IEEE CLOUD COMPUTING

VOLUME 5, NUMBER 4 JULY/AUGUST 2018

Biometrics-as-a-Service

IEEE computer society IEEE ComSoc IEEE Communications Society

www.computer.org/cloud

2018

CSI-IEEE CS Joint Education Award

CALL FOR 2018 AWARD NOMINATIONS
Deadline 1 October 2018

CRITERIA

The Computer Society of India (CSI) / IEEE Computer Society Education Award is a joint CSI-IEEE national society award that recognizes educators who have made significant contributions to computer science and engineering education. The contributions can be at undergraduate or graduate teaching (BTech/BCA/MCA,MTech or PhD). Contributions may include writing influential texts, course materials, and papers on Research or education; inspirational teaching; and innovative teaching and/or development of curriculum or methodology.

ELIGIBILITY

- Nominees should be teaching in India, preferably for a least the past three years.
- Nominators must be either CSI, IEEE or IEEE CS members.
- Self-nominations are not accepted.

NOMINATION REQUIREMENTS

The nomination form must include at least two endorsements.

AWARD

The CSI/IEEE-CS Education Award will consist of a certificate and US $500 honorarium.

PRESENTATION

The award will be presented at the CSI: Annual Convention of CSI.

SUBMISSIONS

Nomination forms should be sent to csiieee @csi-india.org by the 1 Oct. deadline.

COMPUTER SOCIETY OF INDIA™ ESTD. 1965 सर्वे भवन्तु सुखिनः

IEEE computer society

IEEE

IEEE CLOUD COMPUTING

www.computer.org/cloudcomputing

EDITORIAL OFFICE
Publications Coordinator:
cloudreview@allenpress.com
Authors: www.computer.org/web/peer-review/magazines
Letters to the Editors: bbrannon@computer.org
Subscribe: www.computer.org/subscribe
Subscription change of address:
address.change@ieee.org
Missing or damaged copies:
help@computer.org
Reprints of articles: cloud@computer.org
IEEE Cloud Computing
c/o IEEE Computer Society
10662 Los Vaqueros Circle,
Los Alamitos, CA 90720 USA
Phone +1 714 821 8380; Fax +1 714 821 4010
www.computer.org/cloud-computing

IEEE ComSoc
IEEE Communications Society

SUSTAINABLE FORESTRY INITIATIVE Certified Sourcing www.sfiprogram.org SFI-01681

IEEE Cloud Computing: (ISSN 2325-6095) is published bimonthly by the IEEE Computer Society. IEEE headquarters: Three Park Ave., 17th Floor, New York, NY 10016-5997. IEEE Computer Society Publications Office: 10662 Los Vaqueros Cir., Los Alamitos, CA 90720; +1 714 821 8380; fax +1 714 821 4010. IEEE Computer Society headquarters: 2001 L St., Ste. 700, Washington, DC 20036. Subscribe: Go to www.computer.org/subscribe for more information on subscribing.

IEEE CLOUD COMPUTING

July/August 2018
Vol. 5, No. 4

www.computer.org/cloud

TABLE OF CONTENTS

Biometrics-as-a-Service

Feature Articles

Columns and Departments

Also in This Issue

ISSN: 2325-6095 Copublished by the IEEE CS and IEEE ComSoc

Biometrics-as-a-Service – The Final Frontier of Security

Mazin Yousif
T-Systems, International

Editor-in-Chief Mazin Yousif discusses the prospects and challenges of biometrics-as-a-service.

Information and Communication Technologies (ICT) have become integral to everything we do. No one can imagine life today without some form of ICT. With such ubiquitous dependence on ICT, there are those who will abuse it and try to find ways to penetrate our lives through it, whether stealing our private data, robbing our valuables and money or just finding ways to annoy us. To protect businesses and consumers from such bad actors, whether state-sponsored, criminals, or the 14-year-old next door, companies have made billions of dollars introducing and selling all types of security products, services, and features—firewalls, antivirus, hard disk encryption, the list goes on. Biometric-based security products, services, and features, however, are in a class of their own. Their main use is to authenticate a user. When such services are hosted in the cloud, they become biometrics-as-a-service (BaaS)—the theme of this issue.

Biometric services can be very diverse. In fact, I can go out on a limb (pun intended) and say that virtually all of a person's body parts (fingerprints, retina, teeth/dental, hair, facial structure, etc.), components (blood type, microbiome, DNA, etc.), or characteristics (height, weight, eye color, voice, etc.) can be included in a biometric service. The reason is that the constellation of elements that make up a person's body is unique to that person. In other words, they don't rely on the hardware the person carries (e.g., dongles), which can be stolen, or what they know (e.g., birthday, pet's name), which can not only be stolen or guessed by brute-force attacks or social engineering, but worse yet are widely available through social media.

Instead, biometric security relies on hard to reproduce characteristics of people themselves. After all, why rely on a 4-digit PIN when you could use the multi-billion base pair long sequence of human DNA. The simplest biometric service that millions of us are already using is the fingerprint readers in our smart phones, which are becoming ubiquitous (note that fingerprints are typically stored on the phone and not in the cloud, otherwise it could be a challenge to unlock your phone in a place where there is no connectivity). Other types of biometric services are voice recognition, facial recognition, gait recognition, finger vein recognition, and iris recognition. The way this typically works is that the digitized image is encrypted and stored in a secure location in a device such as a smart phone. For BaaS, these are services where the digitized record is stored in the cloud instead of the device the person carries or uses.

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

It is expected that biometric services will see large-scale adoption in the next few years as they are becoming associated with the typical consumers' mobile experience. They are getting adopted in various industry verticals such as banking, transportation, and retail. For example, airlines are experimenting with facial recognition for their boarding process. Banks are rolling out fingerprint technologies to identify clients. Biometrics have also been adopted by governments. In fact, countries are considering using them as national identity methods.

Biometric services do not come without challenges. For example, facial recognition may require a similar amount of lighting and background as when the original images were taken, and security holes have already been identified in commercially available products. Voice recognition in a noisy environment may be difficult. Fingerprints have been "stolen" from broadcast HD video of celebrities. But it is clear that we should be able to combine several biometrics to deal with whatever challenging environment the user is in and also to strengthen the identification process. Additionally, clients need to be given a variety of biometrics options to identify themselves. For example, if someone is driving and needs to access their bank account, they may not be able to easily use a retinal image or perhaps even their fingerprints, but they can easily use voice recognition to identify themselves. An additional potential challenge with biometrics is that they do not change—unlike passwords where we can change them whenever we lose them. Will there be a way in the future to change human biometrics? Anything is possible. Even today, people can wear masks to fool facial recognition or fake fingerprints. As I mentioned earlier, for BaaS, the digitized record, against which authentication takes place, is stored in the cloud and is typically encrypted. But if there are ways to steal it and unencrypt it, then game over—we will not be able to rely on them. That said, there are other ways to reduce or eliminate this possibility further and that is by splitting the digitized record into many bits and pieces and storing them encrypted in various disparate disconnected locations and devices, and to use multi-factor authentication: something you know (e.g., password), something you have (e.g., USB security device), and something you are (e.g., DNA).

An interesting angle to biometric services is to see how to use them in new markets and for new audiences, such as using biometric services in IoT deployments. This will depend on the use case, but it is worthwhile to see how to incorporate them as basic or additional security measures, especially for the things at the edge. Note that things can be people carrying devices such as smart phones, or could be unlocking the wine fridge.

This theme of this issue is a special one: "Biometrics-as-a-Service: Cloud-Based Technology, Systems, and Applications" with Guest Editors Silvio Barra, Kim-Kwang Raymond Choo, Michele Nappi, Arcangelo Castiglione, Fabio Narducci, and Rajiv Ranjan. I would like to thank them for all their efforts in producing this issue. I also urge the readers to read their guest editors introduction.

