Distributed Computation

Secret Agents and Jealous Amazons

Linda Pagli

Distributed Computation

How several autonomous computers can communicate and cooperate to solve problems sometimes too big to be solved by a single one.

K. Kelly, *We are the Web*

Wired, August 2005

- *In the years roughly coincidental with the Netscape IPO, humans began animating inert objects with tiny slivers of intelligence, connecting them into a global field, and linking their own minds into a single thing. This will be recognized as the largest, most complex, and most surprising event on the planet.*

Entity : computer, processor, process, agent, switch…

Each entity is able to

- execute algorithms with great accuracy and very high speed, independently of any other entity.
- transmit messanges

Connected as a network

- are able to agree on a common strategy
- cooperate to the solution of a problem
- work together as a whole.

Computational Power

The intrinsic computational power, that is the ability of solving problems is the same for a single or a set of entities.

What does change is the size of solvable problems and here is the hope, Example: Search Engines

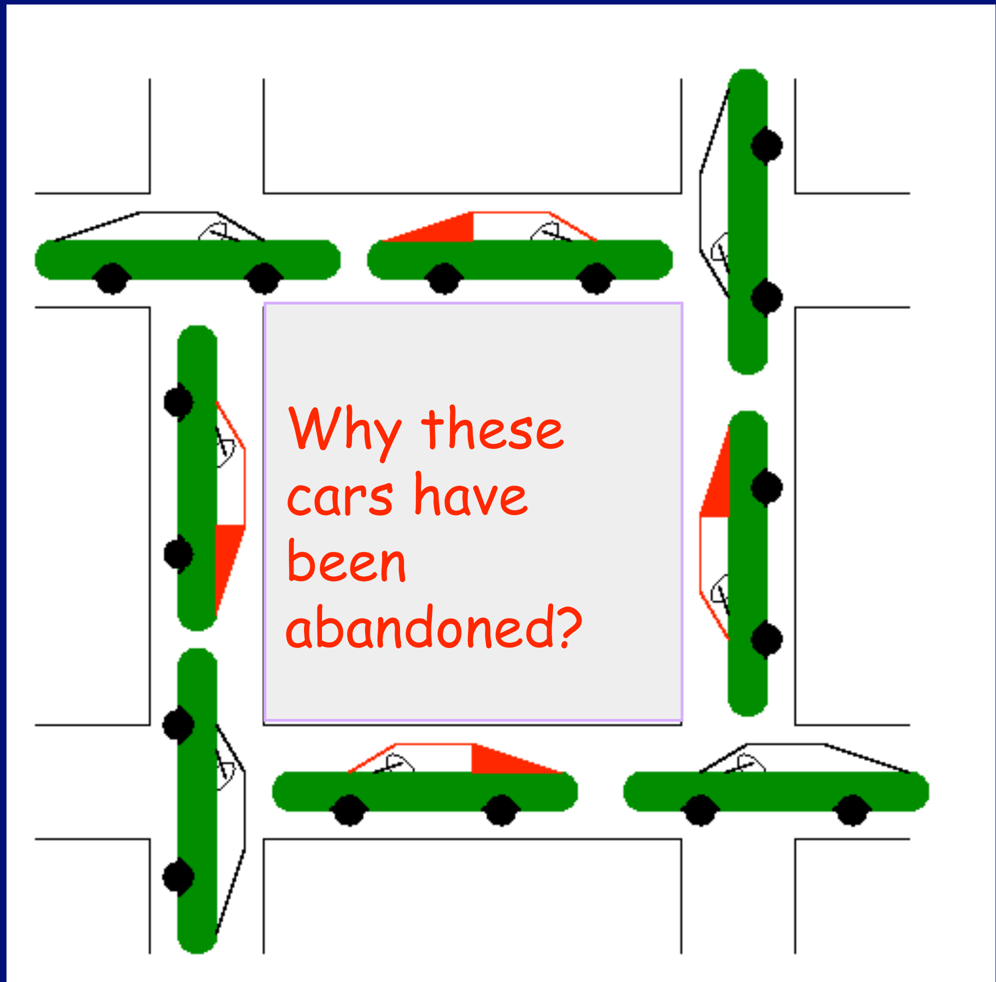
Does cooperative working help?

