Open Science Grid

2011 OSG Summer School

UCSD

Single sign-on in Open Science Grid

by Igor Sfiligoi
University of California San Diego

Summary of past lessons

- HTC is maximizing CPU use over long periods
  - And getting lots of computation done
- DHTC is HTC over many sites
  - Using an overlay system hides most of the change
- Grid and Cloud resources are really very similar
  - Especially is you use them from inside an overlay system

Single sign-on in OSG

Introducing the

Open Science Grid

Scientific Grids

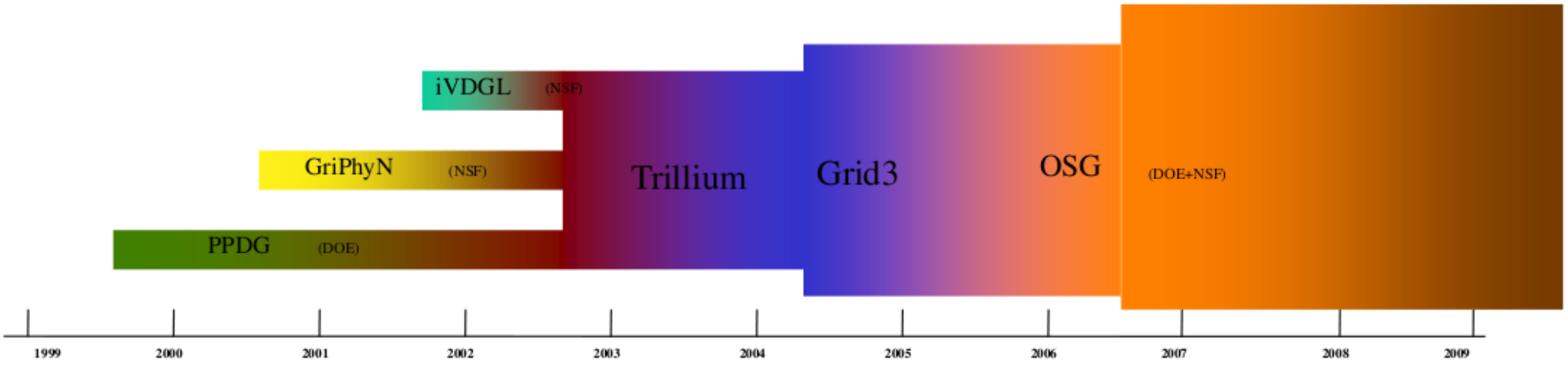
- Just a reminder
  - Widely distributed (continent wide)
  - Many participants O(1k+)
    - Just moderate trust (no way everybody knows everybody else)
  - Many local HTC technologies
    - Joining sites may have existing infrastructure

The Open Science Grid

- US+ wide scientific Grid
  - With partner Grids from around the world (e.g. EGI)
- Sponsored jointly by NSF and DOE

Some history

- Built on experience dating to the previous Millennium



Who is part of OSG

- Heavily HEP dominated (LHC, Tevatron)
  - They **need it** to do their science
- Followed by other physics communities (e.g. LIGO, RHIC)
- But also a healthy mix of other sciences
  - Biology and chemistry related fields
  - Math and CS related fields
  - Engineering related fields

OSG in numbers

- No sites: ~50
  - Ranging in size from ~10 cores to ~7k cores
  - Small sites may have a 10% of a grad student running the system
  - Large ones may have several dedicated sysadmins
- No of users: >10k
  - Not all active at the same time
  - But most of them have used OSG at least once

Scale problem

Is it really my problem?

But I will be using an overlay system.
I will only see my local submit system.

Well, someone still needs to send the pilots.
You think it works **by magic**?

But why should I care?

Why again did you come here???