ABOUT THE AUTHOR

Mazin Yousif is the editor in chief of *IEEE Cloud Computing* magazine. Contact him at mazin@computer.org.

Blockchaining the Cloud

Christine Miyachi
Xerox Corporation

During the rise of cloud computing, activist programmers created blockchain. Now it has the potential to transform the existing cloud applications.

When I was in business school, we studied how to solve problems creatively. One technique introduced was TRIZ,[1] which involves coming up with a contradictory solution. For example, if a faster engine produces too much heat, propose a faster engine that cools—and then figure out how to do it. Let's apply that to one of the thorniest problems in cloud computing—trust. A centralized authority is typically the basis of trust today, but that authority can be spoofed, or be untrustworthy themselves. If I were to use TRIZ, I would propose to create trust established by decentralized authority and assume that no one can be trusted. That solution already exists: blockchain.

Mike Gault, co-founder and CEO of Guardtime claims CIOs (Chief Information Officers) require that cloud suppliers provide "a secure supply chain and that they can verify every step in that supply chain in real-time; when things go wrong it is possible to figure out what went wrong and that there is someone who can be held accountable."[2] But he claims that not a single cloud provider can meet that demand. Blockchain has the promise to deliver, and there are many proposed applications to improve verification. This technology has the potential to revolutionize the financial and legal sectors as well as a wide variety of other industries.

In essence, blockchain will enable untrusted users to work together—to exchange currency, to make agreements, to validate personal records—without centralized authority. Centralized authorities—for example, banks—are expensive. But using blockchain is not free, and the energy and computing power create a transaction is also expensive. Also, security concerns still exist. In this column, I will explore blockchain applications that will alter cloud computing and some of the hazards of blockchain.

DEFINITIONS

In 2009, Satoshi Nakumoto[3] created Bitcoin, a digital currency and one of the first implementations of blockchain technology in response to the 2008 banking meltdown. A centralized authority had failed its users and bitcoin would rid the world of that authority. The technology was made possible by the wide use of cloud computing and underlying internet technologies. A blockchain is a "database encompassing a physical chain of fixed-length blocks that include 1 to N transactions, where each transaction added to a new block is validated and then inserted into the block. When the block is completed, it is added to the end of the existing chain of blocks. Moreover, the only two operations—as opposed to the classic CRUD (create, read, update, delete)—are add-transaction and view-transaction."[4] A thorough introduction to blockchain was written by Morgan E. Peck in *IEEE Spectrum*, which he titled "Blockchains:

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

How They Work and Why They'll Change the World: The technology behind Bitcoin could touch every transaction you ever make."[5] Blockchain is a distributed ledger and the transactions between parties in the ledger are recorded permanently and independently verified by a majority of verifiers. More than currencies, imagine contracts as a blockchain where the code to execute them is embedded. Blockchain may not be disruptive, but more a foundational technology. Karin R. Lakhani of Harvard Business School says, "It has the potential to create new foundations for our economic and social systems. But while the impact will be enormous, it will take decades for blockchain to seep into our economic and social infrastructure."[6]

WHAT MAKES BLOCKCHAIN SECURE

Since blocks don't change after being added, hackers have difficulty tampering with the chain. The entire blockchain is shared among networked computers called nodes. With each new transaction, each copy of the blockchain is updated. But before that transaction is added, it must be verified. In the case of Bitcoin, the verification determines if that entity has a Bitcoin to spend. Nodes called miners validate transactions. The validation involves hashing and cryptography that is explained well in the *IEEE Spectrum* article mentioned above. What makes it tamperproof is that the nodes in the network do the validation and the majority have to agree that the block is valid. One downside is that the cryptographic operations uses a lot of computing and therefore a lot of energy (see Figure 1). The blockchain is a list, and each block has a link to the previous block. If someone wants to change the block, that hash of that block will conflict with the current chain, and the miners will not validate it. The blockchain is difficult to change and therein lies its security.

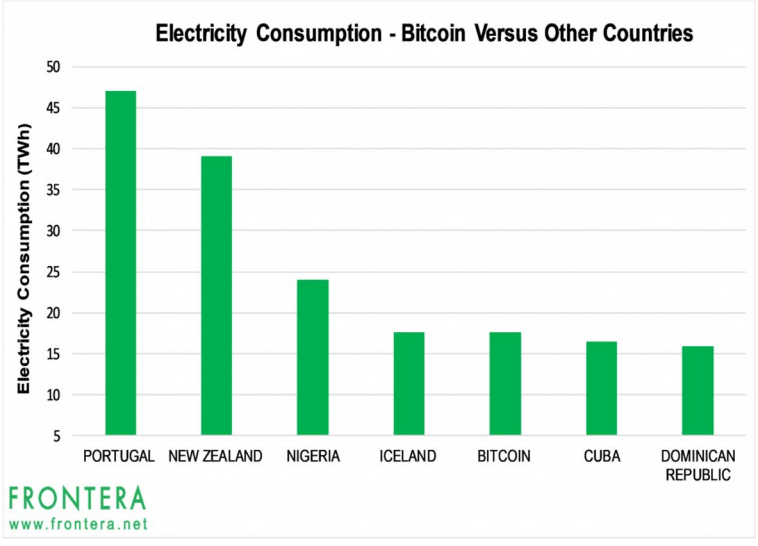


Figure 1. The energy consumption of all of today's Bitcoin processing is on the order of some smaller countries.[7]

THE DECENTRALIZATION OF THE INTERNET

The idea of decentralization was part of the introduction of the Internet. In 1996, John Perry Barlow created "A Declaration of the Independence of Cyberspace" where he made the plea for decentralization by saying to governments, "I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather."[8] While the Internet is still decentralized and governed by authorities like ICANN, a large amount of traffic passes through a few large corporations like Google, Amazon, Netflix and Facebook.[9]

Blockchain applications infused with this decentralization strategy are launching and institutions are rising to the challenge. The European Union put into effect a Payment Services Directive

(PSD2) in 2018.[10] Banks must now provide open programmable interfaces to third parties to manage customer finances. Blockchain is poised to take advantage of these APIs and provide secure transactions and eventually reduce the services of banks and the fees they charge.

With institutions getting more involved with blockchain, the decentralization that the Satoshi Nakamoto group envisioned may not occur. Vili Lehdonvirta, a professor at the University of Oxford, disagrees that blockchain will transform the economy and he calls this "the blockchain paradox."[11] His main point is that while blockchain can enforce rules, we need people to make the rules. Using Bitcoin as an example, he claims that the development team made the initial rules for Bitcoin. He says, "Humans are still very much in charge of setting the rules that the network enforces." Even with his insights, the blockchain was near the top of the hype cycle in 2017 (see Figure 2). The number of applications now available using blockchain has risen dramatically.

