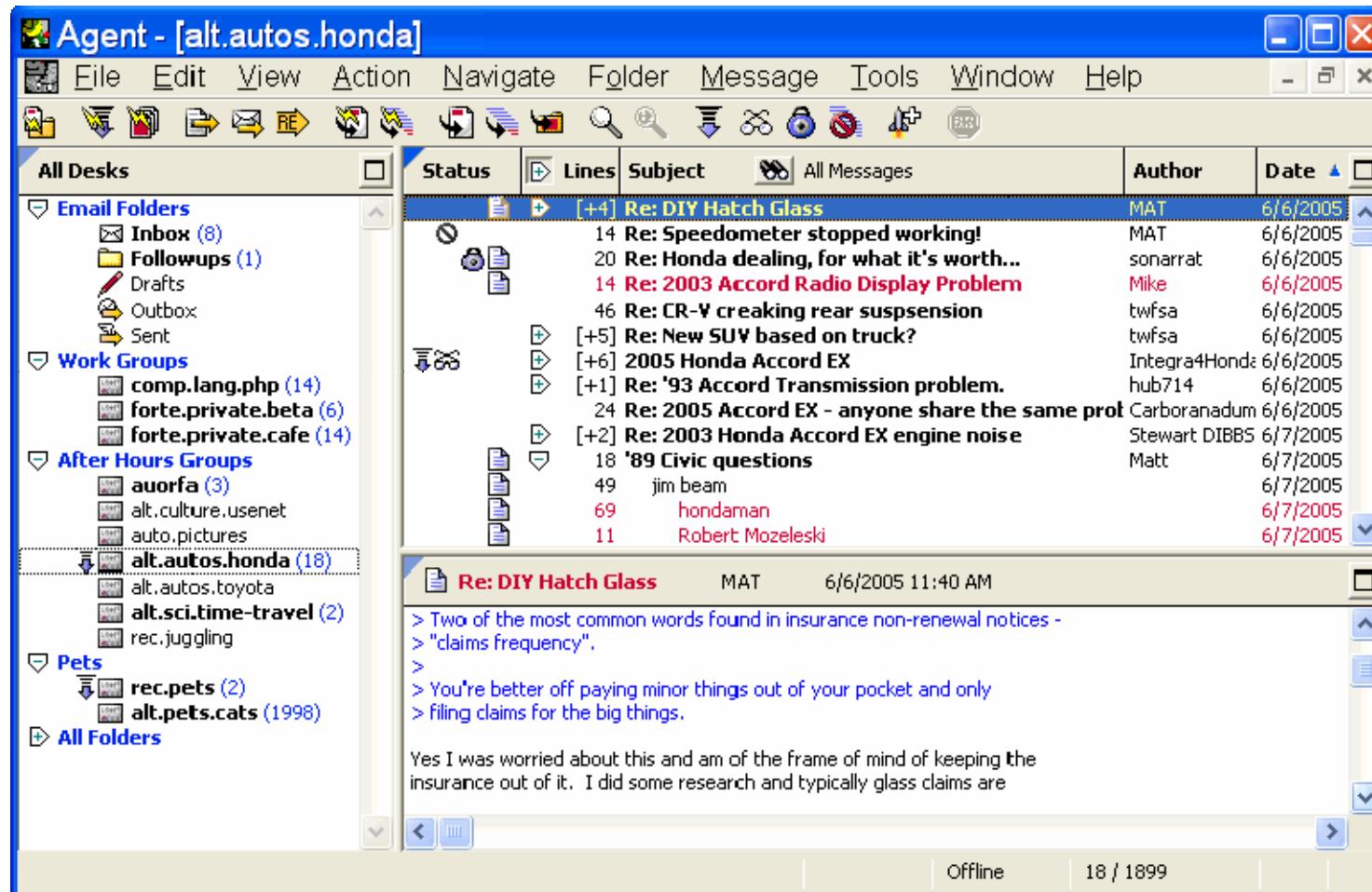
- Phone (good laws for privacy)
 - Only the IPS and Govt
 - Voice for authentication and privacy.
- Email
 - Still works well with SSL
 - Local disk and cloud
 - Accessible to others
 - Some spam.

