# Poster : Collaborative Trustworthy Security and Privacy Framework for Social Media

Hardik A. Gohel, Ph.D.
Applied Research Center
Florida International University
Miami, United States
hgohel@fiu.edu

Himanshu Upadhyay, Ph.D.
Applied Research Center
Florida International University
Miami, United States
upadhyay@fiu.edu

Leonel Lagos, Ph.D.
Applied Research Center
Florida International University
Miami, United States
lagosl@fiu.edu

***Abstract*— Social media technology provides a novel platform for faster, dynamic, manageable, cost-effective and adaptable professional network service provisioning. As such, social media is ideal for online outreach and getting traction with individuals, government organizations and business enterprises. However, as social media technology continues to provide an increasing number of functionalities to expand a social network, there is a growing need to design, develop and evaluate the security and privacy of personal and professional social media services. The proposed collaborative trustworthy security and privacy framework for social media provides a new avenue to develop improved security and privacy that can both verify and validate social media content before being made available online. It also monitors, assesses and reacts when online security and privacy is compromised.**

***Keywords*— *Social Media, Privacy, Security, NLP, RL***

## I. Introduction

Due to the growth of online information storage and sharing, social media network security and privacy is more important than ever. Social media networks have made the world a more connected place. Twitter, Facebook, Instagram, Pinterest and Google+ are the most popular current platforms that connect others on a large scale. However, all social media connections are more vulnerable when they provide unauthorized information access. The increase in cyber hacking and online scams only heighten the need for solutions. According to social media statistics for 2018[1], 400 new users sign up for Facebook every minute. One-hundred-million hours of video content are watched on Facebook daily and 22% of the world's total population use Facebook with 2.01 billion monthly active users. Additionally, 40% of marketers use Google+ as a social platform. Recently reported security and privacy breaches add to the alarm. For example, Twitter estimates that 23 million of its active users are actually bots and, in 2018, a Facebook security breach exposed accounts of 50 million users and at least 87 million records [2].

The content of social media is in the form of text, photos and videos. This data often includes sensitive information of individuals, government agencies and businesses. When anyone gains unauthorized access to the data, security and privacy is compromised. Thus, security and privacy are always a major concern in the study of social media [3, 4].

## II. System Model

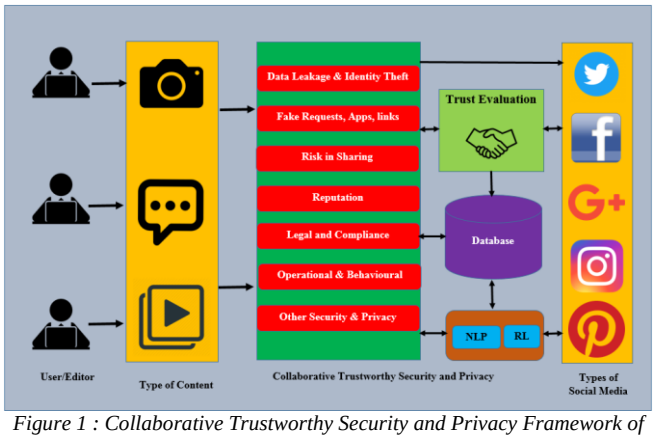



*Figure 1 : Collaborative Trustworthy Security and Privacy Framework of Social Media*

The overall collaborative trustworthy cyberspace architecture is shown in figure 1. At the core of this solution is security and privacy management, trust evaluation and a learning agent. It can verify and validate the security and privacy of social media content available online and also evaluate, improve and manage the online content. When someone from a government agency or business organization creates new content for social media, it first goes to the collaborative trustworthy cyberspace. Here, it undergoes verification and validation via security and privacy management based on the content as well as trust evaluation. After a successful trust evaluation, the content will then post on social media. The interactions of users with the content is collected and stored in the database and reinforcement learning is performed for continuous evaluation, improvement, and management of content security and privacy.

### a) Security and Privacy Management

The process of the security and privacy management task is to identify, analyze, and manage security and privacy of social media content before it is available online. This is a continuous process to be performed on each and every element of content before it goes online. If the content looks vulnerable and exploits the security or privacy of an individual, government agency or business organization based on the above discussed criteria, then it must be edited or rewritten.

### b) Trust Evaluation

Based on various behavioral science aspects, we have formulated and evaluated trust. The value of trust is between 0 and 1. If the evaluation result is greater than 0.75, the content is considered to be at the highest level of trust; less

than 0.25, is considered to be at the lowest level trust. For values between 0.25 and 0.75, various levels of security and privacy checks and evaluations is performed, as discussed in section 2.2, by other trusted users or experts who have trust evaluation results greater than 0.75.

c) Natural Language Processing and Reinforcement Learning

Recent artificial intelligence techniques have attracted much attention for solving complex problems that are otherwise difficult to address using conventional methods. In this research, natural language processing and reinforcement agent learning used for continuous evaluation, improvement and management of social media security and privacy. Online social media content will be stored in a database periodically to learn, and evaluate content using NLP & RL for continuous security and privacy. The key challenge of this research is to evaluate current online content and identify security and privacy issues and then perform steps to resolve the issues. Along these lines, simulated social media traffic is used to replicate a wide range of cyber-attack threats. In this research, data from isolated, clean social media profile accounts is also traced and captured to train the NLP and RL model. The general pseudocode of the Q Learning algorithm with security and privacy iteration of social media content, is written as algorithm 1.

