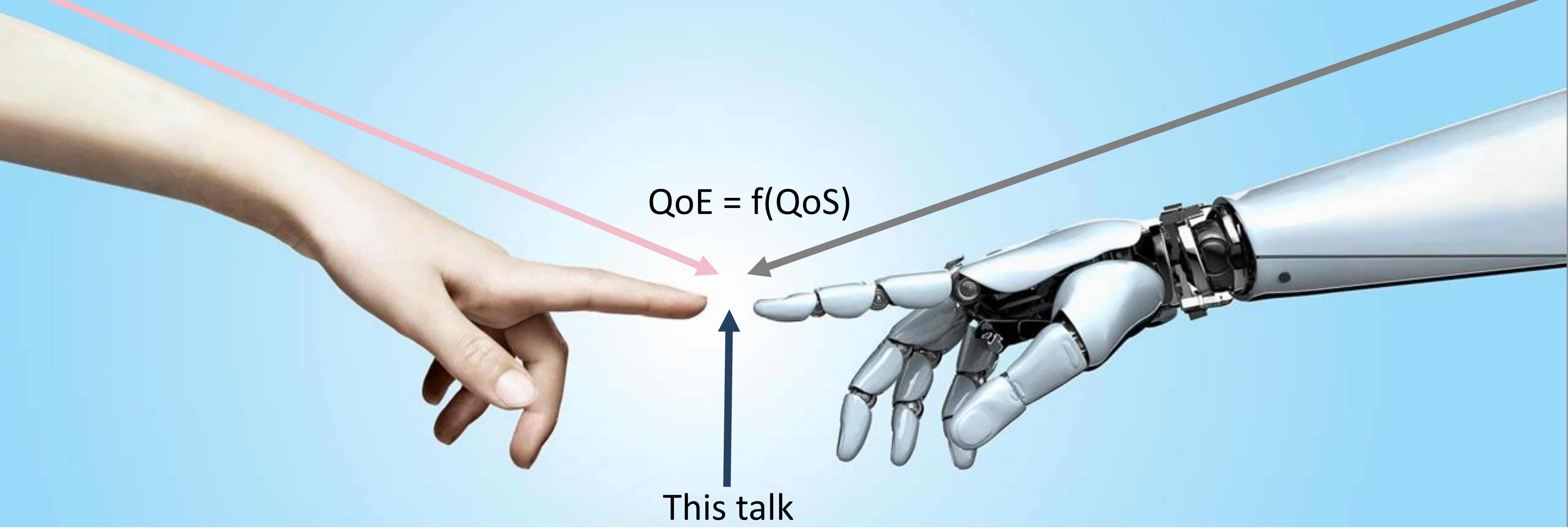
Metrics and models for Web performance evaluation

or, How to measure SpeedIndex from raw encrypted packets, and why it matters



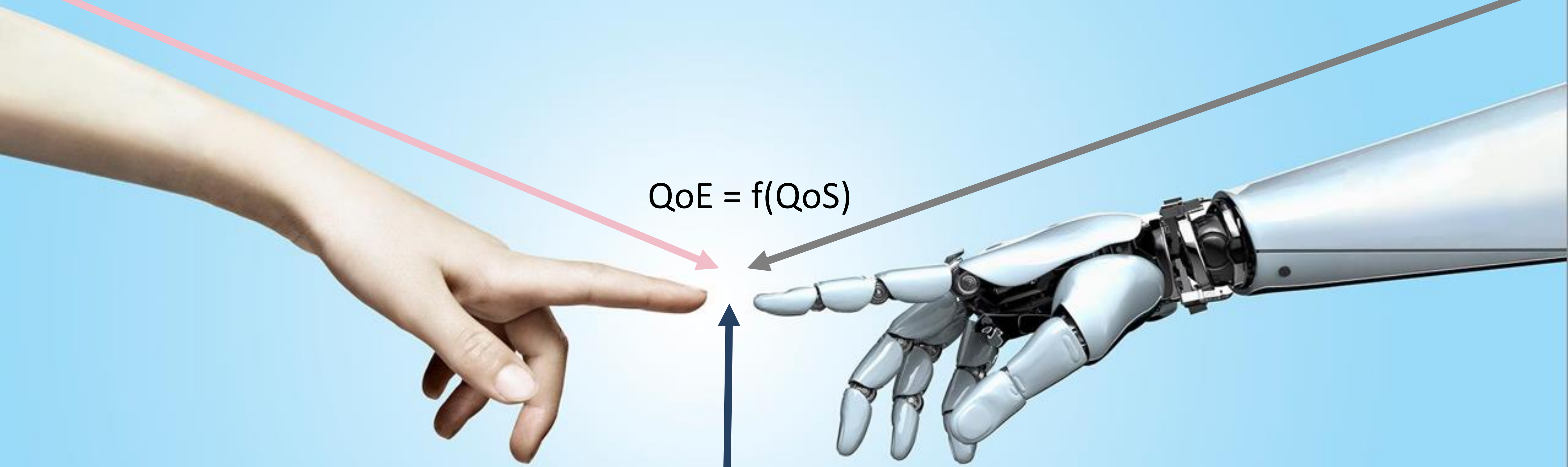
Huawei

Dario Rossi
dario.rossi@huawei.com
Director, DataCom (*) Lab
(*) Data Communication Network Algorithm and Measurement Technology Laboratory

FOSDEM'20
Brussels, Feb 1st 2020

Metrics and models for Web performance evaluation

or, How to measure SpeedIndex from raw encrypted packets, and why it matters

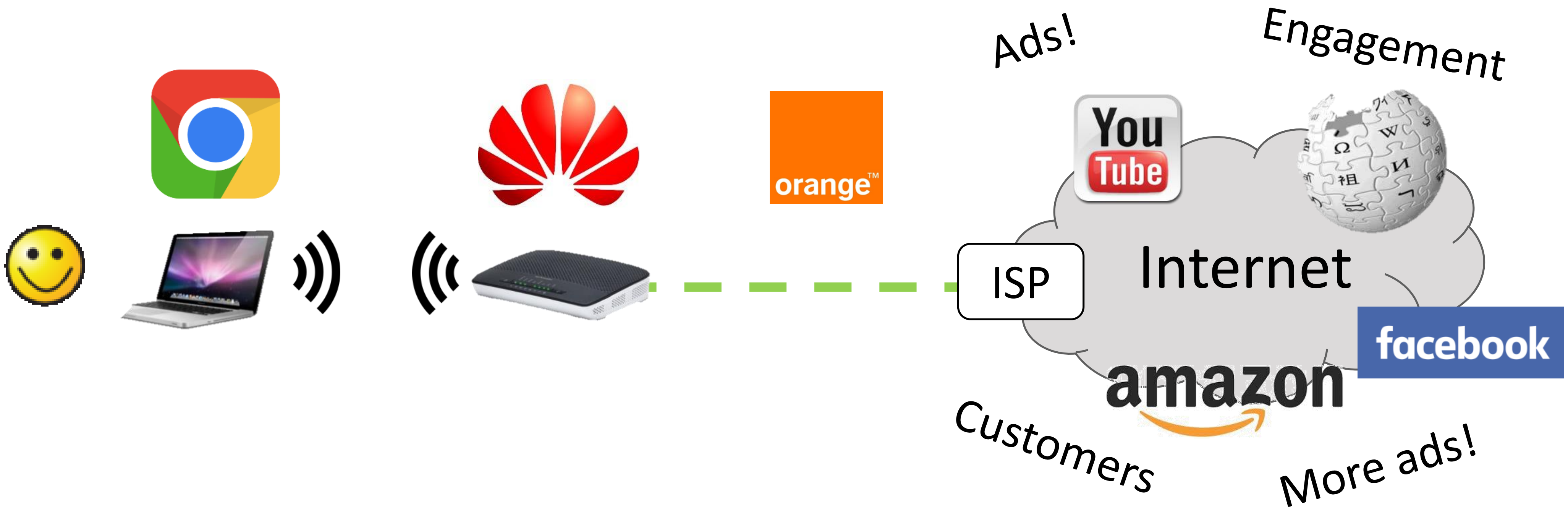


QoE = f(QoS)

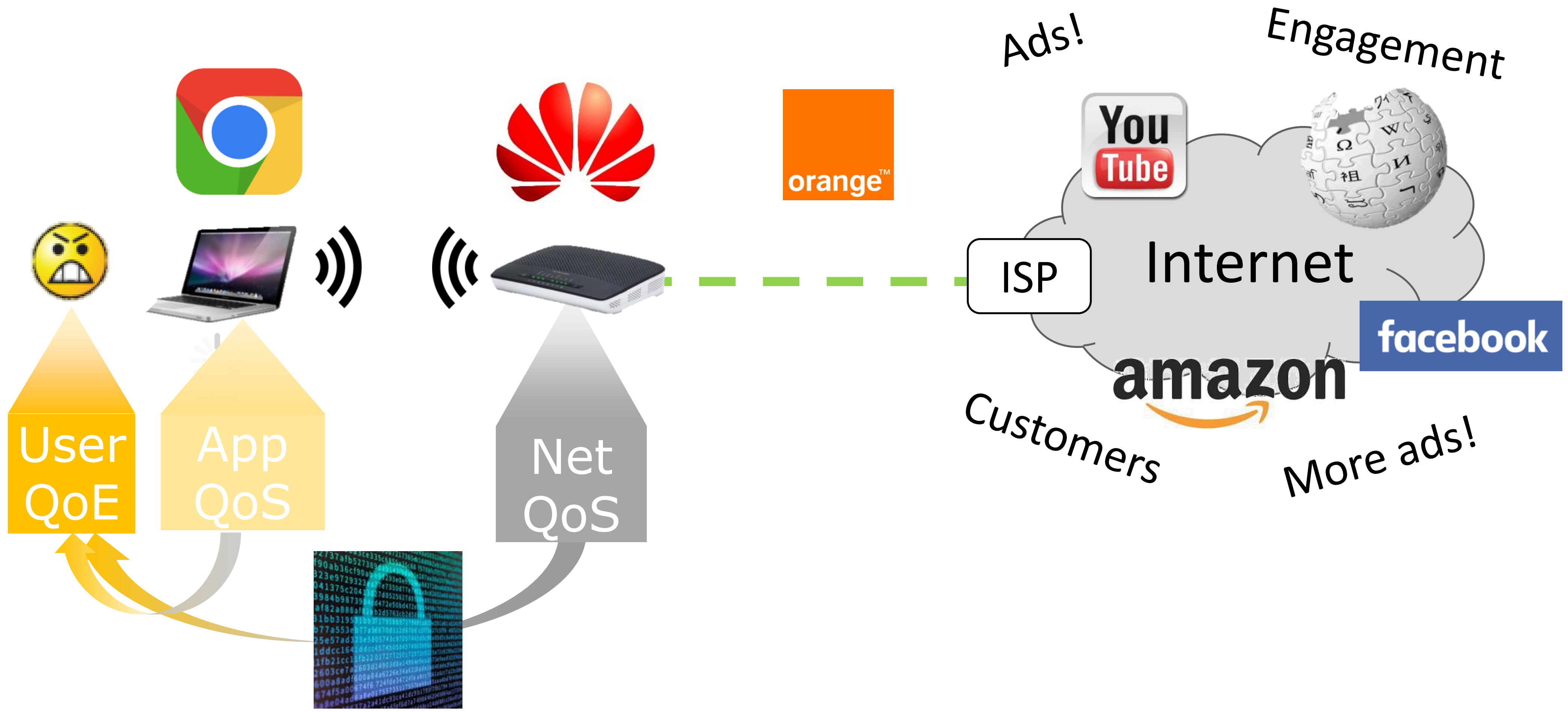
This talk

Dario Rossi

and, in alphabetical order, Alemnew Asrese, Alexis Huet, Diego Da Hora, Enrico Bocchi, Flavia Salutari, Florian Metzger, Gilles Dubuc, Hao Shi, Jinchun Xu, Luca De Cicco, Marco Mellia, Matteo Varvello, Renata Teixeira, Tobias Hossfeld, Shengming Cai, Vassillis Christophides, Zied Ben Houidi



