

The Insecurity of Sharing MY Life on the Web

The Butterfly Effect of Images and Videos in Social Networks

Raphaël Troncy
EURECOM
Sophia Antipolis, France
raphael.troncy@eurecom.fr

Jean-Luc Dugelay
EURECOM
Sophia Antipolis, France
jean-luc.dugelay@eurecom.fr

ABSTRACT

People are more and more interconnected online and active in publishing and sharing their thoughts, feelings, activity, and recorded experiences in media items within their social networks. This new massive amount of data is however usually locked into proprietary platforms with evolving privacy policies. We are interested in indexing and analyzing photos and videos shared on the Web and in studying their diffusion as well as the correlations we can deduce based on subject matters depicted. We are advocating the use of semantic web and forensic imaging technologies for providing new tools to protect users privacy. We use linked data technologies to represent and expose all metadata gathered during these analysis processes. We are actively working in the W3C Media Fragments Working Group for standardizing how to express media fragments in a URI. Finally, we are developing environments that enable contextualized exploration of multimedia content within social networks revealing unexpected connections between objects and people.

Keywords

Image Forensic, Linked Data, Social Networks

1. INTRODUCTION

The Web is traditionally seen as a highly heterogeneous collection of information including structured documents, databases and audiovisual media items. Social network sites have developed new ways of sharing personal content and of actually determining how you trust friends and make new friends producing new social mechanisms. Hence, the Web now hosts activity streams and online conversations of millions of users. These social network sites have emphasized one of the original characteristics of the early Web: despite its technical infrastructure, the Web links people mediated by computers.

People are more and more interconnected online and active in publishing and sharing their thoughts, feelings, activity, and recorded experiences in media items within their social networks. This new, massive amount of data is however usually locked into proprietary platforms with evolving privacy policies. Not only is this data not owned by users anymore, but it is more and more processed by intelligent algorithms that can recommend users new products to buy,

new events to attend and new connections to make. For the good or the bad, the life of people is surfacing more and more to the rest of the world without necessarily the consent and awareness of the user and with little means to regain the control of the data.

2. IMAGE FORENSIC AND SOCIAL MEDIA

We are interested in indexing and analyzing photos and videos shared on the Web and in studying their diffusion as well as the correlations we can deduce based on subject matters depicted. Photo and video sharing is one of the most requested features by users of social networks. The development of rich web applications powered by the evolution of web technologies such as HTML5 will emphasize even more this trend. For example, the number of photos uploaded to Facebook per year is 30 billion at the current rate¹ while YouTube is serving more than 1 billion videos per day². The immediate consequence is a potential malicious use of this media content that can be stolen, transformed and re-purposed. These threats include:

- **Impossibility to assert the right to forget.** Some of *your* pictures can be copied and republished without *your* knowledge leading to a situation where an unknown number of copies of *your* photos live elsewhere on the Web. At worst, this prevents the permanent deletion of photos upon user request as per the new digital laws which stipulate the *right to forget*.
- **Uncontrolled modifications.** Photos can be maliciously transformed in order to derive fake images. Though technologies can be used to trace the image manipulations, people need to be educated that in the digital world, seeing is not believing [1].
- **Unexpected connections between people revealed.** The combination of multimedia and statistical analysis and information retrieval techniques allow to establish *a priori* non explicit connections between people such as common events attended or geo-location coincidence at specific times. We are working on a distance similar to the *Six Degrees of Kevin Bacon*³ applied to the faces depicted on photos shared on Facebook.

¹<http://www.facebook.com/press/info.php?statistics>

²<http://youtube-global.blogspot.com/2009/10/y000000000outube.html>

³http://en.wikipedia.org/wiki/Bacon_number

3. LINKED DATA TECHNOLOGIES

We are advocating the use of semantic web [5] and forensic imaging [2] technologies for providing new tools to protect user privacy. Watermarking technologies can be used to check image integrity [4]. Forensic imaging technologies can help to detect if images have been modified to contain artificial objects or pieces from other images, if they have been tampered or if they contain hidden data, if videos have been recorded without the consent of people depicted (stolen video), etc. We use linked data technologies to represent and expose all metadata gathered during these analysis processes. Multimedia content needs to be addressed and linked at the fragment granularity similarly to HTML documents [6]. We are actively working in the W3C Media Fragments Working Group⁴ for standardizing how to express media fragments in a URL. Finally, we are developing environments that enable contextualized exploration of multimedia content within social networks revealing unexpected connections between objects and people [3]. We are confident that the use of these technologies will contribute to both unlocking current closed platforms and to improving the security of user activity on the Web.

4. REFERENCES

- [1] H. Farid. Seeing Is Not Believing. *IEEE Spectrum*, 46(8):44–48, 2009.
- [2] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme. Can we trust digital image forensics? In *15th ACM International Conference on Multimedia*, pages 78–86, Augsburg, Germany, 2007.
- [3] L. Hardman, J. van Ossenbruggen, R. Troncy, A. Amin, and M. Hildebrand. Interactive Information Access on the Web of Data. In *Web Science Conference - Society On-Line (WebSci'09)*, Athens, Greece, 2009.
- [4] C. Rey and J.-L. Dugelay. A survey of watermarking algorithms for image authentication. *EURASIP Journal on applied Signal Processing*, 2002(1):613–621, 2002.
- [5] N. Shadbolt, T. Berners-Lee, and W. Hall. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.
- [6] R. Troncy, L. Hardman, J. v. Ossenbruggen, and M. Hausenblas. Identifying Spatial and Temporal Media Fragments on the Web. In *W3C Video on the Web Workshop*, 2007.

⁴<http://www.w3.org/2008/WebVideo/Fragments/>