*Algorithm 1: Security and Privacy iterating using Q Learning*

1. Initialization
   V (*sp*) ∈ ℝ and π (*sp*) ∈ Ą (*sp*) arbitrarily for all *sp* ∈ Ș

2. Security and Privacy Evaluation
   Repeat
   Δ ← 0
   For each *sp* ∈ Ș:
   *v* ← V (*sp*)
   V (*sp*) ← Σśṕ, r Ƥ (ś, r|s, Ⅱ (sp)) [r + ɣ V (śṕ)]
   Δ ← max (Δ, | *v* - V (*sp*)|)
   until Δ < θ

3. Security and Privacy Improvement & Management
   *Security and Privacy-Stable* ← *true*
   For each *sp* ∈ Ș:
   *a* ← Ⅱ (sp)
   Ⅱ (sp)← $\text{argmax}_a$ Σśṕ, r Ƥ (ś, r|s, *a*) [r + ɣ V (śṕ)]
   If *a* ≠ Ⅱ (sp), then *Security and Privacy-Stable* ← *false*
   If *Security and Privacy-Stable*, then stop and return V and Ⅱ; else go to 2

## III. Results & Outcomes

To date, NLP algorithm and Q-Learning as a part of reinforcement learning on textual data has been designed to evaluate and analyze the privacy and security of social media, in real time. However, live data trace experiments of developed research is needed to monitor and track time-varying content. We are still investigating other security and privacy issues of online social media content and ways to address those issues. In addition, other key reinforcement learning algorithms like SARS, DQN and DDPG are also in our list of research exploration.

Figure 2 shows sample designed result of security and privacy evaluation of the given social media content. Presently, the displayed evaluation with three results. Checked is to assure that there is no security and privacy issues with online social media content. If there are any security and privacy issues with social media content, then the result is compromised or probably compromised.

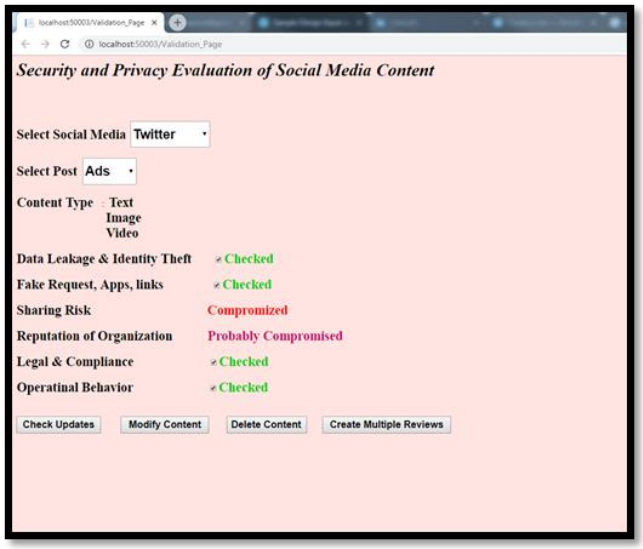



*Figure 2 : Security and Privacy Evaluation of Online Social Media Content*

## IV. Conclusion

The combination of security privacy management, trust evaluation, NLP and reinforcement learning presents a novel and largely unexplored area. This is a robust, reliable and adaptive solution to provide security and privacy for online social media content. It leverages advanced social media capabilities to address security and privacy, particularly in the most popular social media currently available in the market. It identifies the most effective security and privacy policy evaluation, improvement and management using NLP and reinforcement learning. Furthermore, this novel approach will also be integrated into Facebook, Google+, Twitter, and Instagram to evaluate their security and privacy. This will change the future of social media security and privacy management.