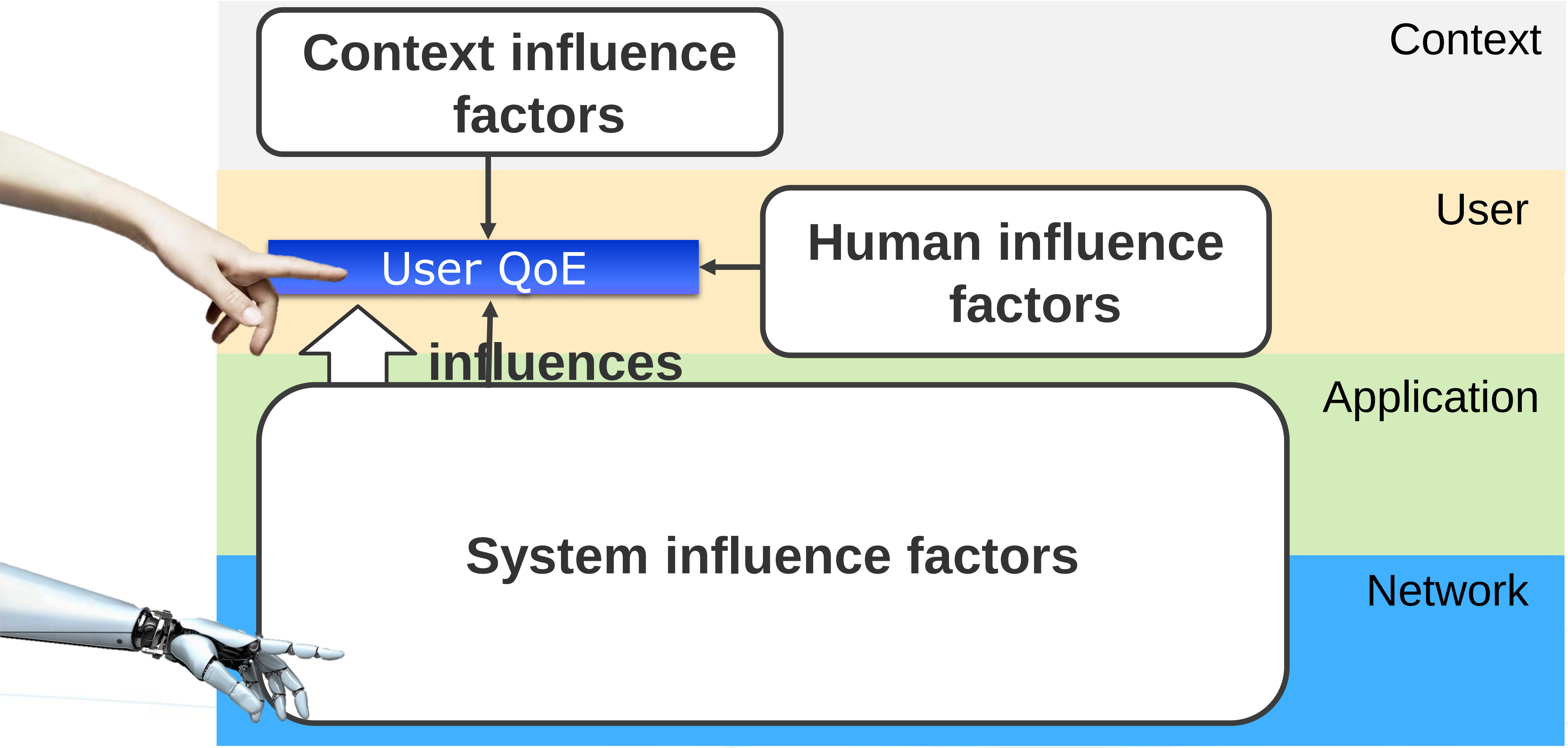
Offering Good user QoE is a common goal



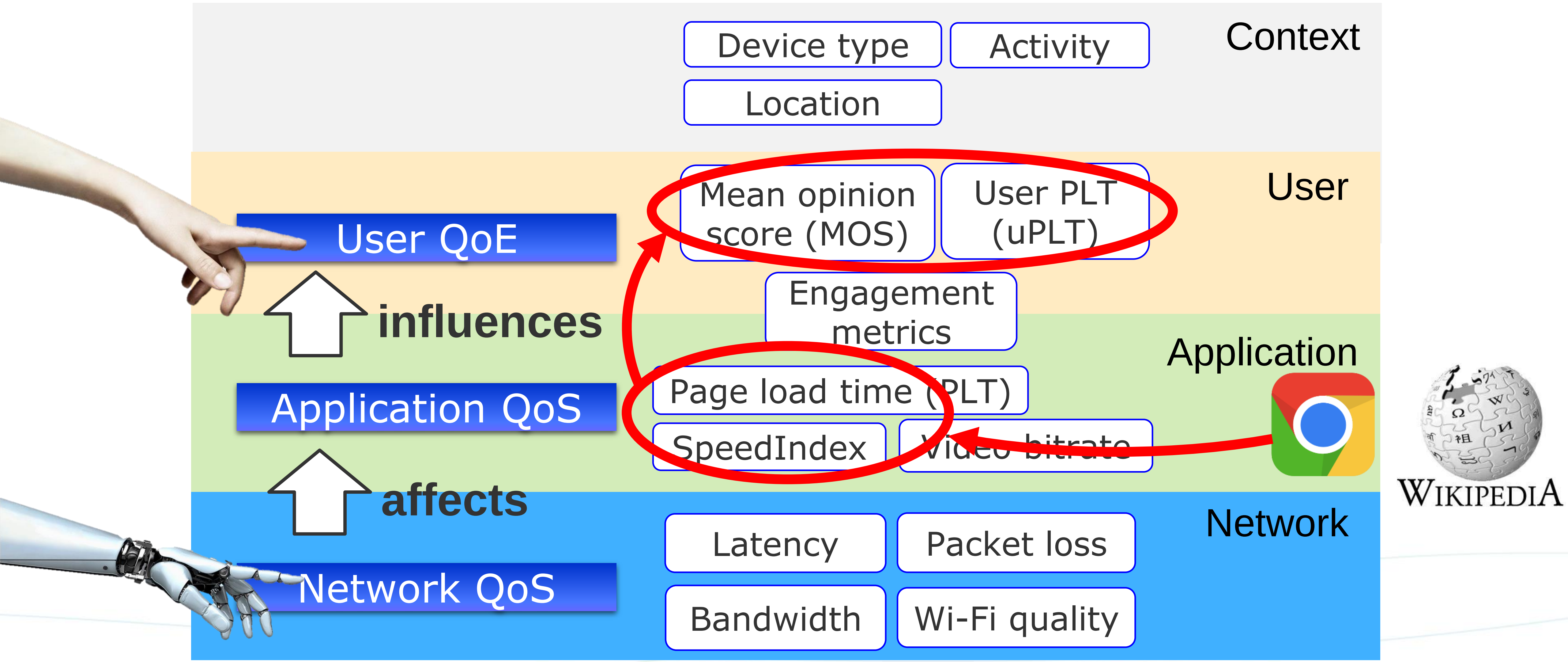
For ISPs/vendors encryption makes the inference harder

Detect/forecast/prevent Q😬E degradation is important!

