Towards a Blockchain-enabled Crowdsourcing platform

Dimitrios G. Kogias
University of West Attica

Helen C. Leligou
University of West Attica

Michael Xevgenis
University of West Attica

Maria Polychronaki
University of West Attica

Evangelos Katsadouros
University of West Attica

George Loukas
University of Greenwich

Ryan Heartfield
University of Greenwich

Charalampos Z. Patrikakis
University of West Attica

Crowdsourcing has been pursued as a way to leverage the power of the crowd for many different purposes in diverse sectors from collecting information, aggregating funds and gathering employees to perform tasks of different sizes among other targets. Data Integrity and non-repudiation is of utmost importance in these systems and is currently not guaranteed. Blockchain technology has been proven to improve on these aspects. In this article, we investigate the benefits that the adoption of blockchain technology can bring in crowdsourcing systems. To this end, we provide examples of real-life crowdsourcing use cases and explore the benefits of using blockchain, mainly as a database.

Crowdsourcing has gained a lot of attention, since it was first defined in 2006 by Howe [1], taking advantage of the developments in communications, the Social Media and the Internet. Originally considered as the combination of the words *crowd* and *outsourcing*, it was perceived as a process where a person or company managed to extend their expertise by using services, knowledge or goods provided by a group of Internet users. To enable the announcement of the requests for tasks by possible employers and their matching with expertise offered by possible targeted employees (claiming a certain fee), various Internet-based platforms have been created. Those platforms were, mainly, centralised systems where both the employer and the employee connect to a server to access the provided service. Later, decentralised approaches were developed, as an effort to address the drawbacks of the existing platforms (see Figure 1). The main problem was that attacks on the server could destabilize the system, allowing the attackers to gain access not only to the personal data of the registered users, but also to the work that has been delivered as part of the agreements that were reached through the platform.

In this paper, we are studying the improvements that the adoption of the Blockchain technology and concepts can bring to the crowdsourcing solutions from the security perspective. Blockchain uses a digital Distributed Ledger Technology (DLT), where no central authority is present, which is shared amongst all the nodes of the network (i.e., Peer-to-Peer (P2P) networks are supported). In this ledger, transactions between peers can be found that either follow certain agreed rules (i.e., smart contracts) or have been approved as legit by the majority of the network nodes (i.e., consesus between peers). When a transaction is inserted it cannot be deleted or erased.

The rest of the paper is organised as follows: a description of the centralised and decentralised crowdsourcing platforms is presented in the next section, followed by a description of known attacks that have been attempted against them. The benefits that blockchain technology brings in the crowdsourcing systems is analyzed and three specific use cases and the brought advantages are presented. Finally, conclusions summarize this work.

CROWDSOURCING SYSTEMS' OPERATION

The traditional crowdsourcing model follows a centralised structure, with "centralised" mostly referring to the task handling process which is an imperative part of the overall crowdsourcing system (and not to its communication structure). Centralised crowdsourcing models inherit the advantage of simplicity in both developing an application and managing the system. Many crowdsourcing platforms have been developed that adopt a centralised architecture the most interesting of them being: Upwork [2], Amazon's Mechanical Turk [3] and Waze Carpool [4], that will be briefly presented here.

Upwork is a platform where numerous projects are posted by people who seek to hire freelancers in order to complete the project's requirements. When an employer hires the freelancer, the platform provides the necessary tools for their communication and the payment. A similar centralised approach is also adopted by Amazon Mechanical Turk (MTurk), while the Waze Carpool is a crowdsourcing navigation platform with a quite different purpose. It receives the data regarding the traffic and the road conditions as those have been provided by the drivers. This application aims to inform about the traffic by establishing a real time network among drivers. It is a community-oriented solution that is based on the logic that nothing can beat real people working together.

However, any centralised solution comes with the Single Point of Failure (SPOF) vulnerability, with Single Point being the central authority or operator of the platform in the crowdsourcing paradigm, since attacking them the service becomes unavailable. In the effort to continuously evolve and improve crowdsourcing solutions, there have been various attempts to create decentralised crowdsourcing systems, mainly to address the previously described SPOF problem. In Figure 1, the main differences between a centralised and a decentralised system are illustrated. In fact, task offloading, along with handling, are the most important procedures in a decentralised crowdsourcing system, due to the randomness that may characterize the mobility of users in a network and which, if not addressed properly, will bring great delays in the successful completion of the tasks. Some solutions to address these issues are presented in [5-6]. In each of these papers, researchers and specialists propose specific algorithms and protocols aiming for faster and fairly task dissemination between nodes. These algorithms are designed so that the nodes can share the workload of any given task, considering the availability of each node. For instance, it has been proven that social relationship data can be exploited by crowdsourcing systems for mobile users and increase efficiency in terms of task completion time [5].