Chat

- Problem of identity, bots
- Spam, no backend algorithms.
- No business model



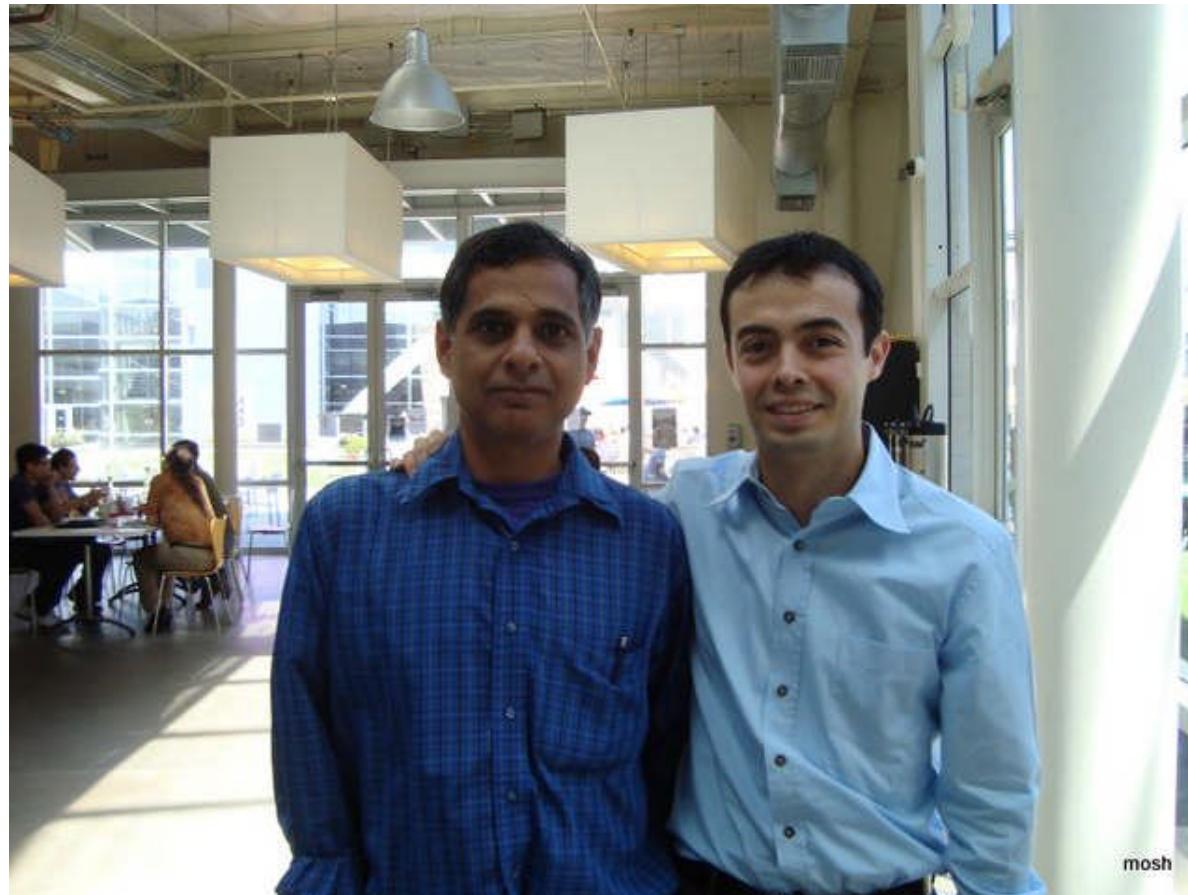
NNTP newsreader



Social communication 1-N

- Newsgroups
 - Centered around topics
 - N people, open or closed.

Orkut





Welcome, Eduardo

scraps photos fans messages
1 0 4 0

Your profile views: Since Feb '06: 361 , Last week: 69 , Yesterday: 34

Your recent visitors: , Dê (1.4%), Han-Wen Nienhuys, Super Homem ~doidão~, Pereira (Rodrigo Pereira Braga)

Today's fortune: You will be fortunate in everything

Tip: When you visit a profile, hover over the "more" link to see all your options. You can do much more than just write a testimonial!

A community has been transferred to you.
[Accept or deny this transfer](#)

upcoming birthdays

Tiago
August 24

my friends (46)

Lucas (79)	Bruno (355)	Manu (283)
Marcelo (239)	Torsten (110)	DECA (570)
Daniel (179)	Luciano (489)	Isabella (237)

[view all](#) [manage](#)

my communities (18)

Too Many Hobbies (3,079)	Eu tenho Rinite (195,799)	Super Master Comunidade XPTO (9)
Google Developer Day 2007 (187)	Spock's Beard (1)	Formula 1 (62,926)

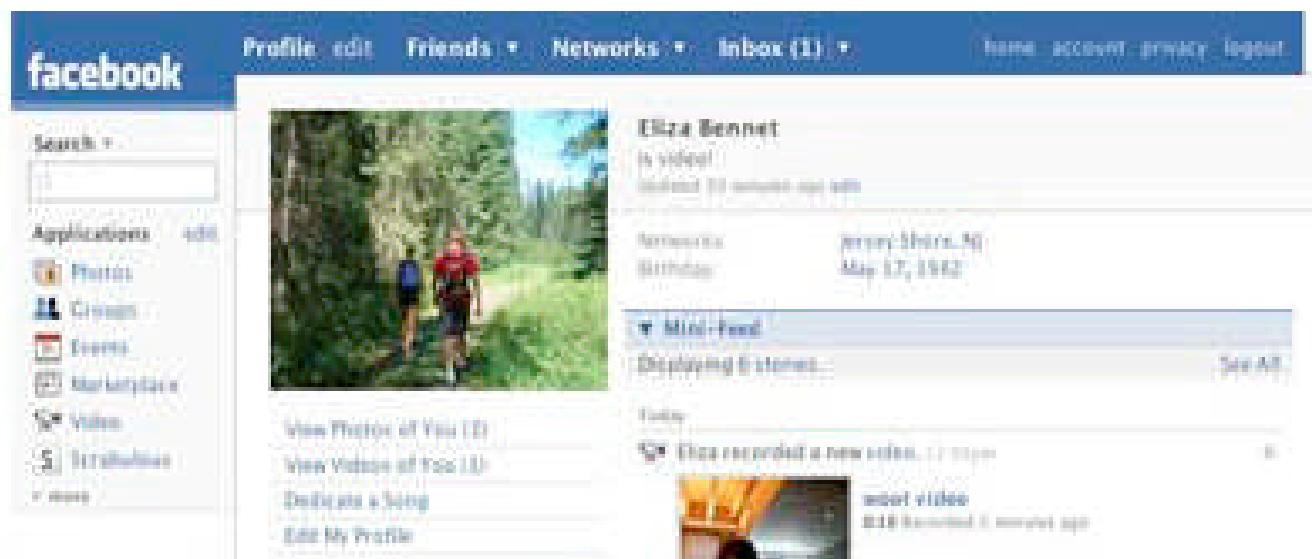
[view all](#) [manage](#)

Orkut

- Orkut
 - Centered around people
 - Personal web page + friends + communities
 - Subgroups of friends
 - Symmetric ($\text{friend}(x,y) = \text{friend}(y,x)$).
 - Post and photos are Private to groups

Facebook - FB

- Centered around wall posts, status updates
- Semi private, inconsistent.