New problems will arise, sometimes difficult to solve, sometimes unexpected....

When several connected entities work together is quite difficult to understand what's going on…as an animated conversation if everybody talks at the same time.

Deadlock Problems

resources used in exclusive mode



Other difficulties

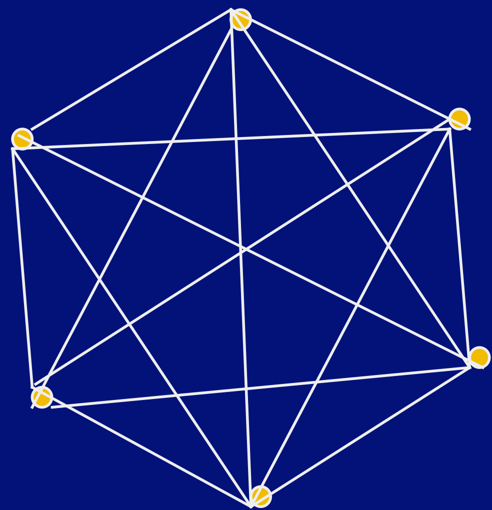
- Description
- Partial knowledge of the network
- Understanding a distribuited algorithm
- The computation depends on the communication delays
- A single algorithm can correspond to several executions.
- Evaluation parameters are different from standard ones

Internet: Popolarity of a protocol

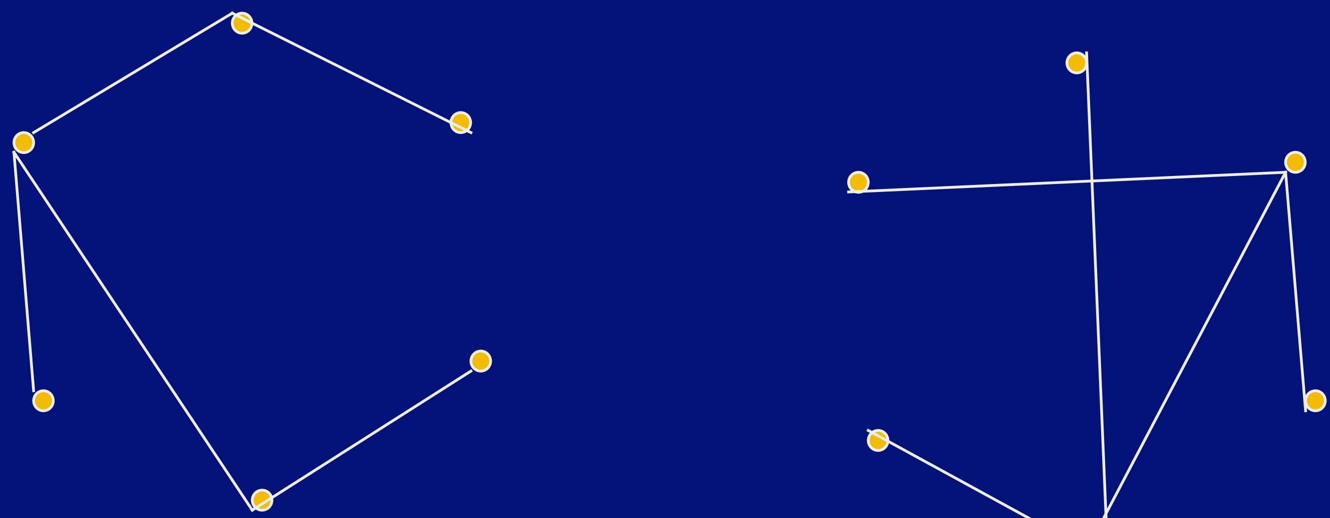
secrets agents

The agents must execute the task of selecting the subset of safe lines

Only 5 among the 15 can be now used.



5 lines are sufficient to keep the complete connection among agents



Safe lines selection protocol

- Initial state: all lines can be used.
- Final state: five lines must be selected in order that all agents are connected.
- One agent is elected leader by the central command.
- The leader starts the protocol.

Safe lines protocol

Algorithmo for agent x

The agent can be idle or leader. At the end each agent is done.

