A Privacy-Preserving Architecture for the Protection of Adolescents in Online Social Networks

Markos Charalambous, Petros Papagiannis, Antonis Papasavva, Pantelitsa Leonidou, Rafael Constaninou, Lia Terzidou, Theodoros Christophides, Pantelis Nicolaou, Orfeas Theofanis, George Kalatzantonakis, Michael Sirivianos



Markos Charalambous Cyprus University of Technology

marcos.charalambous@cut.ac.cy







- Education
 - 2014 2018: BSc in Computer Science, Frederick University
 - 2018 2019: Private Pilot License, Aviator Flight Center
 - 2019 Present: MSc in Data Science and Engineering, Cyprus University of Technology
- Experience
 - 2019 Present: Research assistant at Cyprus University of Technology







Workgroup





Michalis Sirivianos Assistant Professor

Fragkiskos Papadopoulos Assistant Professor



Costas Iordanou Post-doctoral Researcher



Nikos Salamanos Post-doctoral Researcher



Kostantinos Papadamou Phd Student



etSvSci Network Systems Science and

INCOGNITO

SECONDO





Theodoros Christophides Project Manager

Aristodemos Paphitis Phd Student

Petros Papagiannis Post-graduate Researcher



Pantelitsa Leonidou Post-graduate Researcher

Research Interests:

- trust-aware design of distributed systems
- device-centric authentication
- federated identity management
- discrimination based on personal data
- transactional workload scalability

- cybersafety (cyberbullying detection, cybergrooming detection, characterization and detection of hate speech, detection of inappropriate videos targeting young children, characterization and suppression of false information)
- measurement of blockchain systems



Online Social Media (OSN)

- (OSN) constitute an integral part of people's every day social activity
- Specifically, mainstream OSNs, such as Twitter, YouTube, and Facebook are especially prominent in adolescents' lives for communicating with other people online, expressing and entertain themselves, and finding information





Emerging problems













Markos Charalambous - marcos.charalambous@cut.ac.cy

Existing parental control tools

Currently they violate the privacy of adolescents, leading them to use other communication channels to avoid moderation



Cybersafety Family Advice Suite (CFAS)

 We design and implement a user- centric CFAS with Guardian Avatars aiming at preserving the privacy of the individuals towards their custodians and towards the advice tool itself



Architectural Overview



Guardian Avatar Approach

- The Guardian Avatar "follows" the user in their online-activities as a virtual friend.
- When the IWP detects any malicious behavior or incidents, the notifications (warnings, advice, etc.) appear as chat bubbles of the avatar, in a friendly and encouraging text

Demo: <u>link</u>



Design

Instead of simple rule-based filters, our architecture utilizes advanced machine learning algorithms



Implementation

- We employ classifiers created in previous work for the detection of threats in OSNs
- These classifiers are generated on the Back-End and hosted on the IWP
- In case the classifiers detect suspicious activity, the IWP pushes notifications to the browser add-on of the user, and the Parental Console

- A) Detection of Abusive Users on Twitter
- B) Fake and Bot user detection on Twitter
- C) Detection of Hateful and Racist memes on Facebook
- D) Sexual Predator Detection on Facebook

- E) Cyberbullying Detection on Facebook
- F) Personal Information Leakage Detection on Facebook
- G) Watermarking and Steganography
- H) Disturbing videos on YouTube

EVALUATION I

• Performance Evaluation



EVALUATION II

• User Experience (Minors)



Would you allow CFAS to send notifications to your custodian regarding suspicious detection? (1: Totally Disagree, 5: Totally Agree)



Do you believe CFAS would improve your safety when using OSNs? (1: Totally Disagree, 5: Totally Agree)



Do you believe CFAS would improve your safety when using OSNs? (1: Totally Disagree, 5: Totally Agree)

Have you ever experienced the following online-threats?
Select all that apply to you: (a) I prefer not to say; (b) None;
(c) Personal data misused; (d) Personal photo misused;
(e) Cyberbullying; (f) Inappropriate speech and racism; and (g) Sexual grooming

EVALUATION II



Has your child ever reported to you being a victim of the following? (a) I prefer not to say; (b) None; (c) Personal data misused; (d) Personal photo misused; (e) Cyberbullying; (f) Inappropriate speech and racism; and (g) Sexual grooming



Do you think that CFAS would improve the safety of minors when using OSNs? (1: Totally Disagree, 5: Totally Agree)



Would you install CFAS at home? (1: Totally Disagree, 5: Totally Agree)

CONCLUSION

- This architecture aims to protect minors when using OSNs while preserving their privacy
- We propose Guardian Avatars that interact with, warn, and advise adolescences when they face threats on OSNs
- The custodian can only see the relevant content, which indicated to be suspicious, only if the minor had previously given their explicit consent
- The proposed architecture advertises the collaboration between parents and children and aims at bringing the family to work together to protect the vulnerable groups of the Internet while using OSNs

Questions?

Markos Charalambous - marcos.charalambous@cut.ac.cy®