INCS-CoE *Digital Trust* Forum Day 2

Wednesday, December 1, 2021

7:00AM-8:30AM EST; 12:00PM-1:30PM GMT; 9:00PM-10:30PM JST

Hosted by:

Northeastern University

United States

UMBC
Northeastern University

United Kingdom

Imperial College London
Royal Holloway University of London
University of Cambridge

Japan

Keio University
Kyushu University

France

University of Limoges

Israel

Technion Israel Institute of Technology
Ben-Gurion University

Australia

Edith Cowan University

International Digital Trust Forum

Dr Anabel Gutierrez

About me

Dr Anabel Gutierrez

(anabelsgm)

Anabel.GutierrezMendoza@rhul.ac.uk

PhD, Information Systems, Brunel University, UK (2009)

I am a Senior Lecturer in Digital Marketing at Royal Holloway University of London with over 25 years of academic experience which I have balanced with industrial practice gained from consultancy work in IT projects for private and public sectors.

My research interest areas are in innovation and adoption of emerging technologies for the digital economy with particular interest on data privacy concerns, the use of data to understand consumer behaviour and how to improve data-driven decision making. I have published several articles in international peer-reviewed journals and conferences as well as acting as a reviewer since 2006. Currently, I am a member of the SAS UK & Ireland Academic Advisory Board, Co-Chair of the Digital Marketing and Analytics SIG at the Academy of Marketing and member of the International Editorial Review Board (IERB) of International Journal of Information Management (IJIM).

Current Project: "The Ethical Implications & Unintended Consequences of aiding Digital Collaboration through the use of Cutting-Edge Technologies in the Food Sector"

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Working Group

Dr Samantha Kanza
University of Southampton

Mr Samuel Munday
University of Southampton

Professor Louise Manning
Royal Agricultural University

Dr Anabel Gutierrez Mendoza
Royal Holloway, University of London

Dr Peter Craigon
University of Nottingham

Dr Naomi Jacobs
Lancaster University

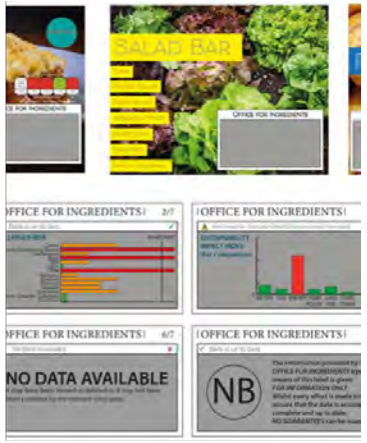
Mr Justin Sacks
Lancaster University

Current Project: "The Ethical Implications & Unintended Consequences of aiding Digital Collaboration through the use of Cutting-Edge Technologies in the Food Sector"

ROYAL HOLLOWAY UNIVERSITY OF LONDON

The working group assessing data trusts, funded by the Internet of Food Things project and the AI3SD project, is made up of experts from many different specialities including computer science, law and design.

- We used as research method a "design fiction" approach, where we created real objects representing a fictional future. Some of the things we've created include minutes from a board meeting that never took place, a clip from a documentary reporting on something that hasn't happened, a website showcasing a non-existent allergy alert app, and smart packaging that simulates the shopping experience depicted earlier.
- We used a set of cards called the Moral-IT deck, which were developed to evaluate the ethics of technology. These cards supported the discussion of ethical issues for the objects created.



Relevant publications on IT, privacy and ethical implications.

Jacobs, N., Brewer, S., Craigon, P., Frey, J., Gutierrez, A., Kanza, s., Manning, L., Munday, S., Pearson, S. and Sacks, J. **(Under revision) Enabling artificial intelligence use in the food sector: Developing a common stakeholder vocabulary**

Jacobs, N., Brewer, S., Craigon, P., Frey, J., Gutierrez, A., Kanza, s., Manning, L., Munday, S., Pearson, S. and Sacks, J. (2021) **Considering the ethical implications of digital collaboration in the Food Sector**. Patterns, ISSN: 2666-3899, Vol: 2, Issue: 11, Page: 100335. Available at: https://doi.org/10.1016/j.patter.2021.100335

Gutierrez, A., O'Leary, S., Nripendra, P.R., Dwivedi, Y.K. and Calle, T. (2019) **Using privacy calculus theory to explore for entrepreneurial directions in mobile location-based advertising: Identifying intrusiveness as the critical risk factor.** Computers in Human Behavior, vol. 95. pp. 295-306. Available at: https://doi.org/10.1016/j.chb.2018.09.015.

Gutierrez, A., Boukrami, E. and Lumsden, R. (2015) **Technological, Organisational and Environmental factors influencing managers' decision to adopt cloud computing in the UK**. *Journal of Enterprise Information Management,* 28(6), pp. 788-807. Available at: https://doi.org/10.1108/JEIM-01-2015-0001.