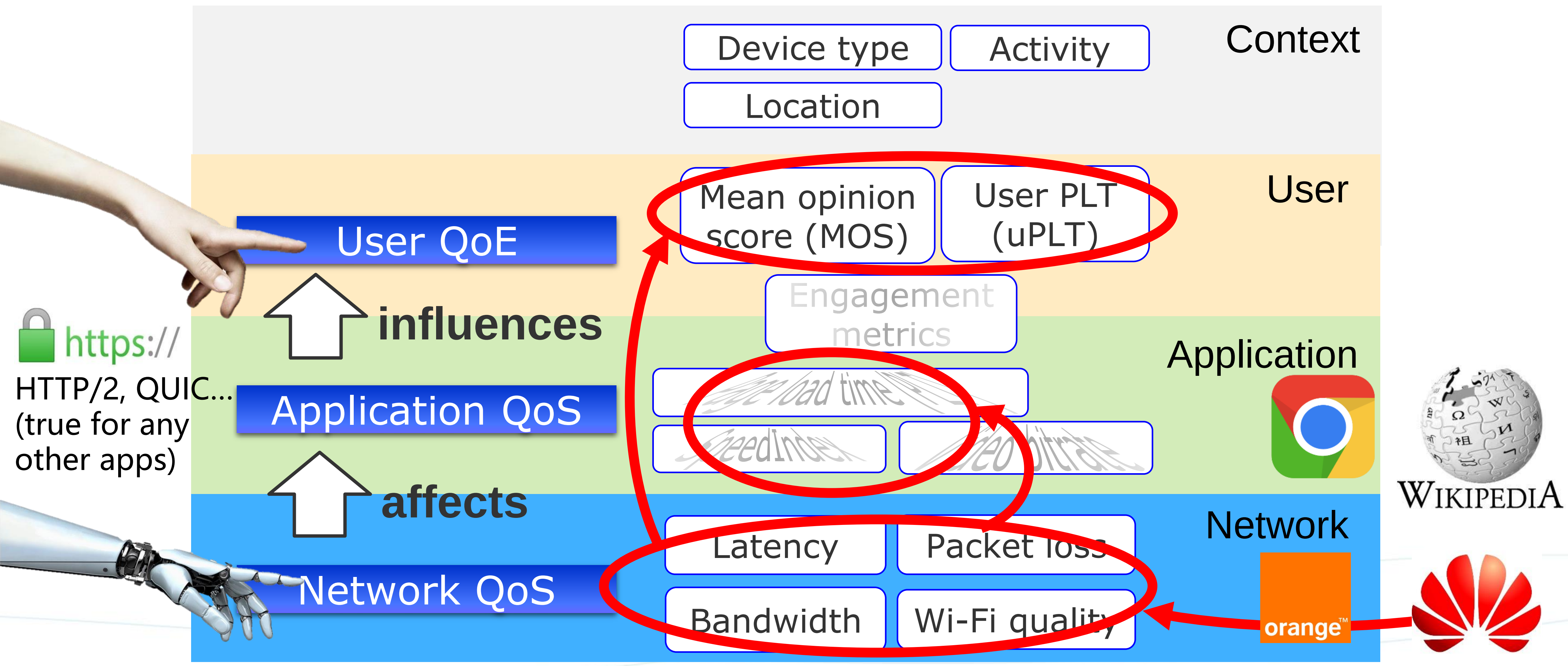
Quality at different layers



Quality at different layers



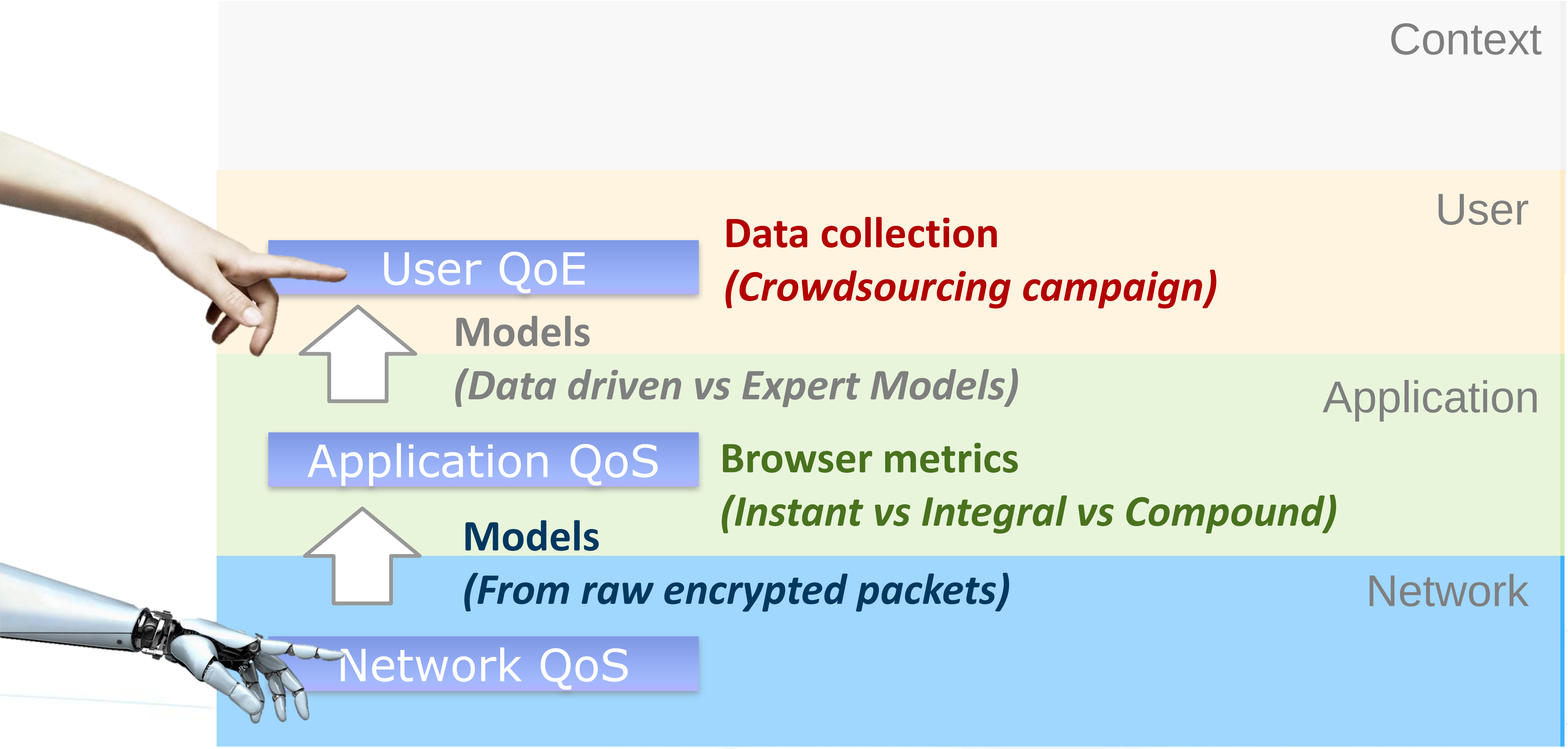
Quality at different layers



Agenda

Metrics and models for Web performance evaluation

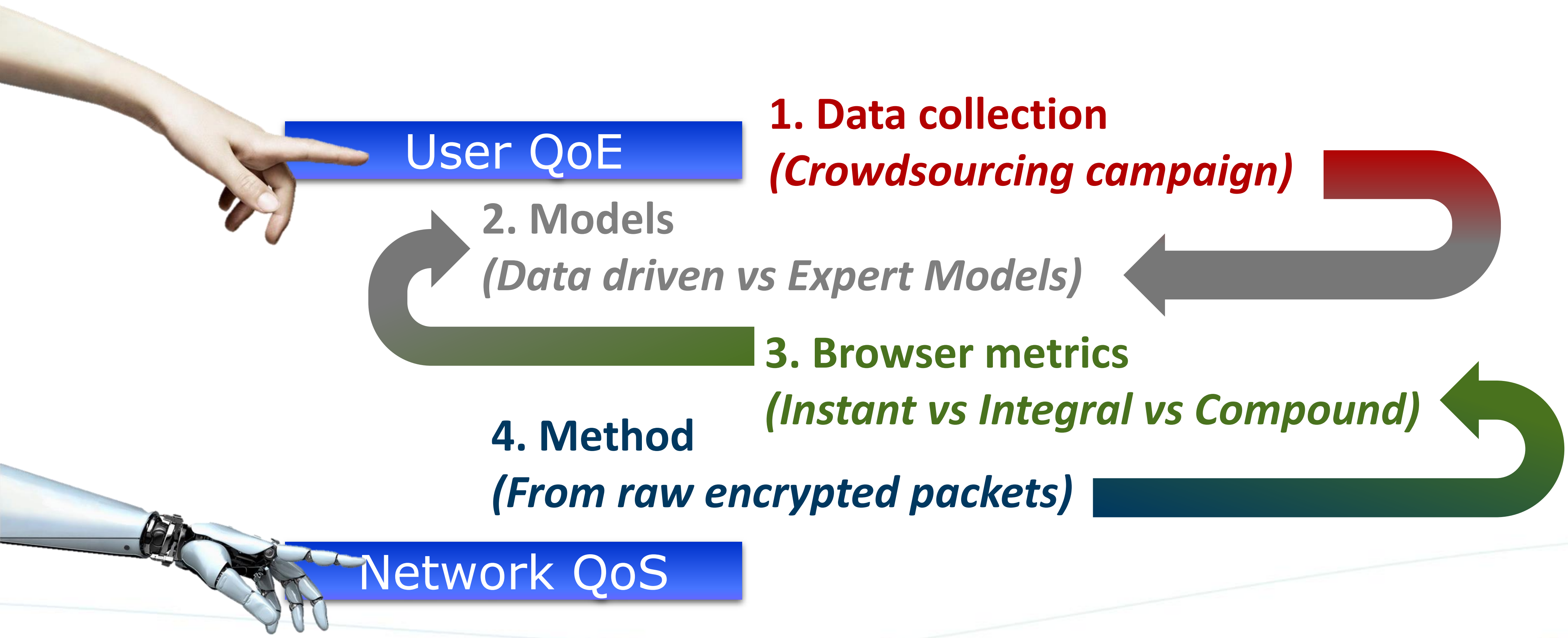
or, How to measure SpeedIndex from raw encrypted packets, and why it matters



Agenda

Metrics and models for Web performance evaluation

or, How to measure SpeedIndex from raw encrypted packets, and why it matters



Data collection: Crowdsourcing campaigns

https://webqoe.telecom-paristech.fr/data

- Mean opinion score (MOS)

"Rate your experience from 1-poor to 5-excellent"

Lab experiments
- − Small user diversity, volounteers
- + Web browsing, but artificial websites
- = Artificial controlled conditions

(Award winning) dataset [PAM18]

- User perceived PLT (uPLT)

"Which of these two pages finished first ?"

Crowdsourcing (payed crowdworkers)
- + Larger userbase, but higher noise
- − Side-to-side videos ≠ Web browsing!
- = Artificial controlled conditions

Ongoing, with eyeorg

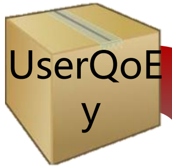
- User acceptance

"Did the page load fast enough ?" (Yes/No)

Experiments from operational website
- + Actual service users
- + Browsing in typical user conditions
- − Huge heterogeneity (devices/browsers/nets)

Collab with [WWW19]

Models: Data driven vs Expert models



https://webqoe.telecom-paristech.fr/models

Expert models

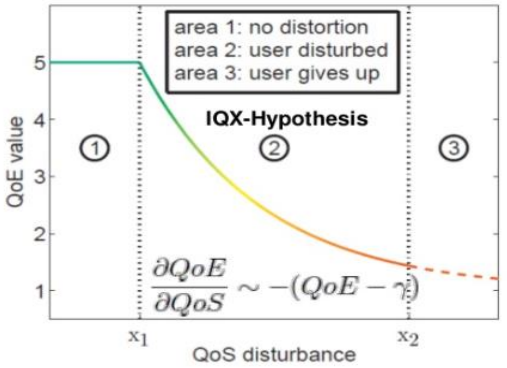
Fit predetermined y=f(x)

x=single scalar metric, generally Page Load Time (PLT)
f(.) = pre-selected by the expert

IQX Hypotesis

$$QoE(x) = \alpha e^{-\beta x} + \gamma$$

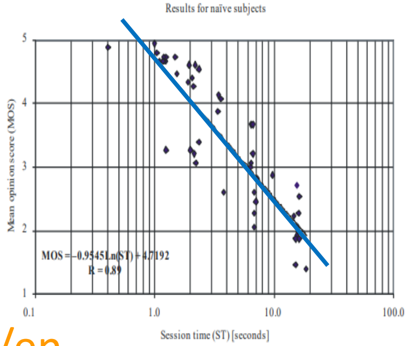
[1] M. Fiedler et al. "A generic quantitative relationship between quality of experience and quality of service." *IEEE Network*, 2010



Comparison of the two models in [QoMEX-18]

Weber Fechner

$$MOS = \frac{4}{\ln(Min / Max)} \cdot \left(\ln(SessionTime) - \ln(Min) \right) + 5$$