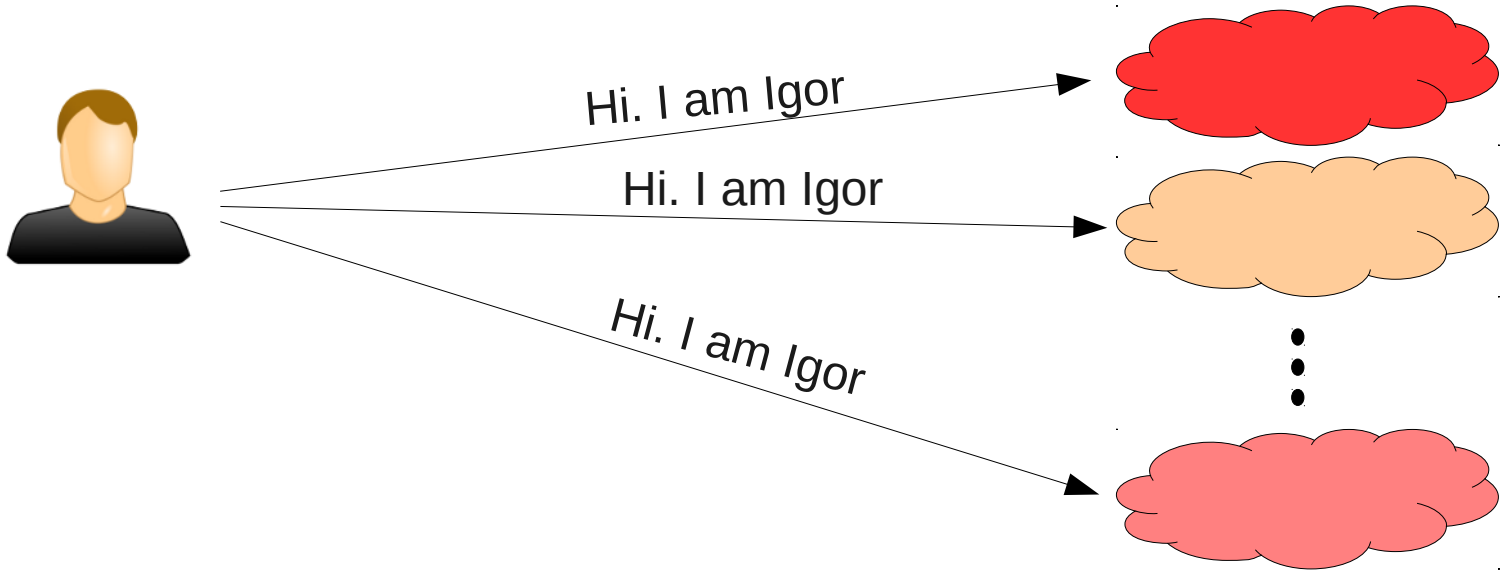
You will need it for storage.

Single sign-on in OSG

Single sign-on
using
X.509 PKI

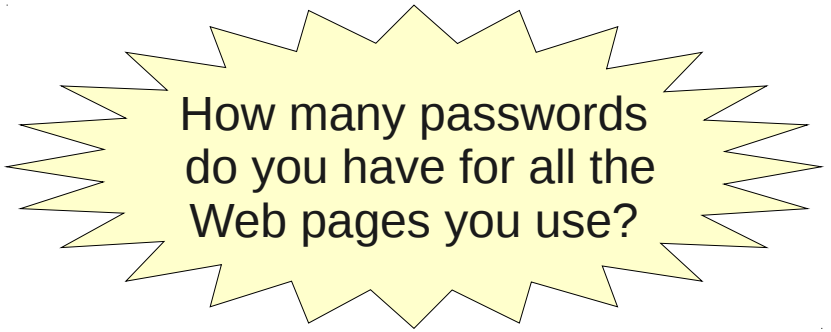
Single sign-on

- The idea is simple
  - The user should use the same mechanism to submit jobs to all O(100) sites