Towards Trustworthy AI-based Systems

Alessio Lomuscio
a.lomuscio@imperial.ac.uk

Verification of Autonomous Systems Lab
Imperial College London, UK

1 December 2021

Problems with systems based on neural networks

Increasingly used in Cyber-Physical Systems, Security, Personal Assistants, and beyond.

- Good on ID data, but may perform poorly on OOD data.
- Fragile on ID data.
- Often strong confidence on incorrect classifications.
- Lack of explanations.

High-profile failures (Autonomous vehicles, ...)

Leading to insufficient trust in many applications.

Work at VAS@Imperial on Verification of Neural Systems

- Verification of NN-based Perception Systems (2018-present).
- Verification of (closed-loop) Neural-Symbolic Multi-Agent Systems (2018-present).

Increasingly conquering scalability from hundreds of parameters to millions of parameters.

Pilot usecases with aircraft manufacturer and car makers.

Verification increases reliability, thereby enhancing trustworthiness.

Planned follow up work (2022-onwards)

- Verification of **security systems** based on recurrent networks.
- Verification-based **explainability**: generation and validation of *robust* explanations for neural systems.
- Derivation of provably-safe **runtime monitors** for neural systems.

Overall objective: methods and tools to enhance trust in neural-based cyber-physical systems, security systems, and personal assistants/robotics.

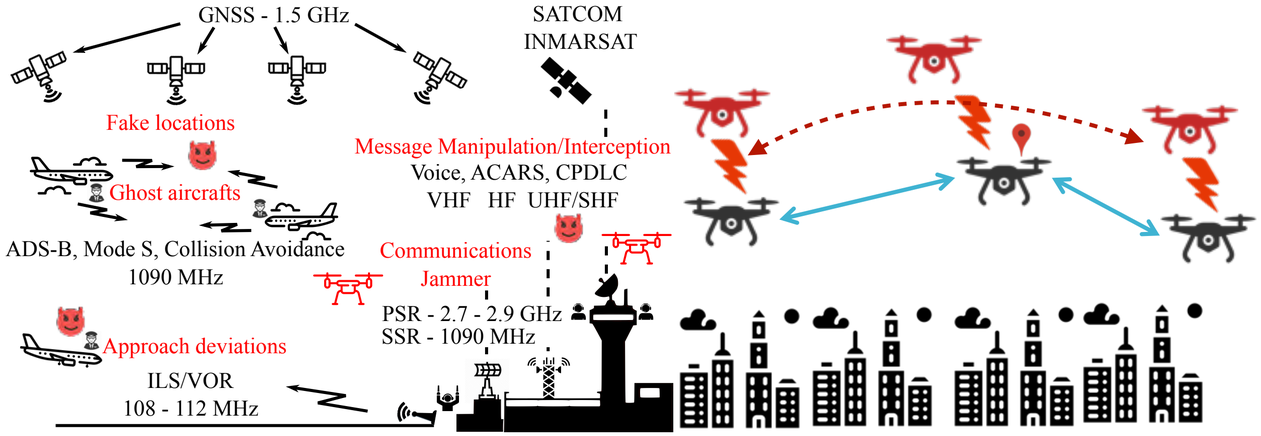
Selected References

- Kouvaros, Kyono, Leofante, Lomuscio, Margineantu, Osipychev, Zheng. Formal Analysis of Neural Network-based Systems in the Aircraft Domain. FM21.
- Henriksen, Lomuscio. DEEPSPLIT: An Efficient Splitting Method for Neural Network Verification via Indirect Effect Analysis. IJCAI21.
- Kouvaros, Lomuscio. Towards Scalable Complete Verification of ReLU Neural Networks via Dependency-based Branching. IJCAI21.
- Batten, Kouvaros, Lomuscio, Zheng. Efficient Neural Network Verification via Layer-based Semidefinite Relaxations and Linear Cuts. IJCAI21.
- M. Akintunde, E. Botoeva, P. Kouvaros, A. Lomuscio. Verifying Strategic Abilities of Neural-symbolic Multi-agent Systems. KR20.
- Henriksen, Lomuscio. Efficient Neural Network Verification via Adaptive Refinement and Adversarial Search. ECAI20.
- Akintunde, Botoeva, Kouvaros, Lomuscio. Formal Verification of Neural Agents in non-Deterministic Environments. Proceedings of AAMAS20.
- E. Botoeva, P. Kouvaros, J. Kronqvist, A. Lomuscio, R. Misener. Efficient Verification of Neural Networks via Dependency Analysis. AAAI20.
- Akintunde, Kevorchian, Lomuscio, Pirovano. Verification of RNN-Based Neural Agent-Environment Systems. AAAI19.
- Akintunde, Lomuscio, Maganti, Pirovano. Reachability Analysis for Neural Agent-Environment Systems. KR18.

