Sarah Anne Wassermann

Personal information

Female Born 17th April 1993 Luxembourgish

Contact

sarah@wassermann.lu www.wassermann.lu www.linkedin.com/in/sarahwassermann

Research in Internet measurements and machine learning
4 international internships in France, Austria, and the United States
24 publications

Research statement

My research lies at the intersection of computer networks and machine learning, and I develop systems which significantly enhance the experience of Internet users. Today, we perform more and more daily tasks and entertaining activities through the Internet and it is thus crucial that we can do so without being hindered by poor performance. My ultimate goal is to conceive intelligent systems which make the Internet smarter and able to face demanding users and an ever-growing amount of heterogeneous network traffic.

Education

PhD Telecommunications

Multiple institutions in France and Austria

- Thesis subject: Machine Learning for Internet-QoE Monitoring and Analysis
- Supervised by Dr. Pedro Casas (AIT Austria)
- In the context of my PhD, my goal is to deliver algorithms and software systems to measure and efficiently retrieve Internet QoE metrics in today's encrypted Internet.
- I was a PhD student at Inria Paris affiliated with Sorbonne Université from 2017 until 2019. Since 2019, I continue my PhD with AIT Austria affiliated with Technische Universität Wien. So far, I have been working on several research projects related to QoE and machine learning.
- Since I have been at AIT Austria, I have been working on the following projects
 - I am conceiving active-learning techniques based on reinforcement learning to gather the users' feedback about their experience in an efficient way. Asking the users too often for feedback is annoying and discourages them from answering accurately. First results are published in [W6, C5, P3].
 - o I am analysing the impact of user interactions when watching YouTube videos (for instance, the effect on the network when the user is seeking, changing the quality or the speed). Additionally, I aim at predicting user interactions (when they occur and what type of interaction it is) using machine-learning models relying on network-related features. This project is carried out in collaboration with Julius-Maximilians-Universität Würzburg.
 - o I am designing a machine-learning-based framework to infer key video-QoE indicators from the analysis of the encrypted traffic in real-time, using exclusively on network-level features. This project is carried out in collaboration with Julius-Maximilians-Universität Würzburg. Early results are presented in [W7, P4, P5].
 - I am exploring web QoE. I implement techniques based on machine-learning to infer web-QoE metrics from encrypted network traffic and design new performance metrics.
- While I was at Inria in the MiMove group, I developed techniques to infer video QoE metrics with a novel lightweight system that analyses traffic generated by DASH on-demand and live video streams while

November 2019

running at the home-network gateway of the users. This project was carried out in collaboration with Princeton University.

- Expected graduation beginning of 2021

MSc. Computer Science

Université de Liège (ULiège) – Belgium

- Specialised in computer systems and networks with additional electives in machine learning
- Master's thesis entitled "Anycast-based DNS in Mobile Networks", supervised by Prof. Fabián Bustamante (Northwestern University) and Prof. Benoit Donnet (University of Liège); with high honours http://orbi.ulg.ac.be/handle/2268/215008
- **Conducted research** under the supervision of Dr. Pedro Casas, Prof. Fabián Bustamante, and Prof. Benoit Donnet in the field of **Internet measurements**. **Topics:** Internet path dynamics and performance, machine learning for networking, anycast in cellular networks, malware detection in smartphones
- Active as a student representative during my whole Master's studies
- Graduated with honours in September 2017

BSc. Computer Science

Université de Liège (ULiège) - Belgium

- Ranked among Top 3 students every year: 2 out of 50 1st year, 1 out of 10 2nd year, and 3 out of 9 3rd year
- Active as a student representative during the last two years of my Bachelor's degree
- Graduated with honours in June 2015

General Certificate of Secondary Education

Lycée Michel Rodange – Luxembourg

- Main subject areas: mathematics and informatics
- Graduated with grade excellent in July 2012

