

NEWSLETTER

DECEMBER 2022 | ISSUE 3

Real-time Analytics for Internet of Sports

| Marie Curie European Training Network

OBJECTIVES

RAIS aspires to provide for 14 Early Stage Researchers (PhD students) a world-class training within a broad spectrum of subjects establishing a fertile inter-disciplinary research and innovation community that will advance:

Wearable Technology

Wearable Sports Sensing and Quantified-self Devices and Accompanying Middleware

Block-chain Powered IoT

Decentralized Block-chain Powered IoT Platforms (generating hundreds of billions of transactions per day) for Big Data Mining

Real-time Edge Analytics

Real-time Edge Analytics And Predictive Modelling To Capture A Broad Range Of Sports-related Data And Trends (e.g., activities and contextual information), Critical In A Variety Of Application Settings

RAIS fellows receive a thorough “hands-on” research training as well as significant exposure to nonacademic environments through industrial secondments. Our rich set of network-wide events, including Interactive Online Seminars, entrepreneurship events, hackathons, workshops and conferences, will safeguard both fellows work as a solid team and individuals development as experts.



This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Innovative Training Networks (ITN) - RAIS No 813162





THE RAIS PROJECT



TRAIN

YOUNG SCIENTISTS



TURN IDEAS INTO PRODUCTS



FOR SOCIAL GOOD



INTERNET OF SPORTS

CONTENTS

Latest News

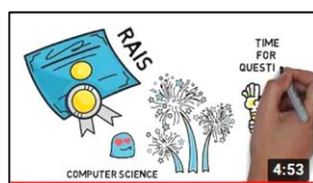
RAIS Research Meeting	3
Rais Summer School and Workshop	4
RAIS Online Seminars and Webinars	6
Rais Fellow PHD Defense Ahmed Lekssays	6

Articles Published in 2022

LifeSnaps, a 4 month multi modal dataset capturing unobtrusive snapshots of our lives in the wild	7
A sneak peak behind the scenes of a Nature Scientific data user study	
Ubiquitous Computing for Wellbeing	14
Privacy of Wearable Devices	17
Publishing data collected by wearables while protecting user privacy	19
MalRec: A Blockchain-based Malware Recovery Framework for Internet of Things	21
Towards a decentralized infrastructure for data marketplaces: narrowing the gap between academia and industry	23
Out of distribution in human activity recognition	25
Mixing temporal experts for human activity recognition	
BenchPilot: Repeatable & Reproducible Benchmarking for Edge Micro-DCs	26

Publications

27



RAIS LATEST NEWS

<https://rais-itn.eu/>

RAIS RESEARCH MEETING

Stockholm, December 16-17, 2021

The third Rais Research meeting was held on the 16th and 17th December 2021 at The Grand Hotel Saltsjöbaden, in Stockholm Archipelago, Sweden.

During the two-day event, Fellows had the opportunity to present their research activities and academic developments as well as to learn the details of future activities and management of the project.

The **Agenda** included very interesting research presentations. You may find details of the meeting at: <https://rais-itn.eu/news-post/rais-research-meeting-stockholm-december-2021>



Fellows presenting their research activities during the 3rd Research Meeting in Stockholm

RAIS SUMMER SCHOOL AND WORKSHOP

Stockholm, September 11-17, 2022

The KTH University hosted physically the
RAIS SUMMER SCHOOL III on
“Linked Data Analytics and Big Data Management”
and
RAIS WORKSHOP III on
“Data Analytics and Machine Learning”

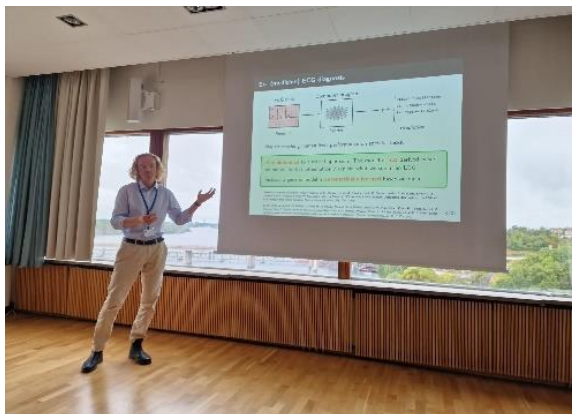


During the event many keynotes and tutorials were given by world-renowned scholars within the areas of Data Science, Machine Learning, and AI.

The **Agenda** included the following:

Keynotes:

- I. ‘The arguments and the vision for a [Personal] Data Internetwork (PDI), Nikos Laoutaris (IMDEA Networks Institute)
- II. ‘Responsible AI’, Ricardo Baeza-Yates (Institute for Experiential AI)
- III. ‘Big Data for Big Challenges: Data Analysis for Sustainable Development and Humanitarian Crises’, Ingmar Weber (Qatar Computing Research Institute (QCRI))
- IV. ‘Long ties: Formation, social contagion, and economic outcomes’, Dean Eckles (MIT)
- V. ‘Formulating flexible probabilistic models’, Thomas B. Schön (Uppsala University)



Thomas B. Schön 'Formulating Flexible Probabilistic Models'



Summer School and Workshop Participants

Tutorials:

- I. 'Opinion Formation in Social Networks: Models and Computational Problems', Aristides Gionis and Stefan Neumann (KTH Royal Institute of Technology)
- II. 'Data Democratisation with Deep Learning: An Analysis of Text-to-SQL Systems', Georgia Koutrika (Athena Research Center)
- III. 'Technological Advances in Real-World Recommendation', Flavian Vasile (Criteo AI Lab)
- IV. 'Assessing Research Impact by Leveraging Open Scholarly Knowledge Graphs', Dimitris Sacharidis (Athena Research Center)
- V. 'Introduction to Conformal Prediction', Lars Carlsson (Royal Holloway University)
- VI. 'Practical data processing in Python: a use case of sleep staging with wearable devices', Joao Palotti (Qatar Computing Research Institute (QCRI))



Georgia Koutrika Tutorial on 'Data Democratisation with Deep Learning'



Aristides Gionis during the Tutorial presentation 'Opinion Formation in Social Networks: Models and Computational Problems'



Rais Summer School and Workshop

For more details, please visit: <https://rais-itn.eu/news-post/summer-school-and-workshop>

RAIS ONLINE SEMINARS and WEBINARS

RAIS Consortium runs periodic Online Seminars and Webinars during which the ESRs give research presentations. During 2022 the following seminars and webinars were presented:

- I. Self-Supervised Learning – A high level Introduction, Lodovico Giarretta (KTH), [Video](#)
- II. Variational Autoencoders, Vangjush Komini (KTH), [Video](#)
- III. Privacy Violations in the Android Platform, Ha Xuan Son (INSUM), [Video](#)
- IV. Mixture of experts for Human Activity Recognition, Debaditya Roy (KTH), [Video](#)
- V. Graph Clustering Performance Algorithms, Susanna Pozzoli (KTH), [Video](#)
- VI. Identifying Complementarities & Substitutabilities of Physical Exercise, George Ioannou (UCY), [Video](#)
- VII. Taking Stock & Thinking Ahead, Sofia Yfantidou (AUTH), [Video](#)
- VIII. Human Activity Recognition, Debaditya Roy (KTH), [Video](#)
- IX. Machine Learning and Data Privacy, Lodovico Giarretta (KTH), [Video](#)
- X. Where's the center of the world?, Susanna Pozzoli (KTH), [Video](#)

RAIS FELLOW PHD DEFENSE

On the 16th December 2022, **Ahmed Lekssays** (RAIS-INSUB-ESR7) successfully defended his Ph.D. thesis entitled '**Blockchain-based Malware Incident Response for Internet of Things**'. The thesis was conducted under the mentorship of Prof. Elena Ferrari and Barbara Carminati. **Rais Consortium** congratulates Dr. Ahmed Lekssays and wishes him lots of success in the future.