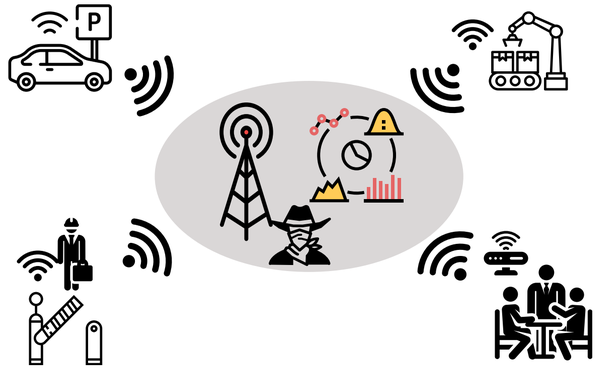
Signal Intelligence Lab @ Northeastern

Security and privacy of wireless networks with a strong focus on *autonomous cyber-physical systems and smart ecosystems*.

Secure and Private Wide-area Positioning



Securing the Skies



Security and Privacy of xIoT

Faculty

Aanjhan Ranganathan
Assistant Professor

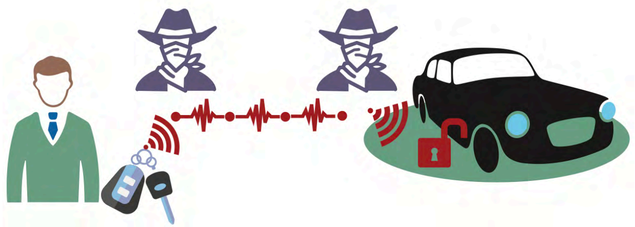
PhD Students




Secure Proximity and Location Verification

Physical location as a Digital Trust Factor

Attacks on Location



Selected Research

Are We Really Close? Verifying Proximity in Wireless Systems, Aanjhan Ranganathan, Srdjan Capkun

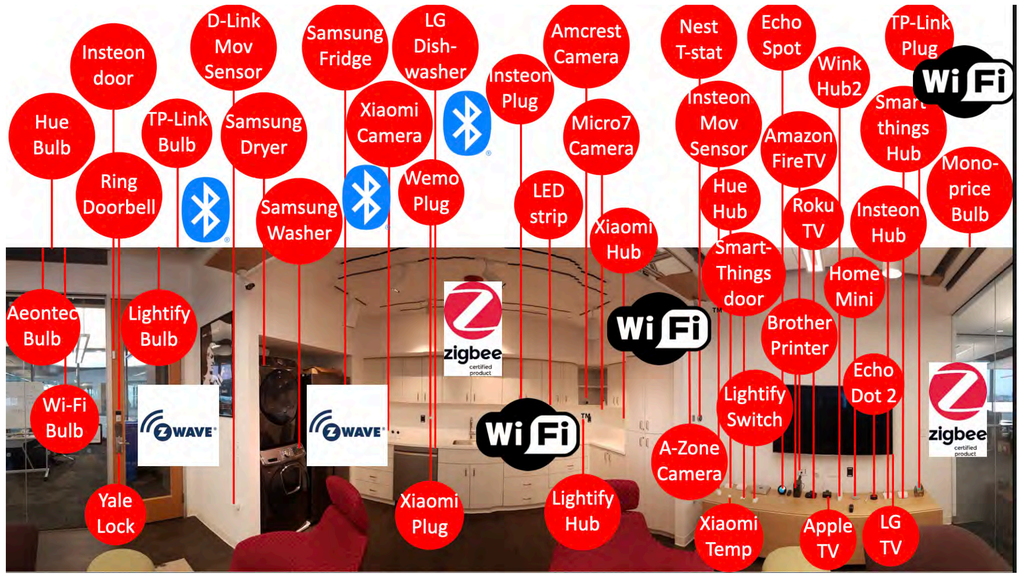
Here, There, and Everywhere: Security Analysis of WiFi 802.11mc Fine Timing Measurement, Domien Schepers, Mridula Singh, Aanjhan Ranganathan

VRange: Enabling Secure Ranging in 5G-NR Wireless Networks, Mridula Singh, Marc Roeschlin, Aanjhan Ranganathan, Srdjan Capkun

SemperFi: Anti-spoofing GPS receiver for UAVs, Harshad Sathaye, Gerald LaMountain, Pau Closas, Aanjhan Ranganathan