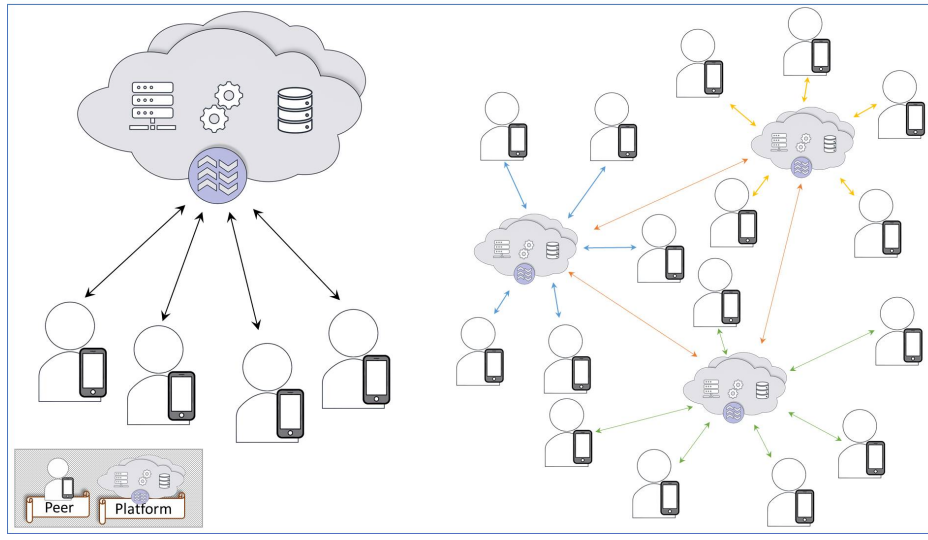

Figure 1: Centralised vs Decentralised Crowdsourcing Systems

However, almost all these models that have been proposed for implementing a decentralised crowdsourcing system rely on one or more central services that all peers have access to (e.g., social network platform). Those peers can retrieve data provided by the central services to optimize their algorithms and improve the system's efficiency and performance. To this end, the described decentralised crowdsourcing applications vulnerabilities range from losing a subset of the functionality (e.g. a subset of tasks and relevant responses) to losing the service that coordinates the different nodes among them.

CROWDSOURCING SYSTEM ATTACKS

In the last few years, Crowdsourcing systems have been victim to a number of cyber-attacks, most of which aimed to compromise and steal data or render systems unavailable. For example, in March 2014 the freelancer platform *Elance* experienced a Distributed Denial-of-Service (DDoS) attack [7] which kept the systems unavailable for more than a day. More precisely, attackers employed the use of a Network Time Protocol (NTP) reflection attack. Another large-scale cyber-attack against a well-known crowdsourcing system happened in October 2016 affecting *UBER*. According to Bloomberg [8], hackers were able to steal personal data by gaining access to *Uber*'s private Github account which contained developers' credentials for their Amazon Web Services (AWS) platform. This ultimately provides access to *Uber*'s AWS databases containing driver's personal data.

Furthermore, free-riding (e.g., benefiting from crowd-sourced task output without having contributed to its production) and false-reporting (e.g., to avoid the payment the employer lies regarding the task's status) are common attacks on crowdsourcing platforms, therefore, the need to propose and apply countermeasures is of great importance for maintaining the data integrity and utility of crowdsource platforms. For example, the use of Eliminating Free-riding and False-Reporting with arbitration (EFF) and Discouraging Free-riding and False-Reporting with arbitration (DFF) auction-based mechanisms and the development of reputation protocols in those untrusted environments are solutions able to prevent these problems [9, 10]. The EFF and the DFF are based on any existing truthful double auction scheme for winner selection and pricing. The auction winner is required to deposit a warranty and then submit a report regarding the status of the corresponding task. The payment is determined by the platform and is based on these reports. While these mechanisms are useful tools, they don't use any kind of encryption to guarantee the integrity of the process.