FB metrics

The math of Facebook:

$$\frac{\text{\# of birthday wall posts}}{\text{total \# of wall posts}} = x$$

The higher x is, the more
of a loser you are.

[View Photos of Bill \(1\)](#)[View Videos of Bill](#)[Send Bill a Message](#)[Poke Bill](#)

Steve, I'm better than you and I have 40 billion reasons why.

Bill Gates just bought Azerbaijani!

[Wall](#)[Info](#)[Photos](#)[Boxes](#)[Notes](#)

RECENT ACTIONS

Bill and Ashton Kutcher are now friends. - Comment · Like

Bill is now a fan of Tool Academy and Project Runway. - Com



Steve Jobs Remember that OS you made that was awesome?
Yeah, neither do I. at 4:45pm March 26 - Comment · Like

Steve Wozniak liked this.



Bill Gates at 4:48pm March 26

I'll mention that to the 88.9% market share I have.
BTW, saw the new iPod shuffle. It looks like a tampon.

facebook

Mark Zuckerberg's Profile

[home](#) [search](#) [browse](#) [invite](#) [help](#) [logout](#)

Harvard

[View More Photos of Mark \(181\)](#)[View All Mark's Friends](#)[Send Mark a Message](#)[Polls Here!](#)[Add Mark as a Friend](#)[Report this Person](#)

Status

Mark isn't receiving Facebook [posts](#) right now.

Harvard Friends:

[160 Friends at Harvard](#)[See All](#)[Ebonee Radke](#)[Ryan Bayko, Dan Sichtler](#)

Information

Account Info

Name:

Mark Zuckerberg [\(add to friends\)](#)

Affiliation:

Harvard

Facebook:

Facebook

Last Update:

San Francisco, CA

August 14, 2006

Basic Info

Sex:

Male

Relationship Status:

In a Relationship

Residence:

Kirjland

Birthday:

May 14, 1984

Hometown:

Dobbs Ferry, NY

Contact Info

Email:

mzuckerberg@harvard.edu

Personal Info:

Activities:

lots of facebook, information flow, exponential growth, innovation, meditation, driving, writing, making things, social dynamics, domination

Favorite Music:

green day, Franz Ferdinand, weezer, fall out boy, my chemical romance

Favorite Books:

mostly biographies and textbooks

Favorite Quotes:

never run out of ammo.

About Me:

I make things that increase information flow between people.

Education Info

College:

Harvard
Psychology, Computer Science

High School:

Phillip Exeter Academy '02

Work Info

Company:

Facebook

Time Period:

2004 - Present

Description:

I like making things.

Tony Blair's Profile[View All Tony's Friends](#)[Send Tony a Message](#)[Poke Him!](#)[Add Tony as a Friend](#)[Report this Person](#)**▼ Status**1 update this week. [See All](#)Tony is the prime minister.
Updated last Friday**▼ London Friends**

33 friends in London.

Charles
Booth QCGordon
BrownBertie
Edelson**▼ Friends in Other Networks**[Networks with the most friends](#)**Tony Blair****London**[Share](#)

Sex: Male
Interested In: Women
Relationship Status: Married
Birthday: May 6, 1953
Hometown: Sedgefield, England
Political Views: Conservative

► Mini-Feed**▼ Information****Personal Info**

Activities: Running the country, sending the army to war
Interests: war
Favorite Books: War and Peace, Harry Potter

▼ Work**Work Info**

Employer: The Queen
Position: Prime Minister
Location: London, Iraq

▼ The Wall

Displaying the only 2 wall posts.

**God** wrote
at 1:24am on February 10th, 2007Mene
Mene
Tekel
Parsin
[Message](#)**George W Bush** (Washington DC) wrote
at 6:58pm on February 14th, 2007Sir spotted a terrorist threat, pants-down situation, recommend
brown alert[Wall-to-Wall - Write on George's Wall - Message](#)