If x is leader:

call all the neighbors disregarding any call received;

Become done;

If x is idle:

upon the reception of the call from y

add the line [x, y] to the set of safe lines;

call all the other agents, but y;

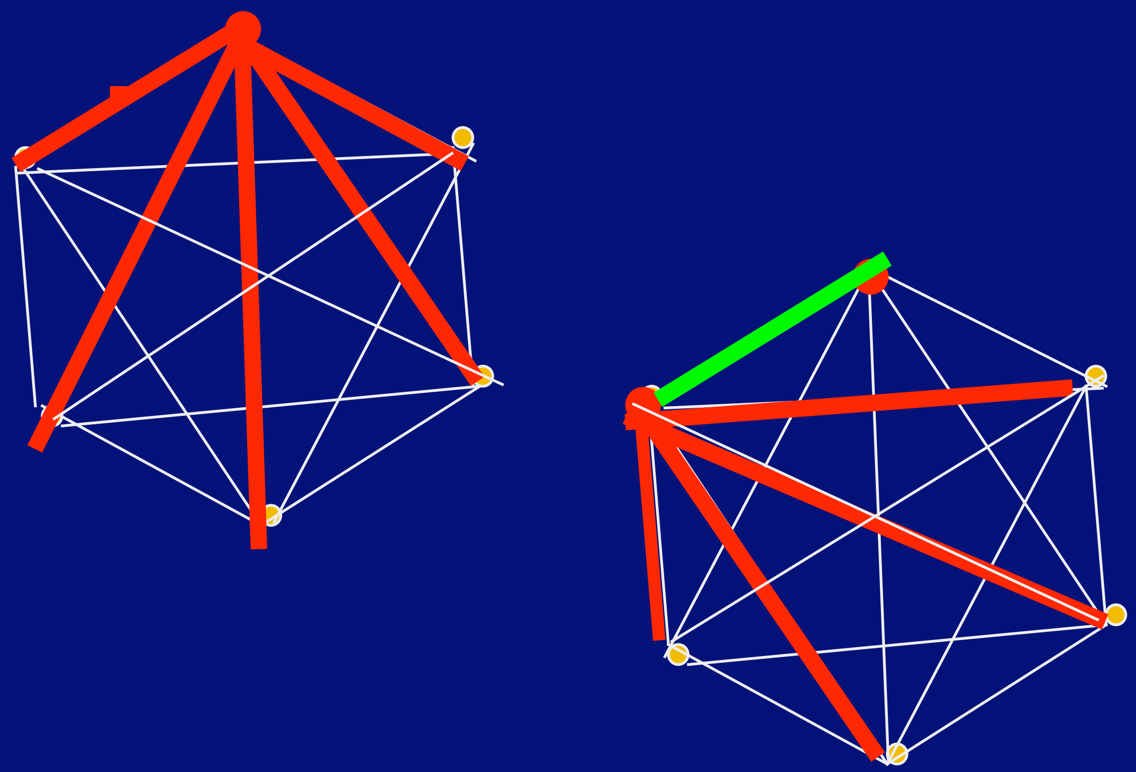
for any neighbor ***z*** accepting the call:

add the line {***x***, ***z***} to the set of safe lines;