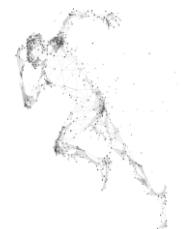


Abstract:

IoT devices have grown drastically in the last decade because of their usefulness in many industrial applications. They have increased from 13.4 billion in 2015 to 38.5 billion in 2020, 285%. Due to this growth, IoT devices have become an attractive target for attackers to perform various attacks, such as DDoS (Distributed Denial of Service). IoT devices are considered the weakest link in companies' security chain since they are not usually well tested and secured against cyber-attacks due to, for example, the adoption of weak passwords and unencrypted network services. In addition, they have low computation power to run sophisticated security solutions. As a result, attackers can easily inject malicious software (malware) into IoT devices to take control of them or steal private information. In this dissertation, we focus on malware threats and leverage the NIST SP 800-83 Malware Incident Prevention and Handling guidelines, one of the leading guidelines in malware incident response. The NIST SP 800-83 has four phases: preparation, detection, containment and eradication, and recovery. The preparation phase is mainly for raising awareness about malware threats in companies, so it is out of scope for this dissertation. More precisely, in this dissertation, we focus on the three last phases of NIST SP 800-83: detection, containment, and recovery. Additionally, we focus on collaboratively mitigating malware attacks since IoT networks are heterogeneous and involve several organizations. Thus, it is important that the malware incident response process is collaborative and based on threat information sharing. Since this collaborative process can involve several organizations that do not trust each other, an effective framework for detecting, containing, and recovering from malware attacks is needed to ensure the traceability and integrity of the shared threat information and the implemented mitigation actions. This dissertation proposes a generic framework based on NIST SP 800-83 guidelines that leverages blockchain. We leverage blockchain to ensure the correct execution of NIST SP 800-83 guidelines to solve the issue of weak trust relationships that might hold among the involved organizations. In addition, blockchain offers smart contracts which are autonomously executed programs where custom logic can be encoded. Blockchain guarantees that the execution of the encoded logic was correct. To this end, it ensures the accountability, integrity, and immutability of the shared data and the execution of smart contracts.



ESR 13 CHRISTINA KARAGIANNI | ARISTOTLE UNIVERSITY OF THESSALONIKI (AUTH) | GREECE
ESR 14 SOFIA YFANTIDIOU | ARISTOTLE UNIVERSITY OF THESSALONIKI (AUTH) | GREECE



LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild

A sneak peek behind the scenes of a Nature Scientific Data user study.

Original article published in Springer's Research Data Community [here](#).

Publication in Nature Scientific Data [here](#).

Background

The newly published data descriptor paper, LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild, introduces a new public dataset empowering future research in different disciplines from diverse perspectives. Pervasive self-tracking devices have penetrated numerous aspects of our lives, from physical and mental health monitoring to fitness and entertainment. Nevertheless, limited data exist on the association between in-the-wild large-scale physical activity patterns, sleep, stress, and overall health, and behavioral and psychological patterns due to challenges in collecting and releasing such datasets, including waning user engagement or privacy considerations. The LifeSnaps dataset is a multi-modal, time, and space-distributed dataset containing a plethora of data collected unobtrusively for more than 4 months by 71 participants. LifeSnaps contains more than 35 different data types totaling more than 71M rows of data. The participants contributed their data through validated self-reported surveys, ecological momentary assessments (EMAs), and a Fitbit Sense smartwatch and consented to make these data available to empower future research. We envision that releasing this large-scale dataset of multi-modal data will open novel research opportunities and potential applications in multiple disciplines.

LifeSnaps contributions

- Privacy and anonymity in accordance with the EU's General Data Protection Regulation (GDPR)
- Large-scale data collected in-the-wild, where participants continued their normal daily routines
- Use of diverse human-centric data modalities with rich data granularity
- Emerging data types rarely studied until now, such as temperature, oxygen saturation, heart rate variability, automatically assessed stress, and sleep phases
- Community code sharing for reproducibility facilitating future research

A sneak peek behind the scenes

Innovative Training Network “Real-time Analytics for the Internet of Sports” (RAIS) aspires to provide 14 Early Stage Researchers a world-class training within a broad spectrum of subjects establishing a fertile interdisciplinary research and innovation community that will advance wearable technology, Block-chain Powered IoT, and Real-time Edge Analytics. The RAIS consortium consists of five partner universities in Europe (KTH, University of Cyprus, University of Insubria, FORTH, AUTH) and seven institutions and companies all around the world.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Innovative Training Networks (ITN) - RAIS No 813162

The RAIS Consortium Experiment officially started on May 25, 2021.

Time has come for the RAIS Consortium Experiment ▷



Sofia Yfantidou <@gmail.com>
to partners, fellows ▾

Tue, 25 May 2021, 18:42 ☆ ↶ ⋮

Hello everybody,

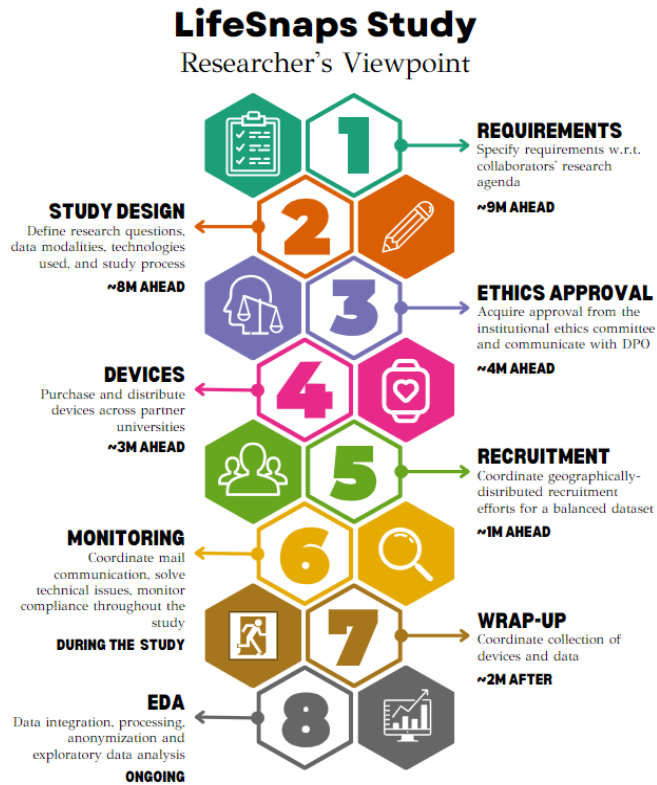
I just wanted to inform you that today marks the start of the RAIS Consortium Experiment! 🎉 Hopefully, our participants will stay engaged throughout the experiment and we will receive a plethora of activity, sleep, mood, and related questionnaire data.

As a side note, along with the RAIS fellows, we've recruited 38 out of 42 participants. There are 2 Fitbit devices vacant at FORTH and 2 at KTH. If anybody at the local labs is interested in joining the experiment, it's still possible to onboard this week. Nevertheless, we have 41 participants in total (we had 3 additional Fitbit devices at AUTH) who will be collecting data for 2 months each, which should be sufficient.