Careful what you say on fb

 **Jane Doe**
If I had a gun I swear to God I would kill so many people! And then myself! :(
 17 hours ago · Share

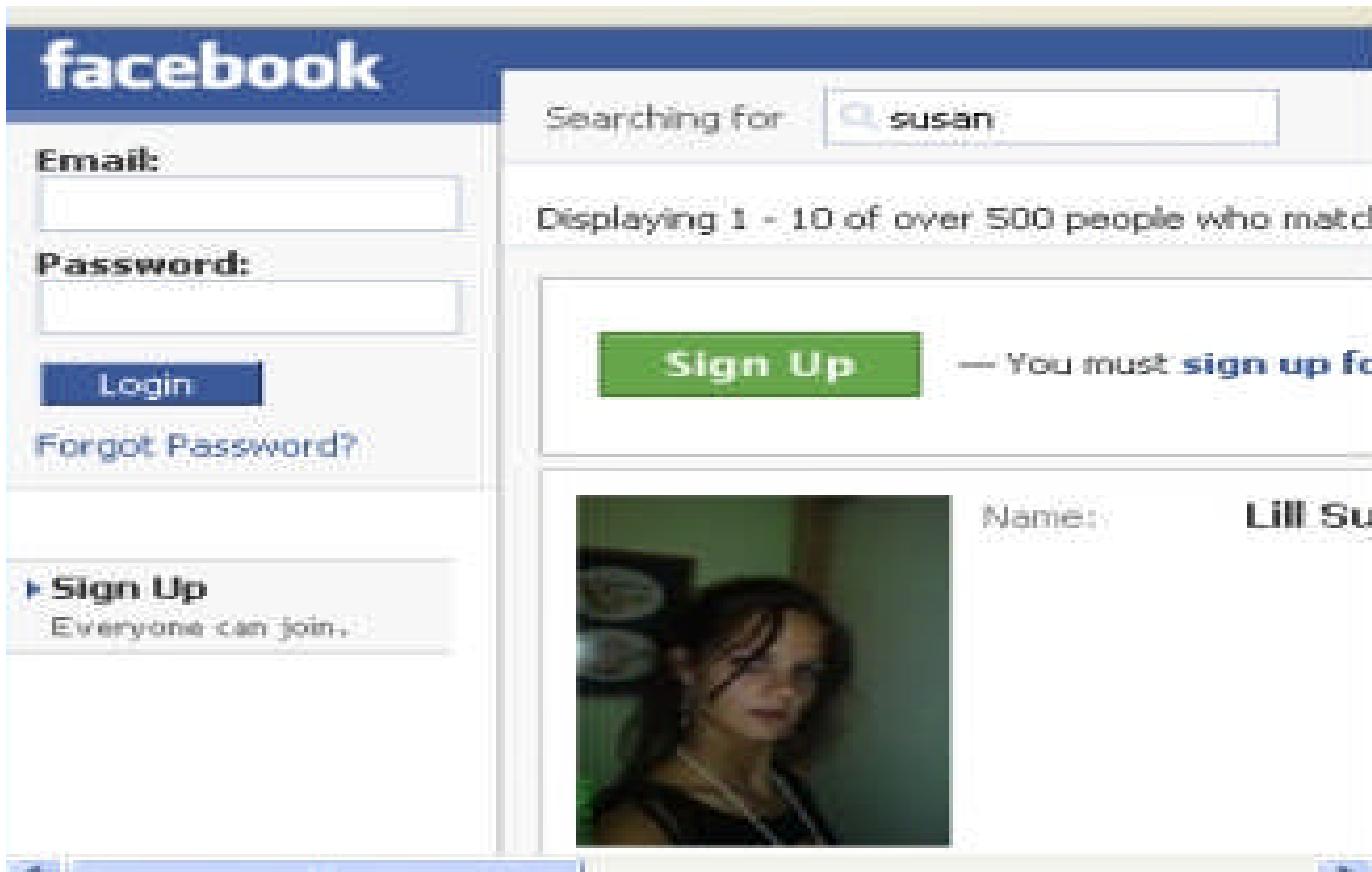
 Be the first of your friends to Like this.

 **Sarah Doe** Jane, people read this shit. Take this down.
12 hours ago · Like

 **Jane Doe** Don't worry Sarah. Wont shoot you. Just maybe
Mom and Jake and my boss.
3 hours ago · Like

 Write a comment

Privacy: FB profiles can be public visible to google search



FB businesses

facebook Home Profile Friends Settings Logout Search

 Add to my Page's Favorites View Update

<http://www.1800flowers.com>
<https://twitter.com/1800flowers>

Information

Founded: 1976

Fans: 6,072,836 fans See All


Shelly Hitchcock


Laurie Ann


Carolyn Dadd


Lori Davis Craig

1-800-FLOWERS.COM Just wanted to let you know that I received your Summer Dunes Bouquet as a result of last week's Twitter contest and the flowers are absolutely beautiful! Thank you so much!

4 hours ago · Report


1-800-FLOWERS.COM Thank you Kimberly! Very nice to hear that, and thank you so much for being such an awesome fan =)

2 hours ago · Report

Lori Davis Craig I am writing to let you know that I received my beautiful flowers from last week's contest. They are absolute gorgeous! Yippeem!!!!!! Thank you so very much. I appreciate this so much. I love this particular bouquet and plan on sending the same one to my mom on her b-day soon. :)

9 hours ago · Report


1-800-FLOWERS.COM Thank you Lori that's great to hear! Thanks for being such an awesome fan :)

9 hours ago · Report

1-800-FLOWERS.COM Hey fans, what type of exotic flowers do you like?

