



# Boosting Human: Cognitive Approach for Anti-Phishing

Daisuke MIYAMOTO, Ph.D.

Assistant Professor

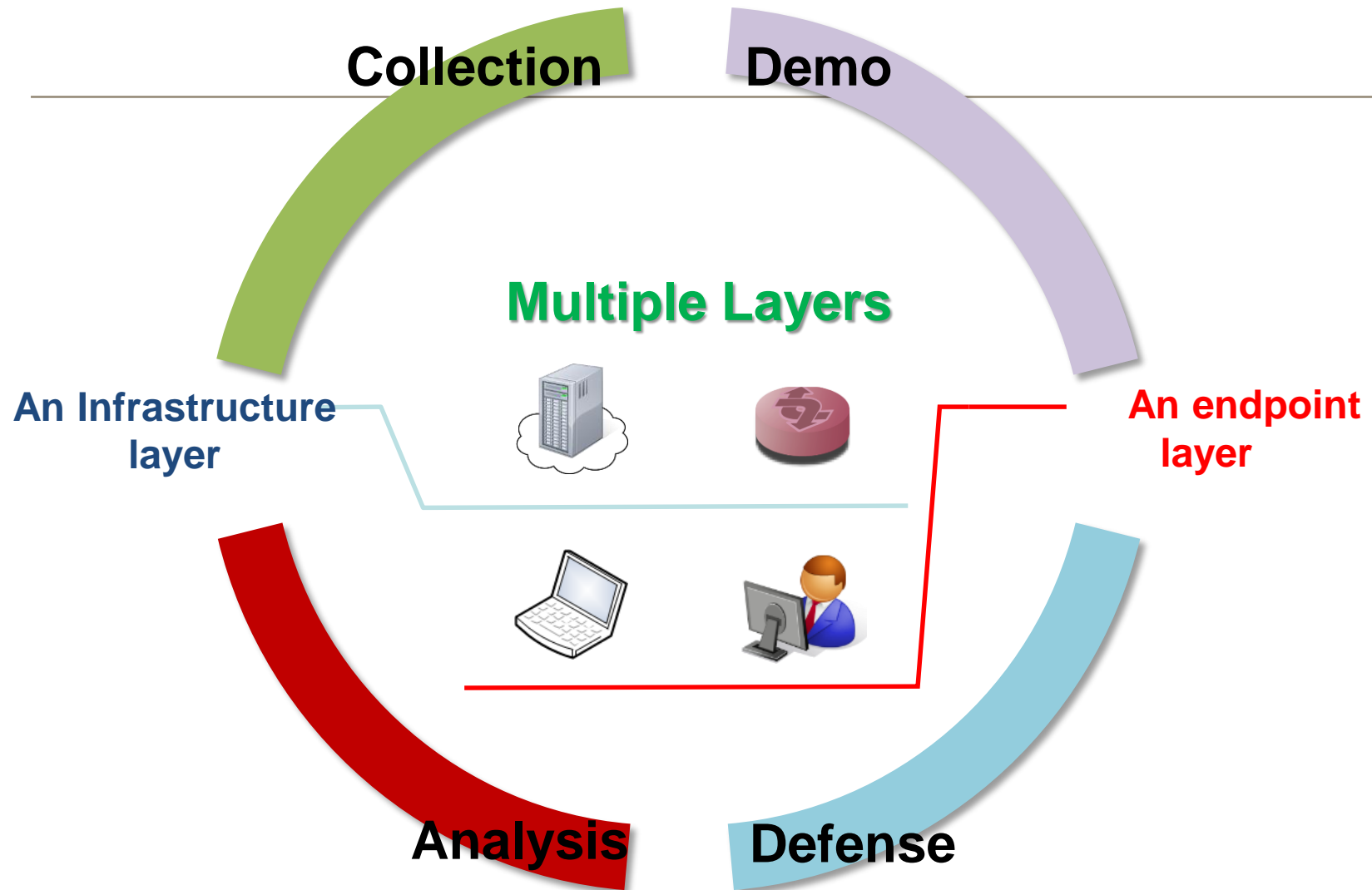
The University of Tokyo

daisu-mi@nc.u-Tokyo.ac.jp

# NECOMA : Project Description

**The NECOMA project aims at providing new means to understand cyber threats and to mitigate their effect on infrastructure and endpoints.**

- Threat data collection
  - NECOMA addresses the aspect of data collection, leveraging past and current work on the topic with the goal to expand these existing mechanisms and orient them towards threat data analysis.
- Threat data analysis
  - it addresses threat data analysis not only from the perspective of understanding attackers and vulnerabilities, but also from the point of view of the target and victim, needing to protect itself in real-time and in the most efficient manner possible.
- Cyber defense for improved resilience
  - it aims to develop and demonstrate new cyberdefense mechanisms that leverage these metrics for deployment and evaluation.



