Walks in Cyberspace: Improving Web Browsing and Network Activity Analysis with 3D Live Graph Rendering

Web Developer @ TheWebConf 2022

Lionel Tailhardat, Raphaël Troncy, Yoan Chabot

Orange & EURECOM

2022-04-28

Who's Who

Lionel Tailhardat AI R&D Engineer

- Dynamic Systems, Dependability and Knowledge Engineering
- genears.github.io

Dr. Raphaël Troncy Associate Professor

- Knowledge Engineering, Knowledge Graphs and Data Science
- www.eurecom.fr/~troncy

Dr. Yoan Chabot AI Researcher

- Knowledge Engineering, Knowledge Graphs and Data Science
- yoanchabot.github.io

Orange

Intl. Telecommunication infrastructure and service provider (and more ...)

- www.orange.com
- hellofuture.orange.com

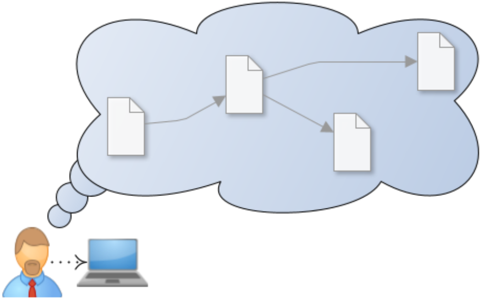
EURECOM

Graduate School and Research Center in Digital Science

- www.eurecom.fr

Our proposition: combine AI and Knowledge Engineering techniques for Complex Networks Resilience and Data Security concerns.

Context and Motivations: from Web Navigation to Structuring Traces



Scenario Web navigation session

Browsing General search loop scheme

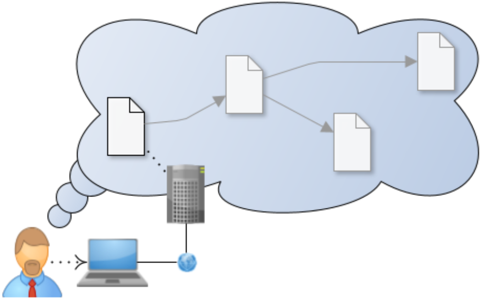
Analytics Persona dependent ⇒

Web user Green deal, privacy

Platform eng. Resource allocation, performance

Security analyst Malicious activity detection

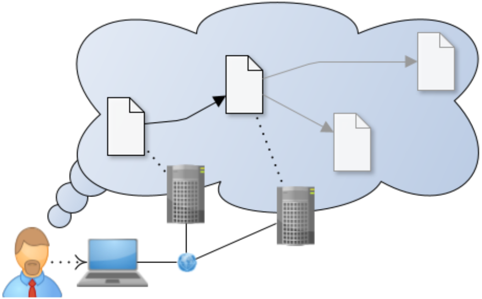
Context and Motivations: from Web Navigation to Structuring Traces



Scenario	Web navigation session
Browsing	General search loop scheme
Analytics	Persona dependent ⇒

Web user	Green deal, privacy
Platform eng.	Resource allocation, performance
Security analyst	Malicious activity detection