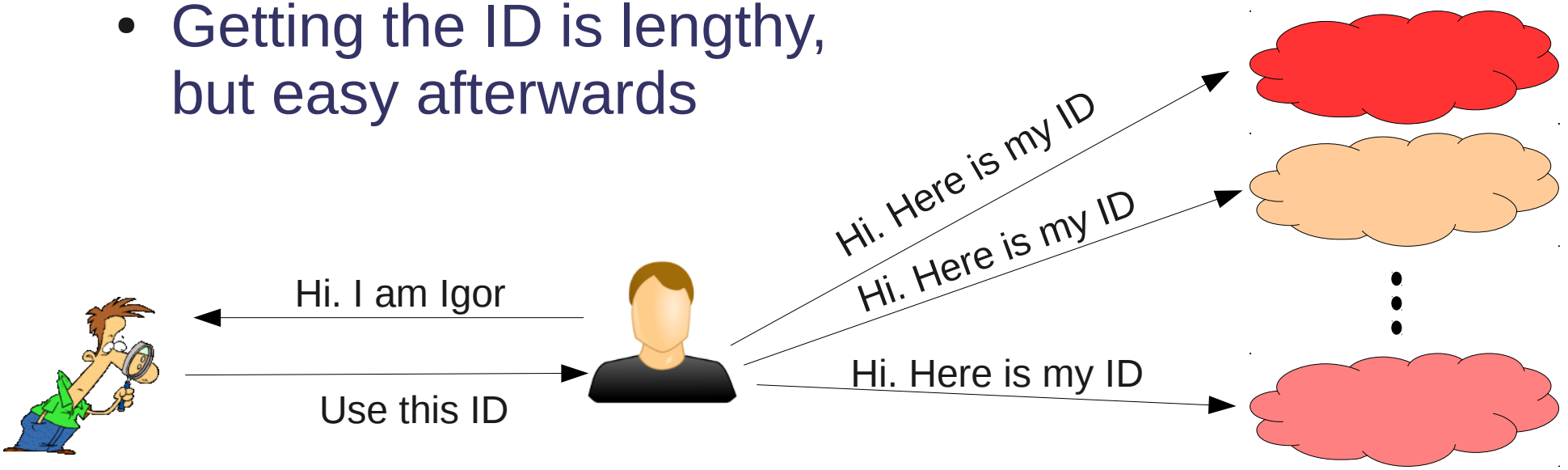
Passwords a non-starter

- We all know username/password is the preferred authentication mechanism
  - Almost everybody use it!
- But not a good solution for distributed systems
  - Uses a **shared secret** between **the user and the service provider**
  - And secrets stay secret only if few entities know it
    - **Sharing passwords between sites a bad idea!**



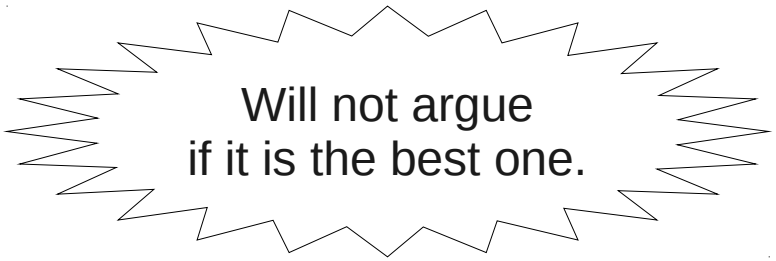
Adding an intermediary

- A better approach is to introduce a highly trusted intermediary
- Have been used in real life for ages
  - e.g. States as issuers of IDs
  - Getting the ID is lengthy, but easy afterwards



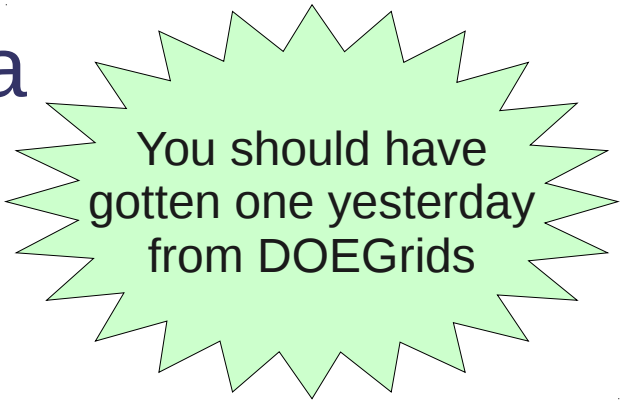
Technical implementations

- Many technical solutions
  - x.509 PKI
  - Kerberos
  - OpenID
  - many more...
- All based on the same basic principle
  - Each has strengths and weaknesses
  - OSG standardized on x.509