# Partners

---

## **EU-side partners**

IMT (France, Coordinator)

Institut Mines-Télécom

ATOS (Spain)

Atos Spain S.A.

FORTH (Greece)

Foundation for Research and Technology  
– Hellas

NASK (Poland)

Research and Academic Computer  
Network

6cure (France)

6cure SAS

## **JP-side partners**

NAIST (Coordinator)

Nara Institute of Science and Technology

IJ-IL

IJ Innovation Institute

NII

National Institute of Informatics

KEIO

Keio University

UTokyo

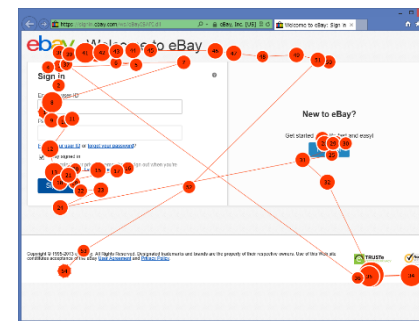
The University of Tokyo

# NECOMA : End-user protection

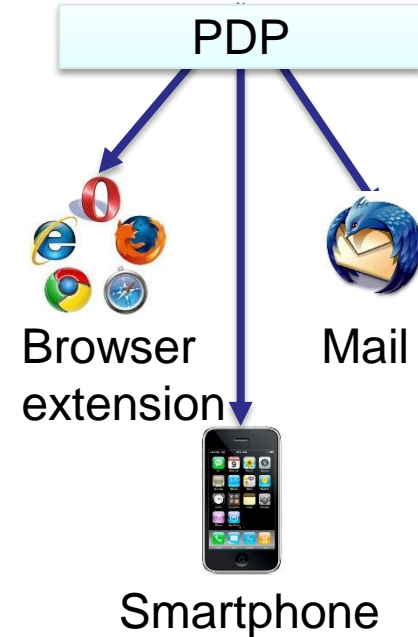
## Eye-Tracking Camera



## Biosensors



## Internal mental mode



## Collection

### End-user behavior

- Users decision (criterion)
- Eye-movement
- Biosensors

## Analysis

### Cognitive anomaly

- Eye-tracking analysis
- Stress detection

## Defense

### End-user protection

- Extension(Browser, MUA)
- Smartphone protection
- Human-error prevention

# Case studies of phishing in Japan



Bank websites in Japan

Legitimate:

[www.smbc-card.com](http://www.smbc-card.com)

Phishing:

[smbc.card.xxx.com](http://smbc.card.xxx.com)



Popular online games in Japan

Legitimate:

[hiroba.dqx.jp](http://hiroba.dqx.jp)

Phishing:

[hiroba.dqx.jp.xxx.xx.xx](http://hiroba.dqx.jp.xxx.xx.xx)



Spear Phishing

Legitimate:

[gateway.u-tokyo.ac.jp](http://gateway.u-tokyo.ac.jp)

Phishing:

[gateway.u-tokyo.ac.jp.xxx.xx](http://gateway.u-tokyo.ac.jp.xxx.xx)

# Approaches to counter phishing



## Education



- Development of educational materials
- Training against social engineering



## Attention

- Development of user interfaces
- Awareness for security information

 PayPal, Inc. [US] <https://www.paypal.com/>

 PayPal, Inc. [US] 

 PayPal, Inc. (US)



## Support making decision

- Detection of phishing sites
- Alerting when visiting suspicious websites

L: 1 (phishing)

## Detection Methods

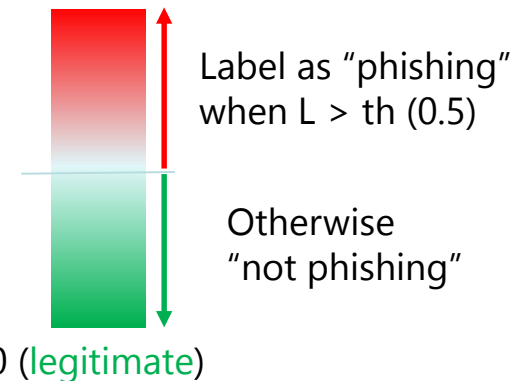
(1) URL filtering : Comparing visiting URL with **database**  
(challenge) lightweight collection and management of phishing URLs

(2) Heuristics-based solution :

Calculating **likelihood** (L) of the site being phishing

Comparing the likelihood with **threshold** (*th*)

*(challenge) development of new heuristics and calculation methods*



# Human Centric Approach (Human Boost)

## Overview

### Users' decision as new heuristics



### Machine learning for personalization

- cover users' weak points
- Personalization : the suitable anti-phishing strategy for novices and security experts might be differentiated.