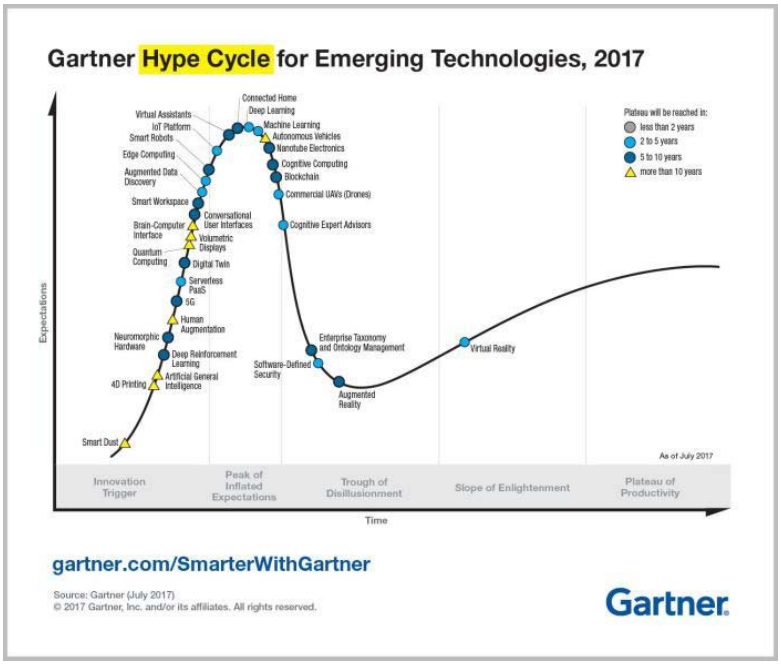


Figure 2. Blockchain was just over the top of the hype cycle in 2017.[12]

APPLICATIONS

Many of the large internet applications mine user data for profit. In the sharing economy, Uber and Airbnb take owner (cars or living quarters) information and match that information to customers who want to get a ride or rent an apartment. The value contributed is not equally distributed back to users as these companies take a share for their service. Just as with banks, a portion of the profits is captured by the intermediary. If people can store their identity on a blockchain, it is possible to reduce the role of this intermediary. La Zooz[13] is an identity management system based on blockchain—it is an open-source decentralized collaborative transportation system. Systems like this promote good behavior because you can't delete accounts and reregister because entries in the blockchain never get deleted. Your identity is verified but can be completely anonymous.

For secure identity, Onename[14] is a startup that allows you to create an ID using Bitcoin and they promise this ID can be used to log into websites without a need for a password. The Onename webapp is built on Blockstack. Blockstack is a decentralized naming and storage system. It's a replacement for traditional internet services like DNS, public-key infrastructure (PKI), and cloud storage and promotes an open internet by putting the users in control.

Real estate titles can be securely recorded and verified and not easily stolen. Many of the transactions of real estate rely on centralized authorities. Factcom[15] is a company dedicated to using blockchain for secure document management in real estate (and other industries), and they remove the dependence on these centralized authorities. Factcom believes "in keeping private data private and securing the world's wealth because privacy and possession of property are basic human rights."

Smart contracts, based on blockchain, are digital contracts that execute automatically and will provide a new business model for legal firms. The contracts are executed digitally and automatically and are irreversible. Lawyers will have to be programmers as well as understand the legal aspects of contracts and mediation. Smart contracts are self-enforcing and execute when a contract is broken or terminate when the terms of the contract complete.

Decentralized file storage based on blockchain promises to reduce costs associated with centralized file storage systems like DropBox or Google Drive. Storj[16] is a file system that rents out storage and bandwidth without using a centralized system. They claim to have the same performance and reliability as a centralized system, at a lower price, but at this time they are not accepting more users until they develop their next generation of software. Interplanetary File System (IPFS)[17] is another decentralized storage system using peer-to-peer methods. Like many of these blockchain applications, it is an open-source project.

Think about decentralized organizations that could work democratically and collaboratively—a company called backfeed.cc[18] provides a framework to do that. They say "Imagine Facebook owned by its users, decentralized transportation networks independent of Uber, markets dominated by open-source communities, where contributors are also shareholders and where the value created is redistributed both fairly and transparently." They promise that backfeed has the infrastructure to do that. Imagine solving the world's most difficult scientific problems by allowing anyone to post a solution and then rating that solution—Matryx.ai[19] provides a framework to perform this service. In the company's whitepaper they explain, "Matryx is composed of a bounty system and a marketplace for digital assets to be bought, sold, and remixed into new assets. Bounties are placed on solutions to specific problems. Submissions to bounty tournaments enter the collection of assets and are available to other users. In this way, collaborators are incentivized to build, distribute, and expand upon each other's work in the pursuit of valuable goals. Matryx reduces friction of collaboration between strangers by providing a common framework and concrete goals."

Blockchain even enables an application that may change the cloud itself. A decentralized cloud platform is available that guarantees privacy. Enigma[20] allows private data to be stored, shared, and analyzed but does not allow the data itself to be fully revealed. They say, "Blockchains without privacy are useless. Smart contracts without privacy are useless. If these technologies cannot work without privacy, then new privacy technologies are the truly useful innovations." And they promise to deliver new solutions to create that privacy.

NO PERFECT SOLUTION

The applications mentioned above have not all been fully implemented and thus are not completely proven. Initial Coin Offerings (ICOs, a blockchain alternative to Initial Public Offerings [IPOs] have had some difficult failures).[21] For many of these applications to work, governments and other institutions will need to go through some revolutionary changes and provide regulatory support. These are the intermediaries—just the type blockchain purports to remove. The blockchain paradox is that this technology needs governance and once the governance is there, blockchain is no long decentralized.

There are also ways to attack a blockchain. One way is an "eclipse attack." Nodes that have copies of the blockchain must communicate and this communication channel can be fooled. Also, a "selfish miner" could fool other nodes into wasting time-solving already-solved blocks. The failures are typically where the blockchain connects to other systems such as "hot wallets" that store the private keys used for the cryptography. There are solutions to these problems but a more difficult issue with bugs. If the blockchain code itself has a bug, the knowledge of this bug could be

used to attack. This is what happened in 2016 when a crowdsourced venture capital platform called The DAO (decentralized autonomous organization) based on the Ethereum blockchain opened for business. Soon after starting, an engineer found a bug in the code, and The DAO was hacked, with hackers stealing 60 million dollars in cryptocurrency.[22]

Private blockchains may be a solution to some of the security issues. The participants would be screened and there would be an immutable log of the participation. Private blockchains can also limit the miners to be trusted sources. Private blockchains are not as widely decentralized but could provide more secure operations. Using private blockchains may make the adoption of this technology more palatable to existing institutions such as governments and banks.

CONCLUSION

Blockchain is a technology that could be applied to solve some cloud security problems. The applications claim to give control back to individuals. But even though the internet is still considered decentralized, large organizations control much of the traffic. The same centralizing could happen with blockchain. Research out of Cornell University found that four miners did 53 percent of the bitcoin mining work in a week and similarly, three Ethereum miners did 61 percent of the work.[23] While the blockchain itself appears to solve some security issues, its widespread use is just starting. Governance will be necessary for blockchain applications to work within current systems. Misuse will always be looming. John Kenneth Galbraith said, "A constant in the history of money is that every remedy is reliably a source of new abuse."[24] And yet it doesn't keep us from reaching for that perfect remedy.

REFERENCES

1. "TRIZ," *Wikipedia*, Wikimedia Foundation, 3 July 2018; https://en.wikipedia.org/wiki/TRIZ.
2. M. Gault, "BlockCloud: Re-Inventing Cloud with Blockchains," *guardtime*, blog; https://guardtime.com/blog/blockcloud-re-inventing-cloud-with-blockchains.
3. "Satoshi Nakamoto," *Wikipedia*, Wikimedia Foundation, 13 July 2018; https://en.wikipedia.org/wiki/Satoshi_Nakamoto.
4. J.J. Bambara et al., *Blockchain: A Practical Guide to Developing Business, Law, and Technology Solutions*, McGraw-Hill Education, 2018.
5. M.E. Peck, "Blockchains: How They Work and Why They'll Change the World," *IEEE Spectrum*, 28 September 2017; https://spectrum.ieee.org/computing/networks/blockchains-how-they-work-and-why-theyll-change-the-world.
6. M. Iansiti and K.R. Lakhani, "The Truth About Blockchain," *Harvard Business Review*, 6 March 2018; https://hbr.org/2017/01/the-truth-about-blockchain.
7. S. Bubna, "Bitcoin Mining Now Consumes As Much Electricity As Iceland," *Frontera*, 17 October 2017; https://frontera.net/news/global-macro/bitcoin-mining-now-consumes-as-much-electricity-as-iceland/.
8. J.P. Barlow, "A Declaration of the Independence of Cyberspace," *Electronic Frontier Foundation*, 8 February 1996; https://www.eff.org/cyberspace-independence.
9. J. Brogan, "A Cheat Sheet Guide to Who Controls the Internet," *Slate*, 1 November 2016; http://www.slate.com/articles/technology/future_tense/2016/11/a_cheat_sheet_guide_to_who_controls_the_internet.html.
10. "Payment Services Directive," *Wikipedia*, Wikimedia Foundation, 3 July 2018; https://en.wikipedia.org/wiki/Payment_Services_Directive.
11. V. Lehdonvirta, "The Blockchain Paradox: Why Distributed Ledger Technologies May Do Little to Transform the Economy," Oxford Internet Institute, 21 November 2016; https://www.oii.ox.ac.uk/blog/the-blockchain-paradox-why-distributed-ledger-technologies-may-do-little-to-transform-the-economy/.
12. F. Van De Ven, "Blockchain, Gartner's Hype Cycle and a local Mexican coin called Túmin: is the age of disillusionment approaching?," *Medium*, 28 February 2018;