Security and Privacy in xIoT

Validating and Building Trustworthy Smart Ecosystems



Mon(Iot)Or Lab at Northeastern University

Selected Research

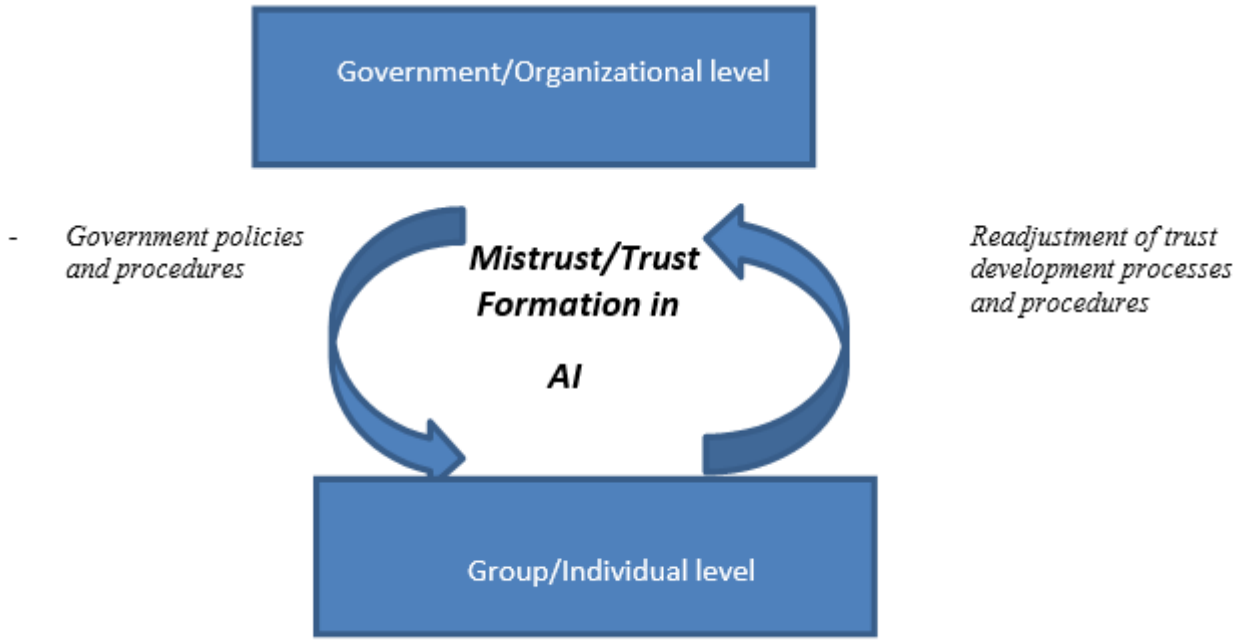
I Send, Therefore I Leak: Information Leakage in Low-Power Wide Area Networks, Patrick Leu, Ivan Puddu, Aanjhan Ranganathan, Srdjan Capkun

ZLeaks: Passive Inference Attacks on Zigbee based Smart Homes, Narmeen Shafqat, Daniel Dubois, Dave Choffnes, Aaron Schulman, Dinesh Bharadia, Aanjhan Ranganathan

Privacy-Preserving Positioning in Wi-Fi Fine Timing Measurements, Domien Schepers, Aanjhan Ranganathan

Niki Panteli – Expertise/Interests/Projects

- *Professor of Digital Business, School of Business & Management, RHUL*
- **Expertise**: Developing Trust Online & technology-mediated settings; leading Virtual Teams and Online Collaborations; qualitative research
- **Interests:** Digital Innovation; Healthcare; Cybersecurity
- **<u>Relevant Current/Recent Projects:</u>**
  - Trust, Identity, Privacy & Security in the Work from Home Covid-19 context
  - Regulatory Aspects of Digital Security – Sociotechnical Perspective
  - AI and Trust: A multi-level perspective



Improving interpretations of trustworthiness

UNCS CoE presentation

Marc Sel

Marc.Sel.2013@live.rhul.ac.uk

marc@marcsel.eu

Royal Holloway University of London — Information Security Group

22 November 2021

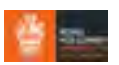
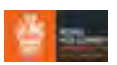


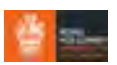
Table of Contents



The subject of the proposed research

- A further elaboration of the trustworthy ecosystem ($\mathcal{TE}$) model, described in M. Sel [2], M. Sel [3] and M. Sel and C. Mitchell [4],
- The model and a sample implementation were proposed in my PhD thesis, M. Sel [5]
- The implementation of the sample rulebook was inspired by the European eIDAS Regulation (EU 910/2014)
- The model is based on four building blocks:
  - a data model, expressed in First Order Logic
  - a rulebook, containing constraints that reflect a particular context for reasoning about trustworthiness
  - trustworthiness evaluation functions
  - real-world instance data

The evaluator corresponds to a potential trustor, and the evaluation subject to a potential trustee.