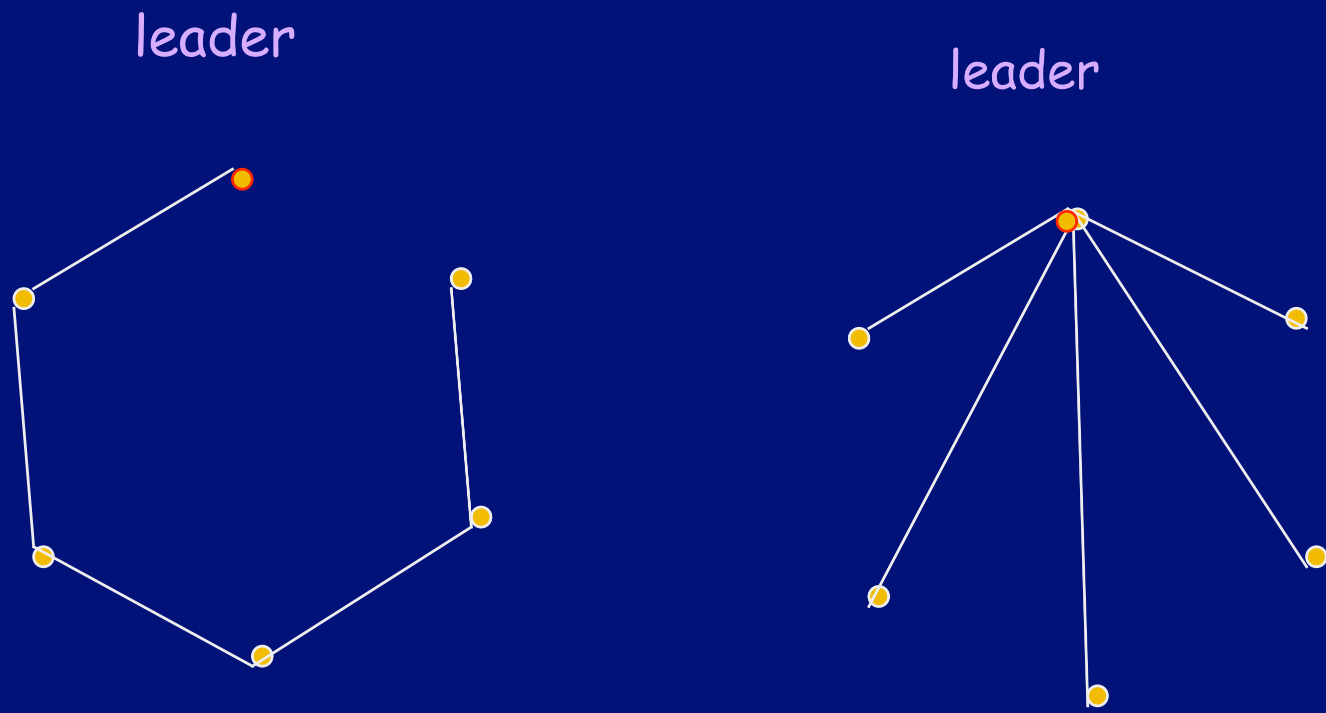
become done;

If x is done…

Safe lines Selection protocol



2 limit executions



The protocol work even if agents does not know the form and size of the network

The 6 agents must decode an encripted message 1000 pages long

1. The secret key is communicated by the central command

The text can be subdivided in 6 parts

The agents take 1/6 of the time taken by a single agent. (+ the time needed for coordination and to recombine the parts)

The 6 agents must decode an encripted message 1000 pages long

2. the key is a s long as a page.
   For each page, but the first, the key is the prevoius page, before encryption.

   The central command knows the key of the first page, but not of the following ones.

The agents cannot divide the pages among them now!
They could work all togheter at the first page, then to the following….

- Coperative solution can be sometimes necessary,
- Sometimes convenient,
- or unuseful,
- or also harmful!

(if the coordination operations cost more than the possible gain in time).

Search Engines

Complex distributed system exploiting intensive parallelism

Search Engines

Before answering to the queries Google :

- Search for the information in the web
    - crowler
    - spyder
- Store this information in its own enormous memories.
- Build the indices with the keywords

Search Engines

computer clusters geographically located
Wherever in the world.

A cluster is composed by thousand of
computers (PC's) and contains various
copies of the web.