## Theoretical background

### New heuristic

a users' decision ( phish or not ) can be used as an identifier of phishing, similar to the existing heuristics

### Calculation of Likelihood

boosting, enables to cover users' weakness by adjusting weights on heuristics

$$D_{t+1} = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases}$$

Such heuristic is assigned high weights that can label correct to the website

where users tends to misjudge

Personalization

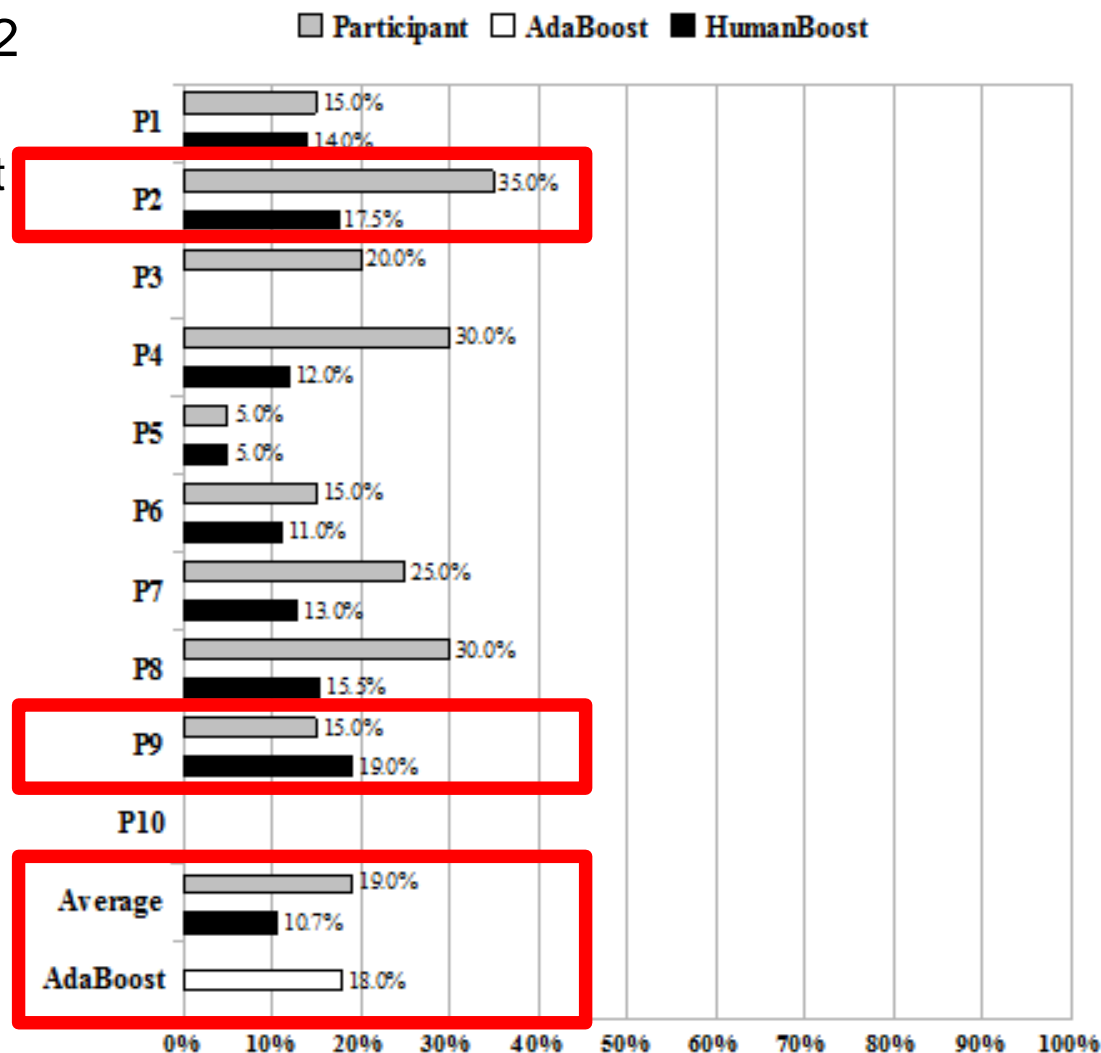


(D. Miyamoto et al, "HumanBoost: Utilization of Users' Past Trust Decision for Identifying Fraudulent Websites", 2009)  
(H. Pareek et al, "Human Boosting", 2013)