If the social media managers want to publish something about the start of the experiment in RAIS social media channels, they can find all the related information [here](#). If you need any further clarifications, let me know.

Best regards,
Sofia

But the real work started way before, around nine months earlier...



Baffled by the lack of open datasets as early-stage researchers, we had the crazy idea of collecting our own, undaunted by the challenges of the feat.

Working within the RAIS Consortium, we collected data requirements from all our RAIS collaborators working on diverse research fields, from artificial intelligence to human-computer interaction to privacy and security. Based on these multifaceted requirements we drafted our study protocol and evaluated different technology products for suitability.

The questions abandoned: Which wearable device covers most of our data requirements? Do they provide sufficient documentation? Are there available APIs? Which option allows us to recruit from a larger participant pool? The answers to these questions led us to Fitbit Sense, Fitbit's flagship smartwatch at the time, which had been released just a few months ago.

But before moving forward with any irrevocable actions, and in accordance with the requirements for ethical research and the GDPR regulation, we submitted our research protocol and consent form to the university's institutional review board (IRB) for review and drafted a data management plan in collaboration with the university's data protection officer (DPO), to do right by our users. On February 23rd, 2021 -just four months before the official start of our study-, we received the final green light, and we were delighted to kickstart our project!

**ΑΠΟΦΑΣΗ ΤΗΣ ΕΠΙΤΡΟΠΗΣ ΗΘΙΚΗΣ ΚΑΙ ΔΕΟΝΤΟΛΟΓΙΑΣ ΤΗΣ ΕΡΕΥΝΑΣ ΤΟΥ ΑΠΘ
ΓΙΑ ΕΓΚΡΙΣΗ ΕΡΕΥΝΗΤΙΚΟΥ ΕΡΓΟΥ**

Τίτλος Ερευνητικού έργου: «Ανάλυση Δεδομένων από Φορητές Συσκευές - Real-time Analytics for the Internet of Sports»

Η Επιτροπή Ηθικής και Δεοντολογίας της Έρευνας του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης (ΕΗΔΕ ΑΠΘ), στη Συνεδρίασή της με αριθμό 25/22-02-2021, αφού έλαβε υπόψη:

- α. την αίτηση της Αθηνάς Βακάλη, Καθηγήτριας του Τμήματος Πληροφορικής της Σχολής Θετικών Επιστημών του ΑΠΘ (αρ. πρωτ. 36132/2021 & 39941/2021),
- β. τη Δήλωσή της ότι έλαβε γνώση του Κανονισμού Αρχών και Λειτουργίας της ΕΗΔΕ ΑΠΘ και ότι αναλαμβάνει την υποχρέωση συμμόρφωσης και τήρησής του,
- γ. το υποβληθέν Ερευνητικό πρωτόκολλο,
- δ. το Ερωτηματολόγιο – έκθεση αυτοαξιολόγησης που προσκομίστηκε,
- ε. το έντυπο ενημέρωσης και συγκατάθεσης που θα δοθεί στους συμμετέχοντες και
- στ. τις διατάξεις του Κανονισμού Αρχών και Λειτουργίας της ΕΗΔΕ ΑΠΘ και της κείμενης νομοθεσίας

αποφάσισε

την έγκριση διεξαγωγής του ερευνητικού έργου με τίτλο «**Ανάλυση Δεδομένων από Φορητές Συσκευές - Real-time Analytics for the Internet of Sports**», καθώς κατά την κρίση της, τηρούνται οι παραδεκτοί κανόνες ηθικής και δεοντολογίας και ακεραιότητας της έρευνας ως προς το περιεχόμενο και ως προς τον τρόπο διεξαγωγής της, καθώς και οι εκ του νόμου προβλεπόμενες προϋποθέσεις.

Η παρούσα απόφαση της ΕΗΔΕ ΑΠΘ σε καμία περίπτωση δεν υποκαθιστά άλλη απαιτούμενη έγκριση ή αδειοδότηση από άλλη αρμόδια δημόσια υπηρεσία, διοικητικό όργανο ή ανεξάρτητη διοικητική Αρχή, που δύναται επιπλέον να απαιτείται εκ του νόμου.

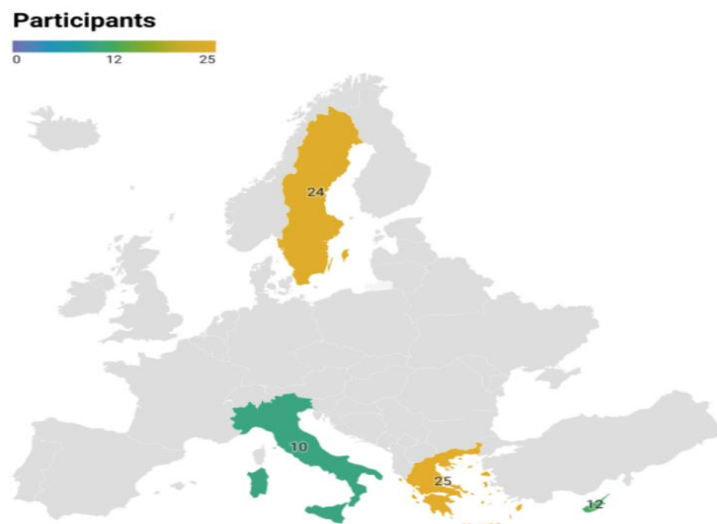
Ο Πρόεδρος της ΕΗΔΕ ΑΠΘ
**Dimitrios
Stamovlasis**
 Δημήτριος Σταμοβλάσης

Αν. Καθηγητής Τμήματος Φιλοσοφίας και Παιδαγωγικής ΑΠΘ

Our potential participant pools were located in our partner universities in four different countries, namely Greece, Cyprus, Italy, and Sweden. Yet, physical meetings were impossible at the time due to the first wave of the COVID-19 pandemic and the strict lockdowns it introduced. That's why, after the purchase of our Fitbit Sense devices we had to distribute them across our partners via registered post.



Each partner was then responsible for disseminating the call for volunteers and recruiting participants for the RAIS Consortium Experiment. The process was overseen by fellows at the Aristotle University of Thessaloniki and special care was given to a gender-balanced recruitment. In total, across both study rounds, we recruited 25 participants in Greece, 24 in Sweden, 12 in Cyprus, and 10 in Italy.



Throughout the duration of the study, all participants received weekly hand-drafted emails (with useful Fitbit tips!) encouraging them to continue contributing to the study and asking them to complete certain surveys.



Welcome to the RAIS Consortium Experiment XXX

Thank you for joining the RAIS Consortium Experiment. Today marks the official start of the experiment which will last till the 1st of August. Below you will find all the information you need to know.


- **First things first:** [Here](#) you'll find all the necessary information about this experiment. don't hesitate to contact us if you need any further clarifications. Also, if you haven't signed our [consent form](#), please do so as soon as possible. 🙌
- **SEMA3 App:** Today, we sent out the SEMA IDs for each participant to connect to the SEMA3 app. If you haven't received the e-mail (please check your spam folder), contact us to resolve the issue. Through this app you'll receive a few short questions per day regarding your mood and step goals. Try to answer as many as possible. 🙌
- **Initial Questionnaires:** To get to know you better, you will need to answer a few short questionnaires during this experiment. This week, you need to complete a [Demographics Questionnaire](#) and a [Physical Activity Readiness Questionnaire \(PAR-Q\)](#). Please complete them by Sunday 30th of May. ✍️

That's all for now! We hope that you will enjoy your participation and that you will get to know more about your physical activity habits through your Fitbit.