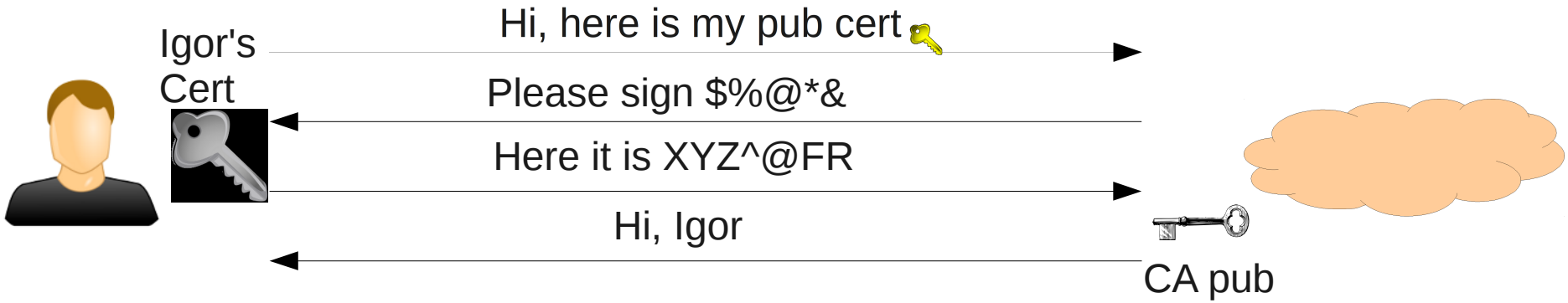
x.509 PKI

- Based on public key cryptography
  - A user has a (public,private) key pair
    - one encrypts, the other decrypts
    - similarly, one signs, the other verifies
- The highly trusted entity is called a **Certification Authority** (CA)
  - The user is given a **certificate**
  - Cert. has user name in it
  - Cert. also contains the (pub,priv) key pair
  - Cert. is signed by the CA private key



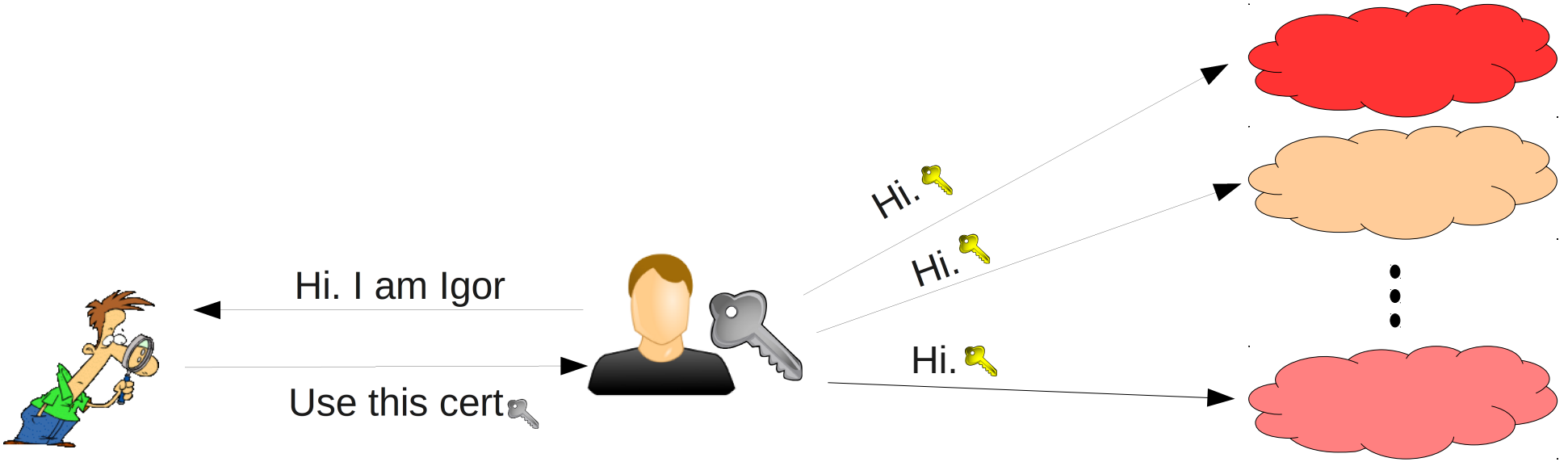
x.509 authentication

- Sites have CA public key pre-installed
- User authenticates by signing a site provided string and providing the public part of the cert