Google™

Web Images Groups News Froogle Local new Scholar **more »**

Miserable Failure

Advanced Search
Preferences
Language Tools

Google Search | I'm Feeling Lucky

Advertising Programs - Business Solutions - About Google

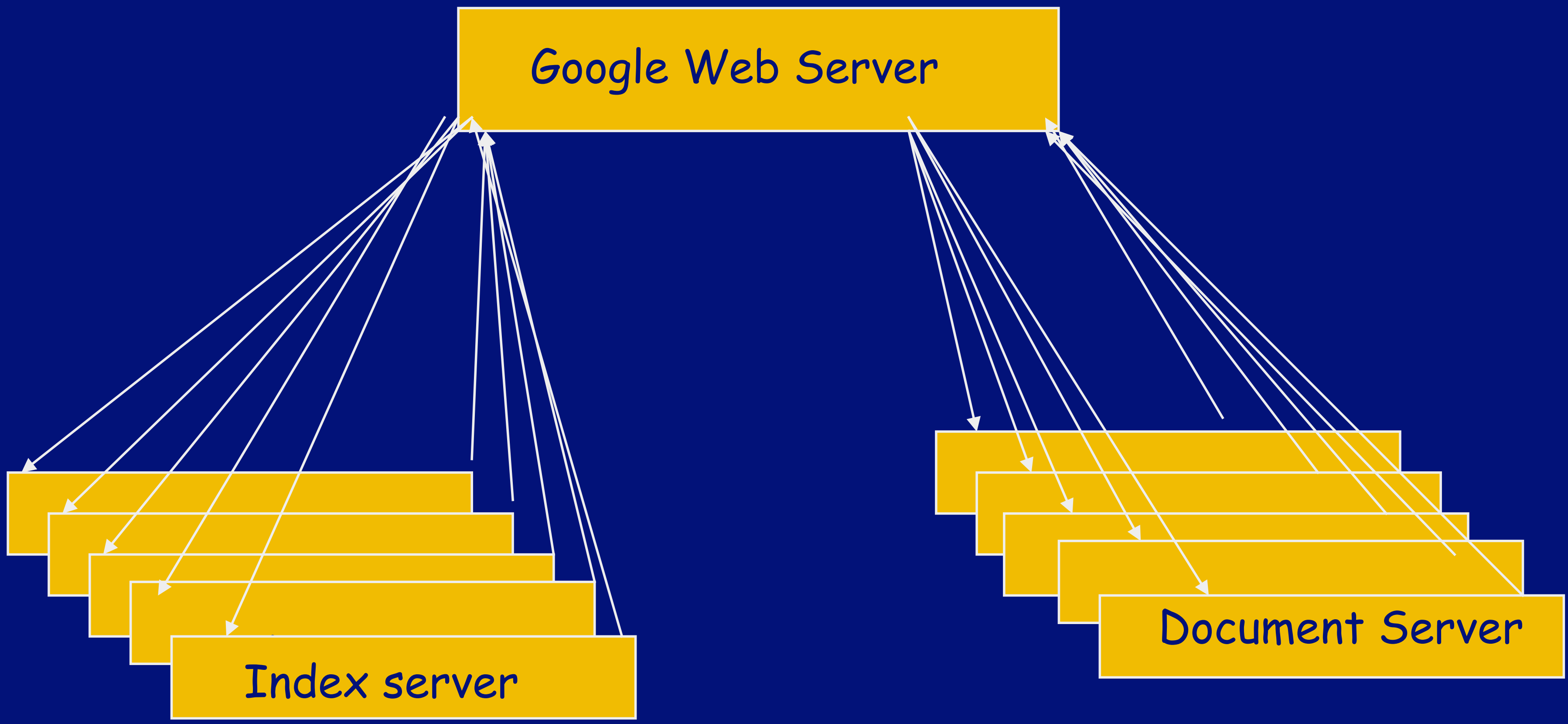
Make Google Your Homepage!

Search Engines

- a query to Google in average:
  - reads hundreds of Megabytes of data
  - executes billions of CPU cycles
- Google handles millions of queries/sec
- Our query is delivered to geographically closest, or to the less busy cluster.

Search Engines

- Into the cluster the query is assigned to a specific machine (GWS) that will prepare the answer.



- Search: 2 phases
  1. The keywords are searched in the index servers

  The index maps each keyword to a matching list of relevant documents (hit list). The result is the intersection of the hit lists of each Keyword, Then list rank is computed by the index servers.

  Each index is divided into pieces, each one storing a randomly chosen subset of the full index.
  A pull of machines serves requests for each piece of the index, (index servers) in parallel.

Search: 2 phases

2. The actual documents are fetched form disks containing a copy of the entire Web.

As in phase 1 the document servers search in parallel as before: the request is issued in parallel on the several parts of the archive and several machines search in parallel on a single part

Web Images Groups News Froogle Local new Scholar more »

Misarable Failure

Advanced Search
Preferences
Language Tools

Google Search | I'm Feeling Lucky

Advertising Programs - Business Solutions - About Google

Make Google Your Homepage!


MISERABLE FAILURE
NASCAR

Why parallel searches are so efficient?

- Paralellelism is used extensively: all searches are independent, the coordination work is fixed and can be decided a priori, and the work to combine partial results is minimal.
- Linear Speed-up
- Requests are delivered in parallel to different clusters and handled in parallel on pieces of the index and of the complete archive.