Standard ITU-T G1030

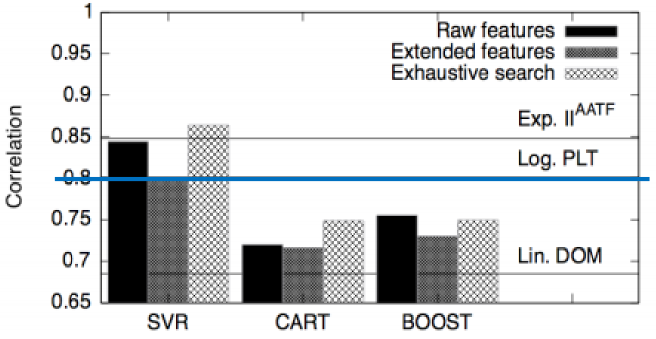
https://www.itu.int/rec/T-REC-G.1030/en

Data-driven

Learn y=f(x)

x=vector of input features
optimal f(.) selected & tuned by machine learning

More flexible and (slightly) more accurate [PAM18]

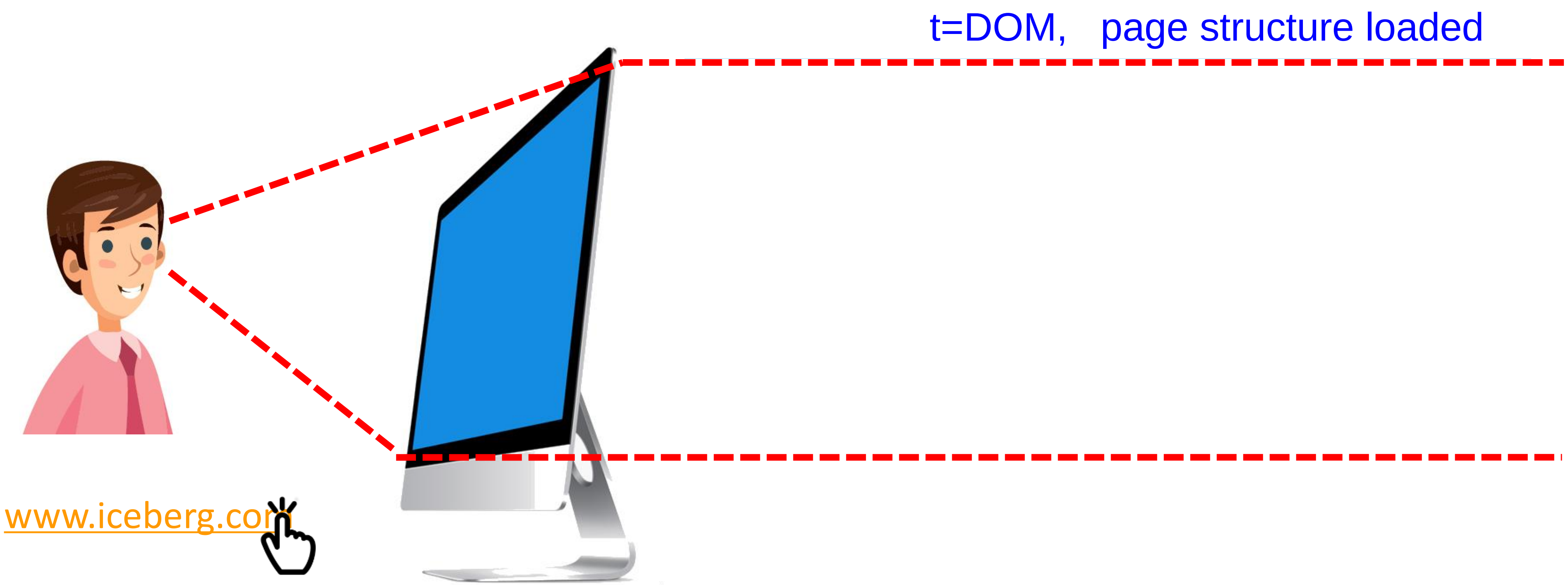


[INFOCOM19]

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{j=1}^{n} w_j x_j + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

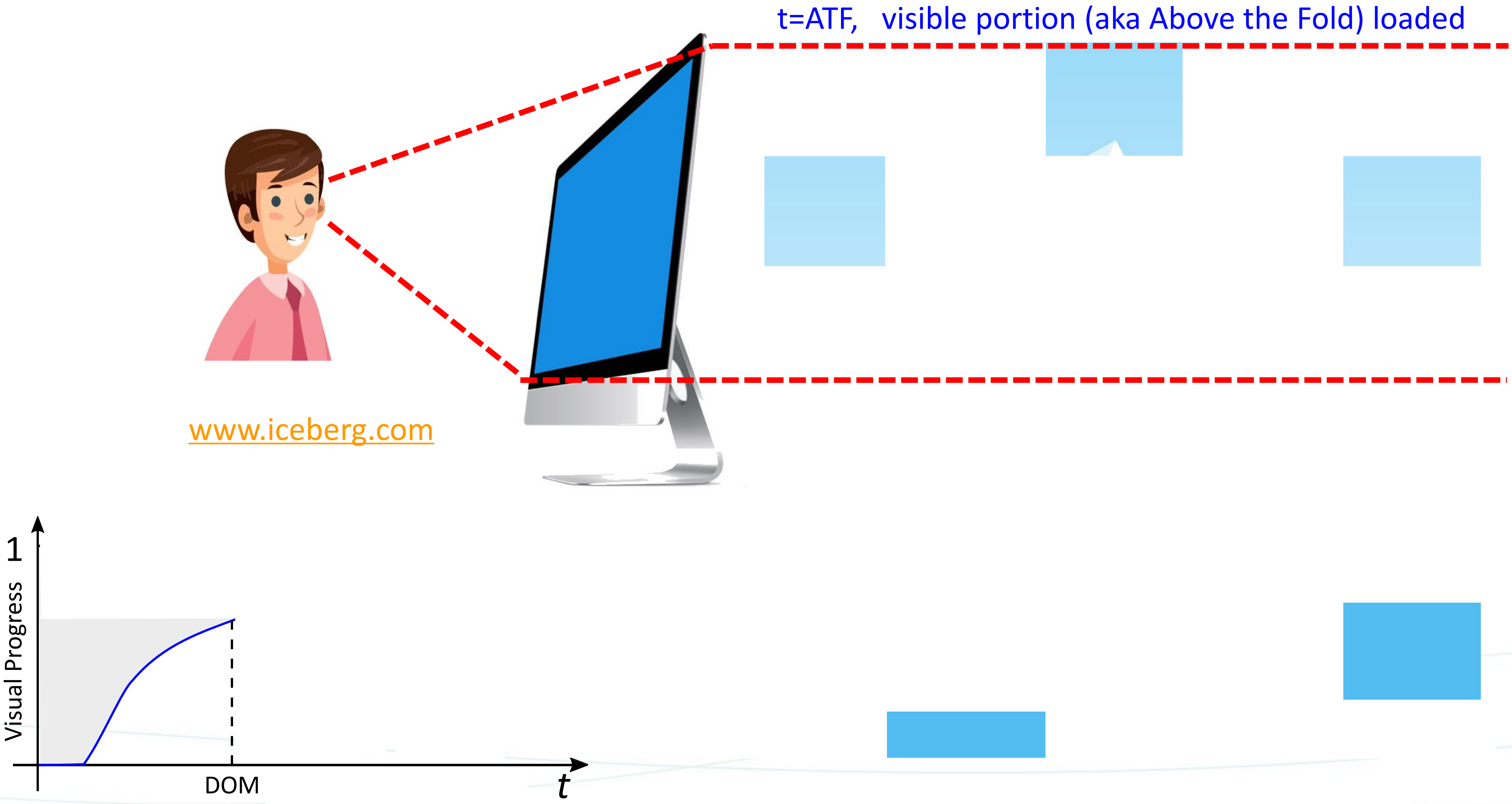
Still room for improvement (see [WWW19])

Browser metrics: Time Instant vs Time Integral (1/2)



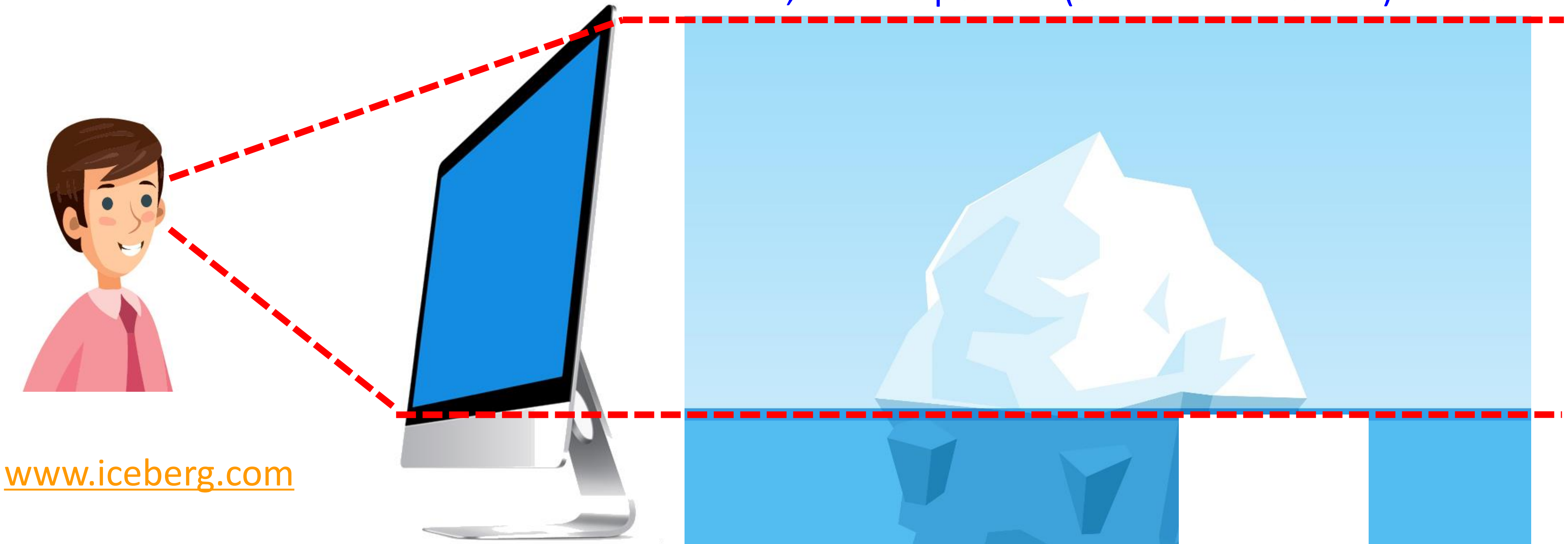
* Images by vvstudio, vectorpocket, Ydlabs / Freepik

Browser metrics: Time Instant vs Time Integral (1/2)