Online courses Coursera

- Statistics with R specialisation (in progress)
 - This specialisation includes 5 courses: Introduction to Probability and Data; Inferential Statistics; Linear Regression and Modeling; Bayesian Statistics; Statistics with R Capstone
- Big Data for Data Engineers specialisation (in progress)
 - This specialisation includes 5 courses: Big Data Essentials: HDFS, MapReduce and Spark RDD; Big Data Analysis: Hive, Spark SQL, DataFrames and GraphFrames; Big Data Applications: Machine Learning at Scale; Big Data Applications: Real-Time Streaming; Big Data Services: Capstone Project
- Deep Learning specialisation (overall score: 100%); March 2018
 - This specialisation includes 5 courses: Neural Networks and Deep Learning; Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization; Structuring Machine Learning Projects; Convolutional Neural Networks; Sequence Models
- Machine Learning (score: 100%); September 2014
- Image and video processing: From Mars to Hollywood with a stop at the hospital (score: 100%);
 March 2014
- Einführung in Computer Vision (score: 65.5%); February 2014
- An Introduction to Interactive Programming in Python (score: 99.6%); January 2014

Awards and honours

- **Best Paper Award Runner Up** for our paper "On the Analysis of YouTube QoE in Cellular Networks through in-Smartphone Measurements" at the 12th IFIP Wireless and Mobile Networking Conference (WMNC); September 2019
- Inria PhD fellowship allowing me to carry out a PhD at Inria Paris (ranked 1st among all applicants by the admission committee); October 2017

- Travel grants to attend ACM CoNEXT 2015, TMA PhD School 2016, IEEE LCN 2016, ACM CONEXT 2016, ACM SIGCOMM 2017 (N2Women fellowship and SIGCOMM travel grant), TMA PhD School 2018, ACM SIGCOMM 2018 (Netflix Diversity travel grant)
- Pisart Grant for pedagogical support; fall 2014
- Best student prize for graduating as the best high-school student of my class (grade: excellent, ranked in Top 1.3% of the Grand Duchy of Luxembourg); July 2012
- Best student prize for having the highest GPA of my class in all the academic years from 2005 to 2009

Scientific publications

Journal papers

[J1] "Unveiling Network and Service Performance Degradation in the Wild with mPlane"
 P. Casas, P. Fiadino, S. Wassermann, S. Traverso, A. D'Alconzo, E. Tego, F. Matera, M. Mellia in IEEE Communications Magazine, Network Testing Series, vol. 54, no. 3, pp. 71-79, 2016.
 Internet-paths monitoring http://orbi.ulg.ac.be/handle/2268/192775

 [J2] "Considering User Behavior in the Quality of Experience Cycle: Towards Proactive QoE-aware Traffic Management"

M. Seufert, S. Wassermann, P. Casas

in IEEE Communications Letters, vol. 23, no. 7, pp. 1145-1148, 2019.

Proactive QoE-aware traffic management, supervised learning for user-interaction prediction

https://hal.inria.fr/hal-02114784

 [Under submission] "ViCrypt 2.0: Real-time, Fine-grained Estimation of Video Quality from Encrypted Streaming Traffic"

S. Wassermann, M. Seufert, P. Casas, L. Gang, K. Li submitted to JSAC Special Issue on Advances in Artificial Intelligence and Machine Learning for Networking, 2019

Supervised learning for video-quality inference

Conference papers

[C1] "Improving QoE Prediction in Mobile Video through Machine Learning"

P. Casas, S. Wassermann

in Proceedings of the 8th International Conference on Network of the Future (NoF), London, United Kingdom, 2017.

Supervised learning for video-QoE prediction

Best Paper Award candidate

http://orbi.ulg.ac.be/handle/2268/214928

- [C2] "Anycast on the Move: A Look at Mobile Anycast Performance"

S. Wassermann, J. P. Rula, F. E. Bustamante, P. Casas

in Proceedings of the Network Traffic Measurement and Analysis Conference (TMA) 2018, Vienna, Austria, 2018.

Analysis of anycast on mobile connections

https://hal.inria.fr/hal-01812440

November 2019

[C3] "Beauty is in the Eye of the Smartphone Holder – A Data Driven Analysis of YouTube Mobile QoE"
 N. Wehner, S. Wassermann, P. Casas, M. Seufert, F. Wamser
 in Proceedings of the 14th International Conference on Network and Service Management (CNSM), Rome, Italy, 2018.

Analysis of the evolution of quality metrics for YouTube Mobile https://hal.inria.fr/hal-01898082

- [C4] "On the Analysis of YouTube QoE in Cellular Networks through in-Smartphone Measurements"

S. Wassermann, P. Casas, M. Seufert, F. Wamser

in Proceedings of the 12th IFIP Wireless and Mobile Networking Conference (WMNC), Paris, France, 2019.