## References


[1] J. Bagadiya, "171 Amazing Social Media Statistics You Should Know in 2018", Social Pilot available at https://www.socialpilot.co/blog/social-media-statistics, 2018

[2] David Bisson, "The 10 Biggest Data Breaches of 2018", available at https://blog.barkly.com/biggest-data-breaches-2018-so-far, 2018

[3] J. Isaak, M.J.Hanna, "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection", Security Corner – IEEE Computer Society, pp. 56-59, 2018

[4] L. Xu, C. Jiang, N. He, Z. Han, A. Benslimane, "Trust-Based Collaborative Privacy Management in Online Social Networks", IEEE Transactions on Information Forensics and Security, Vol. 14, Issue 1, pp. 48-60, 2018

# Collaborative Trustworthy Security and Privacy Framework for Social Media

Hardik Gohel, Himanshu Upadhyay, and Leonel Lagos
Applied Research Center at Florida International University

## Introduction

Privacy

Security

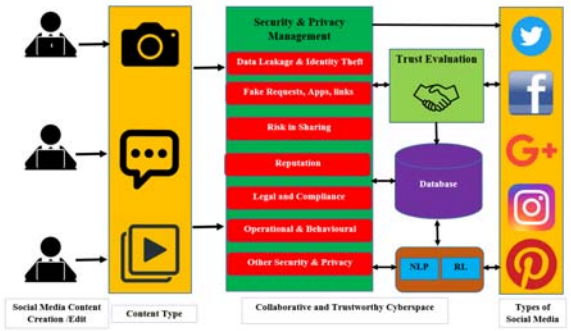
Social Media

**Social Media** platforms are more vulnerable when somebody takes unauthorized access information of individuals, government agencies and business organizations

The content of social media is in the form of text, photos, and videos which contains many sensitive information

**Security and Privacy Framework of Social Media** allows security and privacy management, trust evaluation and natural language processing and reinforcement learning for continuous evaluation of security and privacy – Real-time social media security can be achieved.

## Security and Privacy Framework




1. The different types of content created by editor goes to the collaborative and trustworthy cyberspace
2. It performs verification and validation via security and privacy management
3. After content looks good, it requires to perform trust evaluation

   Once trust is evaluated, the content will then post on social media
4. The interaction of users with the content is collected and stored in the database and natural language processing and reinforcement learning performed
5. Continuous evaluation, improvement and management of content security and privacy is performed by step 4 on content and interaction of social media

## Security and Privacy Management

**Security and Privacy Management** is to identify, analyze, and manage security and privacy of social media content

This is continuous process performed on text, audio, video, graphics, images content of professional social media

If the content looks vulnerable that exploits security and privacy of individual, government or business organization – needs to be rewritten.

The security and privacy criteria are given in figure of an architecture.

## Trust Evaluation

**Trust Evaluation** is based on the various behavior science aspects, formulated to evaluate trust

Value of trust is between 0 and 1.


If the evaluation result is great than 0.75, social media content is considered to be at the highest level of trust; less than 0.25 is lowest level of trust

For values between 0.25 and 0.75 required additional level of evaluations from experts

## Natural Language Processing & Reinforcement Learning

**Natural Language Processing (NLP)** helps to understand, interpret and manipulate social media content languages.

**Reinforcement Learning(RL)**, the concept of an agents, used for continuous evaluation, improvement and management of social media security and privacy.

The key challenge of this research is to evaluate the present online social media content using NLP and identify security and privacy issues and then perform steps to resolve.

Along these lines, simulated social media traffic is used to replicate a wide range of cyber threat attacks.

## RL Q-Learning with Security & Privacy Iteration

1. **Initialization**
   $V(sp) \in \mathbb{R}$ and $\pi(sp) \in A(sp)$ arbitrarily for all $sp \in S$
2. **Security and Privacy Evaluation**
   Repeat
   $\Delta \leftarrow 0$
   For each $sp \in S$:
   $v \leftarrow V(sp)$
   $V(sp) \leftarrow \sum_{\acute{s},\acute{r}} P(\acute{s}, r|s, \pi(sp))[r + \gamma V(\acute{s})]$
   $\Delta \leftarrow \max(\Delta, |v - V(sp)|)$
   until $\Delta < \theta$
3. **Security and Privacy Improvement & Management**
   *Security and Privacy-Stable* $\leftarrow$ *true*
   For each $sp \in S$:
   $Sa \leftarrow \Pi(sp)$
   $\Pi(sp) \leftarrow \text{argmax}_a \sum_{\acute{s},\acute{r}} P(\acute{s}, r|s, a)[r + \gamma V(\acute{s})]$
   If $a \neq \Pi(sp)$, then *Security and Privacy-Stable* $\leftarrow$ *false*
   If *Security and Privacy-Stable*, then stop and return $V$ and $\Pi$; else go to 2

## Implementation

CNTK — Microsoft cognitive toolkit for backend support and NLP as well as RL experiment and research

C# ASP.net — Asp.net with C# used to visualize security and privacy testing results and modification in the content of social media

python — All artificial intelligence algorithms and scripts are developed using the Python scripting and deployed to Microsoft SQL Server.

SQL Server — SQL Server is used to store social media content and run NLP and RL algorithm scripts as a stored procedure

## Summary

- Gives results security and privacy based on given criteria
- Check, validate, modify, delete content and also assign multiple viewers for reviews
- Automate security and privacy checking process using NLP & RL


hgohel@fiu.edu

Research Opportunities: https://arc.fiu.edu