Possible usage outlook

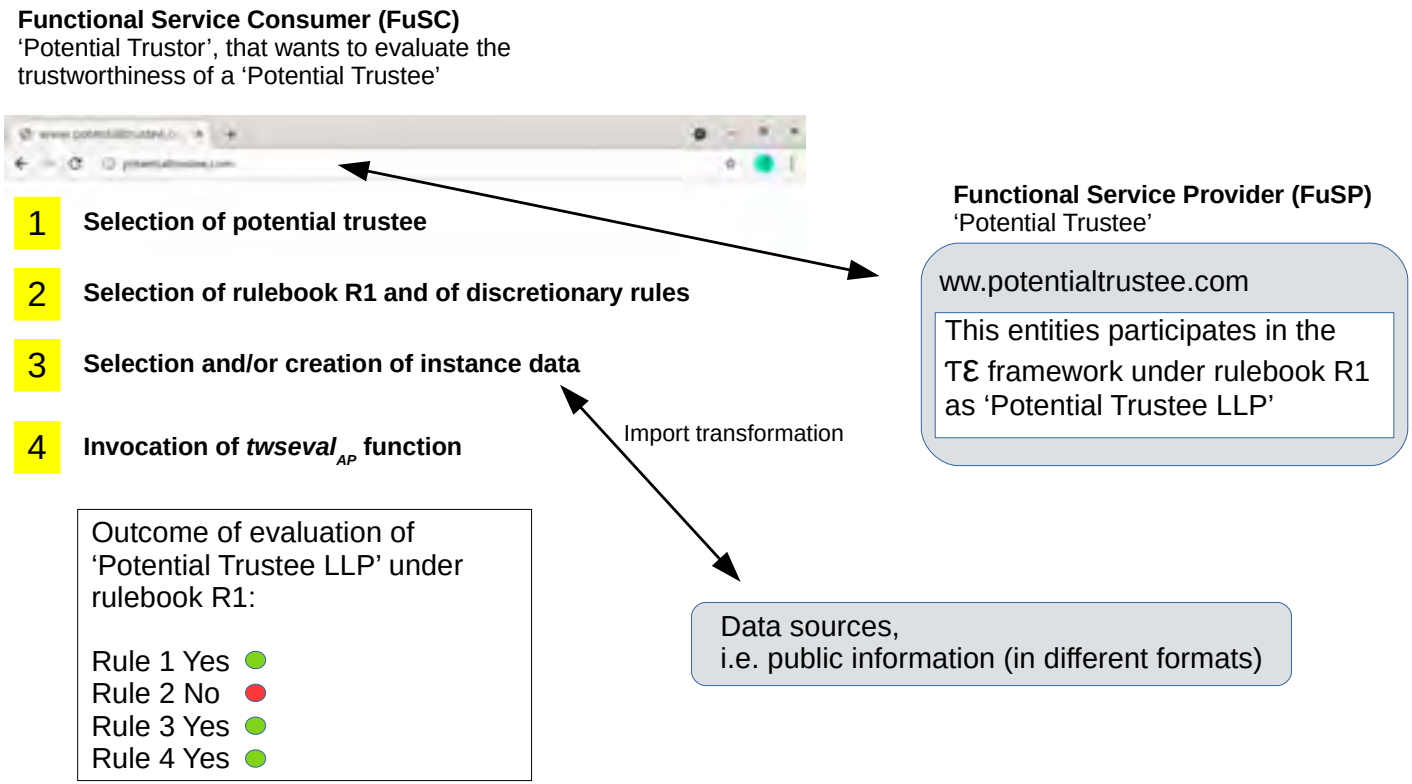
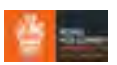


Figure: 1 Possible future use

Research questions

Research questions include the following.

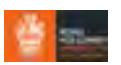
- How can we semantically define trustworthiness?
- How can we reason about trustworthiness?
- On what can reasoning to qualify an entity as trustworthy be based?
- How can we obtain information for use in supporting such reasoning about 'real world' entities?
- How can the above questions be addressed at a global scale?



Bibliography

Publications:

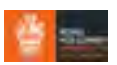
- M. Sel [1], *Using the Semantic Web to generate Trust Indicators*, Securing Business Processes – Proceedings of the ISSE 2014 Conference, Sachar Paulus, Norbert Pohlman and Helmut Reimer (editors), Vieweg+Tuebner, Springer Science+Business Media, ISBN 978-3-658-06707-6, pages 106–119.
- M. Sel [2], *A Comparison of Trust Models*, Securing Business Processes – Proceedings of the ISSE 2015 Conference, Sachar Paulus, Norbert Pohlman and Helmut Reimer (editors), Vieweg+Tuebner, Springer Science+Business Media, ISBN 978-3-658-10933-2, pages 206–215.
- M. Sel [3], *Improving Interpretations of Trust Claims*, IFIPTM 2016, Darmstadt, Germany, July 18-22, 2016, Proceedings, published in Trust Management X — 10th IFIP WG 11.11 International Conference, pages 164–173.



Operation of the model in a nutshell

Verifying whether the constraints are satisfied over a set of instance data allows an evaluator to evaluate the trustworthiness of an evaluation subject.