1. The two generals problem
   (coordinated attack)

2. The jealous amazons problem

Apparently:

Problem 1 easy

problem 2 impossible.....

But....

The two generals problem

A coordinated attack of A and B will defeat C and the fortress will be conquered

A solitary attack will result a disastrous failure

Generale C Army

General A army

General B army

If the communication lines are safe . . . The problem is solved in obvious way

The leader general (A) sends a messanger to agree on the time to attack.

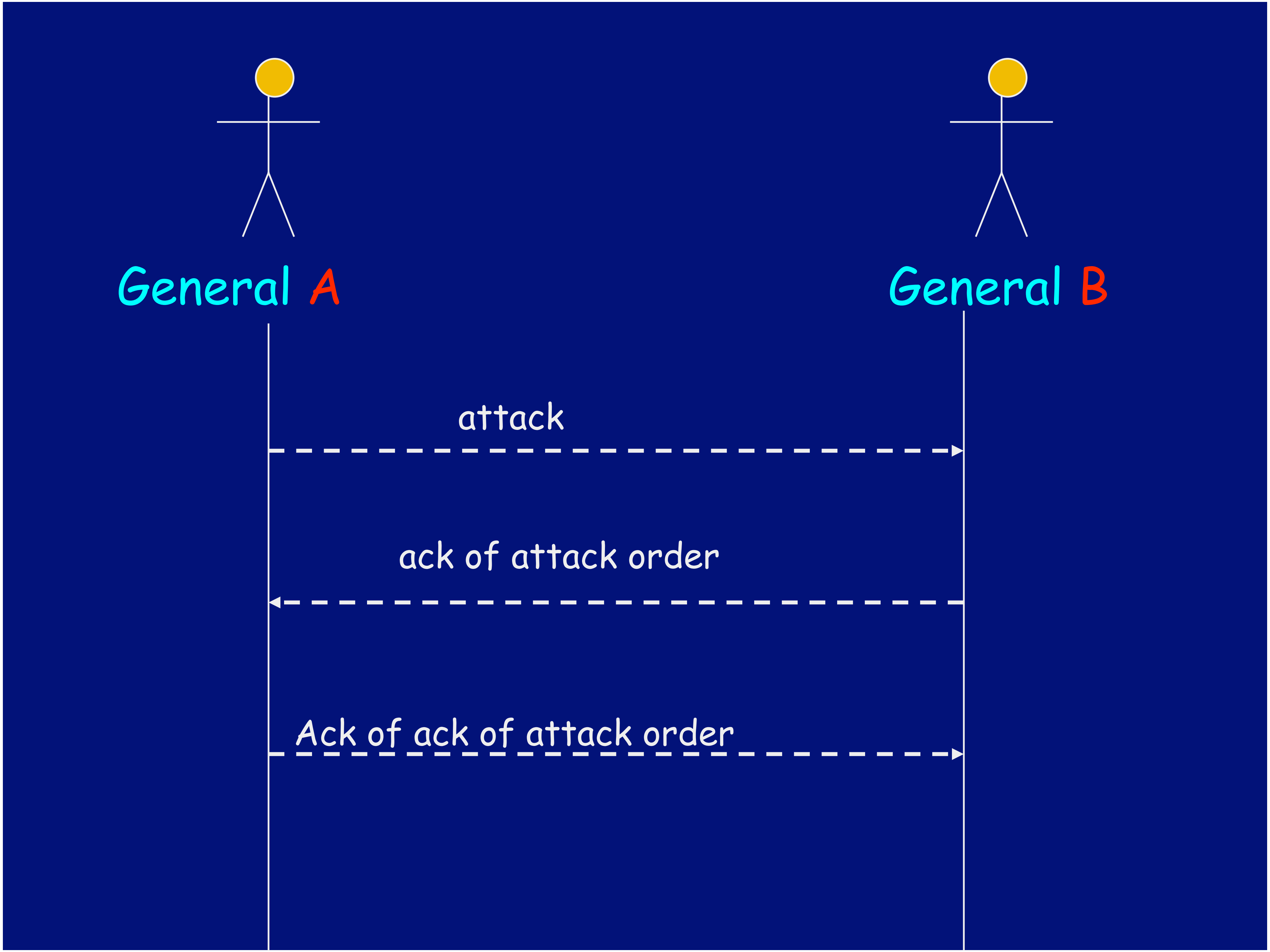
But what happens if the generals fear that a messanger can be hambushed by the enemy?