Context and Motivations: from Web Navigation to Structuring Traces



Scenario Web navigation session

Browsing General search loop scheme

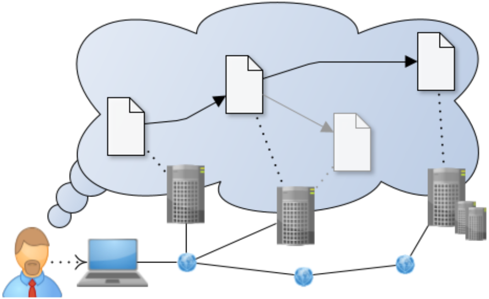
Analytics Persona dependent ⇒

Web user Green deal, privacy

Platform eng. Resource allocation, performance

Security analyst Malicious activity detection

Context and Motivations: from Web Navigation to Structuring Traces



Scenario Web navigation session

Browsing General search loop scheme

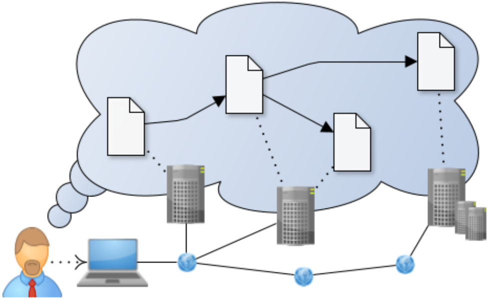
Analytics Persona dependent ⇒

Web user Green deal, privacy

Platform eng. Resource allocation, performance

Security analyst Malicious activity detection

Context and Motivations: from Web Navigation to Structuring Traces



Scenario Web navigation session

Browsing General search loop scheme

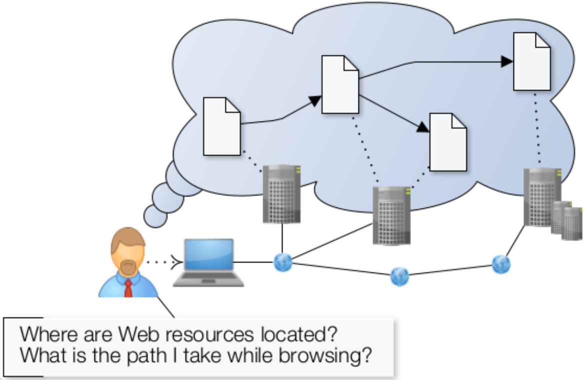
Analytics Persona dependent ⇒

Web user Green deal, privacy

Platform eng. Resource allocation, performance

Security analyst Malicious activity detection

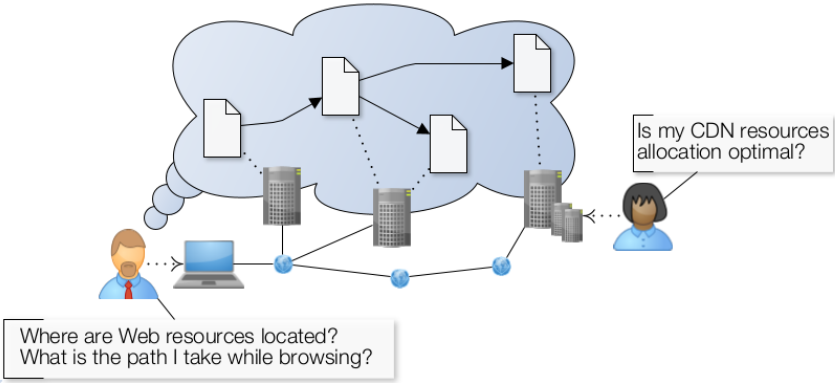
Context and Motivations: from Web Navigation to Structuring Traces



Where are Web resources located?
What is the path I take while browsing?

Scenario	Web navigation session	Web user	Green deal, privacy
Browsing	General search loop scheme	Platform eng.	Resource allocation, performance
Analytics	Persona dependent ⇒	Security analyst	Malicious activity detection

Context and Motivations: from Web Navigation to Structuring Traces

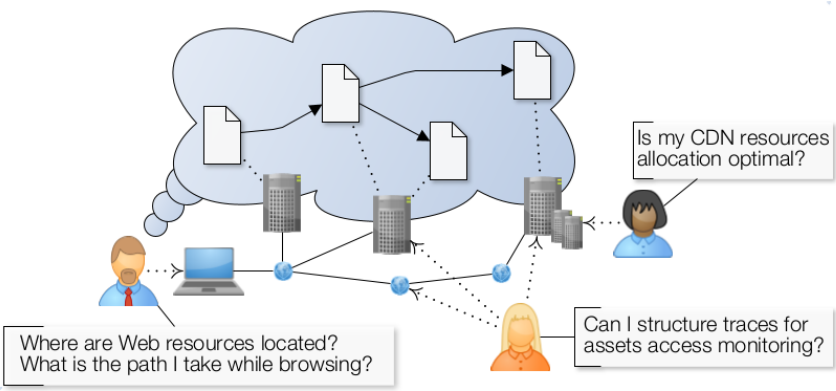


Is my CDN resources allocation optimal?

Where are Web resources located?
What is the path I take while browsing?

Scenario	Web navigation session	Web user	Green deal, privacy
Browsing	General search loop scheme	Platform eng.	Resource allocation, performance
Analytics	Persona dependent ⇒	Security analyst	Malicious activity detection

Context and Motivations: from Web Navigation to Structuring Traces



Is my CDN resources allocation optimal?

Where are Web resources located? What is the path I take while browsing?

Can I structure traces for assets access monitoring?

Scenario	Web navigation session	Web user	Green deal, privacy
Browsing	General search loop scheme	Platform eng.	Resource allocation, performance
Analytics	Persona dependent ⇒	Security analyst	Malicious activity detection

Problem Statement: Data Collection and Representation

Web navigation traces How do we collect them?