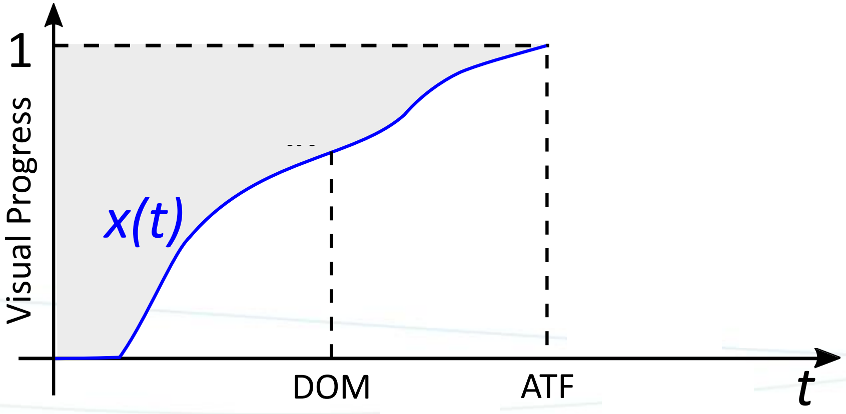


* Images by vvstudio, vectorpocket, Ydlabs / Freepik

Browser metrics: Time Instant vs Time Integral (1/2)

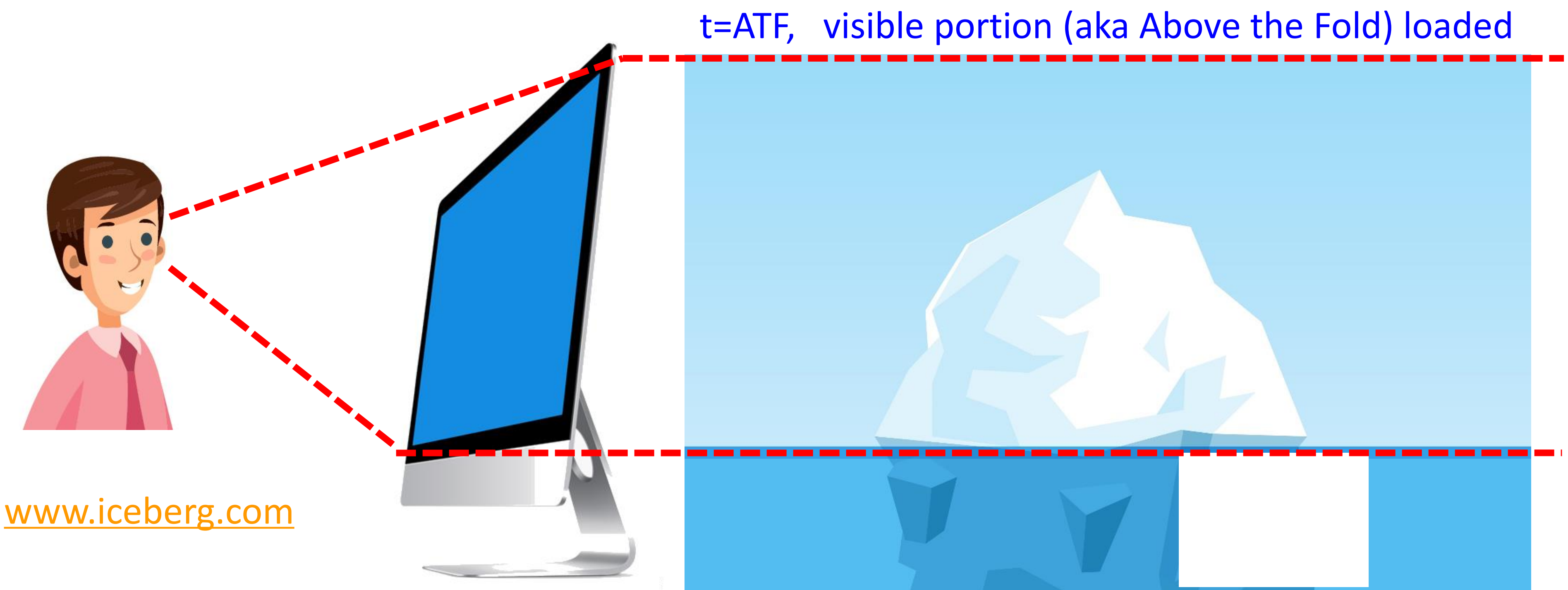


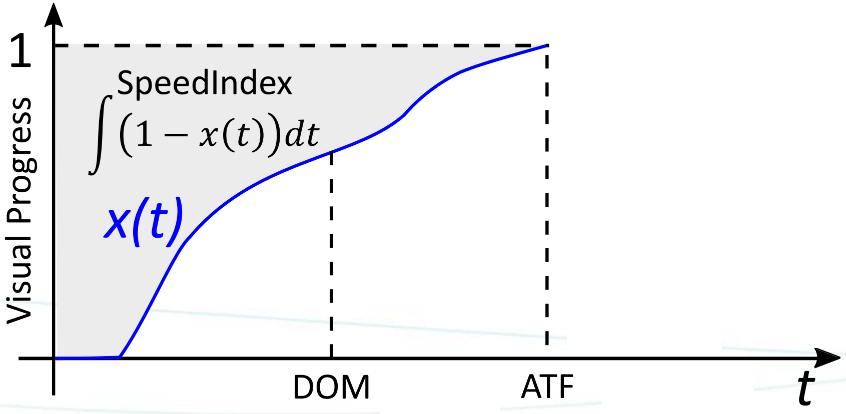
www.iceberg.com



* Images by vvstudio, vectorpocket, Ydlabs / Freepik

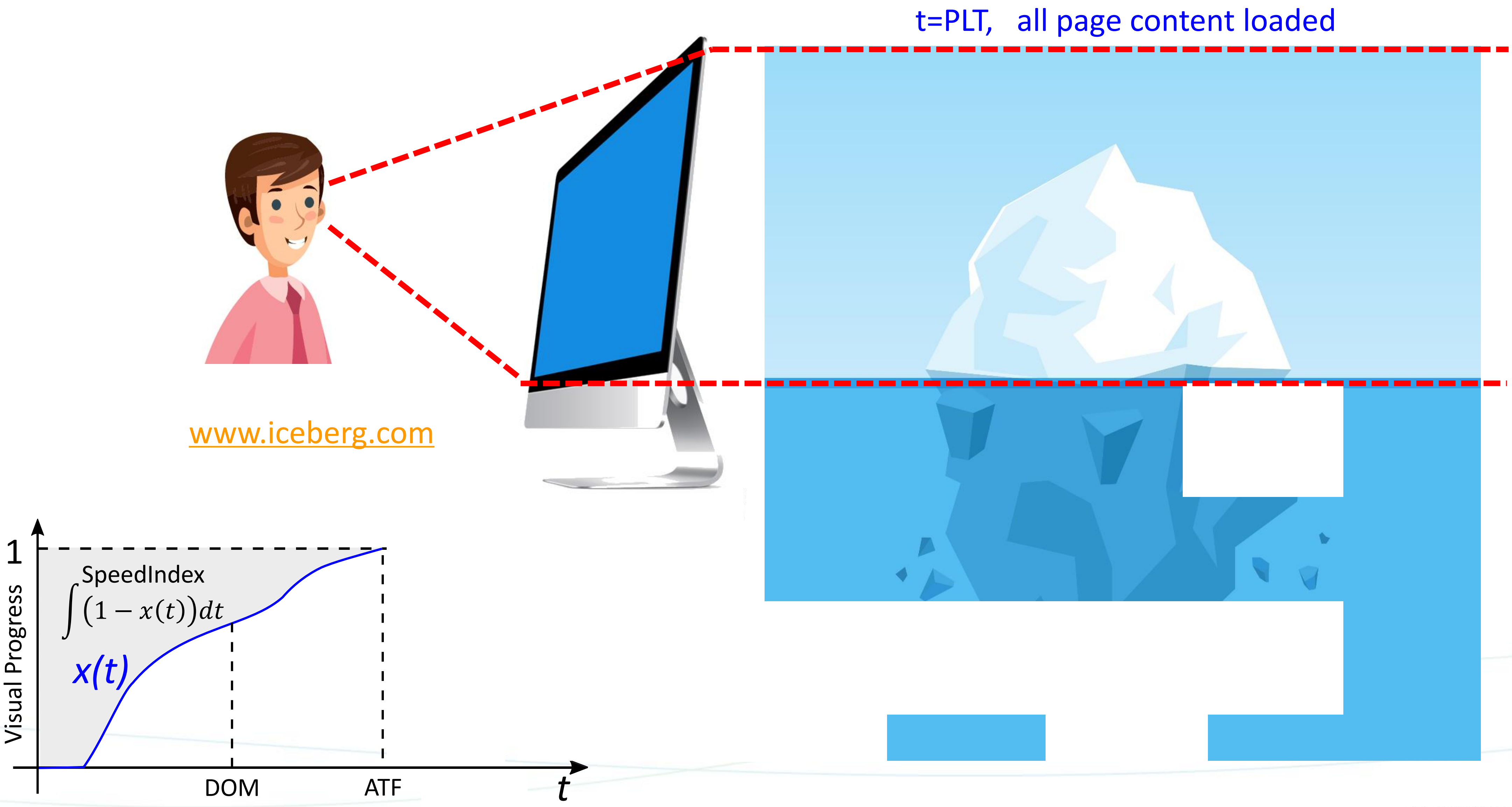
Browser metrics: Time Instant vs Time Integral (1/2)





* Images by vvstudio, vectorpocket, Ydlabs / Freepik

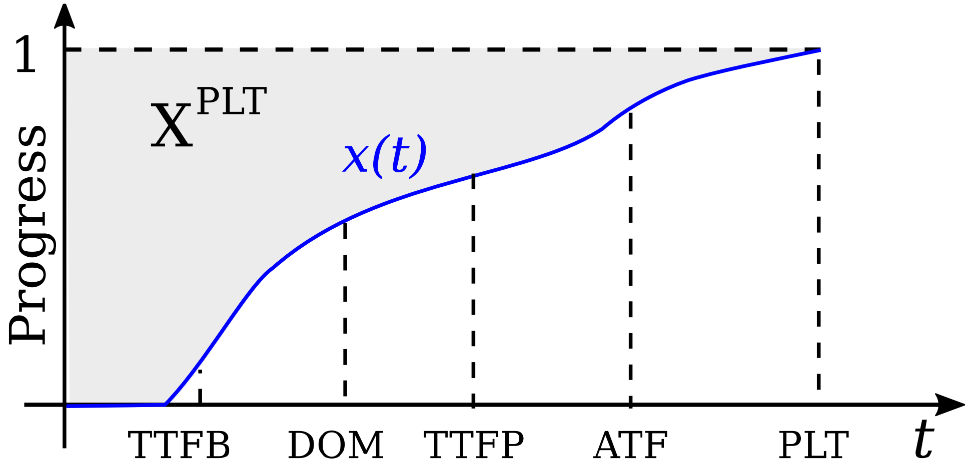
Browser metrics: Time Instant vs Time Integral (1/2)



* Images by vvstudio, vectorpocket, Ydlabs / Freepik

Browser metrics: Time Instant vs Time Integral (2/2)

- SpeedIndex, RUMSI, PSSI
  - Processing intensive ✗
  - Only at L7 (in browser) ✗
  - Visual progress metric ✓
- ObjectIndex, ByteIndex and ImageIndex
  - Lightweight ✓
  - ByteIndex also at L3 (in network) ✓
  - Higly correlated with SpeedIndex ✓
  - Possibly far from user QoE ? ?