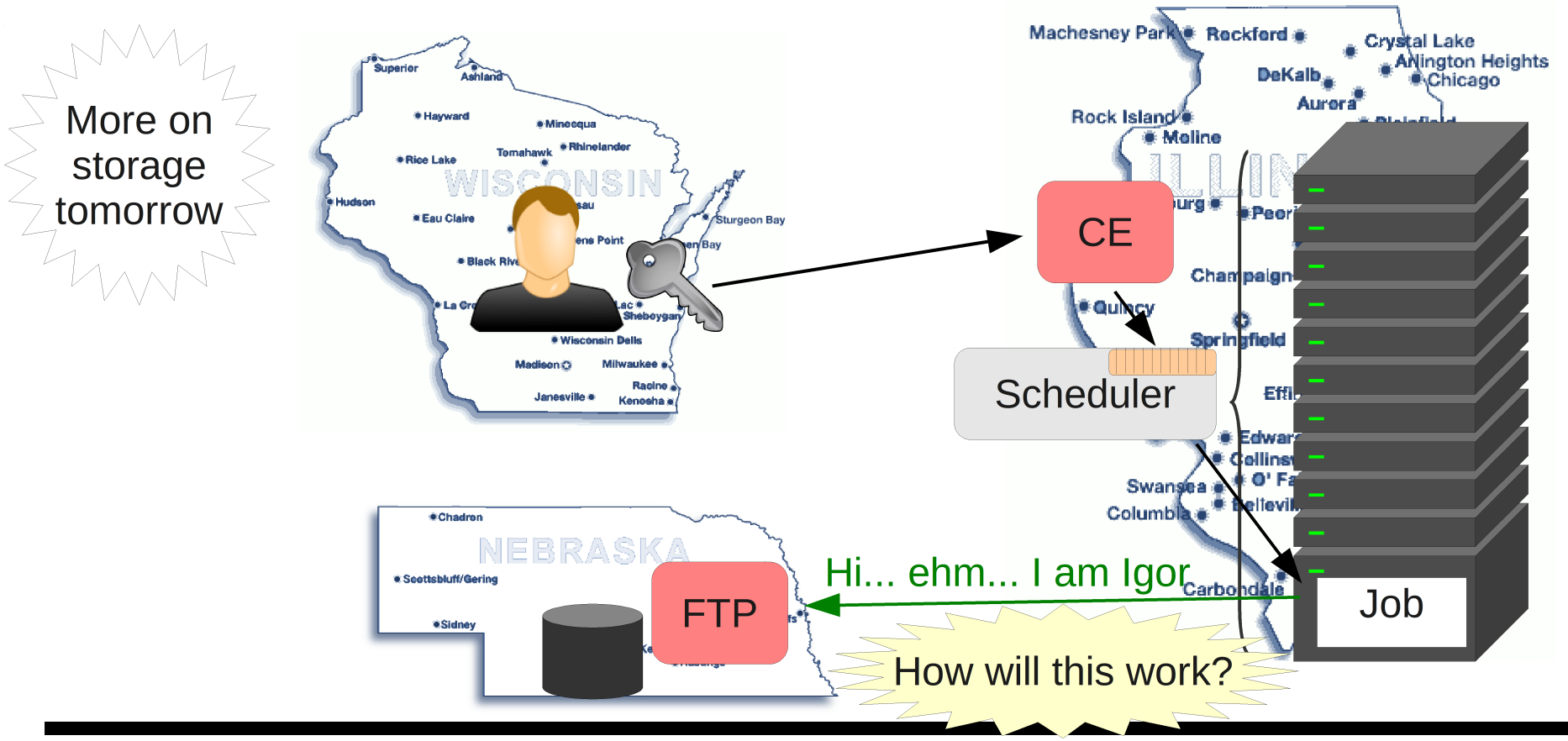
x.509 as single sign-on

- Use the same cert for all the sites



Impersonation

- Sometimes your jobs need to impersonate you
  - For example to access an FTP server

