Multimedia Privacy

[ACM MM 2016 Tutorial]

Gerald Friedland
Department of EECS
University of California,
Berkeley, CA, USA
fractor@eecs.berkeley.edu

Symeon Papadopoulos
Information Technologies
Institute
CERTH, Thessaloniki, Greece
papadop@iti.gr

Julia Bernd
International Computer
Science Institute
Berkeley, CA, USA
jbernd@icsi.berkeley.edu

Yiannis Kompatsiaris
Information Technologies
Institute
CERTH, Thessaloniki, Greece
ikom@iti.gr

ABSTRACT

This tutorial brings together a number of recent advances at the nexus of multimedia analysis, online privacy, and social media mining. Our goal is to offer a multidisciplinary view of the emerging field of Multimedia Privacy: the study of privacy issues arising in the context of multimedia sharing in online platforms, and the pursuit of new approaches to mitigating those issues within multimedia computer science.

CCS Concepts

•Networks → Network privacy and anonymity; •Social and professional topics → Privacy policies;

Keywords

Multimedia; Privacy; Social Media

1. MOTIVATION

The growth of online multimedia, especially due to social networking sites such as Facebook and YouTube, combines with advances in multimedia retrieval (geo-tagging, web search, face recognition, speaker verification, location estimation, etc.) to provide novel opportunities for the malicious and unethical use of multimedia. Multimedia analytics have always had the potential to be a powerful privacy threat, but in isolation or at small scale, that threat has been reasonably contained. However, when analysis systems are linked together and used on an Internet scale, the threat can be enormous and pervasive. Many in the multimedia-analysis community therefore believe that we have an obligation to understand and attempt to mitigate these risks.

Engineering and computer science curricula usually include an abundance of material on multimedia retrieval methods that rely on analysis of the underlying content, but only rarely do they talk about the negative impacts of these technologies. Privacy research per se often focuses more on securing vs. co-opting communication channels—encryption, steganography, and the like—than on content privacy. Though there is a growing community of concern around content privacy, non-specialists may view it as "out of scope" for both education/training and for research and development. Because of this gap in education and organizational culture, many multimedia researchers lack knowledge about how to react to concerns and mitigate potential privacy risks.

In the drive to make the next big advance, it can be easy to say, "We'll deal with privacy and ethics later—right now we need to focus on making it work." But the truth is that it can be quite difficult to "add on" privacy when a system or technique has already developed momentum—even when the researchers and developers have the time and will to do it. For example, if privacy and security had been more of a concern in the earliest stages of developing the Internet, many current issues, such as spam, phishing email, and man-in-the-middle attacks, would probably be much less of a problem.

At the same time, some solutions to security and privacy concerns are quite simple, following a limited set of basic principles that are fairly well-known in the privacy and security communities—but less so in multimedia. If these guidelines are followed in the early stages, developers and researchers can avoid creating larger problems down the line.

2. OBJECTIVES

The Multimedia Privacy tutorial will introduce interested multimedia researchers, engineers, and students to a multidisciplinary perspective on privacy. The tutorial will provide a vivid overview, with many examples based on material developed for the CS10 "Beauty and Joy of Computing" course at UC Berkeley, for the NSF-sponsored Teaching Privacy project (http://teachingprivacy.org), and for the EC-funded USEMP project (User Empowerment for enhanced online

MM '16 October 15-19, 2016, Amsterdam, Netherlands

ACM ISBN 978-1-4503-3603-1/16/10.

DOI: http://dx.doi.org/10.1145/2964284.2986915

management, http://www.usemp-project.eu/). These materials are based in turn on current research from top-notch privacy, multimedia, and security groups and conferences.

Using real-world examples and their consequences, the tutorial will focus on threats related to multimedia retrieval in the context of modern social networking structures. While it will cover a number of technical parameters that should be considered, the tutorial's main thread of argument is that user awareness and empowerment are two crucial requirements for improving privacy management through the design of multimedia sharing and social networking applications.