Analysis and prediction of quality metrics for YouTube Mobile

Best Paper Award runner up

https://hal.archives-ouvertes.fr/hal-02159716

 [C5] "ADAM & RAL: Adaptive Memory Learning and Reinforcement Active Learning for Network Monitoring"

S. Wassermann, T. Cuvelier, P. Mulinka, P. Casas

in Proceedings of the 15th International Conference on Network and Service Management (CNSM), Halifax, Canada, 2019.

Stream-based machine learning, reinforcement learning https://hal.archives-ouvertes.fr/hal-02301393

Workshop papers

Emirates, 2016.

[W1] "On the Analysis of Internet Paths with DisNETPerf, a Distributed Paths Performance Analyzer"
 S. Wassermann, P. Casas, B. Donnet, G. Leduc, M. Mellia
 in Proceedings of the 10th IEEE Workshop on Network Measurements (WNM), Dubai, United Arab

Internet-paths monitoring http://orbi.ulg.ac.be/handle/2268/200967

- [W2] "NETPerfTrace – Predicting Internet Path Dynamics and Performance with Machine Learning"

S. Wassermann, P. Casas, T. Cuvelier, B. Donnet

in Proceedings of the ACM SIGCOMM 2017 Workshop on Big Data Analytics and Machine Learning for Data Communication (Big-DAMA), Los Angeles (CA), United States, 2017.

Supervised learning for Internet-path-performance prediction http://orbi.ulg.ac.be/handle/2268/211667

[W3] "BIGMOMAL – Big Data Analytics for Mobile Malware Detection"

S. Wassermann, P. Casas

in Proceedings of the ACM SIGCOMM 2018 Workshop on Traffic Measurements for Cybersecurity (WTMC), Budapest, Hungary, 2018.

Supervised learning for malware detection https://hal.inria.fr/hal-01812448

[W4] "Machine Learning Models for YouTube QoE and User Engagement Prediction in Smartphones"

S. Wassermann, N. Wehner, P. Casas

in Proceedings of the Workshop on AI in Networks (WAIN) 2018, Toulouse, France, 2018.

Supervised learning for QoE and user-engagement prediction

https://hal.inria.fr/hal-01898083

[W5] "Remember the Good, Forget the Bad, do it Fast: Continuous Learning over Streaming Data"
P. Mulinka, S. Wassermann, G. Marín, P. Casas
in Proceedings of the Continual Learning Workshop at NeurIPS 2018, Montreal, Canada, 2018.
Stream-based supervised learning, adaptive learning under concept drifts
https://hal.inria.fr/hal-01952211

[W6] "RAL – Improving Stream-Based Active Learning by Reinforcement Learning"

S. Wassermann, T. Cuvelier, P. Casas

in Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), Workshop on Interactive Adaptive Learning (IAL), Würzburg, Germany, 2019.

Stream-based active learning, reinforcement learning https://hal.archives-ouvertes.fr/hal-02265426

- [W7] "I See What you See: Real Time Prediction of Video Quality from Encrypted Streaming Traffic"

S. Wassermann, M. Seufert, P. Casas, L. Gang, K. Li

in Proceedings of the 4th ACM MOBICOM Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE), Los Cabos, Mexico, 2019.

Supervised learning for video-quality inference https://hal.archives-ouvertes.fr/hal-02268814

Extended abstracts

- [A1] "Towards DisNETPerf: a Distributed Internet Paths Performance Analyzer"

S. Wassermann, P. Casas, B. Donnet

in Proceedings of the ACM CoNEXT Student Workshop, Heidelberg, Germany, 2015.

Internet-paths monitoring

http://orbi.ulg.ac.be/handle/2268/187290

[A2] "Machine Learning based Prediction of Internet Path Dynamics"

S. Wassermann, P. Casas, B. Donnet

in Proceedings of the ACM CoNEXT Student Workshop, Irvine (CA), United States, 2016.