Have a nice week,
Sofia - RAIS Consortium Experiment Team



We also adopted many monitoring processes to have a first look at **statistics on compliance in real-time** to encourage less engaged participants to contribute to the study.

	RAIS Consortium Experiment - Welcome	Sent	82.5%	67.5%	View Report
	Regular · RAIS Consortium		Opens	Clicks	
	Tags: rais.experiment				
	Sent Mon, May 24th, 2021 7:04 AM to 41 recipients by you				

Once the study’s nine weeks had passed, we had to, unfortunately, bid our participants goodbye. We asked them to complete their last surveys and share their data with us, again reminding them of the importance of **data privacy** and **security** for the RAIS Consortium and the scientific community. And we did hold true to our promise; a dedicated data anonymization team was working for months after the end of the user studies to **verify the anonymity of the LifeSnaps dataset before publication**.



That’s it, XXX ! You’ve made it to the end!

This week came to an end, summer is hitting hard, and you have been an awesome participant. Kudos! This is the last week of the **experiment** and here are the final steps you have to take:

- **Your weekly questionnaires:** You’ve done great work answering all our questionnaires so far! As usual, you’ll need to complete your weekly questionnaires: the **State-Trait Anxiety Inventory (STAI)** and the **Positive Affect Negative Affect Scale (PANAS)**. Please complete them by *Sunday 1st of August*. 📄

- **Your data:** Please export and share your Fitbit data with the **RAIS Consortium** as per [this document](#) by *Sunday, 1st of August*. If you are facing any difficulties, please contact your recruiting researcher. They’ll be happy to help! ❤️

We take all the necessary precautions to keep your data safe and private in a secure server at the Aristotle University of Thessaloniki, Greece, a **RAIS** partner. We do not and will not share your data with any unauthorized third-party service. Eventually, your data will be anonymized -so your identity remains a secret- and shared with the scientific community to help advance research in the mHealth domain, as specified in the [FAQ page of our website](#). 🙏

- **Your final questionnaires:** There are two more short surveys we’d like you to answer to report your progress during this **experiment**. Specifically, the **Stages and Processes of Behavior Change scale** and the **Behavioral Regulation in Exercise scale**. Please complete them by *Sunday 1st of August*. 📄
- **Your watch:** We know you might love your Fitbit by now, but unfortunately, it’s time to say goodbye. Please get in touch with your recruiting researcher to return the watch. 🕒

We can’t thank you enough for your participation! We hope you enjoyed your time with the **RAIS Consortium Experiment** and got to know more about your physical activity habits through your Fitbit.

Have a nice summer,
Sofia - **RAIS Consortium Experiment** Team



A real-world use case

An indicative real-world use case of LifeSnaps is the mental healthcare sector. Finding out the trajectories of a patient's psychological traits often requires repeated verbal interactions. Traditional in-person interviews are not always preferable because of the economic burden on clinicians and patients. Additionally, self-reported questionnaires can be problematic as they are based on patient recollections and self-representation. These two tools can be combined and enriched with passively and objectively collected behavioral data. That's exactly what LifeSnaps is, and hopefully, it will facilitate research in the mental health domain, advancing prevention and treatment.

Other future work scenarios





ESR 14 SOFIA YFANTIDIOU | ARISTOTLE UNIVERSITY OF THESSALONIKI (AUTH) | GREECE

Ubiquitous Computing for Wellbeing

Ubiquitous computing is not a specific technology but a scenario in which computers become more numerous and fade into the background, providing information to human users and embedding intelligence and computing capabilities in seemingly everyday objects, such as personal digital devices. Their prevalence in our society is growing, with more than 6.6B smartphones and 1B connected wearable devices worldwide in 2022, and almost 70M users in the US alone, or, in other words, 1 in 5 US citizens. Such a transition has been called the “Era of Digital Phenotyping”, inspired by the notion of biological phenotype. Digital phenotyping is the moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices.

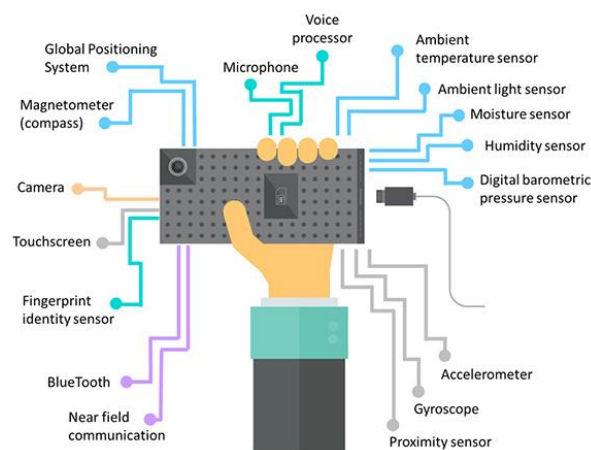


Figure 1 The endless possibilities of digital phenotyping captured through the variety of data modalities

My work within RAIS surrounds the study of this emerging era, focusing on exploiting the user’s digital phenotype to improve their physical and mental wellbeing. Specifically, I have worked with four distinct domains concerning ubiquitous computing: Human-Computer Interaction, User Studies, Machine Learning, and Computational Fairness. In 2022, I published, in collaboration with my colleagues’ various papers touching upon these domains.

At PerCom 2022 conference, we presented WearMerge [1], an interoperable digital phenotype data integration, and standardization framework. Due to the rapid increase in volume, variety, and variability of generated sensor data, integrating data from different personal digital devices for further exploration and analysis has become extremely time-consuming. In addition, it requires advanced technical skills, hindering their widespread adoption in interdisciplinary scientific and industrial research. In our work, we have introduced an extensible, open-source framework and tool called WearMerge that automates the integration and transformation of heterogeneous wearable devices’ data into a common standard across different brands and models. We have accompanied our theoretical contributions with a practical tool to help and ease practitioners and researchers on digital phenotype data analysis.

However, we did not stop there. At the same conference, we presented WeMoD [2], a machine-learning approach for physical activity prediction with personal digital device data. It is indisputable that physical activity is vital for an individual's health and wellbeing. However, one in four adults globally does not meet the recommended activity levels, with substantial personal and socioeconomic implications. In recent years, a significant amount of work has explored the potential of ubiquitous computing for increasing physical activity. Adaptive and personalized goal-setting has proven to be one of the most efficient methods in this direction. To this end, we have proposed a machine learning approach, WeMoD, to predict a user's future daily step count for setting challenging yet achievable goals. To develop WeMoD, we have utilized heterogeneous, multimodal human data collected unobtrusively in the wild. Additionally, we have used a novel fusion of physiological, behavioral, and contextual features, which according to the experimental results, has had a positive effect on the predictive capacity of the models.

As a continuation of the work above, at HealthCom 2022 conference, we proposed UBIWEAR, an end-to-end framework for intelligent physical activity prediction, aiming to empower data-driven goal-setting interventions. To achieve this, we have experimented with numerous machine learning and deep learning paradigms as robust benchmarks for physical activity prediction tasks. To train our models, we have utilized "MyHeart Counts", the largest open dataset collected in the wild from thousands of users. We also proposed a prescriptive framework for digital phenotype aggregated data preprocessing to facilitate data wrangling of real-world, noisy data. Our best model has achieved a MAE of 1087 steps, 65% lower than the state of the art in terms of absolute error, proving the feasibility of the physical activity prediction task and paving the way for future research.