https://medium.com/@frankvandeven/blockchain-gartners-hype-cycle-and-a-local-mexican-coin-called-t%C3%BAmin-is-the-age-of-c3f77de9cc6d.
13. *La'Zooz*; http://lazooz.org/.
14. *Onename*; https://onename.com/.
15. *Factom — Making the World's Systems Honest*; https://www.factom.com/.
16. *Decentralized Cloud Storage - Storj*; https://storj.io/.
17. "InterPlanetary File System," *Wikipedia*, Wikimedia Foundation; https://en.wikipedia.org/wiki/InterPlanetary_File_System.
18. *Spreading Consensus*; http://backfeed.cc/.
19. *Matryx — Tackle Science's Greatest Challenges with VR and Blockchain-Based Collaboration*, Nanome; https://matryx.ai/.
20. *Project Overview ‹ Enigma – MIT Media Lab*, MIT Media Lab; https://www.media.mit.edu/projects/enigma/overview/.
21. G. Lewis-Kraus, "Inside the Crypto World's Biggest Scandal," *Wired*, 19 June 2018; https://www.wired.com/story/tezos-blockchain-love-story-horror-story/.
22. "The DAO (Organization)," *Wikipedia*, Wikimedia Foundation, 20 July 2018; https://en.wikipedia.org/wiki/The_DAO_(organization).
23. M. Orcutt, "How Secure Is Blockchain Really?," *MIT Technology Review*, 25 April 2018; https://www.technologyreview.com/s/610836/how-secure-is-blockchain-really/.
24. "John Kenneth Galbraith," *Wikiquote*, Wikimedia Foundation; https://en.wikiquote.org/wiki/John_Kenneth_Galbraith.

ABOUT THE AUTHOR

Christine Miyachi is a systems engineer at Xerox Corporation and holds several patents. She works on Xerox's Extensible Interface Platform, which enables developers to create applications that work with Xerox devices by using standard web-based tools. Miyachi has two MIT degrees: an MS in technology and policy/electrical engineering and computer science and an MS in system design and management. Contact her at cmiyachi@alum.mit.edu.

TrustChain: Establishing Trust in the IoT-based Applications Ecosystem Using Blockchain

Bin Yu
Monash University,
CSIRO Data61

Jarod Wright, Surya Nepal, Liming Zhu
CSIRO Data61

Joseph Liu
Monash University

Rajiv Ranjan
Newcastle University

The Internet of Things (IoT) has already reshaped and transformed our lives in many ways, ranging from how we communicate with people or manage our health to how we drive our cars and manage our homes. With the rapid development of the IoT ecosystem in a wide range of applications, IoT devices and data are going to be traded as commodities in the marketplace in the near future, similar to cloud services or physical objects. Developing such a trading platform has previously been identified as one of the key grand challenges in the integration of IoT and data science. Deployment of such a platform raises concerns about the security and privacy of data and devices since their ownership is hard to trace and manage without a central trusted authority. A central trusted authority is not a viable solution for a fully decentralized and distributed IoT ecosystem with a large number of distributed device vendors and consumers. Blockchain, as a decentralized system, removes the requirement for a trusted third-party by allowing participants to verify data correctness and ensure its immutability. IoT devices can use blockchain to register themselves and organize, store, and share streams of data effectively and reliably. We demonstrate the applicability of blockchain to IoT devices and data management with an aim of providing end-to-end trust for trading. We also give a brief introduction to the topics and challenges for future research toward developing a trustworthy trading platform for IoT ecosystems.

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

The number of Internet of Things (IoT) devices has already exceeded the world population. With rapid advancement in hardware technologies, these smart devices have been applied in almost every aspect of our daily lives. A large amount of data is generated every second and data science research is actively defining algorithms to process such data to make and enact better decisions for us in our daily activities. For example, wearable smart devices such as smartwatches sense our heartbeat and blood pressure continuously to monitor our health condition; a smart fridge enables us to control the fridge remotely and plan a healthier diet; a smart air conditioner can track our living preferences and adjust the temperature automatically; an autonomous vehicle frees our hands and minds while making our journey safe.

One of the key grand challenges is how we ensure that users trust the IoT ecosystem to make the right decisions and act on them. This involves trusting devices, data and analytics, as previously identified in this column.[1] The focus of this article is to analyze different research and technical issues related to managing trust using blockchain in a fully decentralized IoT ecosystem.

Though these smart devices bring great convenience to us in our daily life, news such as US cell carriers (including AT&T, T-Mobile. and Sprint) selling access to customers' real-time phone location data to a little known company called Securus[2] raises public concerns about the risk of personal data leakage and abuse. Such news prompts a debate on whether these IoT devices are our friends or enemies. Trust is not a one-way street in the IoT ecosystem. Data analysts have concerns about the integrity of the data that data owners provide. At the same time, data owners are concerned about whether data analysts only use their data for its declared purposes. Additionally, data owners care about how to protect their own data (sometime captured by manufacturers) when IoT device ownership changes during its life. For example, what happens to the data of a car owner when an autonomous car is sold or the ownership of a car is changed?

Users find difficulties in enjoying the services provided by these smart devices if they don't meet the high security expectations from them. Some key challenges for building a trustworthy trading platform for IoT devices and data are outlined below:

Lack of trust among participating entities. Trust is hard to achieve among different entities involved in IoT data processing due to the lack of a governance framework. As defined by NIST in its Network-of-Things (NoT)[3] report, which aims to define IoT formally, five key primitives are involved in real IoT applications: sensors (IoT devices for generating data), aggregators (edge, fog or mist infrastructure for aggregating data), communication channels (wired and wireless communication provided by communication service providers), eUtility (SaaS, PaaS, IaaS provided by clouds), and decision triggers (data analysis pipelines, decision making and enacting processes). Each of these primitives is likely to be supported by different service providers. How can they trust to each other? For example, in many cases, IoT device owners would not know who are the data processing entities or cloud service providers. Unless they have a mechanism to trust them to handle the data properly, they cannot use the services they provide to support the five primitives. More seriously, there is no standard agreement among different entities to define the data usage policy and it is hard to supervise the usage of personal data. The reliability and security of all entities providing five primitives is important to establish the trust.

Lack of data supervision and management. In many applications, data collected by IoT devices is mostly maintained and processed by either the device manufacturer or a trusted third-party. For example, consider an IoT application of monitoring chronic patients at home[4] where a patient is monitored for his activities (like exercises) and health (blood pressure, heart beats) using IoT devices. A service provider may share patient data with health data analytics, general practitioners, and related service providers, including cloud data service providers. Patients have limited knowledge about how their data is processed and used. Additionally, when the data generated by IoT devices is transferred from one party to another, there is often no data integrity verification. The tampered data could result in misleading decisions and the source of the fraud is difficult to identify.