Supervised learning for Internet-path-performance prediction

http://orbi.ulg.ac.be/handle/2268/203086

Demo sessions

[D1] "Reverse Traceroute with DisNETPerf, a Distributed Internet Paths Performance Analyzer"

S. Wassermann, P. Casas

in Proceedings of the Demonstrations of the 41th Annual IEEE Conference on Local Computer Networks (LCN-Demos 2016), Dubai, United Arab Emirates, 2016.

Internet-paths monitoring

http://orbi.ulg.ac.be/handle/2268/201059

- [D2] "Distributed Internet Paths Performance Analysis through Machine Learning"

S. Wassermann, P. Casas

in Proceedings of the Demonstrations of the Network Traffic Measurement and Analysis Conference (TMA) 2018, Vienna, Austria, 2018.

Internet-paths monitoring with supervised learning

Best Demo Award candidate

https://hal.inria.fr/hal-01883815

- [D3] "Let me Decrypt your Beauty: Real-time Prediction of Video Resolution and Bitrate for Encrypted Video Streaming"

S. Wassermann, M. Seufert, P. Casas, L. Gang, K. Li

in Proceedings of the Demonstrations of the Network Traffic Measurement and Analysis Conference (TMA) 2019, Paris, France, 2019.

Video-quality-metric estimation with supervised learning

https://hal.archives-ouvertes.fr/hal-02134851

Posters

- [P1] "Anycast on the Move – A First Look at Mobile Anycast Performance"

S. Wassermann, J. P. Rula, F. E. Bustamante

presented during the poster session at the ACM Internet Measurement Conference (IMC) 2017, London, United Kingdom, 2017.

Analysis of anycast on mobile connections

http://orbi.ulg.ac.be/handle/2268/215141

- [P2] "BIGMOMAL - Big Data Analytics for Mobile Malware Detection"

S. Wassermann, P. Casas

presented during the poster session at the ACM Internet Measurement Conference (IMC) 2017, London, United Kingdom, 2017.

Supervised learning for malware detection

http://orbi.ulg.ac.be/handle/2268/215139

- [P3] "Improving Stream-Based Active Learning with Reinforcement Learning"

S. Wassermann, T. Cuvelier, P. Casas

accepted to the poster session at the Women in Machine Learning (WiML) Workshop co-located with NeurIPS, Vancouver, Canada, 2019.

Stream-based active learning, reinforcement learning

https://hal.archives-ouvertes.fr/hal-02375296

[P4] "Decrypting Video Quality from Encrypted Streaming Traffic"

S. Wassermann, P. Casas

accepted to the poster session at the Women in Machine Learning (WiML) Workshop co-located with NeurIPS, Vancouver, Canada, 2019.

Supervised learning for video-quality inference

https://hal.archives-ouvertes.fr/hal-02375298

[P5] "ViCrypt: Real-time, Fine-grained Prediction of Video Quality from Encrypted Streaming Traffic"
 S. Wassermann, M. Seufert, P. Casas
 presented during the poster session at the Internet Measurement Conference (IMC), Early Work, Tools, and Datasets Track, Amsterdam, Netherlands, 2019.
 Supervised learning for video-quality inference
 https://hal.archives-ouvertes.fr/hal-02375301

Technical reports

[R1] "Predicting Internet Path Dynamics and Performance with Machine Learning"

S. Wassermann, P. Casas, T. Cuvelier, B. Donnet

AIT-Big-DAMA Tech. Rep. A3215, 2017.

Supervised learning for Internet-path-performance prediction

http://orbi.ulg.ac.be/handle/2268/209422

Talks

 [T1] "Decrypting QoE in an Encrypted Internet – AI to the Rescue" RIPE 79, Rotterdam, Netherlands, October 2019.
 Video-quality-metric estimation with supervised learning https://sawassermann.github.io/talks/ripe79 2019.pptx

- [T2] "Active Measurements for Path Performance Diagnosis with DisNETPerf, a Distributed Internet Paths Performance Analyzer"

Luxembourg Internet Days, Luxembourg, Grand Duchy of Luxembourg, November 2019. Internet-paths monitoring https://sawassermann.github.io/talks/luxInternetDays 2019.pptx

Research experience

Junior investigator

Austrian Institute of Technology (AIT) – Austria

- Research internship in the context of the ICT15-129 project Big-DAMA under the supervision of Dr. Pedro Casas: **March 2017**
- The subject of this research stay was "Mobile Malware Detection using Machine Learning and Big-Data Analytics". I worked in the analysis of a large-scale dataset of smartphone measurements, with the aim of automatically discovering malware activity. I developed different machine-learning-based approaches for both supervised and unsupervised detection of malware applications, using the scikit-learn machine-learning library. Our first results are presented in [W3, P2].

