Process Mining and User Privacy in D2D and IoT Networks

Muhammad Usman · Marwa Qaraqe
Muhammad Rizwan Asghar
Imran Shafique Ansari

As the communication industry is moving towards the 5th generation (5G) of cellular networks and beyond, the traffic it carries is also becoming a mixture of higher and lower data rate rate traffic originating from cellular users and Internet-of-Things (IoT) networks, respectively. The main industries where IoT has found its applications include, but are not limited to, automotive, manufacturing, supply chain, agriculture, healthcare, and energy. All the aforementioned industries have their specific quality of service (QoS) requirements in terms of bandwidth, latency, and storage regarding the transmission of the data they generate. For instance, vehicle-to-everything (V2X) communication in the automotive industry requires ultra-low latency without any need for higher bandwidth; however, the IoT networks employed in agriculture generally require lower latency and lower bandwidth communication.

In order to accommodate such a diverse requirement in 5G networks, many solutions have been proposed to tackle the problem from different angles. For instance, network functions virtualization (NFV) and software-defined networking (SDN) have been proposed as possible solutions, designed to decouple network services from the hardware they are executed upon. These technologies are also termed virtualization of network resources, which are meant to accommodate diverse QoS requirements onto a single physical network. Although these solutions are promising, there is a long way to go towards end-to-end virtualization of cellular networks.

In the literature, some researchers [1, 2] propose employing device-to-device (D2D) communication in the cellular access network to accommodate many QoS requirements of IoTs. For the uplink, D2D nodes can act as communication hubs to collect slower data from an IoT network and transfer it to the Cloud over cellular communication for storage and processing. For instance, a D2D node can collect data from various appliances in a smart home and transfer it to a remote cloud server via cellular or WiFi network. Similarly, for the downlink, D2D nodes can act as a caching device for many IoT applications to reduce latency [2]. In addition, D2D communication can also be used as an enabler of many proximity-based applications, such as peer-to-peer communication, proximity-based social networking, and proximity-based advertisement broadcasting.

All the aforementioned applications generate huge amounts of data, which a cellular network

https://doi.org/10.1007/s00287-019-01212-y
© The Author(s) 2019.

Muhammad Usman · Marwa Qaraqe
Information and Computing Technology,
College of Science and Engineering,
Hamad Bin Khalifa University (HBKU),
Education City, 34110 Doha, Qatar
E-Mail: {musman, mqaraqe}@hbku.edu.qa

Muhammad Rizwan Asghar
School of Computer Science, The University of Auckland,
1142 Auckland, New Zealand
E-Mail: r.asghar@auckland.ac.nz

Imran Shafique Ansari
School of Engineering, University of Glasgow,
G12 8QQ, UK
E-Mail: imran.ansari@glasgow.ac.uk

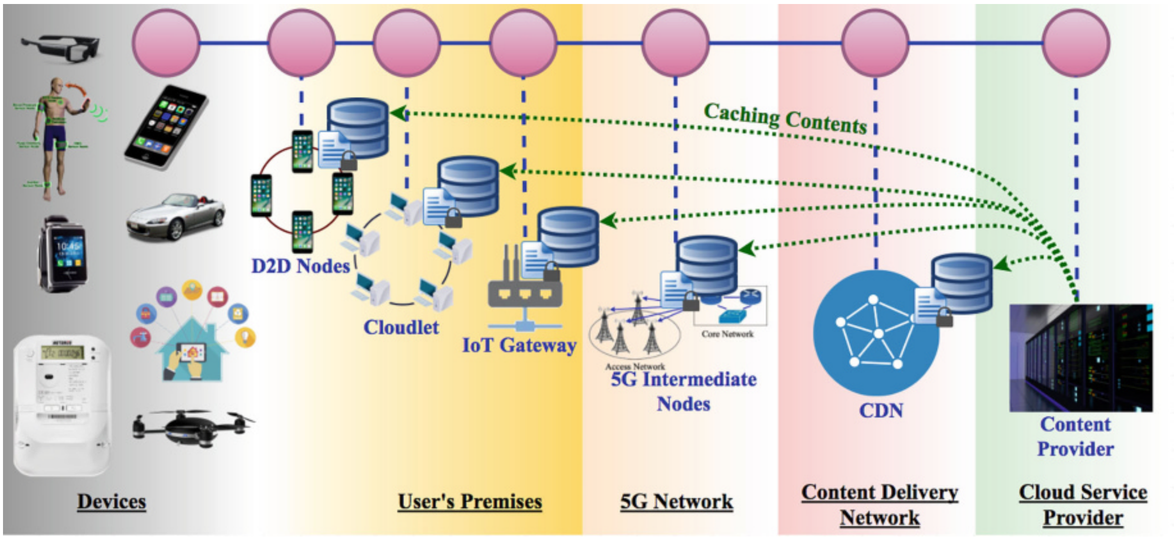

***Fig. 1** Different caching options for a content provider: The contents can be cached at any feasible option between the content provider and the end user depending on the QoS requirement of the user and availability of resources at different caching servers [2]*

and/or D2D node may utilize to prioritize traffic to ensure QoS for different IoT applications. This data may contain some information that may be personal to someone, who may not want to reveal it. For instance, an advertising company may log all the events while a customer visits a particular shop. It also logs the time and date when the customer was near the shop. It can even log the duration a customer stayed near the shop. If the shop is on the way to the office, the events are logged on a daily basis. Carrying out process mining on such events could expose sensitive information that may reveal the personal habits of a person (e. g., going to the office late orbeing absent on a particular day).

Additionally, as D2D communication extends the caching option right at the user's proximity, the cellular network needs to know exactly what a user is looking for at a particular time in order to cache the content in D2D networks. This becomes true for all the caching options presented in Fig. 1. For instance, depending on the requirements, a content provider may place the content across a CDN (content delivery network), a 5G intermediate node, an IoT gateway, or cloudlet or D2D nodes. However, placing the content within cellular network premises (5G intermediate node to D2D nodes) will allow the cellular network to log all the events, i. e., data accesses. Moreover, this type of content caching enables numerous applications that try to save bandwidth across 5G cellular networks. For instance, if two users are accessing the same contents over the Internet, and they are in the immediate proximity, the cellular network will only provide contents to the single user with the better channel conditions (say user 1), and the other user (say user 2) will be served by the first one. It is worth mentioning that user 2 will not know that she is accessing the content from user 1, and, similarly, user 1 will be unaware of her delivery of content to user 2.

It is worth mentioning that a lot of personal information can be exposed to process miners in all these scenarios. Process mining is a technique that takes as an input the event logs and records of the sequence of steps and discovers a process of the model to expose personal information. However, in the light of the above discussion regarding data generation and processing in IoT-based D2D networks, we believe that process mining can be employed to reveal personal information of customers using the services of a particular organization, such as a cellular network or an IoT services provider.

In previous work [2], we proposed a convergent encryption-based solution to ensure security in caching environments. However, we stress that similar approaches can be employed to ensure anonymity in IoT networks that use D2D either as a caching server or as data distribution hub.

Acknowledgements

Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Funding. Open access funding provided by the Qatar National Library.

References

1. Usman M, Asghar MR, Ansari IS, Granelli F (2017) Towards bootstrapping trust in D2D using PGP and reputation mechanism. In: 2017 IEEE International Conference on Communications (ICC). IEEE, pp 1–6
2. Usman M, Asghar MR, Ansari IS, Granelli F, Abbasi QH, Qaraqe K (2018) A marketplace for efficient and secure caching for IoT applications in 5G networks. In: 2018 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, pp 1–6