$$X = \int_0^{t_{\mathrm{end}}} (1 - x(t))\,dt$$

SpeedIndex
%of visual completeness (histogram, rectangles or SSim)

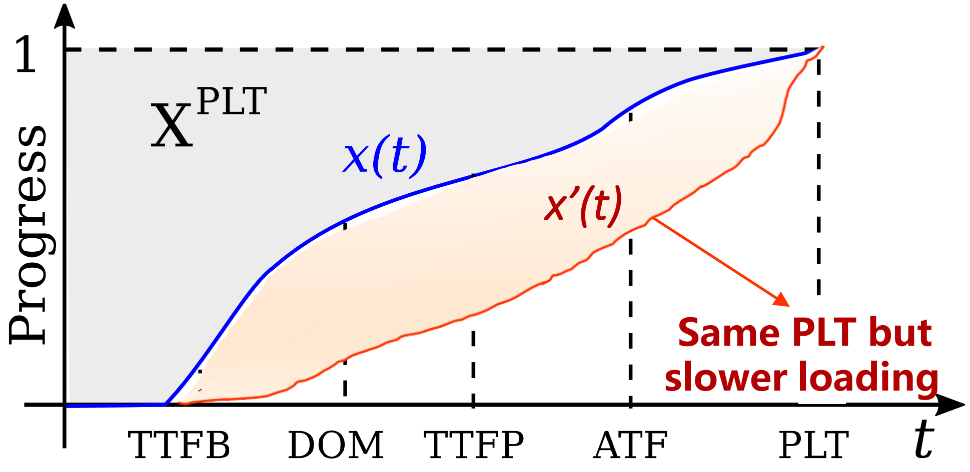
ObjectIndex
% of objects downloaded

ByteIndex
% of bytes downloaded

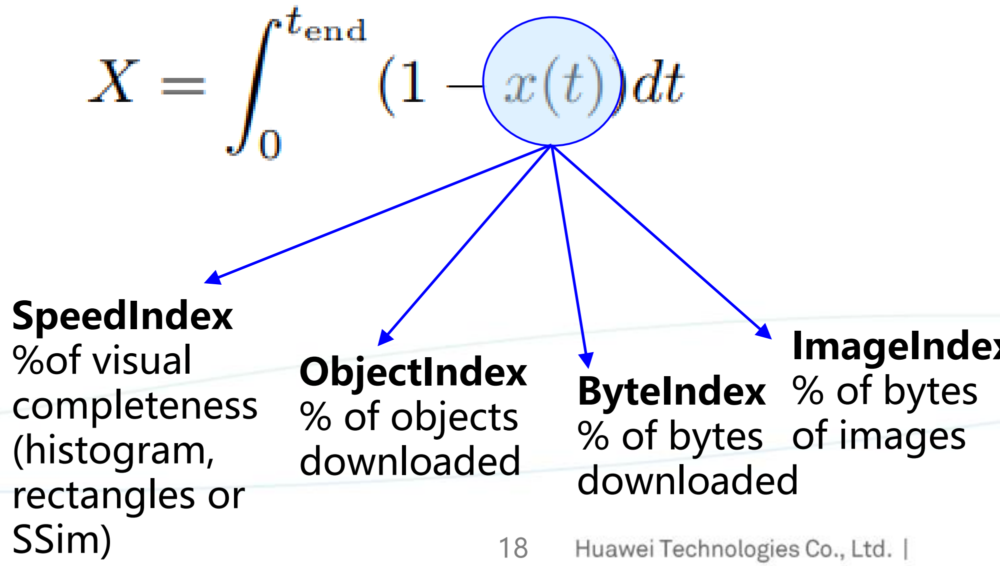
ImageIndex
% of bytes of images

Browser metrics: Time Instant vs Time Integral (2/2)

- SpeedIndex, RUMSI, PSSI
  - Processing intensive ✗
  - Only at L7 (in browser) ✗
  - Visual progress metric ✓



- ObjectIndex, ByteIndex and ImageIndex
  - Lightweight ✓
  - ByteIndex also at L3 (in network) ✓
  - Higly correlated with SpeedIndex ✓
  - Possibly far from user QoE ? ?



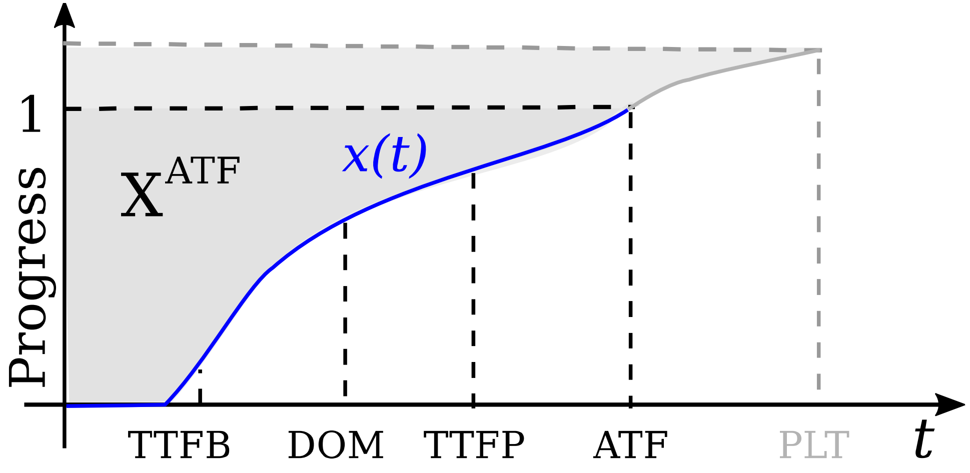
Browser metrics: Time Instant vs Time Integral (2/2)

SpeedIndex, RUMSI, PSSI

- Processing intensive ✗
- Only at L7 (in browser) ✗
- Visual progress metric ✓

ObjectIndex, ByteIndex and ImageIndex

- Lightweight ✓
- ByteIndex also at L3 (in network) ✓
- Higly correlated with SpeedIndex ✓
- Possibly far from user QoE ? ?



$$X = \int_0^{t_{end}} (1 - x(t))dt$$

Different cutoffs

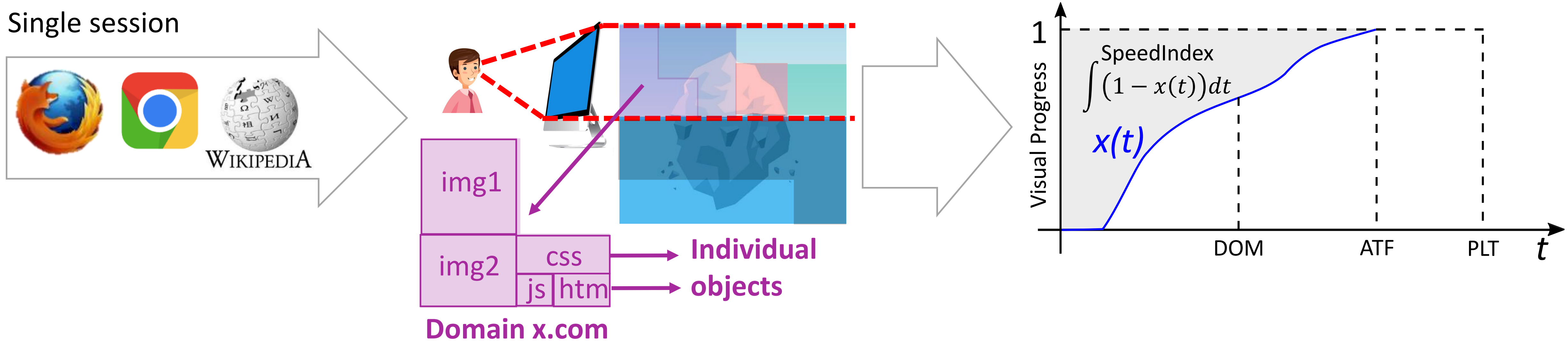
SpeedIndex
%of visual completeness (histogram, rectangles or SSim)

ObjectIndex
% of objects downloaded

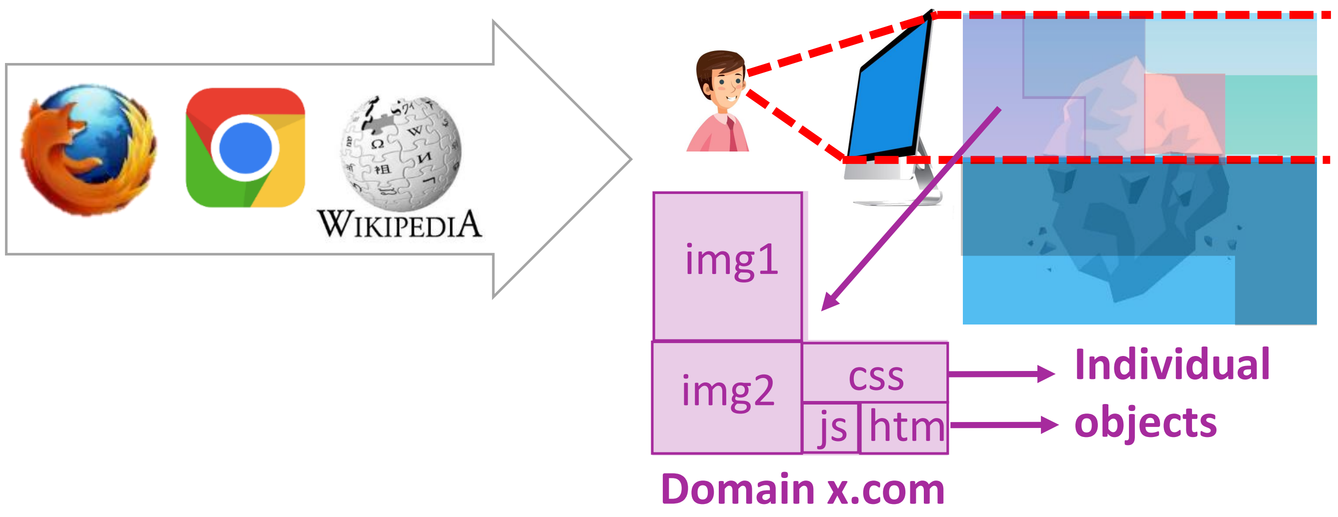
ByteIndex
% of bytes downloaded

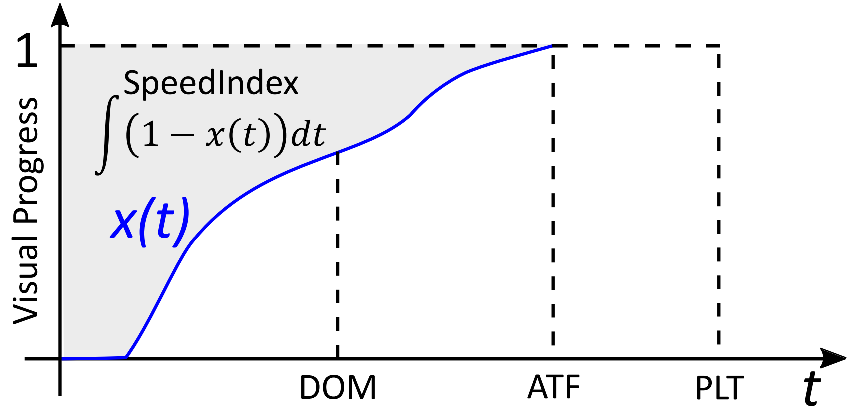
ImageIndex
% of bytes of images

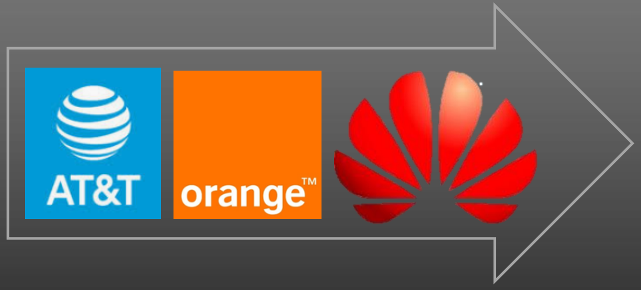
Method: From raw packets to browser metrics (1/2)