# Human Boost in phishing detection

- 10 participants browsed 2 websites
  - Labeling phishing or not
- Error rates were ..
  - AdaBoost-based (18%)
  - Participant (19%)
  - HumanBoost (10.7%)



# Behind Human Boost

---

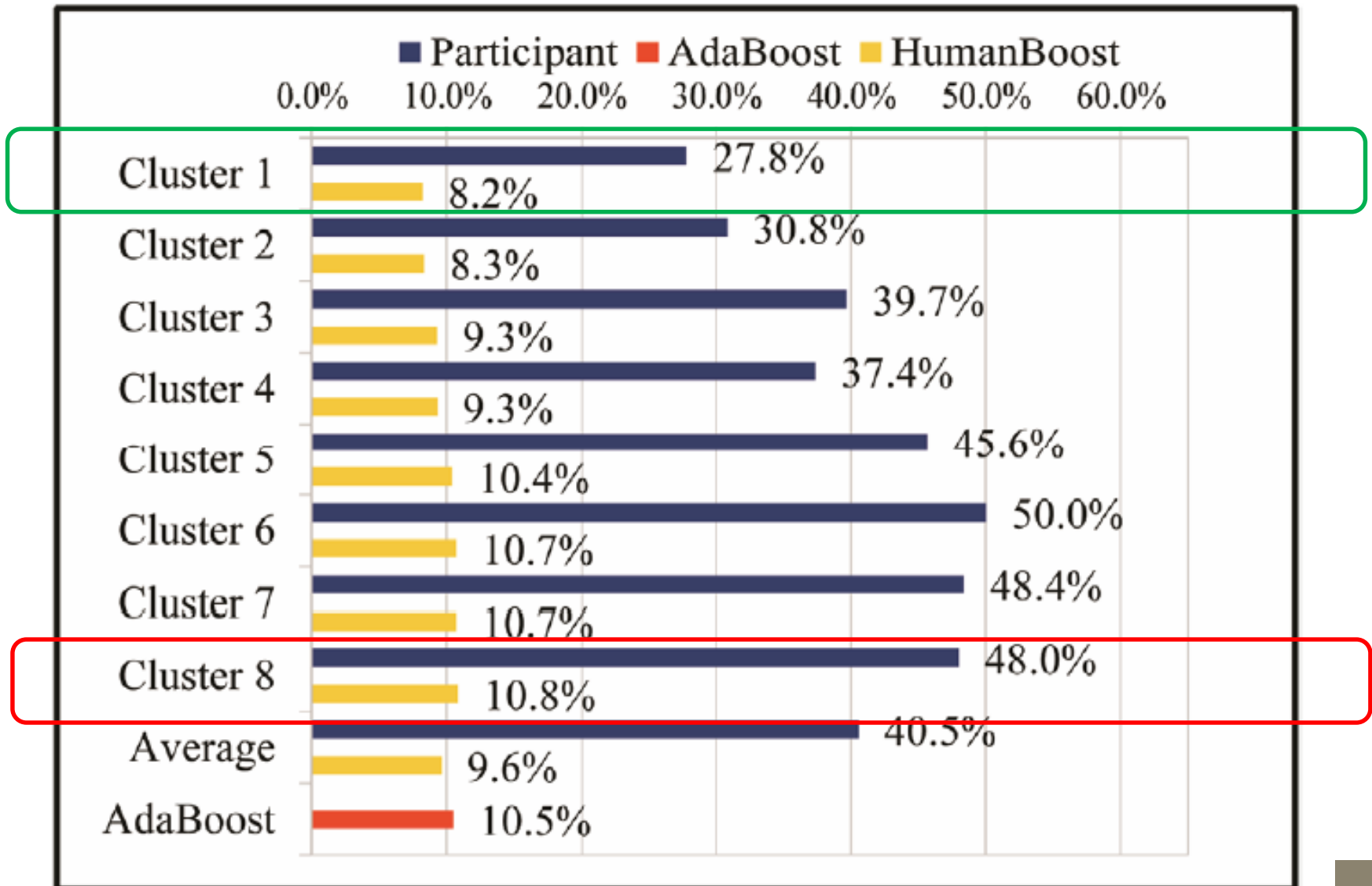
- 309 participants browsed 40 websites
  - 20 for estimating the ability for correct decision
  - 20 for Human Boost
  
- Perception of Website's credibility
  - check of how a participant labeled a site as phishing or not
    - Content of Web page
    - URL of the site
    - Security information of Browser
    - Other Reason

(D. Miyamoto et al, "Behind HumanBoost: Analysis of Users' Trust Decision Patterns for Identifying Fraudulent Websites", 2012)