Lack of devices' lifecycle management. Every product undergoes a series of phases in its lifecycle: design, sourcing of components, manufacturing, distribution, retail, repair, resale, and so on. For IoT devices, the management of devices and related data are critical because the data

generated by devices at different phases should be isolated and well protected. To date, the visibility remains highly siloed and opaque across entities. For instance, patient health data generated by an artificial cardiac pacemaker is fragmented to the manufacturer, doctor, and insurance company. During the device's lifecycle, a patient could change from one doctor to another; in such a scenario, the data accessibility should also be transferred. Currently, the data is fragmented and held by many entities and there is no regulation on auditing the ownership of IoT devices. Similarly, there is no guarantee that the data held by each of them is consistent with others. If the patient is transferred from one doctor to another, there is no mechanism on how the data accessibility is managed.

A lot of effort has been put into resolving the issue of trust among different entities in the IoT application ecosystem. Unfortunately, there is not a single reference scheme that satisfies all stakeholders. The key issue for the traditional solutions is that they all depend on a trusted third-party, which has to be trusted by all stakeholders. Blockchain, as a new data-sharing model, addresses this issue by removing the need for a trusted third-party. It allows all stakeholders to participate in maintaining an immutable ledger in which the data is consistent among all stakeholders. Since the data on the ledger is immutable, we avoid the possibility that any participant tampers with the data by allowing all participants to verify the correctness of the data. In this article, we argue that with the help of the blockchain technology, the management of the IoT device life cycle and the corresponding data privacy can be enhanced in the following ways:

- instead of trusting a third party, IoT devices can exchange data through the blockchain;
- IoT devices and the data generated by IoT devices can be traced to avoid the manipulation of the data by malicious parties;
- different stakeholders can trust the validity and integrity of the data on the chain;
- the communication among different entities can be simplified as they only need to interact with the blockchain to retrieve/upload data;
- the deployment and operation cost of IoT can be reduced through a blockchain since there is no intermediary;
- computation-intensive operations like end-user authentication and access control can be processed on the blockchain instead of IoT devices;
- it is more convenient for the blockchain to maintain device and data ownership; and
- the distributed ledger eliminates a single point of failure within the ecosystem—IoT devices and end users can interact with any blockchain nodes to access the data;

In the following, our aim is to demonstrate the feasibility of a trustworthy trading platform with the above stated capabilities. Before going to a case study of such a platform, we give an overview of the blockchain technology.

BLOCKCHAIN

Trust plays a critical role in information exchanges. It helps different entities deal with each other more effectively and is often a key element in any collaborative system. Traditionally, centralized trusted institutions such as banks or government agencies manage the trust problem. With the help of these centralized institutions, different entities can cooperate with each other with a certain degree of confidence. Blockchain, known as an electronic ledger, tries to replace such centralized institutions by distributing the trust in a decentralized network. In a blockchain system, the ledger is immutable and not held on a single server but among all servers in the network. The openness feature of blockchain allows any participant to modify the ledger under a set of rules dictated by a "consensus protocol." The "consensus protocol" requires the majority of the blockchain participants to agree on the modification of the ledger to ensure the trustworthiness of the blockchain. Once a new consensus is achieved, all participants update their own ledger simultaneously. If any of the participants violates the consensus protocol to propose a new data entry, the network treats that entry as an invalid one.

Practically, transactions are bundled together and submitted to the blockchain as a block. Cryptographic techniques are applied to link all blocks in a deterministic order. The cryptographic algorithm also guarantees that the blocks are immutable, which means that once a block is appended to the chain, it cannot be tampered with.

We take Hyperledger Fabric as an example blockchain platform to illustrate how a smart contract service works in the blockchain. Figure 1 shows how a smart contract is deployed on the blockchain. First, the smart contract administrator needs to compile the smart contract application into a binary code so that it can be executed on the Hyperledger fabric. The administrator then deploys the smart contract on the blockchain. The application receives status notifications when there is any change. For instance, when the smart contract is deployed successfully, the application receives a message saying that the smart contract is now running on the blockchain. Finally, end users can access the service through the interface provided by the blockchain.

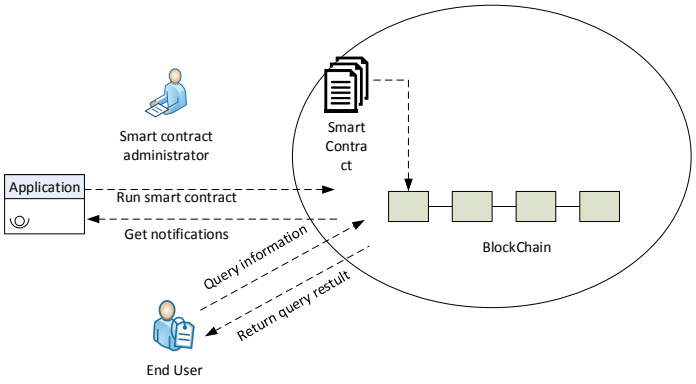


Figure 1. Smart contract on blockchain.

There are typically three different types of consensus protocols. The first is the Proof-of-Work (PoW),[5] adopted by cryptocurrencies like Bitcoin[6] and Zcash.[7] For the PoW protocol, all participants are competing with each other to win the block proposal by solving a specific math puzzle. The first participant to find the solution authorized to propose the transactions in the block and the rest of the participants copy this block to their own chain. A PoW based blockchain can provide an open environment, which allows anyone to join/leave the blockchain freely. However, in this protocol, a high amount of energy is consumed by each participant to solve the math puzzle. As a result, they have a poor throughput. The second type of consensus protocol is based on Byzantine Fault Tolerance (BFT),[8] which is adopted by blockchain systems like Hyperledger Fabric.[9] The size of the network of a BFT-based blockchain is relatively small. As a result, the majority of BFT-based blockchain systems are permissioned blockchains, in which only authorized users can participant in block generation and verification. The last one is the Proof-of-Stake (PoS),[10] which employs a certain number of nodes to generate the blocks on behalf of the whole network. Typical examples of PoS-based blockchain systems are Ouroboros,[11] Neo,[12] and Reddcoin.[13] However, in such systems, rich nodes have more chances to generate blocks. To address such problems, Bitcoin-NG[14] advocates applying a hybrid consensus protocol, which combines the advantages of the PoW and PoS. GHOST,[15] SPECTRE,[16] and MESHcash[17] are recent proposals for increasing the throughput by replacing the underlying chain structure with a tree or a directed acyclic graph (DAG) structure. These protocols still rely on the Nakamoto consensus using PoW. By carefully designing the selection rules between branches of the trees/DAGs, they are able to substantially increase the throughput.[18]

Since the blockchain removes a centralized trusted party and allows a participant to verify the correctness of the data on the blockchain, it is widely applied in two domains—cryptocurrency and the smart contract.[19] The great success of Bitcoin,[6] Litecoin,[20] Zcash,[7] Ripple,[21] and EOS[22] demonstrates the potential market value of the blockchain technology. The smart contract is another application based on blockchain technology. In essence, the smart contract is a service that links different entities together to construct a system to achieve dedicated functions. The approximate turning-complete feature provided by the smart contract allows the majority of existing programs to migrate to the blockchain.

Three key features that the blockchain technology brings to the industry are:

Openness: Trustworthiness is always a key issue for systems that involve multiple parties. All communications between parties are based on a certain kind of trust assumptions. Blockchain technology is so revolutionary that it allows exchanging value directly between parties without them trusting each other. The openness feature allows anyone interested in the system to join and verify the correctness of the data. Because the data on the chain is immutable, it resolves concerns that the data owner might tamper with or modify the data in the future.