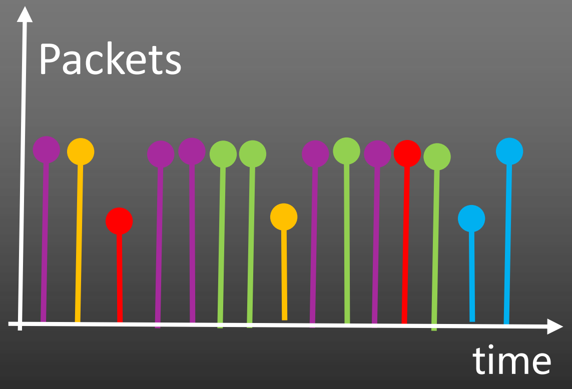
Method: From raw packets to browser metrics (1/2)



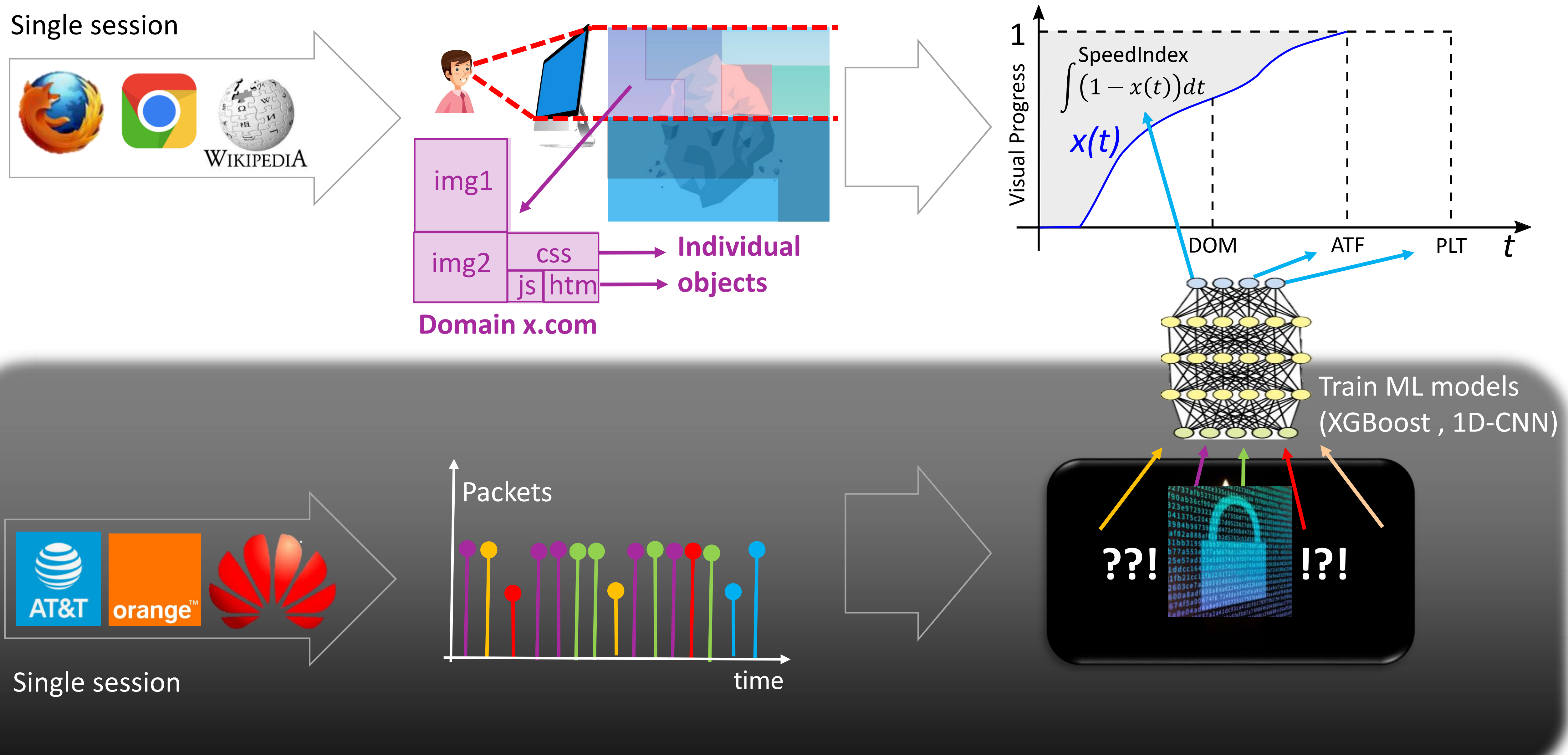




Single session



Method: From raw packets to browser metrics (1/2)



Method: From raw packets to browser metrics (2/2)

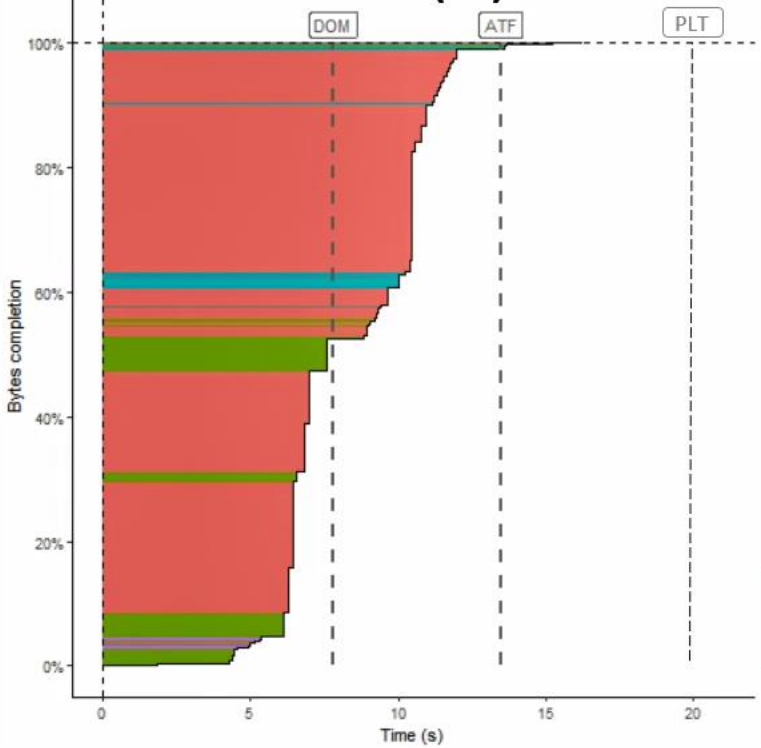
Play

Webpage rendering

User

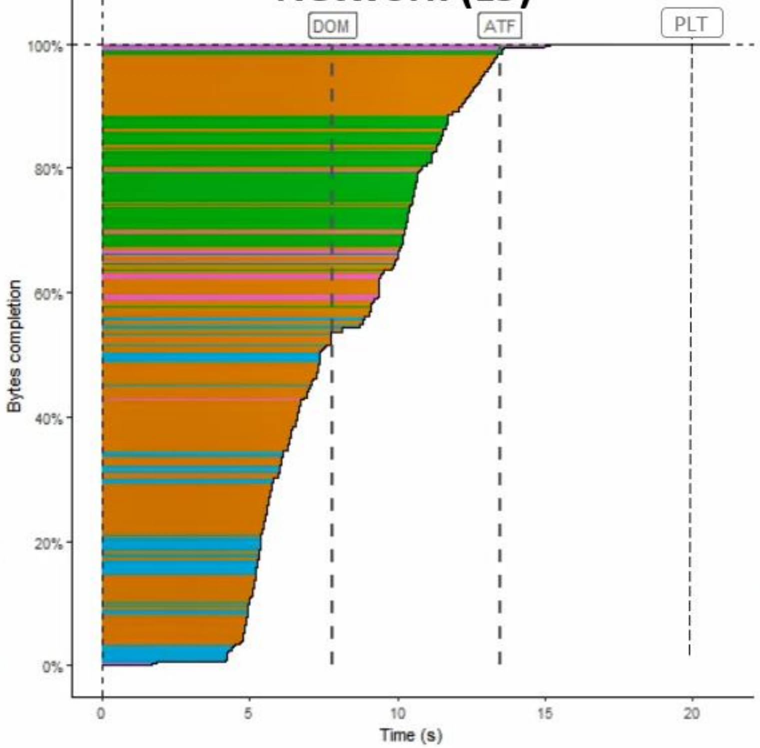
1 burst = 1 object
1 color = 1 domain

Browser (L7)



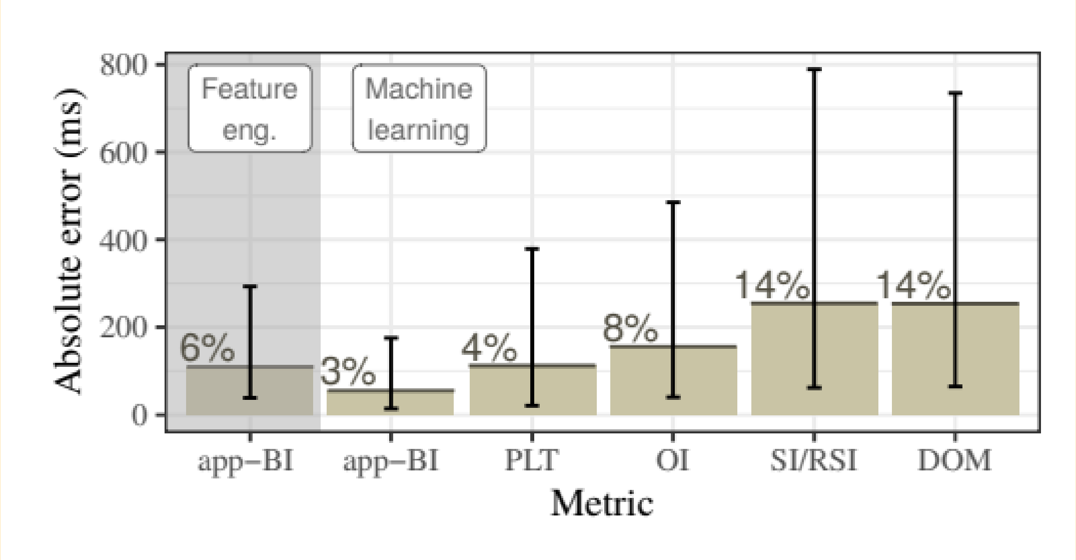
1 burst = 1 packet
1 color = 1 IP server