9 hours ago

Become a Fan

Wall Info Promotions Boxes Shop! Special Off... >

Filter

Create an Ad

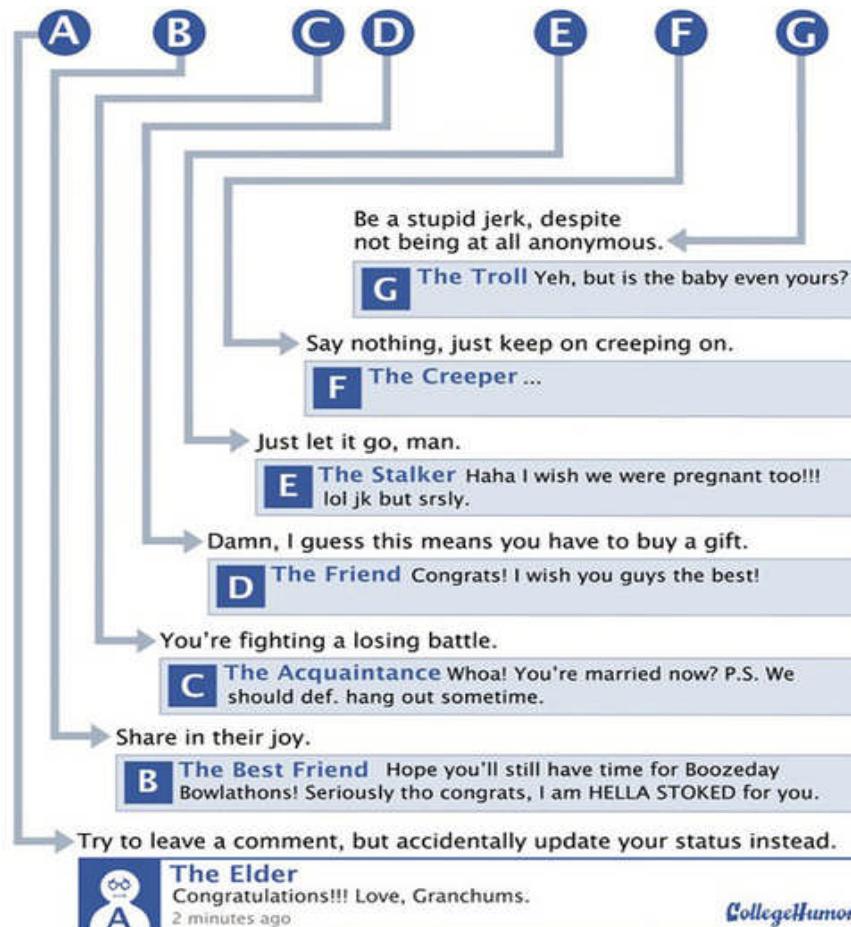
Connect With More Friends

 f

Share the Facebook experience with more of your friends. Use our simple tools to start connecting.

More Ads

FB comment classes



The Elder

Congratulations!!! Love, Granchums.

2 minutes ago

CollegeHumor

Buzz – dated one liners



Buzz and privacy

 [Henry Blodget](#) - 0 connected sites - 8 followers

Share what you're thinking. Post a picture, video, or other link here.

Buzz - [Following 16 people](#) - [Find people](#) - [Refresh](#)

 [Judd Bagley](#) - Buzz - Public
I'll try this, but warning: anybody who buzzes that they're excited about an upcoming Twilight book or movie gets immediately unfollowed.

[Michael Stanley](#) - Oh goodie. I confess I just LOVE those Twilight movies. I can hardly wait for the movie to come out when Kristen gets knocked up. 1:52 pm

[Judd Bagley](#) - Her name is Bella, Mike. 1:57 pm

[Michael Stanley](#) - You would know. I had to look it up. I think Kristen is actresses name. Please remove me from your follow list. My life is none of your business. 2:16 pm

 [Comment](#)  [Like](#)  [Email](#)  Judd is not available to chat

Buzz –moderated comments



[John Carney - Buzz - Public](#)

Buzzz. Is this thing on?

[Matthew Simmons](#) - Nope 12:10 pm

[Melissa Flashman](#) - exactly what i asked myself so, yes, i guess? 12:18 pm

[Matthew Simmons](#) - One thing to try: Connect buzz to your Twitter account. Hit the 'connected sites' link at the top of the buzz page. Then add Twitter. Your tweets will show up in your buzz feed. Useful for people who have gmail but don't use Twitter, and to allow people to comment about your tweets. 12:55 pm

[Vince Veneziani](#) - BUZUZZZZ 1:49 pm

[Comment](#) [Like](#) [Email](#) [Reply by chat to John](#)



Name or location

search

Home Find & Follow Settings Help Sign out

What are you doing?

140

update

Recent Replies Archive Everyone



amygeekgrrl I'm going to have some awesome content while I'm gone. my readers won't want me to come back. :oP less than 20 seconds ago from twirl



amygeekgrrl i have to say, with every guest blogger submission i receive (for posting while I'll be on vacation), the more excited i am about doing it. half a minute ago from twirl



JessicaKnows ate a bunch of jalapenos at my moms tonight and my stomach is now not speaking to me. but my throat is feeling better. go probiotics! 4 minutes ago from web



autismfamily just signed up for google reads, now trying to export amazon wishlist, not really working out, anyone use google reads? 5 minutes ago from web



tiddlytwinks Good evening all! Enjoying my Staycation, spent the day car shopping and drove home with my car of choice. Woot! :) 11 minutes ago from web

Hi,

your profile



momadvice

Currently

@AdInBabywearing Ha! You are a fast responder to my question too. Left you a comment!

Device Updates [add device](#)

phone
 web-only

Stats

Following	169
Followers	173
Favorites	0
Direct Messages	21
Updates	344

People [invite more](#)



Twitter

- One liners announcements to the world
- Asymmetric $\text{read}(X,Y) \neq \text{read}(Y,X)$
- Authors (if any) and followers (if any).
- Govt wants to know, (example?).
- To make bots and fake ids
- Realtime google search for tweets
- Information from the streets.