The two generals must coordinate the attack to C.

Suppose A is the leader general:

1. A sends a msg: "Let's do it tomorrow at 9 o'clock". Before ordering the attack A must make certain that B has receveid the order, so waits for the answer of B.
2. B, if receives the message, replays: " I received the messge and agree to attack at 9 o'clock" Now in turn, B wants to make certain that A received his answer before attacking, so A must confirm once received the message from B.

and so on..


General A
General B
attack
ack of attack order
Ack of ack of attack order

What happens?

Since before attacking each general wants to receive from the other the acknowledgment of the last message sent, the exchange will go on forever.

The attack is impossible!

Why?

- Suppose N messages (forward and backward) are sufficient
- Message (N-1) –th is certified by message N
- the last sender does not know if his message has been received
- He will not attack.

Real world

ATM terminal
in Rome
A

Bank in Tokio
B

A guy in Roma wants to withdraw a big amount using A from his bank account at Japanese bank B.

A and B must organize a "coordinated attack"

The operation must be organized at the same time in Rome and Tokio:

- A pays and B debit or
- A does not pay and B does not debit

Otherwise someone could be in real trouble!

Unsolvable without safe communication lines

Jealous Amazons problem

A. Modigliani "The amazon"

In the country of Amazons (as elsewhere) when somebody has an unfaithful partner, everybody knows it except for herself

Original problem: "Forthy unfaithful wives" Gamos , Stern. Puzzle math, The Viking Press, NY 1958

But enough is enough! One day, the queen of the Amazons proclaims a firm resolution;

"In this country there are unfaithful Amazons. For the safe of social order all further visits abroad are suspended until morality is fully restored in the Kingdom. Is not permitted to communicate on this issue in any way, however, as soon as one of you is certain that her partner has been unfaithful, you shall shoot her at the midnigth of that precise day."

The Amazons went back to their activities.
No one ventured to speak or mention the problem.

There were 17 unfaithful Amazons,
At the midnight of the 17-th day
17 shoots exploded in the night.

Why?

If there are k unfaithful Amzons, there are k homicids at the midnight of the k-th day, starting from the queen speach.

K = 1 :

Only one unfaithful Amazon
Her partner does not know about any case of infidelity. The queen said that there are. The only possible unfaithful partner is her own, at the midnight of the first day shoots.

K = 2:

Two unfaithful Amazons. Their partners know about one infidelity only .

At the midnight of the first day there is no sound of shoots, meaning that there is more than one unfaithful Amazons.

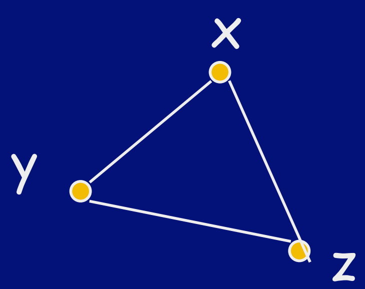
Their partners then understand and at the midnight of the second day shoot.

General case (inductively):

A_1, ..., A_k are partners od unfaithful Amazons, but know only about k-1 case of infidelity. Since at the midnight of the k-1 the day no shoots are heard.....

In a distributed system sometimes is necessary that all entities work with common knowledge that is they know the same things

3 processors



- Each proc. has an ID
- x and y know the ID of z
- z knows its own ID
- (x, y, z) know the ID dof z, but z may not know that x and y know it

In a distributed, it can be uncertainty between what an entity knows and what the other entities know about it.

What the solution of the jaelous Amazons puzzle suggests?

The common knowledge eliminating the ambiguity on the value of an important information can be reached in a fixed number of steps, with no communication, but with the examination of particular events that may occur.

Bibliography

Barroso, Dean Hoelzle: Web Search for a Planet. The Google Cluster Architecture. IEEE Micro 2003.

Fabrizio Luccio, Linda Pagli: The network: A mathematical and Algorthmic of the Internet. Chapmann&Hall, 2010

Nicola Santoro: Design and Analysis of Distributed Algortihms. Whiley 2007