Network (L3)



Method: From raw packets to browser metrics (2/2)

- ✓ Works with encryption
- ✓ Handle multi-sessions (not in this talk)
- ✓ Exact online algorithm for ByteIndex
- ✓ Machine learning for any metric
- ✓ Accurate on joint tests with Orange
- ✓ Accurate for unseen pages & networks
- ✓ Available soon into Huawei products

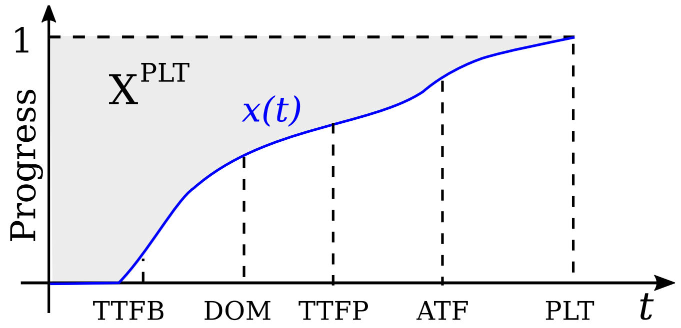


Aftermath (1/3): From raw packets to rough sentiments

- Expert-driven feature engineering
  - Explainable but inherently heuristic approach
  - Hard to keep in sync with application/network change
- Neural Networks
  - Less interpretable but more versatile
  - Downside: requires *lots* of samples....

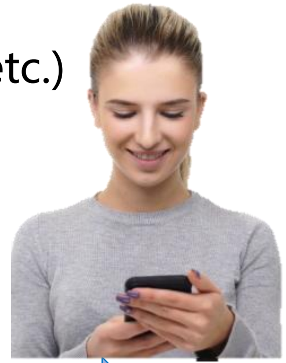
› Feed NN with *x(t) signal*

› Still lightweight

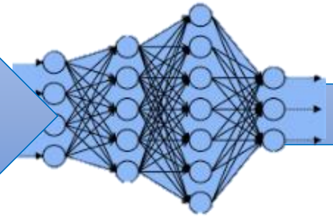


› User feedback (e.g. MOS, user PLT, etc.)

› Smartphone sensors (eg happiness estimation via facial recognition)



Possible inputs



Possible outputs

› Feed NN using a *filmstrip*

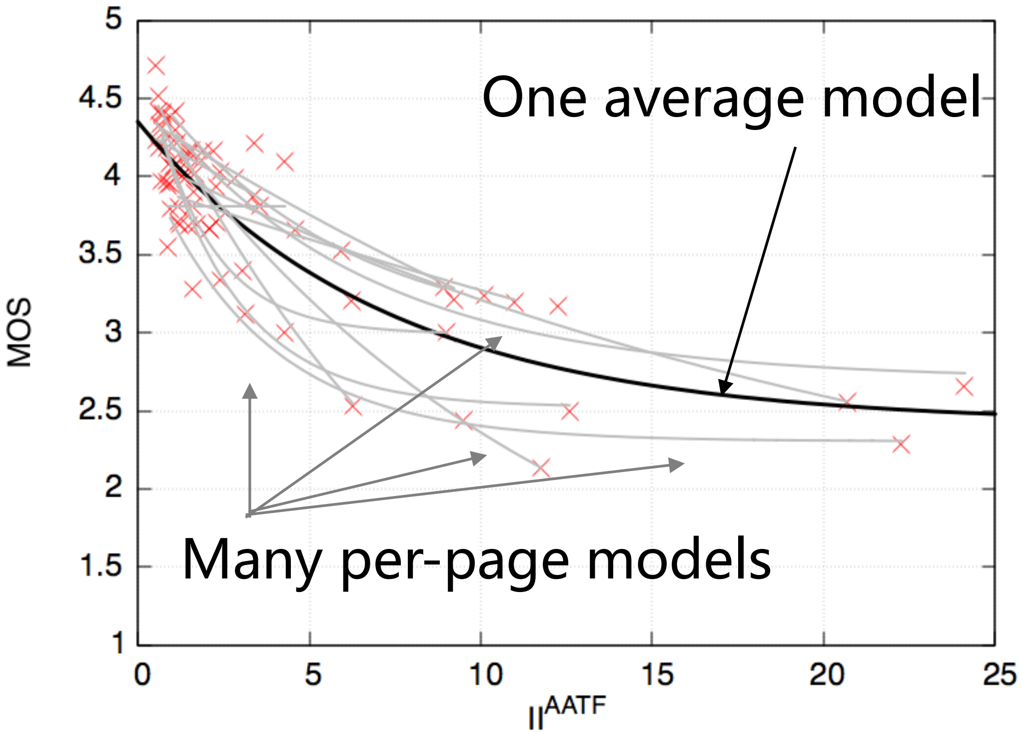
› More complex

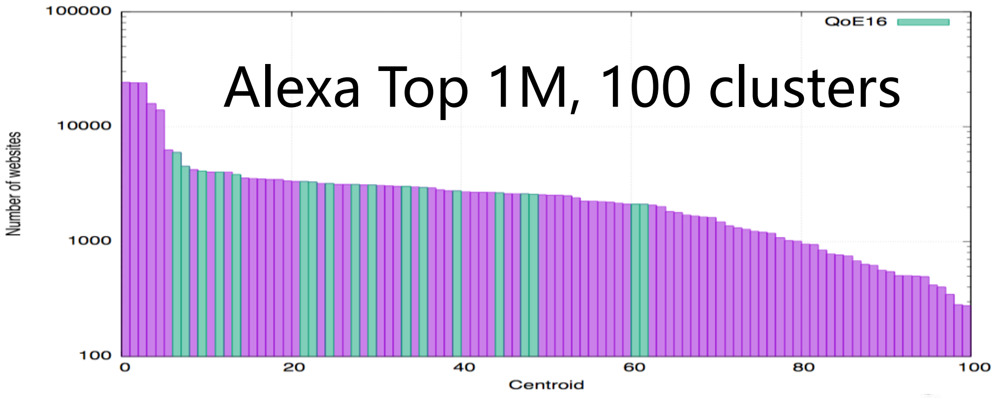
› Brain signals acquired with sensors

› Activity of brain areas correlated with user happiness

Aftermath (2/3): Divide et impera

- World Wild Web
  - Huge diversity, not captured by single model
- Increase accuracy
  - Per-page QoE models
  - Inherently non scalable
- Increase accuracy & scalability
  - Per-page QoE models (eg Alexa top 100 pages)
  - Aggregate QoE models (eg 100 clusters top 1M)
  - Generic QoE model (for the tail up to 1B pages)

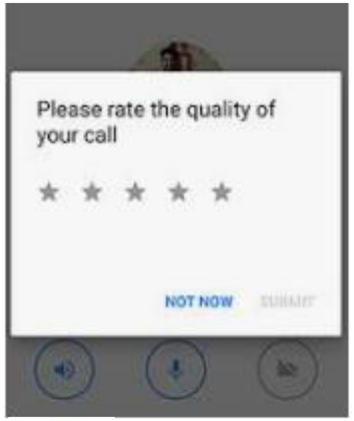




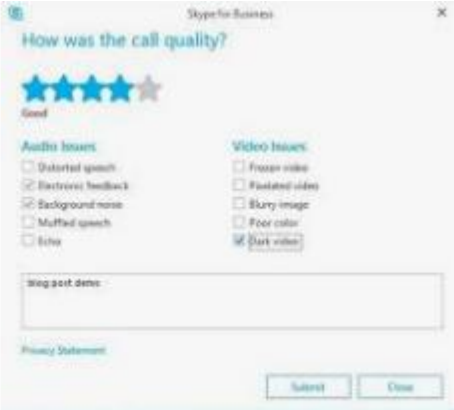
, Ltd. |

Aftermath (3/3): Keep collecting (and sharing) data

- Other applications/players are doing this already!

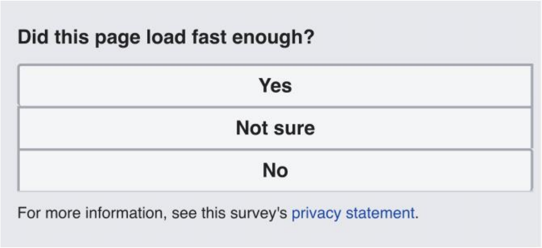


Facebook



Skype

Android



Wikipedia

Physical world

- Sustained continuous user QoE indication benefits
  - Useful samples for QoE management assessment, troubleshooting, regression detection, etc.
  - Get continuous stream of samples for improving QoE = f(QoS) models on the long run
- Very limited downsides (risk of annoying users if leveraging small panels)

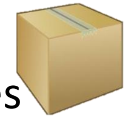
https://webqoe.telecom-paristech.fr/

Documents
Datasets
Code

[SIGCOMM-19] Huet, Alexis and Houidi, Zied Ben and Cai, Shengming and Shi, Hao and Xu, Jinchun and Rossi, Dario, Web Quality of Experience from Encrypted Packets ACM SIGCOMM Demo, aug. 2019

[INFOCOM-19] Huet, Alexis and Rossi, Dario, Explaining Web users QoE with Factorization Machines IEEE INFOCOM Demo apr. 2019

60k+ real user grades



[WWW-19] F. Salutari, D. Da Hora, G. Dubuc and D. Rossi A large scale study of Wikipedia users' quality of Experience Proc. WWW, 2019

Chrome plugin implementation

[SIGCOMM-18] D. da Hora, D. Rossi, V. Christophides, R. Renata, A practical method for measuring Web above-the-fold time, ACM SIGCOMM Demo, aug. 2018,

[QOMEX-18] Hossfeld, Tobias and Metzger, Florian and Rossi, Dario, Speed Index: Relating the Industrial Standard for User Perceived Web Performance to Web QoE 10th International Conference on Quality of Multimedia Experience (QoMEX 2018) jun. 2018

9k real human grades

[PAM-18] D. da Hora, A. Asrese, V. Christophides, R. Teixeira and D. Rossi, Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics Proc. PAM 2018, **Best dataset award ★**

[PAM-17] Bocchi, Enrico and De Cicco, Luca and Mellia, Marco and Rossi, Dario, The Web, the Users, and the MOS: Influence of HTTP/2 on User Experience Proc. PAM 2017

10k automated experiments

[SIGCOMM-QoE-16] E. Bocchi, L. De Cicco, D. Rossi, Measuring the Quality of Experience of Web users, ACM SIGCOMM Internet-QoE worshop 2016, **Best paper award ★**

?? || //

- Thanks for lis

Loading