Twitter – real time broadcast

The screenshot shows a Twitter profile page with several annotations:

- A red circle highlights the "What are you doing?" status update field.
- A red circle highlights the timeline feed, specifically the tweet from **MutaraHobbs** about a course.
- A red circle highlights the sidebar stats section showing Following (23), Followers (12), Favorites (0), Direct Messages (0), and Updates (1).
- A red circle highlights the tweet from **johnreese** about the rise and fall of Twitter.
- A red arrow points from the sidebar stats section to the sidebar promotional text.
- Two gray callout boxes contain promotional text:
 - "You can broadcast **REAL TIME** to your network of "followers" what you're up to, example: just created a movie about my business, just updated my blog with a new article, just created a new promo etc. The opportunity is endless!"
 - "You can see what your mentor's up to **REAL TIME** and what news he's following! Why? So that I can be **FAST** apprised of what's going on in my industry – Internet Marketing, Internet Network Marketing"
- The sidebar promotional text reads:

Connect with people in your industry to share insights & knowledge. Connect with your prospects so that they can learn from you about your business and the value you can bring to them, real time. Be Social! Engage in the conversation! Get noticed!

What are you doing?

140

update

Recent Replies Archive Everyone



amygeekgrrl i'm going to have some awesome content while i'm gone.
my readers won't want me to come back. oP less than 20 seconds ago
from twirl



amygeekgrrl i have to say, with every guest blogger submission i
receive (for posting while i'll be on vacation), the more excited i am
about doing it. half a minute ago from twirl



JessicaKnows ate a bunch of jalapenos at my moms tonight and my
stomach is now not speaking to me. but my throat is feeling better. go
probiotics! 4 minutes ago from web



autismfamily just signed up for google reads; now trying to export
amazon wishlist, not really working out, anyone use google reads? 5
minutes ago from web



tiddlytwinks Good evening all! Enjoying my Staycation; spent the day
car shopping and drove home with my car of choice. Woot! :) 11
minutes ago from web



Public Communication 1-All

- MySpace – web pages with apps
- Blogger – dated web pages, N authors
 - One authors opinion on one topic.
- +1
- Like buttons, clicks with no audience.
- Stars – product ratings
- Reviews – product opinions

Blogs

- Blog one page at a time
- Dated, sequential
- Diary, travel, information opinions
- Photo blogs
- 1 to few authors.
- Lot of spam, fake ids, outdated.



Follow Share Report Abuse Next Blogs

KYM BREEZE ARTIST

OILS, ACRYLICS, PASTELS... SO LONG AS THEY HAVE COLOUR I WILL USE THEM. I CREATE PAINTINGS FROM WITHIN, BECAUSE I LOVE IT. EACH PAINTING HAS ITS OWN LIFE, AND THAT IS THE EMOTION AND FEELING THAT CREATED ITS BEGINNING. I LOVE LIFE IN ALL ITS COLOURS AND I LOVE PUTTING THOSE FEELINGS ONTO CANVAS. A WHITE CANVAS CAN BE CONFRONTING BUT SO CAN LIFE, WHY NOT PLAT, PLAT AT CREATING, AND SEE WHAT HAPPENS ...

KYM BREEZE ARTIST REDBUBBLE

[Buy at >> redbubble](#)

MONDAY, 17 MAY 2010

Talented Artist's of Texas embrace Kym Breeze's art class

On the weekend I had the wonderful opportunity of tutoring an amazing group of talented women artist's in Texas qld.

I went well prepared to be able to either teach traditional or in my mixed media style, not really knowing the background of the artist's. What I found when I arrived was a mixed group of artist's in their styles and experience, the thing they had in common was they love their art and they were there to learn, as well as being a group of friends.

I went with my plan and we created some mix media paintings,

ABOUT ME



KYM BREEZE
PARLING DOWNIE, QLD,
AUSTRALIA

I am an artist, creating images from life's experiences in all it's colours, I work in Acrylics, mixed media, oils and pastels. I am a colourist, keeping my colours pure and bright. This I feel is a reflection of my inner essence. I love life in all its colours, applying this to a white

Math dept announcements blog

The screenshot shows a web browser displaying a blog titled "Math dept announcements blog". The header features a collage of five images related to mathematics: a person working at a computer, a calculator, a swimmer, a clock, and a person writing on a chalkboard.

The main content area has two columns:

- ANNOUNCEMENT**

No more assignments for this week, just concentrate on your review for the exam. Have a good one! Good luck!

Posted by: [Renee](#), [January 18, 2011](#) at 10:00 AM
PM: 12 comments
- ANNOUNCEMENT**

No more assignments for now... prepare yourself for the preliminary exam.
God bless everyone!

On the right side, there is a sidebar for the "COLLEGE OF ARTS AND SCIENCES" featuring a circular logo with a green and yellow design, and a photo of a woman with dark hair. Text in the sidebar includes:

COLLEGE OF ARTS AND SCIENCES

other: [AP Calculus AB](#) [AP Calculus BC](#) [AP Physics](#)

COLLEGE OF ARTS AND SCIENCES

Amanda Averkamp-Greene
Chairwoman of the Dept. of Arts
in Teaching Mathematics
Please see my department website

Thumb created by [MarkosWeb.com](#)

LinkedIn

- Professional social network.
- Little spam, very few fake ids.
- Used by recruiters to find people
- Used to find jobs, consultants.
- Interest groups, announcements.
- Not as interesting data flow as FB.

IDEAL MOBILE SOCIAL NETWORKS

Bring bored friends together

- You are alone Saturday night and decide to search on SN for the latest movie.
- You get recommendation of a good movie
- It tells you 3 more of your friends are bored and in the same neighborhood.
- Its suggests to all 4 of you, watch the movie together.
- SN gets money for renting you the movie.

