2020-02783 - PhD Position F/M PhD Improving online privacy through content blocking and information restriction

Type de contrat : CDD
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 360, including 300 scientists working in sixteen research teams. Recognised for its outstanding contribution to the socio-economic development of the Hauts-de-France region, the Inria Lille - Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects in such fields of computer science as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

Contexte et atouts du poste

Spirals is conducting research activities in the domains of distributed systems and software sciences. We aim at introducing more automation in the adaptation mechanisms of software systems, in particular, transitioning from adaptive systems to self-adaptive systems. We target especially two properties: self-healing and self-optimization. With self-healing, we aim at studying and tailoring data mining and machine learning solutions to the design and implementation of software system, and more especially for software automatic repair. With self-optimization, we aim at sharing, collecting, and analyzing behaviors in a distributed setting to continuously tailor, optimize, and keep under working conditions software systems, and to go toward the goal of obtaining eternal distributed systems.

The PhD student will join the Spirals (https://team.inria.fr/spirals) project-team between the University of Lille and Inria.

Mission confiée

Since its inception, the web has grown substantially and websites have turned into rich client-side experience customized for the user where third parties supply a considerable amount of content. The increasing reliance on third parties has brought a number of privacy issues to the web with web tracking being at the top of that list [7]. With cookies and browser fingerprinting, users are on the losing side of privacy as they can be tracked across the domains they are visiting. To regain control, browser vendors like Mozilla and Apple have added in their own browsers a tracking protection mechanism (called Enhanced Tracking Protection for Firefox [3]) and Intelligent Tracking Protection for Safari [5]) aimed at preventing tracking on the web. Yet, essential functionality of a website is sometimes so intertwined with tracking code that using those protection mechanisms can transitorily 'break' a webpage. We define "page breakage" as an undesirable behavior on a webpage and it includes, but is not limited to, page slowdowns, page freezes, page crashes, page errors and page display issues. In order to push online privacy forward, there is a real need today to identify and block properly tracking entities on the web without the current usability costs associated with it.

Positioned in the context of web security and privacy, this PhD project will focus on the three following topics:

- Blocking non-essential content of a page. Most approaches to block content on the web are based on blacklists. When an element matches a URL or a regular expression present in a blacklist, this element is not loaded no matter its intended functionality. For a lot of scripts, this approach is fine as their only purpose is to put a cookie in the browser's storage. For other scripts, the result can range from display issues to a complete page freeze [1, 2, 4].

In this PhD, the student will investigate how to design a system that will take any third-party script found online and provide a detailed set of information about it. This would go beyond recent studies [6, 7] as it would not be limited to tracking scripts and would use techniques like information flow control, machine learning and de-obfuscation for finer-grained analysis. The ultimate goal would be to differentiate elements that are essential to web browsing from elements that are not. This distinction can help in the design of advanced defense mechanisms that could go much further than the tracking protection we see today.

- Automatic detection of page breakage. Following directly on the previous topic, one big challenge on the web is to identify page breakage. A page that does not load may be easy to identify but one where one crucial component does not trigger the right action is a very different problem.

In this PhD, the student will investigate how to automatically detect page breakage. This includes detecting the type of breakage (e.g., errors or freezers) and its severity (e.g., is the whole page unresponsive or only one button)?

- Restricting information sent to a webserver. To improve user experience on the web, browsers send information like the version of the browser or the OS being used. The problem is that the diversity of modern devices is so great today that it opened the door to a technique called browser fingerprinting. By collecting a set of information related to a user's device, any third-party can build a fingerprint of the device and use it to track an individual online [8]. This technique is particularly dangerous as it is completely stateless and requires no permission to be executed. One way to mitigate the effects of fingerprinting is to send less information or, at least, less precise information so that not a single user has a unique fingerprint.

Informations générales

- Thème/Domaine : Systèmes distribués et intercalés
- Système & réseaux (BAP E)
- Ville : Villeneuve d’Ascq
- Centre Inria : Lille - Nord Europe
- Date de prise de fonction souhaitée : 2020-10-01
- Durée de contrat : 3 ans
- Date limite pour postuler : 2020-07-31

Contacts

- Equipe Inria : SPIRALS
- Directeur de thèse : Rouvoy Romain / Romain.Rouvoy@inria.fr

A propos d’Inria

Inria is the institute national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en générales communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers differents. 500 personnels d’appui à la recherche et à l’innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L’institut s’efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l’économie. Consignes pour postuler

CV + application letter + recommendation letters

Sécurité défense :
Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1426 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention : Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d’autres canaux n’est pas garanti.
In this PhD, the student will assess the impact of restricting information sent to a server on usability and page breakage. This includes blocking information from APIs that are used for fingerprinting and reducing the quantity of information in legacy headers. This topic goes hand in hand with the previous ones as identifying scripts that use some APIs only to collect device information can be blocked without any impact on the rendered webpage.

References


Compétences

The PhD candidate should have a background in computer science. Knowledge in Web programming is a strong plus. As is a common practice in the Spirals research teams, all source code is expected to be open-sourced. The student should publish high-level academic papers, as well as participate in related open source communities. This should assist in the technological transfer from academic prototypes to industry-ready tools.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving house, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

1st and 2nd year: 1 982€ Gross monthly salary (before taxes)

3rd year: 2 085€ gross monthly salary (before taxes)