Finally, the crowdsourcing platforms should perform regularly security assessments regarding their status. These assessments should be carried on by experienced security officers who will, at the end of the process, provide a report that highlights the vulnerable points of the system. To this end, the platform should apply the best practices regarding the storage of sensitive information (i.e. encryption) and should be compliant with the General Data Protection Regulation (GDPR).

BLOCKCHAIN AND CROWDSOURCING

Blockchain technology addresses efficiently the weaknesses of crowdsourcing systems, this way boosting their attractiveness to solve several problems and widening their application potential. A blockchain database retains the complete, indelible and immutable history of all transactions, assets, and instructions executed since the very first one. With this, blockchain allows participating parties—and only those parties—to share accessible, transparent, and trusted information. The main characteristics to remember are: a) decentralised and distributed ledger storage and integrity, b) the ledger is irreversible and immutable, c) its operation is near real time (i.e. transactions verified and settled in minutes vs. days) and in any case satisfies the speed requirements of crowdsourcing which are significantly looser than those of the financial sector initially targeted by blockchain and d) it respects privacy (no personal data need to be registered). Users are identified by digital identities (exactly as credit cards) and only when physical world personal data are linked to those digital identities, is the linkage in place.

Adopting blockchain technology, the ledger of all transactions can be kept in a set of nodes (belonging either to workers or to requesters) obviating the need for a central authority/entity. The node resources are thus contributed by the peers that benefit from the platform and a small reward is granted to them. Such a system is proposed in [11], where a distributed system (entitled *CrowdBC*) is organised into three layers: the application layer, the blockchain layer and the storage layer. The blockchain layer is where the attributes of a transaction are kept (i.e. the ledger) while the storage layer includes the details and the content of the work produced by the workers. The application layer implements the business logic which, in the considered use case, is the user manager, the task manager and the program compiler. An important element of the *CrowdBC* is the use of smart contracts which follow the concept of smart contracts defined in *Ethereum*, [12]. The smart contract [13] is a self-executing digital contract in a secure environment with no intervention, which is verified through network peers. In crowdsourcing systems, a smart contract can be used by the system to describe the request-worker relationship (where the task ID, the task owner, the relevant deposit and task status are kept).

With respect to crowdsourcing, targeting information collection for different purposes ranging from facts (as e.g. Waze Carpool), opinions on events, products and solutions to collection of pieces of evidence and verdicts, blockchain makes possible the involvement of a larger number of people, which increases the quality of the data and thus of the offered service. Blockchain has been proposed for judgement produce to increase the quality of justice in [14]. Blockchain technology is also leveraged to improve other crowd-sourcing cases, like crowdfunding. Equity crowdfunding is considered a new channel of raising money for start-ups encouraging innovation and the adoption of blockchain based solutions has important advantages as reported in [5].

POTENTIAL USE CASES OF BLOCKCHAIN-ENABLED CROWD SOURCING SYSTEMS

As we have seen in the previous section, the crowd sourcing Blockchain-enabled systems, until now, have been comprised of approaches that include the creation of a platform for advertising crowd-working tasks that initiate partnerships between possible “workers” and employers.

In this section, we describe three real-life use cases where a different approach in the form of a decentralised P2P network is adopted. The proposed blockchain-based system could provide access to information on registered users to local news, along with the reputation scores of their publishers, allowing the combination of those information in order to deliver a decision about the truthfulness of a specific story. This way, the result could match the performance of the traditional method (i.e., using specific topic experts to validate a story) in a quicker and efficient process. We assume that the posts from Social Networks (SNs) and the users of the P2P network are used as crowdsourcing input data and are stored in a Blockchain to ensure information integrity and traceability. The users have access to the Blockchain and the information and (meta)data that are covered within, while a trustworthiness scoring system, is proposed in order to provide a validation regarding the integrity and veracity of the stored data.

Three different real-life use-case scenarios are considered: i) Fake news detection, ii) Organization of traditional media, and iii) Access to user's personal information through SNs. In the remaining of this section those use case scenarios will be presented, emphasising the benefits stemming from the adoption of blockchain concepts, as illustrated in Figure 2.