- Trustworthiness evaluation functions take as input a rulebook and a set of data
- A rulebook contains a mandatory and a discretionary part.
  - Mandatory part: constraints that must be satisfied to have the minimal basis for relevant execution of the discretionary rules
  - Discretionary part: allows to specify a trustworthiness evaluation policy
- Data represents real world information about the potential trustee and its context
- Outcome of the evaluation provides evidence for the evaluator to decide to interact with the evaluation subject in the relationship of trustor–trustee



Implementation architecture

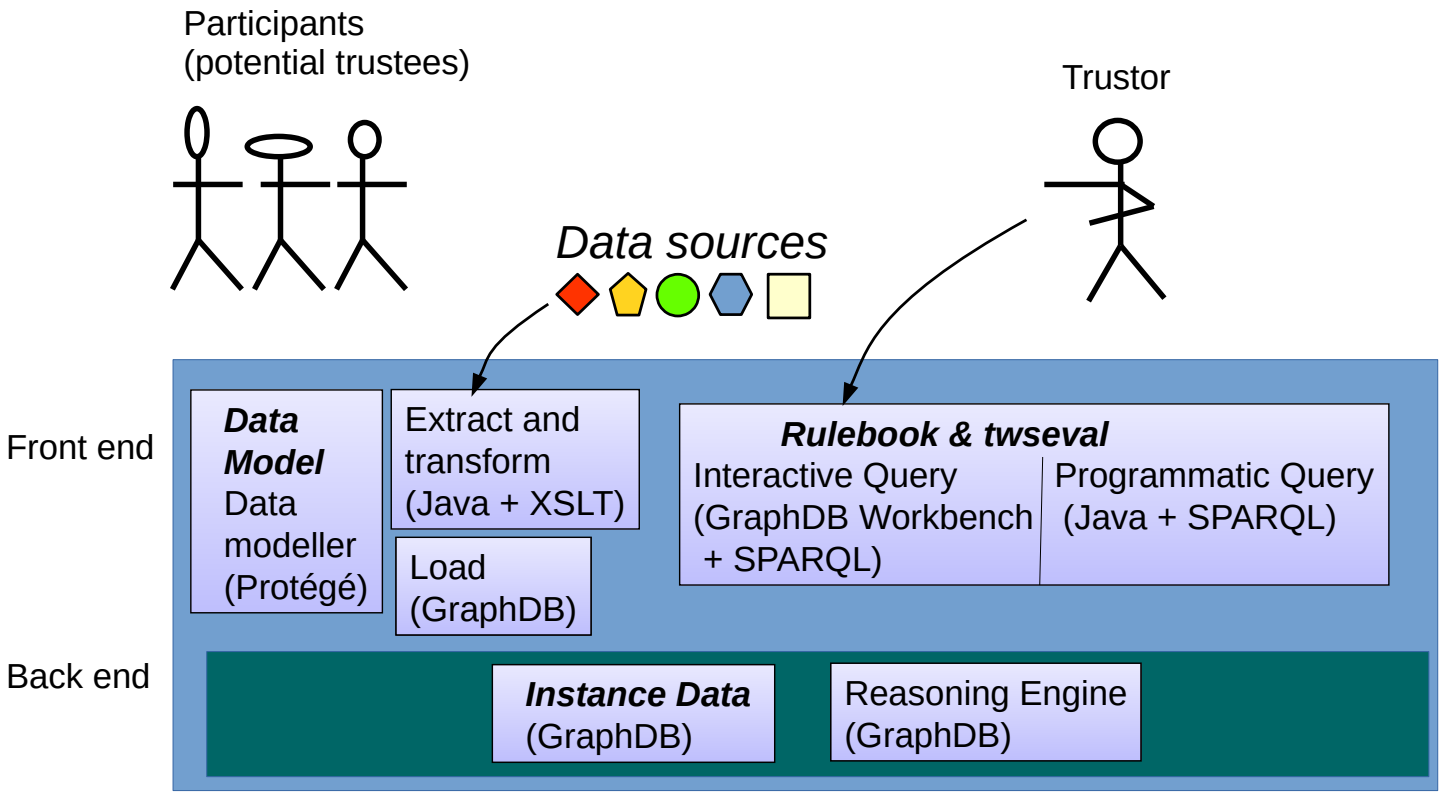
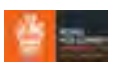


Figure: 2 Implementation architecture


Marc Sel.
Using the semantic web to generate trust indicators.
In Sachar Paulus, Norbert Pohlman, and Helmut Reimer, editors, *Securing Business Processes*, pages 106–119. Vieweg+Tuebner, Springer Science+Business Media, 2014.

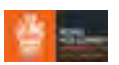

Marc Sel.
A comparison of trust models.
In Sachar Paulus, Norbert Pohlman, and Helmut Reimer, editors, *Securing business processes*, pages 206–215. Vieweg+Tuebner, Springer Science+Business Media, 2015.