Finally, within the domain of Human-Computer Interaction, an emerging wave of research has been exploring the potential of interactive personal digital technology in encouraging positive health behavior change. Numerous findings indicate the benefits of personalization and inclusive design regarding increasing the motivational appeal and overall effectiveness of behavior change systems, with the ultimate goal of empowering and facilitating people to achieve their goals. However, most interventions still adopt a "one-size-fits-all" approach to their design, assuming equal effectiveness for all system features despite individual and collective user differences. To this end, for the UMAP 2022 conference, we have analyzed a corpus of 12 years of research in personal digital technology for health behavior change, focusing on physical activity, to identify those design elements that have proven most effective in inciting desirable behavior across diverse population segments [4]. Our results can be explored through our publicly available tool and corpus. We have then provided actionable recommendations for designing and evaluating behavior change technology based on age, gender, occupation, fitness, and health condition. Through this work, we have engaged in a critical commentary on the diversity of the domain and discussed ethical concerns surrounding tailored interventions and directions for moving forward.

Summing up, the field of ubiquitous computing for wellbeing is still in its infancy, but the era of "Digital Phenotyping" provides exciting opportunities for future work and leaves multiple research questions unanswered.

References:

- [1] Giakatos, D. P., Yfantidou, S., Efstathiou, S., & Vakali, A. (2022, March). WearMerge: An Interoperable Framework for Self-tracking Data Integration and Standardization. In 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) (pp. 76-78). IEEE.
- [2] Vasdekis, D., Yfantidou, S., Efstathiou, S., & Vakali, A. (2022, March). WeMoD: A Machine Learning Approach for Wearable and Mobile Physical Activity Prediction. In 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) (pp. 385-390). IEEE.

[3] Bampakis, A., Yfantidou, S., & Vakali, A. (2022, October). UBIWEAR: An end-to-end, data-driven framework for intelligent physical activity prediction to empower mHealth interventions. In 2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom Main Track). IEEE.

[4] Yfantidou, S., Sermpezis, P., & Vakali, A. (2022, July). 12 Years of Self-tracking for Promoting Physical Activity from a User Diversity Perspective: Taking Stock and Thinking Ahead. In the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP Workshops) (pp. 211-221). ACM.



ESR 5 ANDREI KAZLOUSKI | FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS (FORTH) | GREECE



Privacy of Wearable Devices

During the course of 2022 myself and other colleagues of mine within the RAIS project have published a number of research works. In this newsletter I describe the two principal studies on privacy of wearable devices and the data generated by them.

Kazlouski, A., Marchioro, T. and Markatos, E.P., 2022. **“What your Fitbit says about you: De-anonymizing users in lifelogging datasets.”** In SECURE (pp. 341-348)

Given the recent increase of so-called lifelogging experiments (where the activity of few participants is monitored for a number of days via fitness trackers), we set to investigate whether the existing public data can be considered “safe”. Since data from such experiments are often aggregated in datasets and released to the research community, privacy for participants of the studies may be at risk. Furthermore, given that the RAIS project had been planning to release our own lifelogging dataset, it was of utmost importance to study the currently employed defense mechanisms.

To protect the privacy of the participants, fitness datasets are typically anonymized by removing personal identifiers such as names, e-mail addresses, etc. However, although seemingly correct, such straightforward approaches are not sufficient. In our paper we demonstrate how an adversary can still de-anonymize individuals in lifelogging datasets. We show that for readily available public lifelogging datasets privacy of the participants can be compromised.

We propose two de-anonymization approaches: (i) via the inference of physical attributes, namely gender, height, and weight; and (ii) through the daily routine of participants. In the first threat model the adversary learns the undisclosed parameters of the owners of fitness data from the daily steps, calories, and distance, as recorded by smart fitness trackers. In the second threat model the attacker fingerprints users based on their hourly snippets of data, more specifically tuples of steps, distance, calories, and average heart rate for every hour from 00:00 to 23:00. In both methods the adversary utilizes exclusively fitness and no other information to learn undisclosed insights on the users in readily available public lifelogging datasets.

We train several inference models of various architectures to de-anonymize users in public lifelogging datasets and learn their physical parameters. We achieve up to 93.5% re-identification rate of participants. Furthermore, we show that de-anonymization can be even more effective, reaching 100% success rate for people with distinct physical attributes (e.g., very tall, overweight, etc.).

We consider this work when we make effort to protect privacy of the participants for the RAIS lifelogging dataset.

Kazlouski, A., Marchioro, T. and Markatos, E.P., 2022, November. **“I just wanted to track my steps! Blocking unwanted traffic of Fitbit devices.”** In 12th International Conference on the Internet of Things, (To be published).

While commercial fitness trackers have reached unprecedented market penetration, consumers are left with very little control over what entities are contacted by their wrist-worn devices. Therefore, we set to investigate whether it is feasible for average users of smartbands to limit exposure of their private information to the unknown entities. Prior works have shown that fitness apps associated to wearables often contact unexpected third parties, but it remains unclear whether all of such connections are essential for the correct functioning of the device. In our work we identify the third parties that are being contacted by the official mobile Fitbit application and its partners. Furthermore, we study whether it is feasible to block the unwanted destinations without hindering the essential functionality of the devices. We conduct a 2-rounds experiment, comparing the activity data of 2 simultaneously-worn Fitbit Versa 2 wristwatches. In our setup one of the devices communicates the data via the official companion app that contacts all the default network destinations; for the other smartband we block all unwanted connections. We show that disabling traffic to the domains contained in well-maintained blocklist does not prevent Fitbit devices from correctly synchronizing various fitness parameters, including steps, workouts, duration and quality of sleep, etc.

Moreover, we demonstrate that Fitbit activity data are correctly synchronized for 6 partner apps of Fitbit when utilizing the above blocking rules. Our results suggest that more than 88% of the third parties for the Fitbit-associated apps are contained in credible domain-based blocklists. Furthermore, we find all studied app to contact between 1 and 20 non-required third parties. We establish Facebook (Meta) and Google to be the most contacted entities; and Google and Amazon to have the highest number of entries in the blocklists. Finally, over 50% of the blocked destinations are identified by the default installation of uBlock Origin – universally used content filtering tool (better known as adblockers).

Unlike previous works on blocking unnecessary IoT communications, our methodology can be easily utilized by end-users and does not require specific network equipment.



ESR 6 THOMAS MARCHIORO | FOUNDATION FOR RESEARCH AND TECHNOLOGY – HELLAS (FORTH) | GREECE



Publishing data collected by wearables while protecting user privacy

The privacy risks of fitness data sharing are often underestimated by both users and health researchers. This is typically the case in fitness-related lifelogging experiments, where a small groups of users is recruited and monitored through wearable fitness trackers. These experiments may range from physical activity interventions to patient rehabilitation strategies and involve gathering a large amount of data from the participants.

Indeed, data constantly collected by wearable sensors – known in literature as wearable IoT data – can easily become a fingerprint for the user who produced them. Thus, publishing them may expose users' private information.

Previous papers published by our research group demonstrate that users can actually be identified based on time series records of fitness information such as steps, calories, and distance.

Our current work focuses on mitigating privacy risks via anonymization/sanitization of wearable IoT data in different scenarios.

In [1], my colleagues and I devised a set of guidelines that should be applied on a fitness dataset before publication. These rules were conceived after a thorough examination of existing public datasets, determining which aspects are more likely to hinder the participants' anonymity.

A first necessary step to take consists in applying k -anonymity with respect to the demographic and physiological attributes, such as age, gender, height, and weight. This means that these attributes should be generalized or suppressed until at least k individuals in the dataset have them in common, where k is chosen depending on the sensitivity of the data.