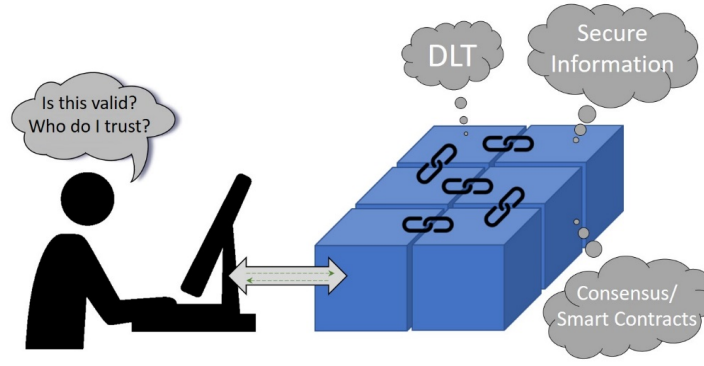

Figure 2: Blockchain - Enabled Crowdsourcing Systems

For the first use case, John, a social journalist in one of the most influential modern web sites is considered. As part of his work, he is supposed to check out on SN platforms (as *Mastodon* [17]) for stories related to a specific issue (e.g., the latest news about a soccer team). Consider the scenario where a certain story is presented in *Mastodon*, with multiple comments on it, regarding breaking news about a potential injury of one of the star players of the team. Trying to identify the truthfulness of the story, John is unable to find any other confirmation about it and, therefore, has to decide whether he should spread it only based on the noticed number of comments (called *toots* in *Mastodon*). Since the number of comments is not a reliable measure to describe a source as reliable, a more efficient one is needed. To achieve this, each user or journalist will be expected to have an account (or wallet in blockchain terms) with an attached reputation score. The post will be inserted as transaction in the Blockchain (using a low energy-hungry consensus algorithm presented later), while the score post will be updated following the reputation score of each of the users that interacted with it. Finally, the user score will be updated periodically following the post's score with which (s)he interacted at each interval. Considering that the storing of the scores will take place in a Blockchain, the truth will both be encouraged and rewarded. Steem blockchain platform [18] is a similar approach, since it combines both a connection to the SNs and the use of an inherent reputation system for the users, but by being based on a crypto-currency system and not being open source (i.e., not able to apply our custom reputation system) leads to develop solutions based on open-source tools like Hyperledger Fabric.

In the second case, traditional journalism dictates the use of subject experts and domain-specific journalists to check for the veracity of a story before this is published by the media. Those experts could use their experience and sources to check whether a story is true or biased and, therefore, prevent the publishing of articles that could harm the reputation of the media. But, since consulting an experienced person is often a time-consuming process, there is a need in modern internet-based journalism to search for validation and confirmation about a story quickly and efficiently. Therefore, a database (or system) that could be used as a source for confirmation would be highly appreciated. To this end, the proposed Blockchain database would be of great help to a journalist that needs confirmation regarding the news about an incident that takes place in the capital city of a foreign country, considering that the reputation score of each post will be visible through the Blockchain just by interacting with it.

The final use case regards the way that a blockchain enabled crowdsourcing storage could prevent users from actions that could, even, prove malicious to their health. In this use case, the collection of user data from SN applications is studied and a way to protect from the targeted prioritization of the news by the SN is suggested. In this scenario, a student, Lucy, is using *Mastodon*, a social platform that offers instances (or groups) where people with similar interests could exchange ideas and experiences. Especially, Lucy is following an instance regarding studying art and, often, posts and discusses her problems and struggles at keeping up with her studies at the local University.

One day, entering her account, Lucy finds a thread on her timeline regarding the use of a legal "smart" drug that will allow her to enhance her performance and, therefore, be able to meet up with the strict schedule of the University. Even though no scientific basis is provided regarding the performance of the drug and its effect on the user's health is not described, the narrative is very persuading and seems to match her needs. By using the blockchain data storage, Lucy could be able to search further and read reports by people that have already used the drug to find whether it is efficient and healthy to use it, based on the reputation score of those reports which stems from the reputation score of the users that interacted with them. That way, Lucy would be able to reach the right decision regarding the use or not of the suggested drug, without hesitation about making the right call.

It is worth stressing that one of the drawbacks attributed to blockchain technology is the energy consumption increase which is caused primarily by the mining and consensus process. While before implementing such a system the energy consumption should be considered as well, we anticipate that the volume of "transactions" in a crowd-sourcing blockchain solution is by far less than in case of a blockchain solution is used for money transactions. Additionally, a consensus algorithm different than Proof of Work (PoW) can be used (e.g., Proof of Stake, PoS) that leads to significantly lower energy consumption. For example, employing a private blockchain solution, the energy consumed is significantly lower due to lower intensity of processing thanks to the lightweight consensus mechanisms (see Hyperledger Fabric solution). On top of it, Hyperledger allows each node to hold *more than one ledger* allowing the creation of different channels to host each ledger. Therefore, each studied SN could use each own ledger to decrease the growth rate of each applied solution.