Robustness: Denial of service and a single point of failure are common issues for existing centralized systems. If the centralized servers are under attack, the quality of service might be affected and system security could be compromised. However, since every participant holds a copy of the data, and the network size could be large, it is impossible for an adversary to attack the blockchain system by compromising the majority of the distributed blockchain servers.

Cooperation: Blockchain enables a new cooperation pattern among multiple parties in which untrusted parties can exchange data more confidently by hosting the servers locally to construct a blockchain network. For example, take an electronic voting system, when the voting is conducted in a traditional way, all voters and stakeholders should agree on a trusted third party to organize the voting. Blockchain removes this trusted party by allowing all stakeholders to participate in the voting administration. That is, all stakeholders can verify the correctness of the voting result by looking up the data on the blockchain node held by themselves.

There are two paradigms for blockchain resource management: permissioned blockchain and permissionless blockchain. For the permissioned blockchain, a membership service exists that asks all parties who want to contribute in the blockchain maintenance to register with the blockchain system. Hence, only authorized users can access the blockchain. In contrast, for the permissionless blockchain, everyone can access the data on the blockchain and participate in the blockchain management without registering. We make a comparison between permissioned and permissionless systems in Table 1. We briefly describe them below.

Table 1. Comparison between permissioned and permissionless blockchain

	Permissioned Blockchain	**Permissionless Blockchain**
Operational costs	Depends on the redundancy requirements	High (Bitcoin estimate $657,000,000 per year in 2017 at $1000/BTC)
Interoperability	Poor	Excellent
Transaction Throughput	Good	Poor
Data Privacy	Good	Poor
Scalability	Poor	Good
System robustness and resilience	Fair	Good

Permissionless blockchain: The advantages of a permissionless blockchain are: 1) it has an open network to enable anyone to join/quit the protocol freely; 2) the network typically has an incentivizing mechanism to encourage more participants to join the network; and 3) it is suitable for cryptocurrency and applications that do not have strict privacy requirements. However, it consumes a lot of power to maintain the distributed ledger at a large scale for PoW based systems and the trust of the blockchain is hard to achieve for PoS based systems. Furthermore, very limited transaction privacy is preserved since any nodes in the permissionless blockchain can have a copy of all transactions.

Permissioned blockchain: The advantages of a permissioned blockchain are: 1) all blockchain participants are registered and verified by the protocol administrator and as a result, it is easy to identify nodes that do not comply with the protocol; 2) since the public has no access to the blockchain, privacy is preserved; and 3) since the blockchain administrator can control the network size by controlling the number of nodes involved in the blockchain, the permissioned blockchain usually has a high transaction throughput. However, the permissioned blockchain has a number of disadvantages: 1) the public may have low confidence in the correctness of the blockchain because they have no access to the verification of the data on the chain; 2) some stakeholders may collude with each other to make some transactions invalid; 3) participants need to follow a series of strict policies to join/quit the protocol (i.e., a membership server should assign/withdraw its access policy) and 4) some protocols (e.g., PBFT) are based on assumptions that two or three of the total nodes are always online.

CASE STUDY: TRUSTCHAIN – A PLATFORM FOR IOT DEVICE AND DATA TRACKING AND TRADING

In this section, we provide a concrete example of how the blockchain is applied to enhance the trust in a practical scenario to track and trade IoT devices and the corresponding data.

The popularity of wearable devices is increasing and people are regularly upgrading their IoT devices, while few of them understand how to dispose of their out-of-date devices. Trade-in or resale to another customer usually means the potential or accidental transfer of all personal data on the device to the new buyer, resulting in personal data leakage. On the other hand, when a device is put up for sale, the manufacturer needs to trace the ownership of the device to provide a warranty to the correct customer and recall the device if any defects are found. Current faulty airbag recalls on vehicles is a good example—many consumers are not aware that their vehicles have been recalled.

Applying the blockchain technology in the above scenario, we demonstrate how our IoT device and related data tracking and trading system resolves the trustworthiness issue in the IoT ecosystem. Four entities are involved in our system: 1) manufactures, who sell the products to retailers; 2) retailers, who buy the products from manufactures and sell them to customers; 3) customers, who consume the service provided by products; and 4) data analysis companies that buy the personal data for analysis.

For the IoT device and data tracking and trading system, we would like to have the following functions: 1) a manufacturer can trace the status and ownership of the device during its lifecycle; 2) a customer can transfer the ownership of his/her devices to another customer; 3) a customer can share/sell the data generated by his/her IoT devices; and 4) a smart contract that only allows the data owner to sell his/her own data; once the IoT device is sold, he/she cannot access the data generated by that device any longer.

The logical architecture of our tracking and trading system is shown in Figure 2. It demonstrates how cell phones, as IoT devices are transferred from a manufacturer to a retailer, a retailer to a customer, and finally, a customer to another customer. With the help of a smart contract, the manufacturer can transfer the ownership of the cell phone to the retailer. When Alice as a customer purchases this item, a record that corresponds to this purchase is appended to the smart contract, which demonstrates that the cell phone is owned by the customer, Alice. If Alice agrees to sell her cell phone as a used device to Bob, the smart contract can inform the manufacturer that the device is owned by Bob. The manufacturer then provides any remaining warranty service to Bob. Since the device is transferred to Bob, Bob can sell the personal data generated by his cell phone to a data analysis company. At the same time, he has no rights to handle the data generated by the same device previously under the ownership of Alice.

Figure 2 shows the interactions between different entities that are carried out through a smart contract. Since the smart contract is regarded as an independent trusted party, no entity can cheat others or modify the existing data related to this device.

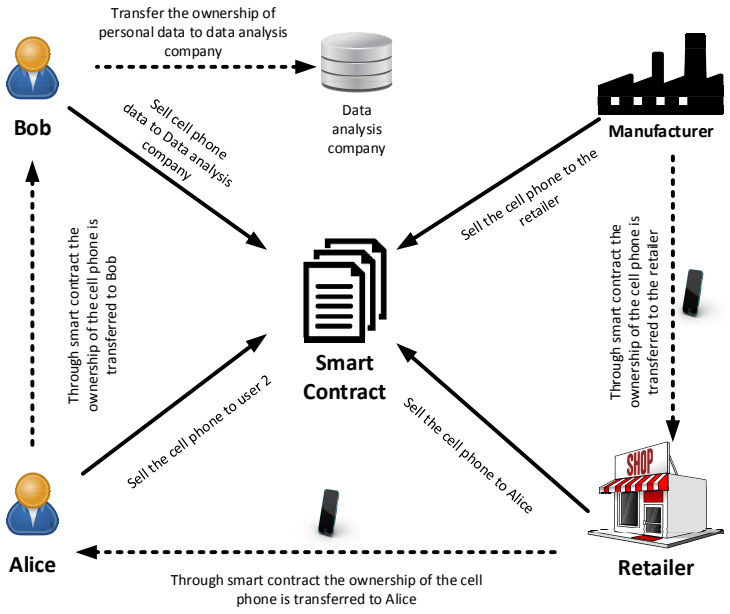


Figure 2. Device tracking and data trading system.

The above case studies can be implemented in both permissioned and permissionless blockchains. Let us first consider the permissioned blockchain. We employ Hyperledger Fabric as a private blockchain that only allows relevant stakeholders to store IoT devices and the data ownership information. Users who own IoT devices can verify the manufacturer of those devices and sell the data generated by them. Since the public does not own the IoT device, they cannot trace any information related to the device. In practice, to preserve the data privacy, we only allow the manufacturer, retailer, and government consumer affairs office to host the blockchain validation nodes.

The logical structure of Hyperledger Fabric is shown in Figure 3. It consists of the following components.

Client: The client represents the entity that acts on behalf of an end-user. It must connect to a peer to communicate with the blockchain. Clients create and thereby invoke transactions.