Concerning time series data (e.g., steps, calories, etc.) k -anonymity cannot be applied without completely hindering their utility. A more conservative strategy could be re-sampling the records at a lower frequency, e.g., aggregating them by day rather than by hour. Another possibility is to reduce the granularity of the data, which is similar to the generalization concept in k -anonymity. For instance, a precise value like 8315 could be binned in the range 8000-8500.

Furthermore, a general rule to be followed is the data minimization principle. When a dataset is collected with a specific objective in mind (e.g., comparing the activity of different demographic groups), the published data should reflect solely that objective. Data that are unrelated to the purpose should not be disclosed.

This work was published in the proceedings of the 32nd Medical Informatics Europe Conference (MIE 2022).

In [2], we designed a decentralized algorithm to compute Naïve Bayes models in a federated learning setting under differential privacy guarantees.

Federated learning has lately been gaining momentum as a solution to collaborative training of machine learning models across different data partitions. For instance, it is used to train text prediction models on personal devices without collecting private text messages.

However, research on federated learning is often focused on complex models such as artificial neural networks. In our work, instead, we aim to facilitate decentralized training of simpler models

such as Naïve Bayes. These models, albeit less powerful, typically require less data and provide more consistent results.

Our implementation of federated Naïve Bayes allows data owners to compute a single local update on their data partitions. The updates are collected by a central server which combines them into the final Naïve Bayes model.

We designed the federated algorithm so that differential privacy can be enforced. Differential privacy is a mathematical models that can be achieved through randomization. It guarantees the presence of an individual data point cannot be inferred from the final model, making it suitable for cases where the training data contain sensitive information.

We evaluated federated Naïve Bayes for different privacy levels and found that a favorable privacy-utility tradeoff can be achieved on different benchmark datasets.

The work was published in the proceedings of the 19th International Conference on Security and Cryptography (SECRYPT 2022).

References:

- [1] Marchioro T, Kazlouski A, Markatos E. How to Publish Wearables' Data: Practical Guidelines to Protect User Privacy. *Studies in Health Technology and Informatics*. 2022 May 1;294:949-50.
- [2] Marchioro T, Giaretta L, Markatos E, Girdzijauskas Š. Federated Naive Bayes under Differential Privacy. In *19th International Conference on Security and Cryptography (SECRYPT)*, JUL 11-13, 2022, Lisbon, Portugal 2022 (pp. 170-180). Scitepress.



ESR 7 AHMED LEKSSAYS | UNIVERSITY OF INSUBRIA (INSUB) | ITALY



MalRec: A Blockchain-based Malware Recovery Framework for Internet of Things

According to PurpleSec [1], cyberattacks had a 600% increase during the COVID-19 pandemic in 2020. Hackers exploit security vulnerabilities to perform malicious activities such as data theft, espionage, etc. Cybercriminals generally attack different targets using malware (i.e., malicious software) such as viruses, worms, spyware, ransomware, etc. In addition, IoT devices are estimated to reach 41.6 billion by 2025 with an overall annual growth rate of 28.7% from 2018 to 2025 [2]. However, they suffer from various security issues such as open ports, system updates not being performed, and weak security mechanisms. This combination of large numbers of IoT devices and low-security measures poses a serious threat to organizations' infrastructures since IoT devices are the weakest link. Moreover, with the development of data processing regulations in different countries, a need for transparent recovery systems that can help organizations present their due diligence arises. One of the main challenges for recovery systems is ransomware attacks. Ransomware are a type of malware that encrypts files and make devices unresponsive. In 2021, only 6 ransomware families were responsible for a loss of \$45 million paid in cryptocurrencies to their developers. They attacked, in the US only, 600 hospitals through medical IoT devices [3]. To reduce the impact of different malware attacks on IoT systems, we introduce a malware recovery framework, called MalRec [4], that allows compromised devices to recover and return to a safe state.

MalRec exploits blockchain to help companies comply with different standards and regulations (e.g., GDPR, ISO/IEC 27040, NIST SP 800-171, etc.) by enforcing backup policies through smart contracts, which are autonomously executed programs hosted on the blockchain. For example, suppose an organization must comply with a backup policy that states that the organization must have daily backups with 3 different replicas in different storage media. In that case, a smart contract can ensure that any submitted backup complies with this policy by verifying the submission time and the locations of the backups. Smart contracts can also be used to retrieve backups information that satisfies a specific filter, such as the submitter and the timeframe when the backups were submitted. Blockchain ensures a transparent and immutable verification of the policies that companies may want to enforce. In addition, it ensures that all interactions with the blockchain are authenticated through cryptographic primitives. To ensure the confidentiality of the stored backups MalRec exploits public-key cryptography to encrypt backups before uploading. Furthermore, it ensures the integrity of the backups via digital signatures. Finally, since blockchain, by design, is transparent, it provides accountability and audit features to the whole process.

References:

[1] <https://purplesec.us/resources/cyber-security-statistics/>

[2] <https://www.zdnet.com/article/iot-devices-to-generate-79-4zb-of-data-in-2025-says-idc/>

[3] <https://illinois.touro.edu/news/the-10-biggest-ransomware-attacks-of-2021.php>

[4] Lekssays, A., Sirigu, G., Carminati, B., & Ferrari, E. (2022, August). MalRec: A Blockchain-based Malware Recovery Framework for Internet of Things. In Proceedings of the 17th International Conference on Availability, Reliability and Security (pp. 1-8).



ESR 3 LODOVICO GIARETTA | ROYAL INSTITUTE OF TECHNOLOGY (KTH) | SWEDEN



Towards a decentralized infrastructure for data marketplaces: narrowing the gap between academia and industry

Authors: Lodovico Giaretta (RAIS ESR3, KTH), Thomas Marchioro (RAIS ESR6, FORTH), Evangelos Markatos (RAIS, FORTH), Sarunas Girdzijauskas (RAIS, KTH)

Extended Abstract

Our society produces larger and larger amounts of data every year, thanks to the increasing usage of digital services and the widespread adoption of IoT and connected devices, from smartwatches to home assistants to electric vehicles.

At the same time, the fields of machine learning and artificial intelligence have seen incredible progress and are now powering advanced digital services in a variety of industries and use-cases. This meteoric rise has been fueled by the availability of larger amounts of data and by advances in both software and hardware.

However, a point of great friction has been reached. Machine learning-focused companies and startups are struggling to expand the use of artificial intelligence into new sectors and use-cases, as that requires large amounts of data from different sources, which are not available in-house and are hard to obtain. On the other hand, large amounts of data sit unused on user devices or in the data lakes of large organizations, who do not have the expertise or use-cases to exploit them. Our society as a whole has both the data and the know-how to exploit it, but these are often separated.

Thus, the so-called “data economy” is quickly gaining importance, centered around the idea of data marketplaces, where entities that need data can obtain it, and entities who own data can monetize it. The key technologies behind data marketplaces have been widely studied by an interdisciplinary scientific community focused on the data economy, and several commercial marketplaces exist in the industry.

However, there is a growing disconnect between academic research and practice. The former has been focused on the many important questions that arise in the data economy: how to maximize data sharing, ensuring data producers and consumers can be efficiently matched and useful data easily discovered? How to ensure that sensitive data remains private and is not harmfully leaked by either data producers or consumers? How to measure the value of the data being traded? On the other hand, current industrial marketplaces behave more like the digital equivalent of brick-and-mortar shops, without specific provisions for openness, transparency, data discoverability, data privacy or data valuation.