Visiting pre-doctoral fellow

Northwestern University (NU) – United States

- Research stay in the Aqualab group under the supervision of Prof. Fabián Bustamante and Dr. John Rula; **July September 2016**
- My research topic was entitled "Analysis of Anycast-based Content Distribution in Cellular Networks". I studied the performance of anycast routing for cellular clients. I conducted and analysed the measurements from distributed mobile vantage points to a live anycast service-, measuring the latencies and routes to server deployments, and investigated the causes of poor performance for cellular clients. This internship lead to my Master's thesis, and we discuss the outcome of our study in [C2, P1].

Research intern

Forschungszentrum Telekommunikation Wien (FTW) - Austria

- Research internship in the context of the of the FP7 ICT European Research Project mPlane under the supervision of Dr. Pedro Casas; July – September 2015
- The subject of this internship was "Tracking the Performance of CDN Servers-to-Customers Internet Paths with Distributed Active Measurements". I worked in the field of geo- and topology-based location of active probes devices capable of launching active measurements at the Internet scale, relying on the well-known RIPE Atlas distributed active-measurement platform. I designed and developed DisNETPerf, an Internet-scale measurement framework to track the performance of Internet paths, relying on distributed probes' location and active measurements. In a nutshell, DisNETPerf can measure path performance using standard traceroute measurements, but can do so for paths connecting arbitrarily selected nodes. I developed different techniques to find the optimal probe in terms of network topology and path latency closest to a desired server, which is then used to masquerade this server and to collect active network-path measurements. DisNETPerf thus allows users to accurately monitor servers through active measurements, even if they do not have access to these servers. An in-depth description and evaluation of my tool can be found in [W1, A1, D1, D2].

Summer intern

Laboratoire d'Informatique de Paris 6 (LIP6) - France

- Internship at the Laboratory of Information, Networking and Communication Sciences (LINCS) under the supervision of Prof. Timur Friedman, Dr. Marc-Olivier Buob, and Dr. Jordan Augé; **July 2014**
- I was involved in the Paris Traceroute and libparistraceroute projects. I developed Paris Ping, a generic ping tool based on the libparistraceroute library which can handle IPv4, IPv6, and TCP, UDP, ICMP probes. Contrary to the standard ping tool, the flow IDs of each sent packet remain constant in order to avoid flow-based load balancing, which enhances the accuracy of the latency measurements. In addition to the implementation of Paris Ping, I also extended the libparistraceroute library.

Editorial boards

Reviewer

- IEEE Communications Letters, 2016
- IEEE/ACM Transactions on Networking, 2019
- Women in Machine Learning (WiML) Workshop, co-located with NeurIPS, 2019

Open-source projects

DisNETPerf

I designed and implemented DisNETPerf – a Distributed Internet Paths Performance Analyzer – during my internship at FTW, with the collaboration of Dr. Pedro Casas and Dr. Pierdomenico Fiadino.

DisNETPerf is a tool that allows one to **locate the closest RIPE Atlas box** (in terms of minimum RTT) to a given IP address. Once the closest RIPE Atlas box has been located, DisNETPerf permits to **launch traceroutes** from this box to a destination IP address provided by the user.

More details about the tool and how the closest probe is actually chosen are explained in [W1, A1, D1, D2]. DisNETPerf is freely available on GitHub: https://github.com/SAWassermann/DisNETPerf

NETPerfTrace

I developed NETPerfTrace — an Internet Path Tracking System — while working on a research project about the prediction of Internet path dynamics and performance, in collaboration with Dr. Pedro Casas.

NETPerfTrace is a tool capable of **forecasting path changes and path latency variations**.

NETPerfTrace aims at predicting three metrics:

- the residual life time of a route (i.e. the remaining life time of the route before it actually changes)
- the number of route changes in the next time window
- the average RTT of the next traceroute sample

The overall idea of this tool and preliminary results are presented in [W2, A2, D2, R1].