CONCLUSIONS

Blockchain's specific inherent characteristics, such as enhanced integrity and tamper proof operation (mainly attributed to the maintenance of the "ledger" in a distributed manner) are studied here, on how they can increase the performance of a novel crowd-sourcing platform. Especially, the use cases of EUNOMIA are presented, where the creation of a P2P network with reputation mechanisms between the peers manages, by using Blockchain as a database, to enhance the performance of the system and to meet the requirements of modern professionals (e.g., social journalist). The described scenarios are indicative on the many aspects of a human's life that can be affected and improved with the use of a Blockchain and how crowd-sourcing data can be efficiently used in this direction.

ACKNOWLEDGEMENT

The work presented in this document was carried out in the framework of H2020- EUNOMIA project preparation which has been accepted for funding from the European Union's Horizon 2020.

REFERENCES

1. Howe, J. 2006. "The rise of crowdsourcing". Wired Magazine, 14(6), p. 1–4.
2. "Upwork", Available link: https://www.upwork.com/, [Online].
3. "Amazon Mechanical Turk", Available link: https://www.mturk.com/, [Online].
4. "Waze Carpool", Available link: https://www.waze.com/, [Online].
5. P. Yang, Q. Li, Y. Yan, X.-Y. Li, Y. Xiong, B. Wang, and X. Sun, "friend is treasure: Exploring and exploiting mobile social contacts for efficient task offloading," IEEE Transactions on Vehicular Technology, vol. 65, no. 7, pp. 5485–5496, 2016.
6. M. H. Cheung, R. Southwell, F. Hou, and J. Huang, "Distributed time-sensitive task selection in mobile crowdsensing," in Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing. ACM, 2015, pp. 157–166.

7. D. Mayer, “Elance and oDesk hit by major DDoS attacks, downing services for many freelancers”, available link: https://gigaom.com/2014/03/18/elance-hit-by-major-ddos-attack-downing-service-for-many-freelancers/, last accessed November 2018.
8. E. Newcomer, “Uber paid hackers to delete stolen data on 57 Million People”, available link: https://www.bloomberg.com/news/articles/2017-11-21/uber-concealed-cyberattack-that-exposed-57-million-people-s-data, last accessed November 2018.
9. X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, “Keep your promise: Mechanism design against free-riding and false-reporting in crowdsourcing,” IEEE Internet of Things Journal, vol. 2, no. 6, pp. 562–572, 2015.
10. M. v. d. S. Yu Zhang, “Reputation-based incentive protocols in crowdsourcing applications,” in 2012 Proceedings IEEE INFOCOM, Florida,USC, 2012, pp. 2140–2148.
11. Li, Ming, et al. "CrowdBC: A Blockchain-based Decentralised Framework for Crowdsourcing." IACR Cryptol. ePrint Arch., Univ. California, Santa Barbara, Santa Barbara, CA, USA, Tech. Rep 444 (2017): 2017.
12. Ethereum, Available link: https://www.ethereum.org/, [Online]
13. K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," in *IEEE Access*, vol. 4, pp. 2292-2303, 2016. doi: 10.1109/ACCESS.2016.2566339
14. A. S. Federico Ast, “The crowdjury, a crowdsourced justice system for the collaboration era,” 2015.
15. H. Zhu and Z. Z. Zhou, “Analysis and outlook of applications of blockchain technology to equity crowdfunding in china,” Financial Innovation, vol. 2, no. 1, p. 29, 2016.
16. M. E. Peck, "Blockchain world - Do you need a blockchain? This chart will tell you if the technology can solve your problem," in *IEEE Spectrum*, vol. 54, no. 10, pp. 38-60, October 2017. doi: 10.1109/MSPEC.2017.8048838.
17. Mastodon, Available link: https://mastodon.social/about, [Online].
18. Steem: Available link: https://steem.com/, [Online].

ABOUT THE AUTHORS

Dimitrios G. Kogias was born in Athens in 1978. He received his diploma in Physics from the National and Kapodistrian University of Athens in 2001. In December 2004 he received his M.Sc. in Electronics and Radioelectrology and in May 2010 his Ph.D degree from the National and Kapodistrian University of Athens on algorithms for dissemination of information in Unstructured Networking Environments. He works as an Adjunct Lecturer in the Department of Electrical & Electronics Engineering at the University of West Attica (UWA). He has participated as a Senior Researcher in various Research projects (e.g., most recently on TRILLION and STORM at HORIZON 2020) funded from National and/or European resources. His current research interests include Blockchain technology, Cloud integration, security in the Internet of Things (IoT), Machine-to-Machine (M2M) communications and privacy issues in these environments. His works have been published in international Journals and Conferences, while he has, also, co-authored scientific book chapters.