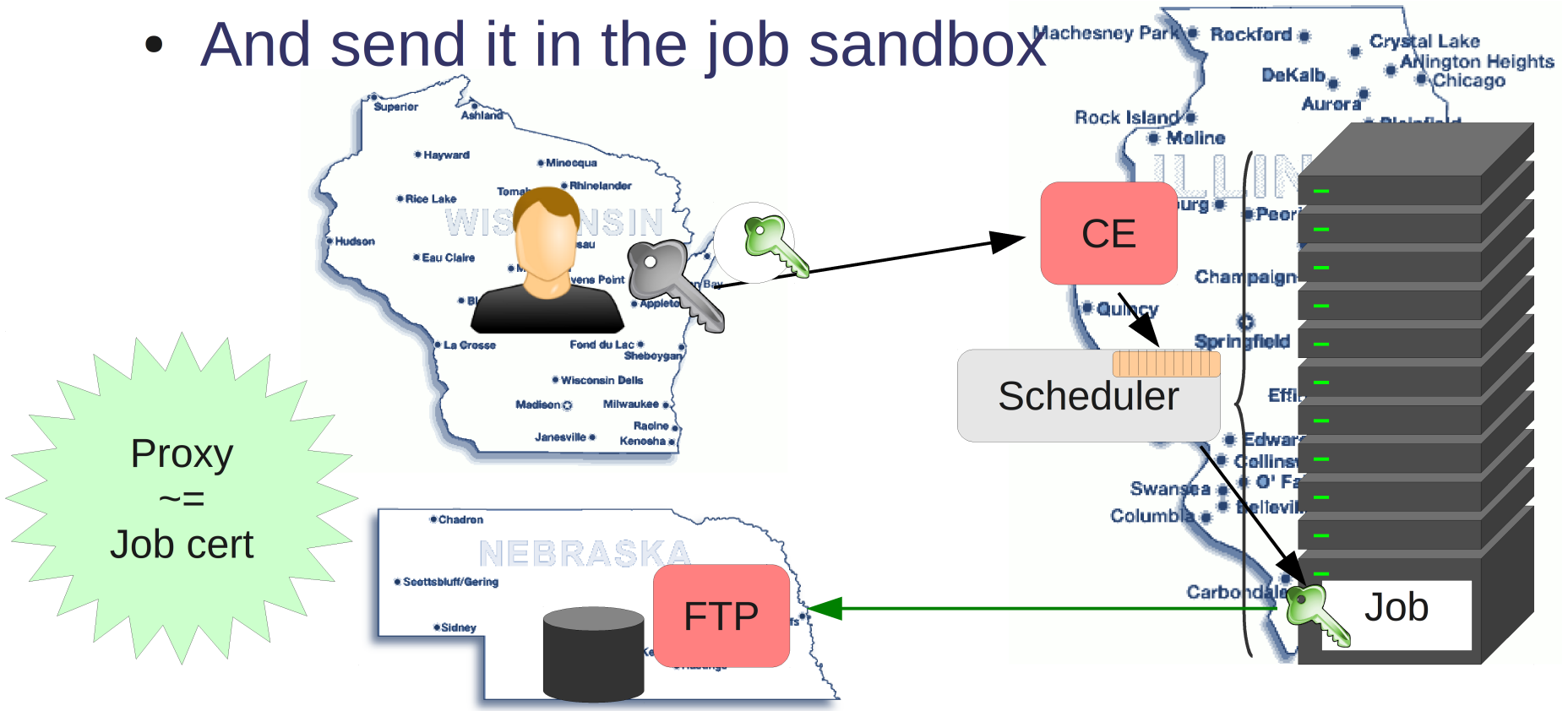


Impersonation problem

- The problem is that the job **does not have** your private key
  - And it should not, if it is to remain “private” (remember, secrets only stay secret if few entities know about it)
  - So **it cannot impersonate** you
- We have similar problems in real life, too
  - e.g. attorney representing you in court
  - Nobody will buy it that he is you, yet he can speak on your behalf

Proxy delegation

- The **job is** indeed **not you**
- Create a certificate for the job
  - And send it in the job sandbox