Peer: The peer receives the ordered state updates in the form of blocks from the ordering service and maintains the state and the ledger. The peer nodes are held by different stakeholders to ensure that the data on the blockchain are verified by all stakeholders to avoid any party tampering with or creating an incorrect block on the chain.

Ordering service: This service provides a communication channel to clients and peers, and offers a broadcast service for messages containing transactions. The channel outputs the same messages to all connected peers in the same logical order.

Certificate Authority (CA) server: The server is responsible for creating user/server certificates and verifying servers' validity in the network. Peer nodes in the blockchain network also ask the CA server to verify the identity of peer nodes.

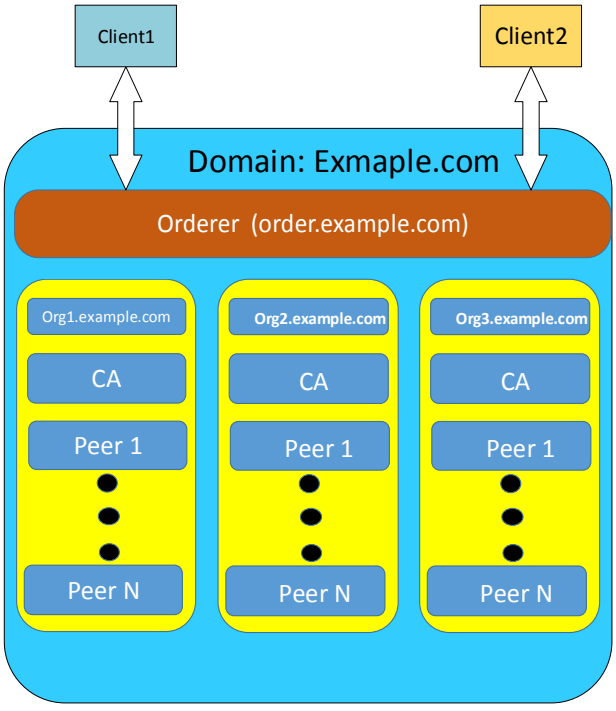


Figure 3. Hyperledger Fabric structure.[4]

SMART CONTRACT LOGIC

In this section, we explain the portal and the interfaces of different entities in our system to illustrate the logic of our smart contract in Figure 4.

The smart contract runs atop the Ethereum blockchain as a form of distributed computation. When the contract is interacted with, either by a retailer or a manufacturer selling/buying a device on the blockchain or by an end user seeking to buy/modify a device, they call one of the functions outlined below. Both inputs and operations of these functions are then computed across the whole blockchain. An individual who invokes the function pays for all nodes in the network to perform that function through the use of "gas." Gas is the execution fee that scales according to the amount of computational power needed to perform the invoked function by all nodes doing the computation. This execution fee is the incentive for nodes in the system to actually perform the necessary computations to ensure the contract is obliged by the blockchain. As the computation is distributed and performed by all nodes in the Ethereum blockchain, the need for a trusted authority to validate the operation is superseded by the use of the consensus protocol. By needing all nodes performing the computation to agree to the output, the ability of bad actors and malicious nodes to tamper with the operation of the contract is entirely negated. However, due to the nature of the contract being executed across the entire network, the contract will not be executed unless the blockchain is sure that the execution will be successful. This prevents wasted operations on the network. The consensus-based nature of this security model means that an attack to the system requires more than 50% of the network to inject any malicious transactions into the blockchain, making it a highly secure decentralized autonomous marketplace (when the continuous expansion and the current size of the Ethereum network are considered).

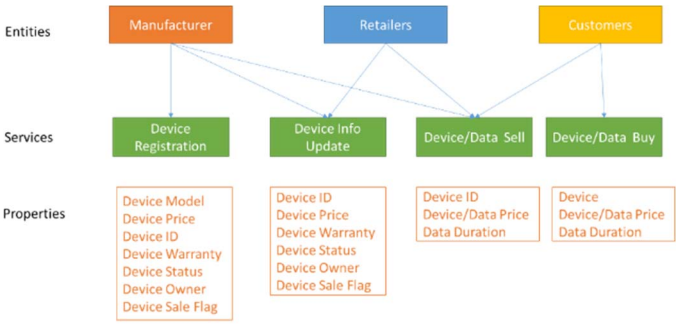


Figure 4. Smart contract logic.

Figure 5 shows an example of how cloud storage can be used. We next describe current implementation of the services in our platform.

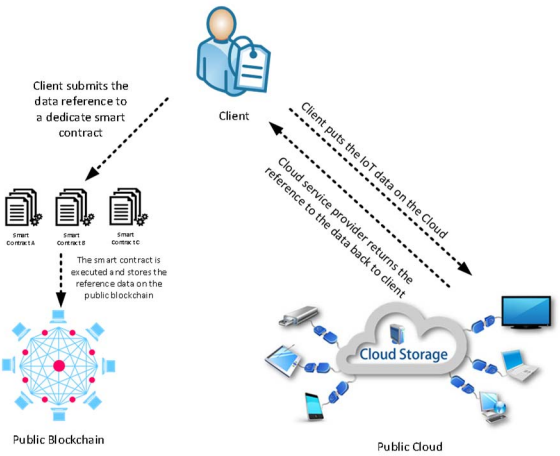


Figure 5. Our current implementation.

Manufacturers are the entity that produce new IoT devices. They have three functions: Registration, Update, and Sale. The registration service is responsible for registering a newly made device onto the blockchain. Actually, it creates device information containing the metadata related to the device including device model, device ID, and device warranty information. Manufacturers also need to update the device state, which indicates the device can be traded on the market. This update service allows manufacturers to update the device information including the device owner, device price, device warranty, and device status. For instance, when a manufacturer finds a specific product has a safety defect and wants to recall the product, it can set the device status to untradeable. Customers need to return them to the manufacturer; otherwise, they lose the rights to trade them in the platform. The sale service allows the manufacturer to put a specific device on the market for sale. A retailer can buy the device from manufacturers. This service can only operate when the manufacturer sets the device for sale indicator as true.

Retailers have the following three functions: Buy, Update and Sell. The buy service enables retailers to buy a specific device from manufacturers with an expected price. Once the agreement is made between a retailer and a manufacturer, the manufacturer transfers the product's ownership to the retailer. The update service allows the retailer to update the device information including the device owner and the device price. This service allows the retailer to set the price for a given device and allows the retailer to transfer the ownership of the device to another retailer or customer when the ownership transferring agreement is achieved. The sell service allows the retailer to put the specific device on the market, enabling customers to buy the device from the retailer. The retailer needs to set a reasonable price for the device before putting it on the market.

Customers also have three functions: Buy, Update and Sell. The buy service enables a customer to buy a specific device or data from a retailer with the expected price. Once the agreement is achieved between the customer and retailer, the retailer transfers the device's or data's ownership to the customer. The update service allows the customer to update the device/data information including the device/data owner, device/data price. The device/data owner can also set the price for a given device/data and allow the device/data owner to transfer the ownership of the device to another customer when the ownership transferring agreement is achieved. The sell service allows the customer to put a specific device/data on the market; thus, customers can buy the device/data from another customer.

FUTURE RESEARCH CHALLENGES

Through the case study, we demonstrated how a blockchain can be applied to develop a trusted device and data trading platform for the IoT ecosystem where different entities can cooperate with each other. However, blockchain is not a panacea to resolve the trust issue in IoT environments. There are some challenges that still need to be studied. We outline some research challenges and potential future works below:

Data Privacy. Currently there is a dilemma/trade-off between public verifiability and privacy. In a trustworthy data trading platform, the trust is established by providing verifiability. The challenge is how to protect privacy of data, device and individuals without losing the verifiability property. One simple solution is to encrypt all data on the blockchain. This might help to address the privacy problem, but the data cannot be verified by other validation nodes. One potential avenue to further this research is use of Zero-Knowledge-Proof (ZKP). However, the computation overhead is often cited as a key problem in using ZKP.