Helen C. Leligou is assistant professor at University of West Attica. She received the Dipl.Ing. and Ph.D. degrees, both in electrical and computer engineering, from the National Technical University of Athens (NTUA), Athens, Greece, in 1995 and 2002, respectively. Her research interests lie in the area of protocol design for communication systems, access control mechanisms in optical access/ metro/ core networks, with emphasis on their implementation in hardware (FPGA, ASIC technologies) as well as in security, routing and application layer protocols for wireless sensor networks and their implementation in embedded systems. Recently, she has been working on technology enhanced learning applications and blockchain technologies. Her research results have been published in more than 100 scientific journals and conferences. She has participated in several EU-funded ACTS, IST and ICT and H2020 research projects in the above areas.

Michael Xevgenis is a Research Associate at the Computer Network and Services Research Team (CONSERT) of the University of West Attica. His research concerns are on the Networking and Cloud Computing sector. The last few years he has participated in two EU projects: the TRILLION and the STORM of H2020. Finally, he holds the Master in Data and Networking Communications of Kingston University in collaboration with TEI of Piraeus.

Maria Polychronaki was born in Athens on 1995 and is currently an Electrical & Electronics Engineering student at University of West Attica. Main focus of her thesis was the development of REST-ful APIs in the context of communication improvement for Internet of Things systems. Her current research interests include Internet of Things and Blockchain technologies.

Evangelos Katsadouros was born on 1994 in Athens, Greece. He received his Computer Systems Engineering degree from the Piraeus University of Applied Sciences in 2017. He is currently attending a Msc degree in Digital Systems Security at University of Piraeus and he is a researcher of the CONSERT team of Computer Networks Lab of Electrical & Electronics Engineering Dept of UNIWA.

George Loukas is an Associate Professor and head of the IoT and Security (ISEC) research group at the University of Greenwich. Over the last five years, he has been his institution's principal investigator in seven research projects. His current projects include H2020 EUNOMIA, where he is the overall coordinator on the development of technologies for thwarting fake news in social media; H2020 TRILLION, where he lead the work on enhancing the cyber trustworthiness of mobile devices used by human sensors of physical crime; British Council's Secure Haįj and Umrah with blockchain-based authentication for smart services for pilgrims; and the EPSRC CHIST-ERA COCOON project, where he leads the development of attack and monitoring techniques, and also studies the performance of human sensors of Internet of Things cyber threats. Dr. Loukas has a PhD in Network Security from Imperial College. He is on the editorial board of BCS's The Computer Journal and Elsevier's Simulation Modelling Practice and Theory. His book "Cyber-physical attacks: a growing invisible threat" has been chosen by ACM in the top 10 in the Computing Milieux category of the 2015 annual list of notable books and articles published in computing.

Ryan Heartfield is a Research Associate in Cyber Security at the University of Greenwich. He is currently involved in multiple UK and European research projects in cybersecurity, ranging from information trustworthiness in social media (H2020 EUNOMIA), the security of autonomous vehicles, measuring the trustworthiness of human sensor platforms (H2020 TRILLION), as well as studying cyber threats and the emotional impact of security breaches in smart home environments (EPSRC CHIST-ERA COCOON). Dr. Heartfield has a Ph.D. in Cyber Security from the University of Greenwich. His research interests include semantic social engineering threats, intrusion detection systems, cyber-physical attacks, software-defined networks, cloud computing and network security.

Charalampos Z. Patrikakis is an Associate Professor at the Dept. of Electrical and Electronics Engineering of the University of West Attica. He has participated in more than 35 National, European and International programs, in 20 of which he has been involved as technical coordinator or principal researcher. He has more than 100 publications in chapters of books, international journals and conferences, and has 2 contributions in national legislation. He is a member of the editorial committee of more than 50 international journals and conferences, and has acted as editor in the publication of special issues of international journals, conference proceedings volumes and coedited three books. He is a senior member of IEEE, a member of the Technical Chamber of Greece, and counselor of the IEEE Student Branch of the University of West Attica.