Marc Sel.
Improving Interpretations of Trust Claims.
In *Trust Management X — 10th IFIP WG 11.11 International Conference, IFIPTM 2016, Darmstadt, Germany, July 18-22, 2016, Proceedings*, pages 164–173, 2016.




Marc Sel and Chris J. Mitchell.
Automating the evaluation of trustworthiness.
In *Proceedings of TrustBUS 2021: September 2021 (forthcoming). Springer-Verlag, 2021 (Lecture Notes in Computer Science)*, 2021.


Marc Louis Sel.
Automating interpretations of trustworthiness.
PhD thesis, Royal Holloway, University of London, 2021.
`https://pure.royalholloway.ac.uk/portal/en/persons/marc-sel(d0b600a4-e99c-456b-a63c-af8b744e97f0).html`.



慶應義塾
Keio University

Mutual Declaration Mechanism of Multi-provider Relationship for Trusted Web Service

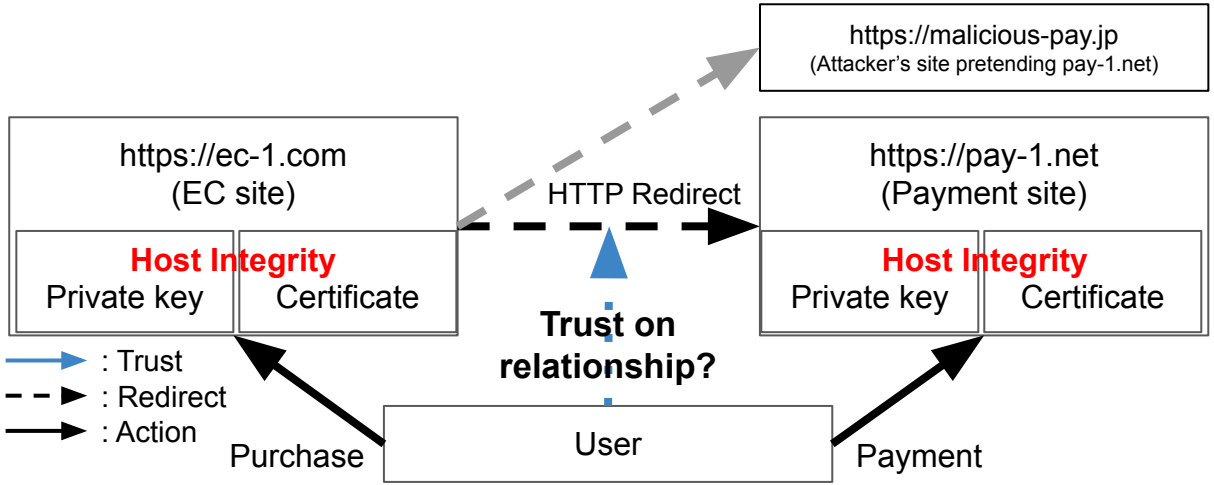
2021/11/30 INCS-CoE Digital Trust Forum

Keio University
Takao Kondo (latte@itc.keio.ac.jp) / Wataru Ohgai (alt@sfc.wide.ad.jp)

Mutual Declaration Mechanism of Multi-provider Relationship for Trusted Web Service

慶應義塾 Keio University

- Web security model is based on host integrity assurance by TLS
- Web backend infrastructure is becoming more complex
  → rerouting traffic among several service providers (SPs)
- TLS cannot ensure integrity (relationship) between SPs



Current threats against rerouting

慶應義塾
Keio University

- Machine-in-the-Middle contents modification attacks
- DNS/ARP spoofing attacks
  → HTTPS → TLS downgrade attacks[1][2]
- HSTS → HTTPS context confusion attacks[3]

- Currently nothing can assure Digital Trust between
  - SPs in a relationship
  - Web services and the end users