3. OUTLINE

Who's concerned about multimedia privacy and why:

- Privacy perception as a function of age, technical awareness, and cultural background
- Social vs. institutional privacy
- Legal frameworks and ethical standpoints in the U.S. and the EU
- Crime potentially enabled by multimedia: cyberstalking, cybercasing, identity theft, etc.
- Non-criminal consequences of privacy leaks

The state of the art in multimedia and privacy:

- Geotagging and its implications for privacy
- Multimedia retrieval and its implications for privacy (location estimation, activity detection, etc.)
- Matching users across accounts with multimedia content analysis
- Facebook & LinkedIn vs. physical-world social networking

Mitigation strategies and tools:

- Attack and defense model and taxonomy
- Personalizing privacy via settings and via inference
- Scoring and visualizing personal information disclosure
- Potential techniques for built-in identity obfuscation
- User education: ten principles for social media privacy

Conclusions:

- Possible implications for research, education, industry
- Opportunities for start-ups, new research
- First steps everybody can take right now

4. TUTORS

Dr. Gerald Friedland is a Principal Scientist responsible for Audio and Multimedia Analysis at Lawrence Livermore National Lab and is also teaching as an adjunct professor at the Electrical Engineering and Computer Sciences department of the University of California, Berkeley. He has published more than 200 peer-reviewed articles in conferences, journals, and books. He led the development of the teachingprivacy.org portal and also co-authored a new textbook on multimedia computing together with Dr. Ramesh Jain. He is associate editor for *ACM Transactions on Multimedia* and *IEEE Multimedia Magazine* and regularly reviews for *IEEE Transactions on Acoustics, Speech, and Language Processing*; Springer's *Machine Vision and Application*; and other journals. Among other recognitions, he won the ACM Multimedia Grand Challenge in 2009 and 2016. Dr. Friedland received his doctorate (summa cum laude) and master's degree in computer science from Freie Universitaet Berlin, Germany, in 2002 and 2006, respectively.

Dr. Symeon Papadopoulos received his Diploma degree in Electrical and Computer Engineering from the Aristotle University of Thessaloniki in 2004. In 2006, he received the Professional Doctorate in Engineering (P.D.Eng.) from the Technical University of Eindhoven, the Netherlands. Since 2006, he has been working with the Information Technologies Institute of CERTH. In 2012, he defended his PhD dissertation on the topic of knowledge discovery from large-scale mining of social media content. He has participated in several EU-funded research projects (WeKnowIt, SocialSensor, REVEAL, USEMP) as technical leader, and has co-authored more than 70 publications in refereed journals and conferences. Dr. Papadopoulos successfully led the organization of the Workshop on Social Multimedia and Storytelling in ICMR 2014 and the Workshop on Web Multimedia Verification in ICME 2015, and co-organized the Social News on the Web (SNOW) workshop in the context of WWW 2014 and 2016. In addition, he was one of the organizers of the 10th European Summer School on Information Retrieval (ESSIR 2015).

Julia Bernd is a staff researcher at the International Computer Science Institute in Berkeley, CA. She coordinates the Teaching Privacy project, which aims to educate people about how online sharing works, what the risks are, and what individuals can and can't do to protect themselves. Teaching Privacy, begun by Friedland and others, grew out of research on multimedia privacy and often focuses on multimedia examples (such as geotagging—and content-based location estimation for untagged images and videos). In other words, the goal is to leverage the expertise of multimedia computer scientists to educate the general public, in particular young people, about potential issues arising out of advances in the field. Bernd is a linguist by training (MA, Stanford, 2002), but her recent academic work focuses largely on multimedia corpus development (in the context of the Multimedia Commons) and on developing CS curricula that integrate potential impacts of advances in computing.

Dr. Yiannis Kompatsiaris is a Senior Researcher (Researcher A') with the Information Technologies Institute of CERTH, Thessaloniki, Greece. His research interests include semantic multimedia analysis, indexing and retrieval, social media and big data analysis, knowledge structures, reasoning and personalization for multimedia applications, eHealth, security, and environmental applications. He received his PhD in 3-D model-based image sequence coding from the Aristotle University of Thessaloniki in 2001. He is the co-author of 90 papers in refereed journals, 38 book chapters, 8 patents, and more than 320 papers in international conferences. He has been the co-organizer of numerous international conferences and workshops and has served as a regular reviewer for a number of journals and conferences. He is a Senior Member of IEEE and of ACM.

5. ACKNOWLEDGMENTS

S. Papadopoulos and Y. Kompatsiaris received support from the European Commission-funded USEMP project, contract #611596. J. Bernd and G. Friedland were supported by US National Science Foundation awards #1637601 and #1514509.