In this work, we look at this growing disconnect, reviewing many of the key enabling technologies proposed by the research community to tackle the aforementioned questions, and asking ourselves which of these could be feasibly be implemented in the short term on top of the current industrial landscape, to narrow the gap between academia and industry.

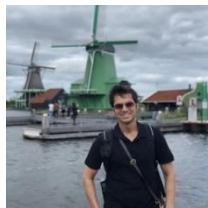
We start by noting that the current industrial landscape is extremely fragmented, with hundreds of small marketplaces catering to different sectors and employing a variety of business models. This makes it very challenging for consumers to discover data potentially useful to them, and also forces producers to put substantial efforts to publish their data in multiple systems, to ensure maximum data exploitation. We can expect consolidation to happen in the future, but believe that such

process would lead to a different set of problems, namely transparency and fairness. We therefore argue in favor of a federation of marketplaces based on a permissioned blockchain, where full transparency together with verifiable identities can ensure openness and trust. Data discovery can be easily built on top of such a federation via third party entities, which can also provide a variety of other value-add services by accessing marketplace metadata through a transparent and standardized interface.

On the topic of privacy, we note that, as soon as any entity “sees” the data, no privacy can be guaranteed. Data can be cheaply copied and transmitted, and anonymization techniques can be easily defeated by combining multiple data sources, which in exactly what an open data economy incentivizes. Thus, data needs to be processed “blindly”: it must be stored and transmitted in encrypted form, and can only be allowed to be decrypted in Trusted Execution Environments (TEEs), specialized hardware devices that are secured to avoid any leaks. But even that is not enough: once a ML model has been trained in a TEE, it needs to be extracted and used by the consumer. But the parameters of the model itself and the outputs of its usage may be leaking some of the data used to build it! Therefore, we argue for the use of an additional technique, called Differential Privacy, to further secure the trained model against these types of leakage. Finally, we argue that scalability to complex model and large data sets from multiple providers can be achieved by employing decentralized training techniques such as Federated Learning and Gossip Learning.

The last key issue we consider is that of data valuation. In an economy fueled by machine learning models and complex data analytics, the value of a dataset is typically not fixed, but depends on the use-case for which it is sought, on the algorithms used to process it, and on the other datasets it is combined with. We therefore split the data valuation issue into two key questions. The first is that of a priori data selection: in a vast ocean of available datasets, how can a consumer choose which datasets to use? The second is that of a posteriori value apportionment: having select a number of datasets, combined them and fed them into an algorithm, how can the consumer establish how much of the business value extracted by the algorithm can be traced back to each of the datasets employed? We argue that both these aspects are crucial and discuss several techniques to achieve precise data valuation.

Overall the gap between academia and industry is currently quite wide. However, we believe that there are feasible paths to bridge this gap. We hope that further engagement by the research community, industrial leaders and the wider public will lead to following these paths to achieve a more sustainable, robust and thriving data economy.



ESR 9 DEBADITYA ROY | ROYAL INSTITUTE OF TECHNOLOGY (KTH) | SWEDEN



Out of distribution in human activity recognition

With the growing interest of the research community in making deep learning (DL) robust and reliable, detecting out-of-distribution (OOD) data has become critical. Detecting OOD inputs during test/prediction allows the model to account for discriminative features unknown to the model. This capability increases the model's reliability since this model provides a class prediction solely at incoming data similar to the training one. OOD detection is well-established in computer vision problems. However, it remains relatively under-explored in other domains, such as time series (i.e., Human Activity Recognition (HAR)).

Since uncertainty has been a critical driver for OOD in vision-based models, the same component has proven effective in time-series applications. We plan to address the OOD detection problem in HAR with time-series data in this work.

To test the capability of the proposed method, we define different types of OOD for HAR that arise from realistic scenarios. We apply an ensemble-based temporal learning framework that incorporates uncertainty and detects OOD for the defined HAR workloads. In particular, we extract OODs from popular benchmark HAR datasets and use the framework to separate those OODs from the in-distribution (ID) data. Across all the datasets, the ensemble framework outperformed the traditional deep-learning method (our baseline) on the OOD detection task.

Furthermore, the method is applied to the practical use case of elderly fall detection, where falls are detected as out-of-distribution cases.

Mixing temporal experts for human activity recognition

Temporal patterns are encoded within the time-series data, and neural networks process those patterns with their unique feature extraction ability to provide a better predictive response.

Ensembles of neural networks have proven to be very effective in Human Activity Recognition (HAR) tasks with time-series data, e.g., wearable sensors. The combination of predictions coming from the individual models in the ensemble helps boost the overall classification metric through efficient temporal pattern recognition. Currently, simple averaging is the most common strategy for combining the predictions coming from individual models. However, since each ensemble model learns different temporal patterns of the time-series classification problem, a simple averaging strategy is sub-optimal.

This paper addresses this sub-optimality through a neural network-based adaptive learning framework. The method's core is training a neural gate that ingests the same input time-series data fed to the other temporal models. The goal of the training process is to adaptively learn scaler values against each temporal model by looking at the input data. These scaler values weigh each temporal model while combining the ensemble. The framework obtains superior predictive performance as compared to the standard ensembling techniques. The framework is evaluated on a benchmark HAR dataset called PAMAP2 with two popular state-of-the-art ensemble architectures, DTE and LSTM-ensemble. In both cases, the classification performance of the framework in HAR tasks surpasses the state-of-the-art models.



ESR 2 MICHALIS KASIOULIS | UNIVERSITY OF CYPRUS (UCY) | CYPRUS



BenchPilot: Repeatable & Reproducible Benchmarking for Edge Micro-DCs

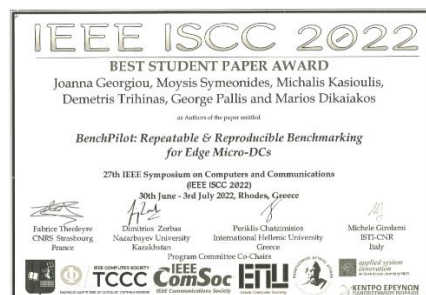
Authors: Joanna Georgiou (UCY), Moysis Symeonides (UCY), Michalis Kasioulis (RAIS ESR2, UCY), Demetris Trihinas (UNIC), George Pallis (UCY), Marios D. Dikaiakos (UCY)

Abstract

Micro-Datacenters (DCs) are emerging as key enablers for Edge computing and 5G mobile networks by providing processing power closer to IoT devices to extract timely analytic insights. However, the performance evaluation of data stream processing on micro-DCs is a daunting task due to difficulties raised by the time-consuming setup, configuration and heterogeneity of the underlying environment. To address these challenges, we introduce BenchPilot, a modular and highly customizable benchmarking framework for edge micro-DCs. BenchPilot provides a high-level declarative model for describing experiment testbeds and scenarios that automates the benchmarking process on Streaming Distributed Processing Engines (SDPEs). The latter enables users to focus on performance analysis instead of dealing with the complex and time-consuming setup. BenchPilot instantiates the underlying cluster, performs repeatable experimentation, and provides a unified monitoring stack in heterogeneous Micro-DCs. To highlight the usability of BenchPilot, we conduct experiments on two popular streaming engines, namely Apache Storm and Flink. Our experiments compare the engines based on performance, CPU utilization, energy consumption, temperature, and network I/O.

Published in: 2022 IEEE Symposium on Computers and Communications (ISCC)

This paper was judged worthy of the **Best Student Paper Award** at IEEE ISCC 2022.