Data representation How do we combine technical and intentional perspectives?

Approach

1. Explore strengths and weaknesses of standard data collection methods
   - Web browser history,
   - Network traffic dumps.
2. Discover technical structures depending on the navigation session with visualization a priori.

Working hypothesis Using multilevel-graphs for data representation.

Problem Statement: Data Collection and Representation

Web navigation traces How do we collect them?

Data representation How do we combine technical and intentional perspectives?

Approach

1. Explore strengths and weaknesses of standard data collection methods
   - Web browser history,
   - Network traffic dumps.
2. Discover technical structures depending on the navigation session with visualization a priori.

Working hypothesis Using multilevel-graphs for data representation.

Experimental Design

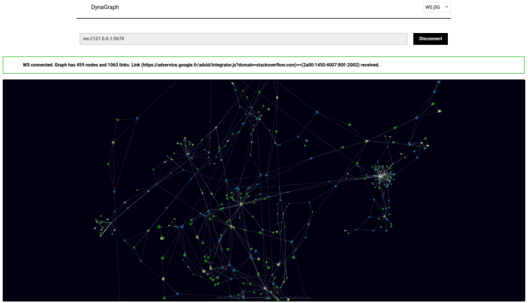
The DynaGraph framework

- Classical traces dumping tools: tshark tool + Firefox’s Network Monitor component.
- Web application: live 3D rendering of graph data derived from traces.

Strategy Comparing “Live Packet” and “Web Browser” capture & render features on arbitrary Web navigation session with the DynaGraph framework.

Demo Short video recorded example...

(available at Orange-OpenSource: DynaGraph_TWC2022)


DynaGraph
WS jSG
Disconnect
Left-click: rotate, Mouse-wheel/middle-click: zoom, Right-click: pan

ws://127.0.0.1:5678

WS connected. Graph has 459 nodes and 1063 links. Link (https://adservice.google.fr/adsid/integrator.js?domain=stackoverflow.com)=>(2a00:1450:4007:80f::2002) received.

Experimental Design

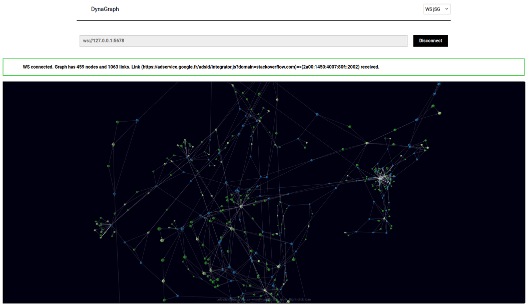
The DynaGraph framework

- Classical traces dumping tools: tshark tool + Firefox's Network Monitor component.
- Web application: live 3D rendering of graph data derived from traces.

Strategy Comparing "Live Packet" and "Web Browser" capture & render features on arbitrary Web navigation session with the DynaGraph framework.

Demo Short video recorded example...

(available at Orange-OpenSource: DynaGraph_TWC2022)



WS connected. Graph has 459 nodes and 1063 links. Link (https://adservice.google.fr/adsid/integrator.js?domain=stackoverflow.com)=>(2a00:1450:4007:80f::2002) received.

Experimental Design

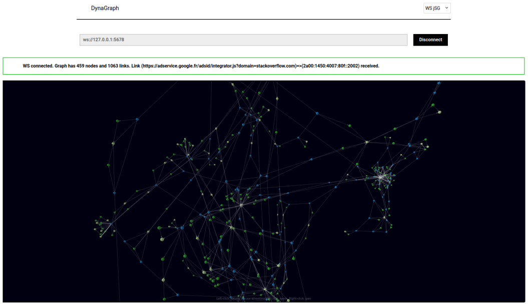
The DynaGraph framework

- Classical traces dumping tools: tshark tool + Firefox's Network Monitor component.
- Web application: live 3D rendering of graph data derived from traces.

Strategy Comparing "Live Packet" and "Web Browser" capture & render features on arbitrary Web navigation session with the DynaGraph framework.

Demo Short video recorded example...

(available at Orange-OpenSource: DynaGraph_TWC2022)


DynaGraph
WS jSG
Disconnect

ws://127.0.0.1:5678

WS connected. Graph has 459 nodes and 1063 links. Link (https://adservice.google.fr/adsid/integrator.js?domain=stackoverflow.com)=>(2a00:1450:4007:80f::2002) received.