Restaurant recommendations

- You are in Labone district and hungry
- You look at your SN in the cell phone.
- The SN says, 5 of your contacts, and 20 of your friend's friends like Churcheese restaurant.
- It also tells you, many of your friends tried and liked the okra fufu.

Social network dinner

- Restaurant pays SN to advertise on it.
- Restaurant offer mobile discount coupons via SMS to SN, to give to new customers in the area.
- You are tempted, when you see a 10% food discount coupon.
- You ask your SN, show me other friends nearby.
- It may find no friends are nearby
- But 10 friends' friends are in churcheese right now (circle of friends).

Friends for dinner

- You phone pings all the friend's friend's cell phone in churcheese.
- If they are open and have time, they wave to you.
- You join them and have a meal recommended by your friends at 10% discount with new friends.
- Without SN: eat alone in some restaurant, while your best friend eats alone in the same restaurant 1 hour later.

Meeting new people

- Your phone talks to all the phones in the vicinity with the SN protocol.
- SN matches your likes/dislikes with those in the vicinity.
- It finds two people in the room like the same football team, and are studying same subjects as you.
- It shows “Do you want meet up new people who like ...”? To the matching profiles in your vicinity.

Blind introduction

- If both sides reply “yes”, both sides see the matching profiles.
- Both sides decide next decide to accept the introduction or not.

Why Mobile SN?

- Always powered on
- Can attract your attention with a ring
- Low cost
- Low maintenance
- Low user training
- Less virus and identity theft
- Real person owner
- Cell phone company can be banker/ broker.
- Voice, Camera
- Real time Location

Localized targeted advertising

- Mon evening, pizza shop has 10 extra pizzas and not enough customers. Should the pizza go waste?
- Advertise via SN to people in the vicinity.
- Offer: 3cd Pizzas to first 10 walkins 7-8pm, mention code p3 to get discounts.
- SN knows you like pizza.

Local targeted ads

- SN knows your location from the cell company,
- You see the ad, you rush to get the offer.
- Pizza shop is happy to get 10 customers in 1 hour and avoid wastage, it pays SN the 1cd/customer who mention SN ad.

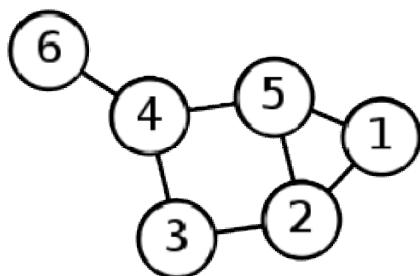
SOCIAL GRAPH

Social graph (from wikipedia)

- The **social graph** is a term coined by scientists working in the social areas of graph theory. It has been described as "the global mapping of everybody and how they're related"

Graph

- A drawing of a graph in which each person is represented by circled number called node and the friendship relationship is represented by a line called edge.



Social graph

- Nodes: people
 - Hundreds of millions of nodes
 - E.g. alice, bob, carol
 - Also topics, events, companies, photos,..
- Edges: are interaction
 - Trillions of edges
 - E.g. alice to carol: I bought a gucci bag.
 - Comments
 - “Like”

Social graph data mining

- Gmail's you got the wrong Bob
- Gmail's you forgot Bob
- Friend suggest in FB.

Did you forget Bob?

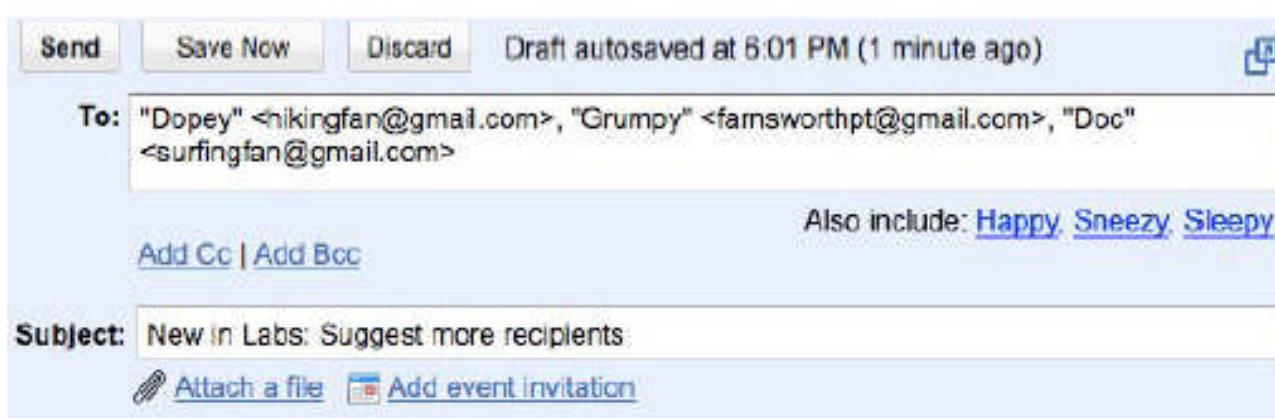


Figure 4: Example screenshot of the "Don't forget Bob!" lab in action. Given the user's initial seed contacts, "Dopey", "Grumpy", and "Doc", the Friend Suggest algorithm suggests additional recipients "Happy", "Sneezy", and "Sleepy".



Size matters

- Small 10k users, messaging systems
 - disorganized, noisy, spammy.
- Medium, 100k users, are niche forums
 - Special interest group.
- Large, 1M users, are social networks
 - All your friends, contacts are there.
 - Machine learning works.
 - Network effect – everyone wants to join.