# Decision criterion and error rates

Content of web page	URL of the site	Security information of browser	The average error rate
v			61.9%
	v		25.5%
		v	36.8%
v	v		51.7%
v		v	60.0%
	v	v	17.9%
v	v	v	49.8%

# Comparison of the average error rates

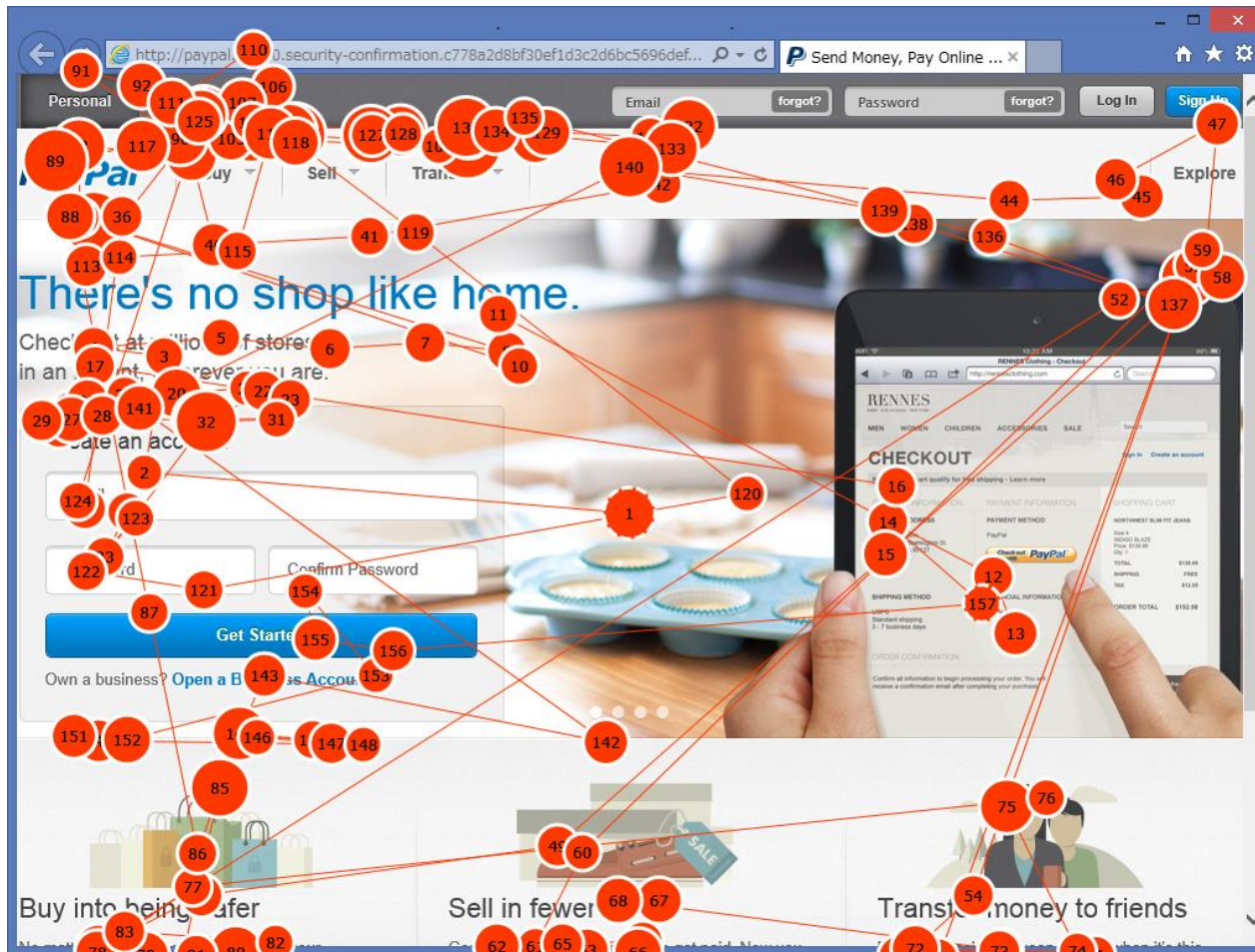


# Observations

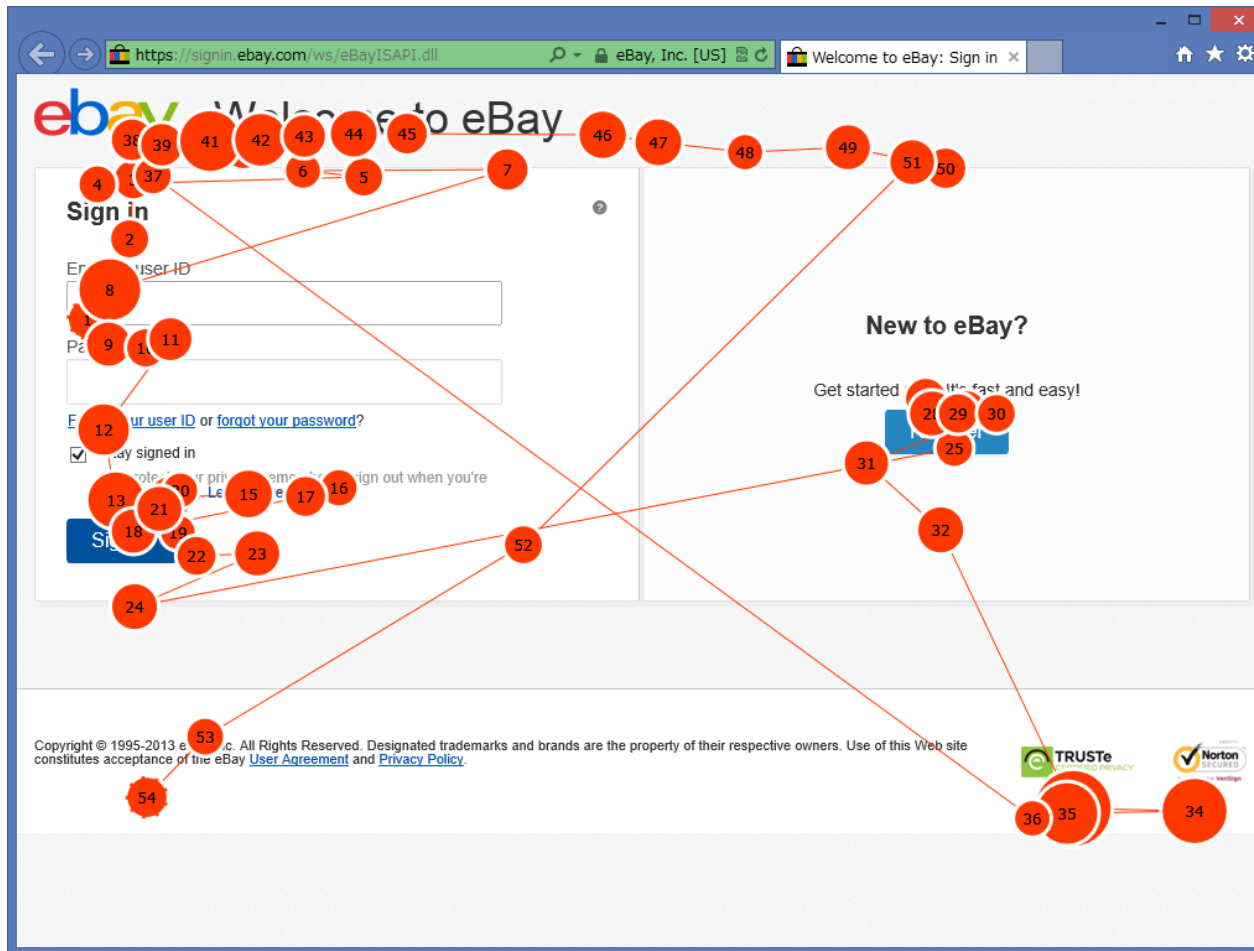
---

- Experts
  - tend to evaluate sites' URL and/or browser's SSL indicator
  - tends to ignore signals from web content
  - have useful decision patterns for Human Boost
  
- Novices
  - receive strong signals from web content
  - do not have enough knowledge about URL, and security information
  - do not aware of the address bar
  - are protected by machine learning algorithm, without taking their decision in Human Boost

