

Web Security: introduzione al Web e HTTP

Leonardo Taccari

`<s1069964@studenti.univpm.it>`

Sommario

World Wide Web

Uniform Resource Locator (URL)

HyperText Transfer Protocol (HTTP)

Conclusioni

Riferimenti

World Wide Web

World Wide Web

- ▶ World Wide Web, AKA WWW, AKA Web ¹
- ▶ È una collezione globale di documenti e altre risorse, collegate da **hyperlinks** e **URL**.
- ▶ I documenti vengono scambiati attraverso il protocollo di comunicazione **HTTP**
- ▶ La maggior parte dei documenti sono degli **ipertesti** scritti nel linguaggio **HTML**

¹Lo chiameremo semplicemente Web a seguire.

World Wide Web

Un po' di storia

- ▶ Proposto da Tim Berners-Lee al CERN nel 1989 e rilasciato pubblicamente nel 1993
- ▶ Nel 1994 viene fondato il World Wide Web Consortium (W3C) che si occupa degli standard che costituiscono il Web
- ▶ Inizialmente nato per scambiare **ipertesti**, poi **ipermedia**
- ▶ **Web 1.0**, **web statico**, l'utente può esclusivamente leggere i contenuti senza possibilità di modificarli
- ▶ **Web 2.0**, **web dinamico**, l'utente può anche modificare i contenuti

World Wide Web

Home page del primo sito web della storia

World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an [executive summary](#) of the project, [Mailing lists](#) , [Policy](#) , November's [W3 news](#) , [Frequently Asked Questions](#) .

[What's out there?](#)

Pointers to the world's online information, [subjects](#) , [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#) ,X11 [Viola](#) , [NeXTStep](#) , [Servers](#) , [Tools](#) , [Mail robot](#) , [Library](#))

[Technical](#)

Details of protocols, formats, program internals etc

[Bibliography](#)

Paper documentation on W3 and references.

[People](#)

A list of some people involved in the project.

[History](#)

A summary of the history of the project.

[How can I help ?](#)

If you would like to support the web..

[Getting code](#)

Getting the code by [anonymous FTP](#) , etc.

Home page del primo sito web della storia, pubblicato il 20 dicembre 1990 su <https://info.cern.ch/hypertext/WWW/TheProject.html> (catture da Wayback Machine dell'Internet Archive)

Uniform Resource Locator (URL)

- ▶ Ogni risorsa nel Web è identificata da un indirizzo detto **Uniform Resource Locator (URL)**
- ▶ Ad esempio, `https://www.example.org/index.html` identifica l'**home page** di **www.example.org**
- ▶ Nel **browser web** l'URL viene mostrato in una barra

Uniform Resource Locator (URL) I

Anatomia di un URL

- ▶ Un **URL** è costituito da diverse parti
- ▶ La sintassi generale è
`scheme://[userinfo@]host[:port][path][?query][#fragment]`

scheme indica il protocollo utilizzato, ad esempio `http` o `https`

userinfo (opzionale) solitamente nel formato `username:password` utilizzata per accedere a risorse che richiedono delle credenziali

host hostname o indirizzo IP

port (opzionale) port, di default 80 per HTTP e 443 per HTTPS ²

path (opzionale) "percorso", ogni componente è prefissato da un `/`

query (opzionale) "query string", costituisce coppie attributi-valori separati da `&`, solitamente utilizzate per sottomettere dati

Uniform Resource Locator (URL) II

Anatomia di un URL

fragment (opzionale) identifica un determinato frammento all'interno della risorsa (ad esempio permette di "evidenziare" una parte di testo)

- ▶ Esempio: `https://www.example.org/index.html`

`https` è lo **schema** seguito da :

`www.example.org` è l'**host** prefissato da //

`/index.html` è il **path** costituito da un singolo componente

²Vedremo meglio ciò in Network Security!

HyperText Transfer Protocol (HTTP)

- ▶ **HyperText Transfer Protocol (HTTP)** è il protocollo di comunicazione utilizzato per scambiarsi risorse nel Web
- ▶ Basato sul modello **client-server**: la comunicazione viene iniziata dal **client** che **richiede** una **risorsa** al **server** che a sua volta fornisce al client la risorsa in una **risposta**
- ▶ L'applicazione client HTTP viene detta **browser web**
- ▶ L'applicazione server HTTP viene detta **web server**

HyperText Transfer Protocol (HTTP)

Un po' di pratica!

- ▶ **curl** è un client **HTTP**
- ▶ Non è un vero e proprio browser: non permette di visualizzare gli ipermedia ma ci permette di effettuare richieste HTTP e visualizzare la risposta HTTP
- ▶ È molto potente ed utile sia per capire HTTP che per risolvere challenge
- ▶ Vediamo e commentiamo insieme
`curl -http1.1 -verbose http://www.example.org/index.html`

Conclusioni

- ▶ Abbiamo visto che cosa è il **World Wide Web** e la sua storia
- ▶ Abbiamo visto che cosa sono gli **URL**, gli identificatori di risorse alla base del Web
- ▶ Abbiamo visto il protocollo di comunicazione **HTTP** e visto in dettaglio - ed anche in pratica! - le richieste e risposte HTTP

Riferimenti

- ▶ Materiale Didattico del Portale di allenamento delle Olimpiadi Italiane di Cybersicurezza