People are just names

- Multiple nodes per people, when they can't be identified uniquely.
- E.g. Bob Smith on linkedin
- And Bob Smith on facebook.com
- Are they the same person?
- Need probabilities:
 - 70% likely to be same.
 - 90% same if same email id.
 - 100% certain if confirmed by same emailid

People have multiple personalities

- Alice the shopper
 - Alice the family lady
 - Alice the manager
 - Alice the student
-
- When Alice is interacting, she is assuming one of the above *personas*.

Are names people?

- Spammers, scammers, phishers create millions of fake accounts per day.
- Real people create fake accounts for
 - Privacy.
 - Forget passwords.
 - Multiple identities.
 - Paid marketers across the globe.

Detecting real people with Ids

- Require email
- Captcha
- Need cell phone number
- SMS verification
- Credit card
- Third party, chain of trust
 - e.g. school certifies for all student email ids.

Real people without Ids

- Global membership without Ids.
- Semantic machine learning
 - Messages sent
 - Messages that got replies
 - Ip address
 - Browser headers
 - Time spent typing message
 - Web habits
 - Original content generated

People database

- 200M users
 - 10 photos, media = 300k
 - 100 message = 100k
 - 10,000 clicks = 100k
- $200M \times .5M = 100 M M = 100T$
- Minimum 100 machines just to store the data.

Who is my friend?

- People I know?
- People I interact with?
- People I want to interact with?
- People who know me?
- People I met?
- People I don't want to meet?
- People who share my interests?
- People I work with

The world is not flat

Alice's contacts are:

Office coworkers

Classmates from school

Classmates from college

Friends from a party she attended

Church members

Family members

Daily Interactions

- Assume 1M people are online daily
- 100 events/person
- Each event is 1k
- Size of data / day = 1G
- Activity 1200 events/second,
- Need 20 http servers.

Processing daily report

- Assume cpu can process 1000 events/second
- Time required = $100 \text{ M}/1000 = 100\text{k}$ seconds = 1.2 day.
- 100 machines in parallel can process this data in 20 minutes.

Interactions

- Edges between people
- Comments
- Conversations
- What happens someone deletes a comments from middle of a conversation?

Real life interactions vs web

- Facial expression is lost
- Nuance of speech lost
- Were you joking?
- Sarcasm lost?
- Offline, without feedback.
- read out of context in different cultures.
- Web never forgets, one bad word and it stays forever in the server logs.

Processing interactions in real time

- Alice writes on Bob's wall: xyz
 - Is 'xyz' in bad word list?
 - Is 'xyz' spam?
 - Is 'xyz' commercial?
 - Is Alice allowed to write to Bob?
 - Has Bob blocked Alice?

Privacy

- What is privacy?
 - Any information?
 - Friends?
 - Birthday?
 - Wall – what I did yesterday?
 - My cell phone number?
 - My email id?
 - What I bought from Amazon?
 - What clothes I wear?
 - My photos?

Private talk?

- Alice writes on Bob's wall: xyz
- Who else can see 'xyz'?
 - Friends of Bob?
 - Friends of Alice?
 - Everyone on the network?
 - Google for alice bob xyx

Privacy and Profit

- How to Profit from social network?
- Fees, like real clubs.
- Advertising for businesses
- Targeted ads
- Recommendations
- Big brother wants to know too.

Elephants don't forget, nor does the internet.

- In real life, people forget with time.
- Information gets diminished as it goes between people. No one really knows what “Charles told Diana” in real life, it is mostly via grape-wine.
- On the other hand internet remembers everything word forever.

Processing a billion interactions

- You can use 1000 machines running the map-reduce package in parallel.
- Split interaction graph data
- data → mapper → reducer → report.
- Report is the recommendations, ratings, trends, suggestions, spam, memes, friend suggestions,..

What is the Social Graph API?

- The Social Graph API makes information about the public connections between people on the web more easily available. Developers can query this public information to offer their users dramatically streamlined "add friends" functionality and other useful features. (from google).

Social protocols

- XFN (XHTML Friends Network)
- FOAF "The Friend of a Friend (FOAF) project is creating a Web of machine-readable pages describing people, the links between them and the things they create and do.
- Google API
- Facebook API
- Java, Python, PHP, Ruby, C#, JavaScript

Securing the social network data

- Clients can talk to SN servers with
 - Ajax, RPC over http
 - JSON data.
 - XML APIs
- Use RSA (public key, private key).
- Verify certificates of servers
- Sign and encrypt data sent.
- Use SSL.

Securing the transactions

- Communication can be secured with Public key (e.g. using RSA)
- Secure with photo and voice recognizers.
- Store keeper can check your photo with the SN signature, when you pay using the SN bank on the mobile phone.
- Loss of privacy.

SN as the digital bank

- With phone SIM id and digital signatures
- Phone can exchange digital money
- Payments over SN.
- SN acts as the bank, verifying all signatures and keeping accounts.
- See facebook payments

SN as the digital store

- Like Amazon, SN knows what your likes are, what you want, what your friends bought.
- It uses that data to generate product recommendations, in its store.
- You buy with one click from the SN store.
- E.g. Apple can setup a ipod store on SN.

SN mall

- Groups of stores on SN, group by categories, ratings, products, and social graph.
- You browse the SN mall and buy as in a real store.
- You see other people in the same store and can chat with them.
- You can see what your friends say about the store.

Shopping in a physical mall with a cell phone

- Point your cell phone to a bar QR code.
- Look up your friends recommendations on SN about the product.
- See competitive offers
- Read about the product on Wikipedia
- Find discount coupons
- Buy product and pay with cell phone