INTERVIEW WITH PROF. MINORU KURIBAYASHI, OKAYAMA UNIVERSITY, JAPAN

Minoru Kuribayashi received B.E., M.E., and D.E degrees from Kobe University, Japan, in 1999, 2001, and 2004. He was a Research Associate and an Assistant Professor at Kobe University from 2002 to 2007 and from 2007 to 2015, respectively. Since 2015, he has been an Associate Professor in the Graduate School of Natural Science and Technology, Okayama University. His research interests include multimedia security, digital watermarking, cryptography, and coding theory. He serves as an associate editor of JISA and IEICE. He is a vice chair of APSIPA TC of Multimedia Security and Forensics, and a TC member of IEEE SPS Information Forensics and Security. He received the Young Professionals Award from IEEE Kansai Section in 2014, and the Best Paper Award from IWDW 2015 and 2019. He is a senior member of IEEE and IEICE.

Could you briefly introduce your research?

My primacy research interest is the protection of multimedia content from illegal copying and ownership infringement. One promising solution is the use of data hiding techniques that insert tiny signals into multimedia content without degrading its perceptual quality. This is a kind of active approach to check the ownership and soundness of the content by slightly distorting it. On the other hand, analyzing distortions caused by editing or modifying content is the passive approach, which is called multimedia forensics. In both cases, the handling of tiny signals involved in multimedia content is important, and a combination of signal processing and machine learning techniques is inevitable in this research

What is your main work?

My main contribution in this research area is to achieve traceability of multimedia content, which is called digital fingerprinting. Uniquely assigned information, called fingerprint, are embedded into multimedia content with the help of data hiding techniques. Fingerprinting requires consideration of two difficult requirements: asymmetric property in buyer-seller protocols and collusion resistance.

Asymmetry refers to the information gap between buyers and sellers. Typically, it is assumed that the buyer will violate the ownership rights of the content by redistributing the illegal copies. However, seller can frame innocent buyers by distributing the fingerprinted content and claiming that the illegal copies were leaked from the buyer. To address this issue, cryptographic protocols between the buyer and seller have been investigated to assure that only the buyer can obtain the content containing his/her fingerprint information.

In a fingerprinting setup, different versions of the same content are distributed to multiple users, and hence, a coalition of illegal users can compare uniquely fingerprinted content and modify or delete the fingerprint information, which is called a collusion attack. Therefore, resistance against collusion attack has been investigated both in terms of encoding fingerprint (approach of coding theory) and modulating signals (approach of communication theory).

What is a challenging topic in multimedia security?

Due to the advance of deep learning (DL) technology, creation and manipulation of multimedia content have progressed to the point where they can now ensure a high degree of realism. In movies, realistic characters and exciting scenes can be created according to the interest of movie director without having them perform dangerous actions. Artificially created newscasters can continue to work on news multicasts in smooth tones. On the other hand, the DL-based signal processing operations of image, video, and audio may be abused to generate fake news like misinformation that mimics famous people. By using DL technology with multiple videos of people as supervised data, it is possible to create fake content that realistically reproduces false statements. One famous fake content is DeepFake, which is hard to distinguish from real or fake. DeepFake is basically reproduced media obtained by injecting or replacing some information into the target content. For the classification of DeepFake, unnatural signals involved in multimedia content are analyzed by using various signal processing operations and the DL techniques, which is called multimedia forensics.

What are the difficult problems in the multimedia forensics?

With the progress of defense techniques, attackers will develop content generator models according to the weaknesses exploited by the defense techniques. The use of generative adversarial network enables the generator to update and improve the performance without any theoretical and mathematical formulation if a reasonable amount of computing resources is available.

DL technology helps us to analyze the traces (tiny signals) in multimedia content for classifying whether it is fake or not. However, the reliability of DL-based system will be dropped due to the vulnerabilities against adversarial attacks such that intentional perturbations are crafted to fool the system by misleading the results. When considering defense techniques, it is necessary to assume defense-specific adversarial attacks. From the attacker’s perspective, such defense techniques can be considered to craft adversarial perturbations.

Could you explain about your current research projects?

- **EIG CONCERT-Japan**

The European Interest Group (EIG) CONCERT-Japan is an international joint initiative to support and enhance science, technology and innovation (STI) cooperation between European countries and Japan. **Detection of fake newS on SocIal MedIa pLAtfoRms (DISSIMILAR)** is a project within **CONCERT-Japan** (Connecting and Coordinating European Research and Techology Development with Japan) programme and 7th Joint Call on **"ICT for Resilient, Safe and Secure Society"** which is realized by the consortium consisting of researchers from: **Okayama University** (Japan), **Fundació per a la Universitat Oberta de Catalunya** (Spain), and **Warsaw University of Technology** (Poland). The project will be realized between **June 2021 and May 2024**. The funding is provided through grants number **PCI2020-120689-2** (Spanish Government), **JPMJSC20C3** (Japanese Government), and **EIG CONCERT-JAPAN/05/2021** (National Centre for Research and Development, Poland).

DISSIMILAR

http://dissimilar.ii.pw.edu.pl/

DISSIMILAR combines research on watermarking and machine learning with a user experience study to develop novel technological tools to help users to distinguish between original and altered media content. With the proposed tools we expect online social media users to be able to identify the authorship of a content to distinguish legitimate from fake multimedia content in an autonomous manner, without the need for the platform manager to control or validate any content. Furthermore, the watermarking tools will also provide content creators with a way to protect their creations against manipulation. Hence, the aim of this project is to provide user centric tools that combat disinformation and that contribute minimizing the redistribution of fake news in online social media.

What do you expect about the changes in the consumer technology?

In the early days of the Internet, anyone was free to enter the network market and offer services such as e-mail, chat, social network service, video streaming. With the proliferation of networks, it becomes necessary to protect against malicious activities such as eavesdropping, malicious software (malware), denial-of-service attack, phishing attack, and spread of fake news.

As for the e-mail, when setting up a new mail server, formal registration with authentication servers on a public network is inevitable to legitimize the services offered by the server.

In the near future, the distribution of multimedia content will be controlled over the network to prevent the spread fake content. Content Authenticity Initiative, which a community of media and tech companies, NGOs, academics, and others working to promote adoption of an open industry standard for content authenticity and provenance, is one of the new trends of services related to multimedia. One of potential framework is the Content Credentials functionality which allows the history of content capture, editing, and publication to be verified, including the provenance and attribution. By making the history of multimedia content transparent, it can ensure the reliability of information distribution on cyberspace.