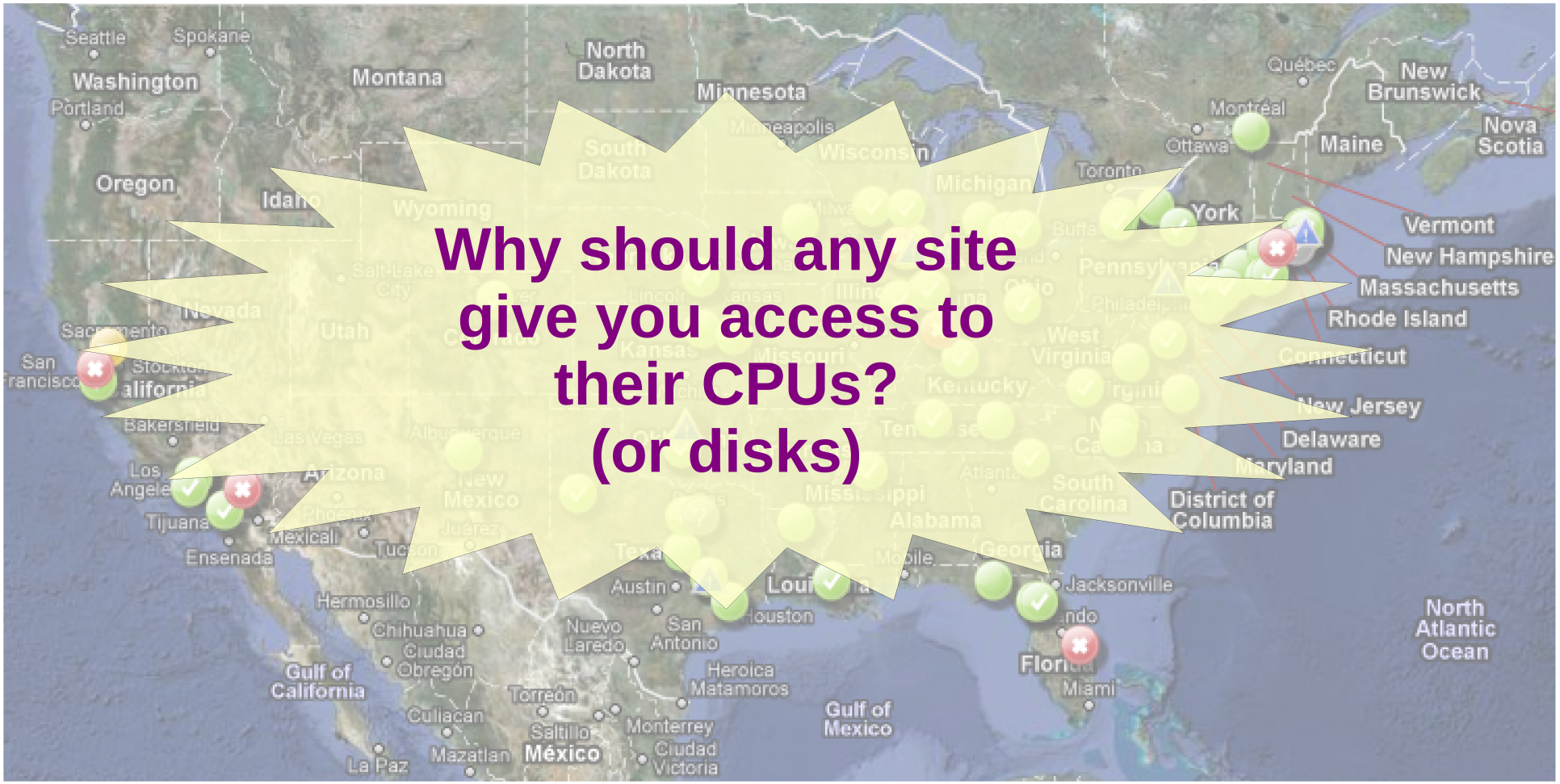
Proxy risks

- You are sending private keys with the job
  - You risk that they be **stolen**
- To mitigate this risk, the proxy lifetime **should be very limited**
  - Say, a few hours
  - But long enough to do the needed work

Delegation and overlays

- Not really any different
  - Still need priv keys on the remote CPU
- But you may exploit
  the additional control you have
  - e.g. Condor will automatically
    shorten proxy lifetime
    and re-delegate as needed

Are we done yet?



Do I really care?

But I will be using an overlay system.
All I need is contacting my local sysadmin.

YOU AGAIN!!!
Someone still needs to send the pilots!

Just kidding

And don't forget about storage.

Single sign-on in OSG

Tiered authorization

or

Introduction to

Virtual Organizations

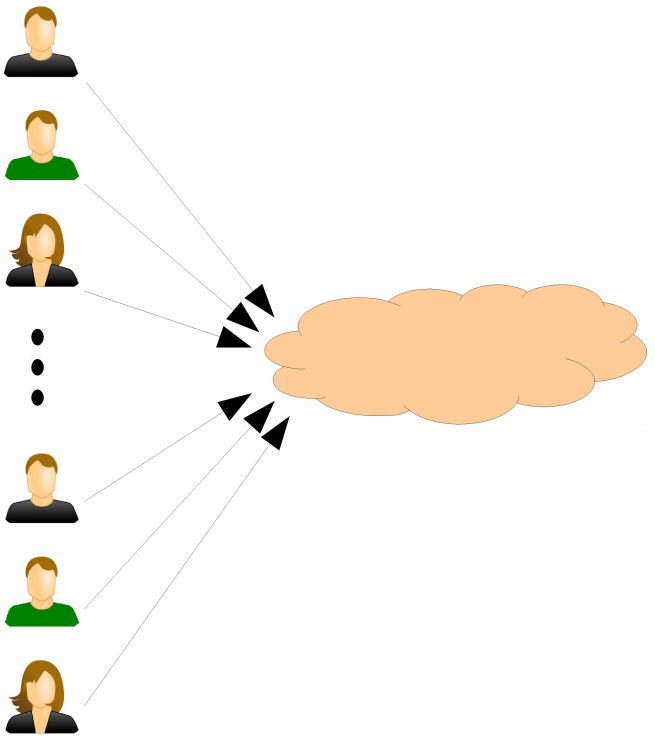
Authentication vs. Authorization

- Just because you can authenticate yourself, it does not mean you are authorized, too
  - e.g. your drivers license tells who you are, but does not allow you to enter a nuclear plant
- x.509 PKI only covers **authentication**
  - Tells the site who you are

Authorization in OSG

- Each site in OSG **decides autonomously** which users to authorize
  - Nobody in OSG can force a site to let a user in (but we can ask)
- The problem, again, is scale
  - Over 10,000 users!