# Recognition of phishing (novice)

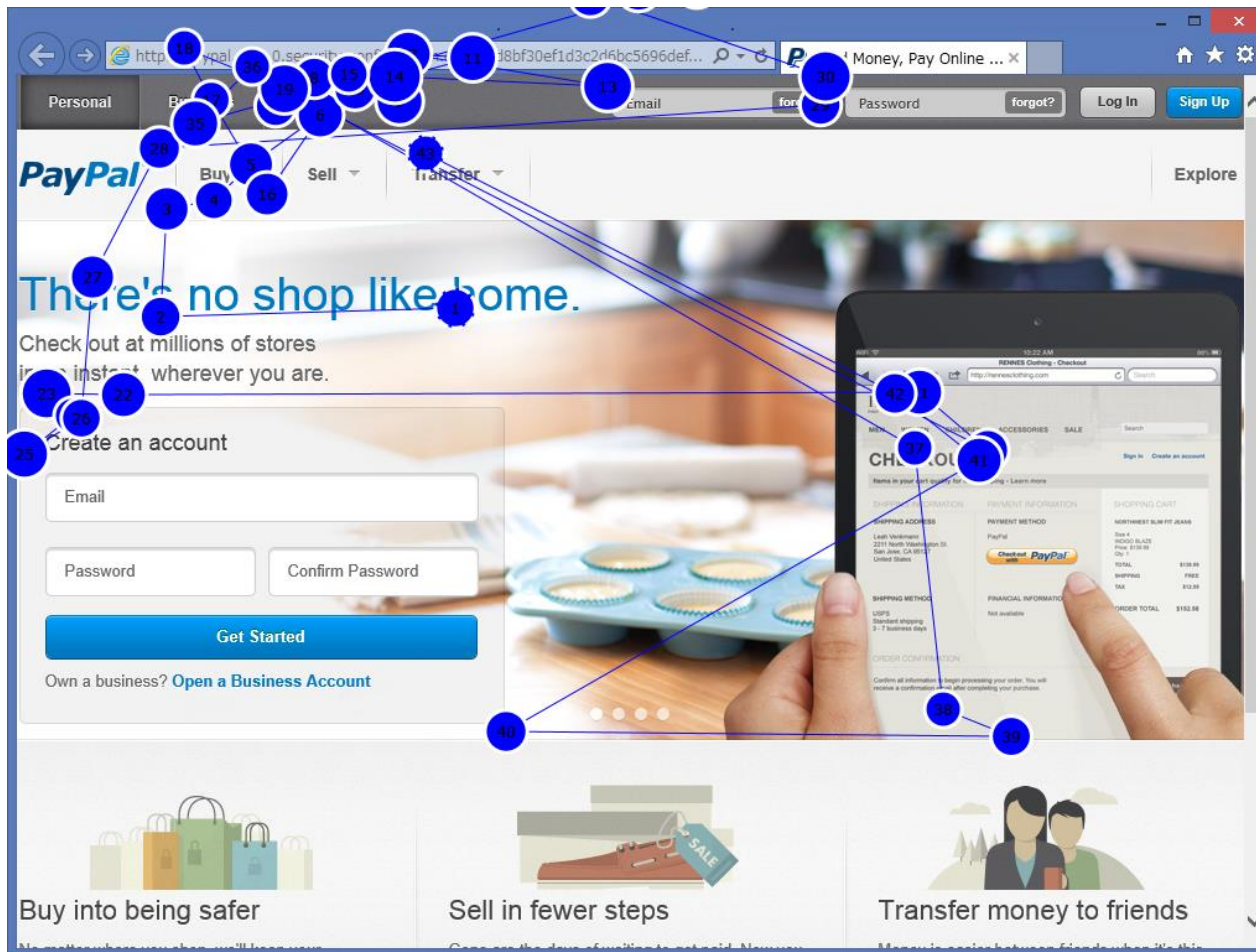


# Recognition of legitimate (novice)



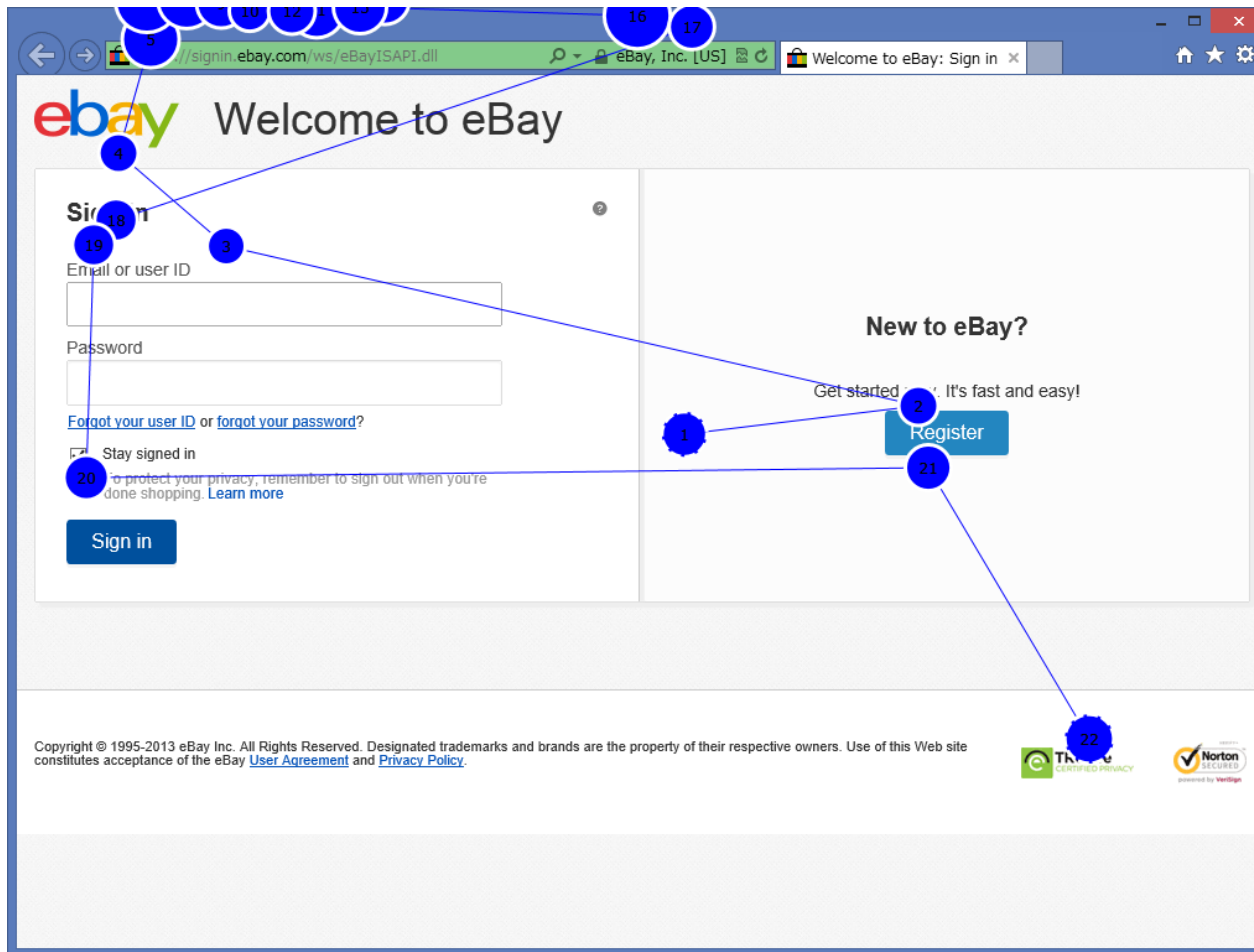


# Recognition of phishing (expert)





# Recognition of legitimate (expert)



# Development of security habits (1)

- Advantage of security habit  
“See address bar before inputting information”

- It maximizes educational effect
  - “knowledge about URL and security information” could not work before seeing the bar
- It assists for end users to be aware of security information, in collaborative manner
- It increases detection accuracy by personalization;
  - Detection system can recognize users' security level
  - Expert : human boost  
(machine learning + user decision)
  - Novice : machine learning
- It would not heavily penalize users' experience

## Approaches to counter phishing



education



attention



detection

# Development of security habits (2)

- Advantage of security habit
  - Habitual actions is often performed under unconscious.
- Reduction of human factors
  - People have limited resources, limited capacity for information processing and routinely multitasks



- People thinks security is an abstract concept
  - To save mental resources, they sometimes used risky shortcuts

Even if people have knowledge, they sometimes could not do right behavior.

(1) Find mental anomaly (2) Establish a way at conditioned reflex manner