Left-click: rotate, Mouse-wheel/middle-click: zoom, Right-click: pan

Evaluation

Network traffic capture techniques

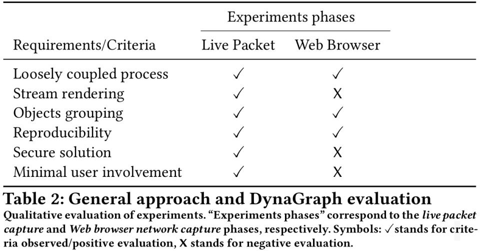
Strength and weaknesses

- Operating system network interface level
  easy to set up, low tracking details whenever data encryption is present.
- Web browser process level
  full tracking details, no live streaming, high user involvement.

Data exploration

3D spatialization and various sampling periods

- Immersing into a navigation path
  force layout spreads nodes over the representation space.
- Access patterns as spatial structures
  prevalent structures: ”various urls from same server”, “several times same url from same server” and “many sub-resources from many servers”.



	Experiments phases	
Requirements/Criteria	Live Packet	Web Browser
Loosely coupled process	✓	✓
Stream rendering	✓	X
Objects grouping	✓	✓
Reproducibility	✓	✓
Secure solution	✓	X
Minimal user involvement	✓	X

Table 2: General approach and DynaGraph evaluation
Qualitative evaluation of experiments. “Experiments phases” correspond to the *live packet capture* and *Web browser network capture* phases, respectively. Symbols: ✓ stands for criteria observed/positive evaluation, X stands for negative evaluation.

Evaluation

Network traffic capture techniques

Strength and weaknesses

- Operating system network interface level

  easy to set up, low tracking details whenever data encryption is present.
- Web browser process level

  full tracking details, no live streaming, high user involvement.

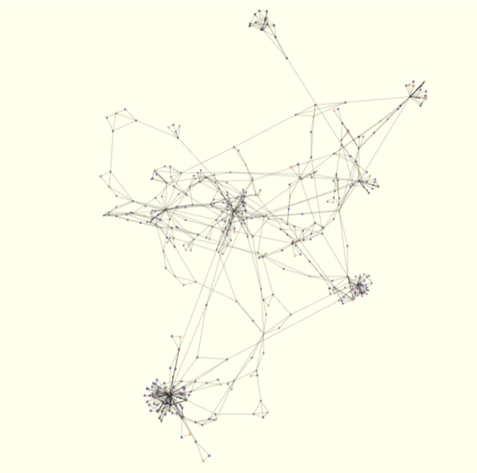
Data exploration

3D spatialization and various sampling periods

- Immersing into a navigation path

  force layout spreads nodes over the representation space.
- Access patterns as spatial structures

  prevalent structures: ”various urls from same server”, “several times same url from same server” and “many sub-resources from many servers”.



Where do I start? The “Live Packet” quick start

1. Git clone the project to your computer

```
git clone https://github.com/Orange-OpenSource/dynagraph.git
cd dynagraph
```

2. Install the DynaGraph tool set, then launch the DynaGraph Web app

```
pip install -r data-collection/requirements.txt
npm install
npm run start
```

3. Install the tshark tool, then start and forward the capture

```
sudo apt install tshark
cd data-collection/ && sh ./start-tshark-basic.sh
```

4. Connect, observe and play with incoming graph data in the DynaGraph Web app
   - Open the DynaGraph Web app UI available at http://localhost:3000/
   - Select the WS jSg source method
   - Set the ws://127.0.0.1:5678 source and click the Connect button

More in https://github.com/Orange-OpenSource/dynagraph/blob/main/data-collection/README.md

Summary and Future Work

The topic Understanding and making use of the form of Web navigation traces.

Our approach User-side data collection, graph streaming, 3D rendering and visual data exploration.

Insights Full approach, methodology, evaluation and future work details... see our paper ⇒

Contribute Web Extension? SemWeb serialization? Graph structures database? Topos analysis?... see the code and just extend it ⇒

Paper

Lionel TAILHARDAT, Raphaël TRONCY, and Yoan CHABOT. 2022.

Walks in Cyberspace: Improving Web Browsing and Network Activity Analysis with 3D Live Graph Rendering.

https://doi.org/10.1145/3487553.3524230

Code repository

Orange-OpenSource / dynagraph public project on GitHub

https://github.com/Orange-OpenSource/dynagraph

Thank you!

lionel.tailhardat@orange.com
raphael.troncy@eurecom.fr
yoan.chabot@orange.com