RAIS PUBLICATIONS

- **"Early-stage Ransomware Detection based on Pre-Attack Internal API Calls"**, Filippo Coglio and Ahmed Lekssays and Barbara Carminati and Elena Ferrari, Proceedings of the 37th International Conference on Advanced Information Networking and Applications (AINA-2023), 2023.
- **"PriApp-Install: Learning User Privacy Preferences on Mobile Apps' Installation"**, Ha Xuan Son, Barbara Carminati, Elena Ferrari, Information Security Practice and Experience - ISPEC 2022, 2022.
- **"MalRec: A Blockchain-based Malware Recovery Framework for Internet of Things"**, Ahmed Lekssays, Giorgia Sirigu, Barbara Carminati, Elena Ferrari, 17th International Conference on Availability, Reliability and Security, 2022.
- **"12 Years of Self-tracking for Promoting Physical Activity from a User Diversity Perspective: Taking Stock & Thinking Ahead"**, Sofia Yfantidou, Pavlos Sermpezis, Athena Vakali, UMAP '22 Adjunct: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022.
- **"BenchPilot: Repeatable & Reproducible Benchmarking for Edge Micro-DCs"**, Joanna Georgiou, Moysis Symeonides, Michalis Kasioulis, Demetris Trihinas, George Pallis, Marios D. Dikaiakos, Proceedings of the 27th IEEE Symposium on Computers and Communications" (ISCC '22), 2022.
- **"Demo: The RAINBOW Analytics Stack for the Fog Continuum"**, Moysis Symeonides, Demetris Trihinas, Joanna Georgiou, Michalis Kasioulis, George Pallis, Marios D. Dikaiakos, Theodoros Toliopoulos, Anna-Valentini Michailidou, Anastasios Gounaris, Proceedings of the 27th IEEE Symposium on Computers and Communications" (ISCC '22), 2022.
- **"Federated Naive Bayes under Differential Privacy"**, Thomas Marchioro, Lodovico Giaretta, Evangelos Markatos, Šarūnas Girdzijauskas, 19th International Conference on Security and Cryptography (SECRYPT), Lisbon, Portugal, 2022.
- **"What your Fitbit says about you: De-anonymizing users in lifelogging datasets"**, Andrei Kazlouski, Thomas Marchioro, Evangelos Markatos, 19th International Conference on Security and Cryptography (SECRYPT), Lisbon, Portugal, 2022.

- **"A Risk Estimation Mechanism for Android Apps based on Hybrid Analysis"**, Ha Xuan Son, Barbara Carminati, Elena Ferrari, Data Science and Engineering, 2022.
- **"How to Publish Wearables' Data: Practical Guidelines to Protect User Privacy"**, Thomas Marchioro, Andrei Kazlouski, Evangelos Markatos, Studies in Health Technology and Informatics, 294, 949-950, 2022.
- **"SchemaWalk: Schema Aware Random Walks for Heterogeneous Graph Embedding"**, Ahmed E. Samy, Lodovico Giarretta, Zekarias T. Kefato, Šarūnas Girdzijauskas, WWW '22: Companion Proceedings of the Web Conference 2022, 2022.
- **"WeMoD: A Machine Learning Approach for Wearable and Mobile Physical Activity Prediction"**, Dimitrios Vasdekis, Sofia Yfantidou, Stefanos Efstathiou, Athena Vakali, 3rd Workshop on Human-Centered Computational Sensing (HCCS'22), Pisa, Italy, 2022.
- **"WearMerge: An Interoperable Framework for Self-tracking Data Integration and Standardization"**, Dimitrios Panteleimon Giakatos, Sofia Yfantidou, Stefanos Efstathiou, Athena Vakali, PerCom Demos 2022, Pisa, Italy, 2022.
- **"PAutoBotCatcher: A blockchain-based privacy-preserving botnet detector for Internet of Things"**, Ahmed Lekssays, Luca Landa, Barbara Carminati, Elena Ferrari, Computer Networks 200 (2021): 108512, 2021.
- **"LiMNet: Early-Stage Detection of IoT Botnets with Lightweight Memory Networks"**, Lodovico Giarretta, Ahmed Lekssays, Barbara Carminati, Elena Ferrari, Sarunas Girdzijauskas, Proc. of the European Symposium on Research in Computer Security (ESORICS 2021).
- **"A Risk Assessment Mechanism for Android Apps"**, Ha Xuan Son, Barbara Carminati, Elena Ferrari, Proc. of the IEEE International Conference on Smart Internet of Things (SmartIoT 2021).
- **"Decentralized Word2Vec Using Gossip Learning"**, Abdul Aziz Alkathiri, Lodovico Giarretta, Sarunas Girdzijauskas, Magnus Sahlgren, 23rd Nordic Conference on Computational Linguistics (NoDaLiDa 2021), Reykjavik, Iceland, 2021.
- **"Self-Tracking Technology for mHealth: A Systematic Review and the PAST SELF Framework"**, Sofia Yfantidou, Pavlos Sermpezis, and Athena Vakali, arXiv, 2021.
- **"PDS2: A user-centered decentralized marketplace for privacy preserving data processing"**, Lodovico Giarretta, Ioannis Savvidis, Thomas Marchioro, Sarunas Girdzijauskas, George Pallis, Marios D. Dikaiakos, Evangelos Markatos, Third International Workshop on Blockchain and Data Management (BlockDM 2021), in conjunction with the 37th IEEE International Conference on Data Engineering (ICDE), Chania, Crete, Greece, 2021.

- **"Federated Word2Vec: Leveraging Federated Learning to Encourage Collaborative Representation Learning"**, Daniel Garcia Bernal, Lodovico Giaretta, Sarunas Girdzijauskas, Magnus Sahlgren, arXiv:2105.00831 [cs], 2021.
- **"Do partner apps offer the same level of privacy protection? The case of wearable applications"**, Andrei Kazlouski, Thomas Marchioro, Harry Manifavas, Evangelos Markatos, In Proceedings of the 7th Workshop on Sensing Systems and Applications using Wrist Worn Smart Devices (WristSense), 2021.
- **"I still See You! Inferring Fitness Data from Encrypted Traffic of Wearables"**, Andrei Kazlouski, Thomas Marchioro, Harry Manifavas, Evangelos Markatos, 14th International Joint Conference on Biomedical Engineering Systems and Technologies HEALTHINF, Vienna, Austria, February, 2021.
- **"Do you know who is talking to your wearable smartband?"**, Andrei Kazlouski, Thomas Marchioro, Harry Manifavas, Evangelos Markatos, European Federation for Medical Informatics - Special Topic Conferences (EFMI-STC), 2020.
- **"Gossip Learning: Off the Beaten Path"**, Lodovico Giaretta and Sarunas Girdzijauskas, IEEE Big Data 2019, Los Angeles, CA, USA, 2019.



Find at: <https://rais-itn.eu/publications>



BENEFICIARIES



PARTNERS



FOLLOW US



WEBSITE



<https://rais-itn.eu/>

CONTACT US

Project Coordinator
Sarūnas Girdzijauskas
Computer Science Dept.
School of Electrical Engineering and
Computer Science (EECS)
KTH - Royal Institute of Technology, Sweden
sarunasg@kth.se

Newsletter Content Editor
Demetra Katziani
Computer Science Dept.
Laboratory for Internet Computing
(LInC)
University of Cyprus (UCY)
katziani.demetra@ucy.ac.cy



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Innovative Training Networks (ITN) - RAIS No 813162