Cognitive analysis

Development of security habits

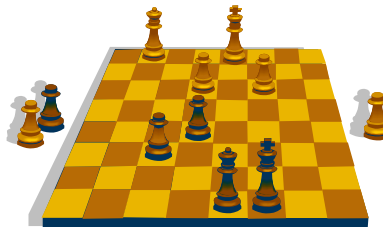
# Cognitive analysis

---

- Eye-movements
  - Strong link between eye movement and mental disorders [Crawford 2005, Noris 2007]
    - Saccades (changing with what they are seeing)
    - Fixation (maintaining of visual gaze)
  - Mental rotation is suppressed during movements [Irwin 2007]
  - A person's mental/cognitive busyness can be estimated with saccadic intrusions [Tokuda 2011]
- Facial information
  - Nose skin temperature [Ora 2010]
- Other information sources
  - Heart measures [Wilson 2003]
  - Blood pressure [Miyake 2000]
  - Brain activity (EEG) [Wilson 2001]
  - Respiratory [Veltman 1998]

# Research direction to counter phishing

- Artificial intelligence and Human intelligence
  - Metaphor from “Advanced Chess”



software for  
supporting  
decision



(you)



(attackers)

# Thank you for your attention !

---

- Collaboration of human and machine intelligence
  - How to protect the (possible) weakest link in cyber spaces ?
- Development of security habits
  - as well as education, attention and detection for supporting making trust decision
- Cognitive analysis
  - Analysis for identifying mental anomaly (tends to misbehave)
  - eye-movements, respiratory, brain activity, facial temperature ... to find an abnormal mental modes