If each site had to worry about every single user, very few users would be authorized!



Adding roles

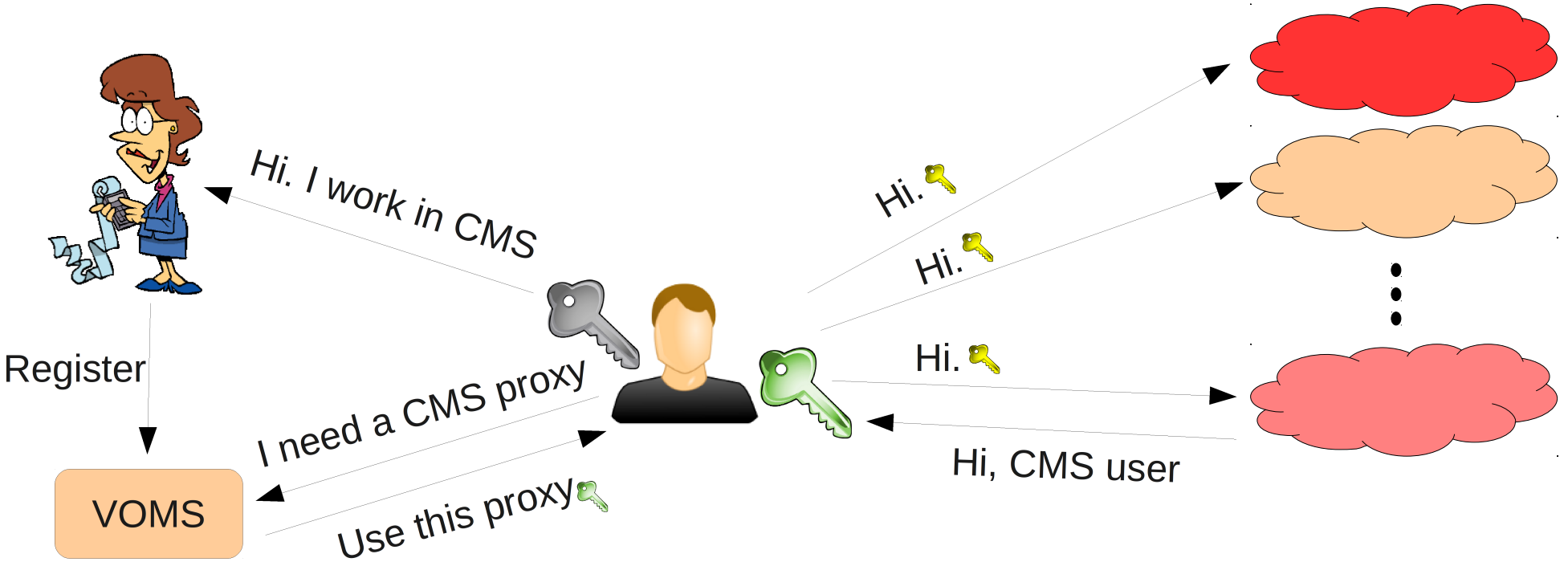
- Sites want to operate on higher level concepts
  - Some kind of attribute
- Like in real life
  - Think about passport vs driver's license
  - Both tell a cop who you are
    (and to 1st approx. are issued by the same entity)
  - But the driver's license tells him
    you are allowed to use a car, too
    - "Class:C"

Attribute authority

- Like before, we need someone trustworthy to issue attributes
  - Sites cannot just trust whatever the user says!
  - In the case of driver's license it was the DMV
- In OSG, the attribute authority is called the **Virtual Organization** (VO)
  - And the service issuing the attributes VOMS
  - Based on issuing augmented x.509 proxies

VO and VOMS

- VO decides who is worthy of an attribute
  - Site decides based on that attribute



OSG VOs in numbers

- O(10) VOs
  - Typically one per scientific domain
    e.g. CMS, ATLAS, SBGrid, LIGO, …
  - But we have regional VOs as well
    e.g. Holland Computing Center (HCC), Fermigrid
  - OSG operates an “Engage VO” for new users until there is an appropriate VO for them
- Sites typically support a set of VOs
  - And all the users inside those VOs
    (although they don't need to)

Are we done yet?

Copyright statement

- This presentation contains images copyrighted by ToonClipart.com
- These images have been licensed to Igor Sfiligoi for use in his presentations
- Any other use of them is prohibited