- This proposal is to realize these Digital Trust

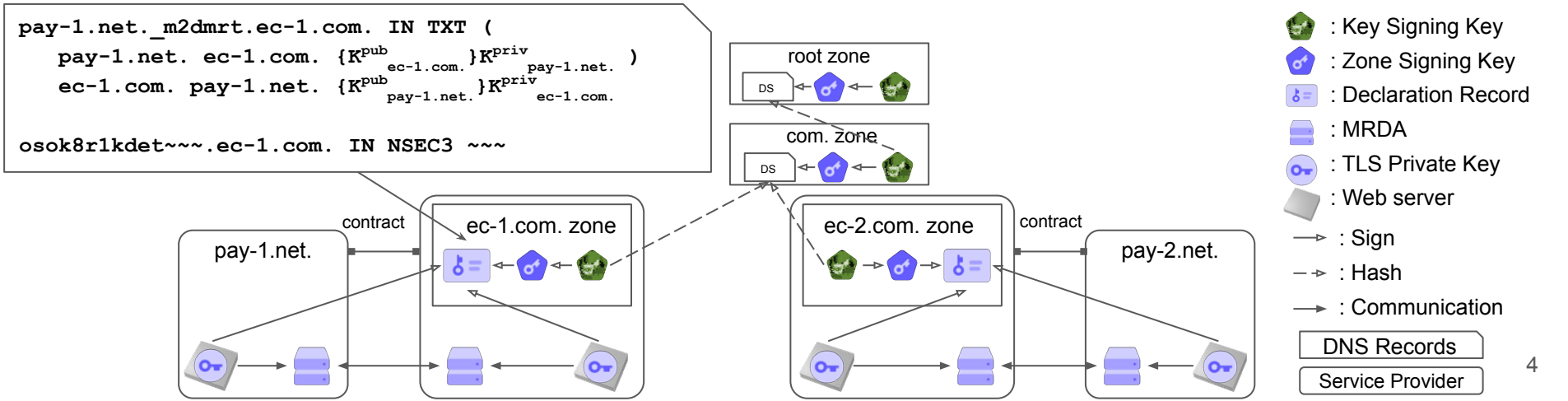
[1] M. Marlinspike. New Tricks For Defeating SSL In Practice. In Proc. of BLACKHAT Europe '09, 2009.
[2] X. Li, C. Wu, S. Ji, and R. Gu, Q.and Beyah. HSTS Measurement and an Enhanced Stripping Attack Against HTTPS. In Security and Privacy in Communication Networks, pages 489–509. Springer Intl. Pub, 2018.
[3] M. Zhang, X. Zheng, K. Shen, Z. Kong, C. Lu, Y. Wang, H. Duan, S. Hao, B. Liu, and M. Yang. Talking with Familiar Strangers: An Empirical Study on HTTPS Context Confusion Attacks. In Proc. of ACM CCS'20, pages 1939–1952, 2020.

Architecture of the solution

慶應義塾 Keio University

M2DMRT:

- Sign related SP's TLS public key by own TLS private key mutually
- Publish the signature in DNSSEC-protected SP's DNS zone
- End users can trust the relationship by verifying digital signature



Trust in AI-enabled Workplace

Amany Elbanna

Reader (Associate Professor) in Information Systems

Royal Holloway University of London, TW20 0EX, Egham, UK

Amany.Elbanna@rhul.ac.uk

Motivation for the Research

- Use of AI to address the information overload and provide assistance or automation of tasks
- use machine learning and natural language processing techniques
- Influences of integrating such digital innovations into the workplace are important areas of inquiry
- Collaborative relationship between employees and chatbots

Research Question

How AI design could impact trust (enhance or reduce)?

Research Context

AI – enabled systems including:

- *intelligent chatbots/ personal assistants*
- *document analysis*
- *Decision making*
- *Monitoring and maintenance*

Theoretical Grounding

- Sociotechnical perspective
- how actors shape technology and being shaped by it.
- Generalized symmetry

Methodology

- Interpretive case study approach
- Qualitative data collection
  - semi-structured interviews
  - participant observations
  - document reviews
- Data Analysis: Qualitative, Inductive and interpretive approaches

International Digital Trust Forum

Northeastern University
December 1, 2021

Dr Konstantinos Mersinas
konstantinos.mersinas@rhul.ac.uk

Research interests / Expertise

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Risk perceptions
Risk attitudes
Rationality

- Cybersecurity professionals: Cognitive biases can have significant effects on decision-making
- Dark web communications: cyber crime related activities; sector- and country-specific analyses.

HIVE

(Hub for research into Intergenerational Exploitation to Vulnerability)

Higher Education Innovation Fund (HEIF)

Met Police, NCSC, UK Charities

Projects:

- Protecting adolescents and the from cyberbullying / cyberstalking
- Financial fraud and abuse targeted to the elderly

AI & human rationality

- Emotions, biases, heuristics and ‘rational’ decision-making

Cyber Security Culture

- Empirical study in UK Higher Education Institutes (NCSC)

Cultural differences

- User security behaviours across 12 countries

Human and societal aspects of digital trust

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Stage 1: Existing International Trust Frameworks Study

- Individual & group level review (e.g. per country, org, group, individual)
- Objective: Review of individual user acceptance based on levels of trust

Stage 2: Usage and Implementation of Trust Frameworks

Objective: identification of gaps and challenges

- Sector-specific study (cross-sector comparisons and maturity level)
- Objective: threat analysis;
- Approach: empirical (interviews) and/or secondary data;
- Justification: pandemic and environmental crisis

Human and societal aspects of digital trust

Stage 3: Identified opportunities & solutions for Trust Frameworks

- Individual-level acceptance
- Objectives: measuring acceptance between countries
- Approach: Empirical study or secondary data
- Possible empirical approach: Student collaboration / joint project across INCS-CoE countries, focused on specific technologies

Thank you!