NETPerfTrace is freely available on GitHub: https://github.com/SAWassermann/NETPerfTrace

RAL

I conceived RAL – Reinforced stream-based Active Learning –, which is an active-learning technique relying on reinforcement-learning principles, using rewards and bandit-like algorithms.

In particular, the rewards are based on the usefulness of RAL's querying behaviour. The intuition behind the different reward values is that we attribute a positive reward in case RAL asks the oracle for ground truth and it was necessary (i.e. the underlying models would have predicted the wrong label), and a negative one otherwise (i.e. querying the oracle was unnecessary as the models predicted the right label anyway).

The system additionally makes use of the prediction certainty of the classification models. We combine the aforementioned reward mechanism with the model's uncertainty to tune the sample-informativeness heuristic to better guide the query decisions.

The technique is described in detail in [W6, C5, P3].

RAL is freely available on GitHub: https://github.com/SAWassermann/RAL

libparistraceroute

I contributed to the libparistraceroute project (which includes the well-known Paris Traceroute) during my internship at LIP6.

My main contribution was the development of **Paris Ping**. I also extended the libparistraceroute library itself by adding probe matching. More precisely, I implemented functions to check whether a probe corresponds to the reply to a given probe (for the IPv4/6, ICMPv4/6, and TCP protocols).

libparistraceroute is freely available on GitHub: https://github.com/libparistraceroute/libparistraceroute

Teaching activities

Undergraduate teaching assistant

Université de Liège (ULiège) – Belgium

I was hired as an undergraduate teaching assistant to supervise and help students during the exercise sessions for the course **Introduction to Computer Programming** (1st year BSc. course in Computer Science, taught by Prof. Benoit Donnet); **fall 2013 and 2014**

Work experience

Active member

IEEE Student Branch Liège

I designed flyers and posters for the events organised by the student branch. I also actively participated in the organisation of conferences. I have been the main organiser for the talk entitled "*Practical Internet-of-Things:* the reality between people, process, data and things" given by Emmanuel Tychon (Cisco) at the University of Liège in November 2016. I participated in IEEEXtreme 8.0 as a competitor; **2013 – 2017**

IT coordinator

Board of European Students of Technology (BEST)

I was responsible for redesigning and reworking the website of the local BEST group of the University of Liège; **2013 – 2015**

November 2019

Event organiser

Board of European Students of Technology (BEST)

I was in charge of organising events in collaboration with the IEEE student branch at the University of Liège. I helped organising the conference about photonic quantum computers given by Dr. André Hautot in February 2014; **2013 – 2015**

Technical skills

Programming languages Python, Java, C, C++

Data analytics scikit-learn, pandas, Weka, Hive

Query languages SQL

Web publishing HTML, CSS

Software development kits Android SDK (Java)

Operating systems Microsoft Windows, Linux

Development environmentsMicrosoft Visual Studio, PyCharm, IntelliJ IDEA, MathWorks MATLAB

Other software Adobe Photoshop, Maxon Cinema 4D

Languages

Luxembourgishnative proficiencyFrenchbilingual proficiency

English full professional proficiency **German** full professional proficiency

Personality and interests

Personality

hard worker, results-driven, detail-minded, creative

Interests

Travelling, reading (mostly novels), photography, listening to music (particularly electronica), playing the piano, graphic design, video games (especially action-adventure and strategy games)

References

Dr. Pedro Casas

Austrian Institute of Technology (AIT) - Austria

Email: <u>pedro.casas@ait.ac.at</u>Homepage: http://pcasas.info/

Prof. Dr. Fabián E. Bustamante

Northwestern University (NU) – United States

- Email: <u>fabianb@cs.northwestern.edu</u>

Homepage: http://www.cs.northwestern.edu/~fabianb/

Prof. Dr. Benoit Donnet

Université de Liège (ULiège) - Belgium

- Email: <u>benoit.donnet@ulg.ac.be</u>

Homepage: http://www.montefiore.ulg.ac.be/personnel.php?op=detail&id=1077

Prof. Dr. Timur Friedman

Laboratoire d'Informatique de Paris 6 (LIP6) – France

Email: <u>timur.friedman@upmc.fr</u>Homepage: <u>http://www-npa.lip6.fr/~timur/</u>