Delegating trust. Another challenge of any effective trust model is trust delegation. In practice, it means how one can practically delegate trust to someone else. For example, Bob brings a device home and he claims/registers it as his device, perhaps with a straightforward method. Bob is the sole person who can control it and is privy to the data it collects. In certain circumstances, Bob may want to give others access to his device. There needs to be a scheme to ensure that operation can be done reliably and Bob has a full understanding of the implications. The challenge is how to support a trust delegation function without violating underlying security and privacy.

Dynamic Access Control. The access control mechanism is widely used to control access to the data. These methods have been directly applied to IoT environments. However, the IoT system is very dynamic and it operates in a context. For example, an emergency doctor might be able to access the IoT health data or the IoT health data becomes accessible from the IoT devices worn by patients when they are in the emergency room. In essence, it is difficult to predefine all potential access control rules. The device itself should have a mechanism to generate access control rules dynamically based on the contextual information. For example, a drunken driver would not be able to start the car. Though there has been some progress in this area, further research is needed to build a reliable dynamic access control mechanism for IoT applications.

IoT device identification. For the data and device trading platform to function properly, IoT devices that are trusted need to be identifiable. This requires an easy to use identity management system to be made available for all IoT devices at all times. However, the identity management systems (such as username/password pairs, and X.509) are invented for general purpose identification and are thus inadequate and rarely address many known issues that exist in the IoT environment. There is no defined and accepted standard for device identification management in the IoT environment. One of the key features of such a device identification mechanism should be automatic discovery of devices. The challenge is as soon as the device is powered and in operation, it would be discoverable (without violating the underlying security and privacy).

Human-centric trust model. Human-centric trust models are another research topic, which means a trust model aimed at giving effective administration of security and privacy not to computing professionals, but to average users. This is a cross-cutting topic to all of the challenges stated above. For example, a human-centric trust model can be designed for people to sensibly delegate the controls of data and device to others with full understanding of security and privacy

implications. The goal of a human-centric trust model is to let the service itself evaluate the security risks and apply the security policies according to the potential attack; thus, an average person can enjoy the same device security levels as security professionals.

Developing a holistic benchmarking kernel. Understanding performance bottlenecks of blockchain-based large scale IoT application systems remains a challenge, hence it is useful to identify benchmark kernels that are relevant for testing particular aspects (e.g., overlay networking, consensus protocol, querying) of the blockchain. Existing benchmarking literature in the context of IoT systems is limited as they are largely focused on studying the scalability of data processing programming models. For instance, the benchmark (kernels) that are available in context IoT systems focus on following data processing programming model aspects (not applicable to benchmarking performance of blockchain):

- **Edge layer.** TPCx-IoT (for data aggregation, real-time analytics & persistent storage), Google ROADEF & Linear Road benchmarks (for stream processing).
- **Cloud layer.** TeraGen, TeraSort, TeraValidate, and BigDataBench (for batch-oriented processing).

Hence, creating benchmark kernels that can test different aspects of the blockchain, and more importantly, identify performance bottlenecks and dependencies need more attention from the research community.

Scalable Search and Communication. Developing a scalable protocol for searching data blocks and smart contracts within a large scale blockchain network remains a challenge. Existing search and consensus communication protocol adopted by the state of the art blockchain networks are based on a complete broadcast routing algorithm. Drawbacks for broadcast-based routing include high network communication overhead and non-scalability. Hence, the future investigation will need to develop scalable methods for search and consensus communication protocol. To this end, one possible research direction can be to interconnect peer nodes in the blockchain network based on Distributed Hash Tables (DHTs) overlay. In contrast, to broadcast based peer-to-peer systems, DHT overlays (e.g., Chord, Pastry, and Tapestry) have been proven to be more scalable as the size of the network grows.

Solutions to these research challenges will form a building blocks and contribute to components required for building a trustworthy data and device trading platform. We believe the blockchain technology is key to finding appropriate solutions to these challenges as it provides a basic foundation for trust in an untrusted, distributed IoT environment.

CONCLUSION

Many our lives are influenced by IoT-driven data science applications, ranging from precision health/medicine to self-driving cars. While enjoying the convenience and benefits brought by IoT devices, we must also face the challenge of trusting such IoT ecosystems. Traditionally, a trusted party performs a supervisory role in data collection and analysis. Blockchain, which is designed to remove the trusted third-party in a decentralized system, is an ideal solution to resolve the trust issue in IoT ecosystems. With the help of blockchain, different parties can trust and verify the data. Additionally, the ownership of IoT devices and their related data can also be traced. Though the blockchain can resolve the trust issue, it is not the panacea to every IoT challenge. A number of research challenges need to be addressed including data security and privacy on the blockchain, trust delegation, device identification, discovery, and authentication.

REFERENCES

1. R. Ranjan et al., “The Next Grand Challenges: Integrating the Internet of Things and Data Science,” *IEEE Cloud Computing*, vol. 5, no. 3, 2018, pp. 12–26.

2. Z. Whittaker, "US cell carriers are selling access to your real-time phone location data," *ZDNet*, 14 May 2018; www.zdnet.com/article/us-cell-carriers-selling-access-to-real-time-location-data.
3. J. Voas, *Network of 'Things'*, NIST Special Publication 800-183, NIST, 2018; https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-183.pdf.
4. B. Celler et al., "Impact of At-Home Telemonitoring on Health Services Expenditure and Hospital Admissions in Patients With Chronic Conditions: Before and After Control Intervention," *Journal of Medical Internet Research*, vol. 5, no. 3, 2017, p. e29.
5. A. Gervais et al., "On the security and performance of proof of work blockchains," *Proc. ACM SIGSAC Conference on Computer and Communications Security* (CCS 16), 2016, pp. 3–16.
6. S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008; https://bitcoin.org/en/bitcoin-paper.
7. D. Hopwood et al., *Zcash protocol specification*, technical report Tech. rep. 2016-1.10, Zerocoin Electric Coin Company, 2016.
8. M. Castro and B. Liskov, "Practical Byzantine fault tolerance," *Proc. 3rd Sym. Operating Systems Design and Implementation* (OSDI 99), 1999, pp. 173–186.
9. C. Cachin, "Architecture of the Hyperledger blockchain fabric," *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, 2016.
10. S. King and S. Nadal, "Ppcoin: Peer-to-peer crypto-currency with proof-of-stake," August 2012.
11. A. Kiayias et al., "Ouroboros: A provably secure proof-of-stake blockchain protocol," *Advances in Cryptology – CRYPTO 2017*, Katz J., Shacham H., Lecture Notes in Computer Science, vol. 10401, Springer, 2017.
12. *NEO white paper*, white paper, Neo, 18 May 2018; http://docs.neo.org/en-us/.
13. *REDDCoin*; https://reddcoin.com/.
14. I. Eyal et al., "Bitcoin-NG: A Scalable Blockchain Protocol," *Proc. 13th Usenix Conf. Networked Systems Design and Implementation* (NSDI 16), 2016, pp. 45–59.
15. Y. Sompolinsky and A. Zohar, "Secure high-rate transaction processing in bitcoin," *Financial Cryptography and Data Security*, Böhme R., Okamoto T., Lecture Notes in Computer Science, vol. 8975, Springer, 2015.
16. Y. Sompolinsky, Y. Lewenberg, and A. Zohar, "SPECTRE: A Fast and Scalable Cryptocurrency Protocol," *IACR Cryptology ePrint Archive*, 2016, p. 1159.
17. I. Bentov et al., "Tortoise and Hares Consensus: the Meshcash Framework for Incentive-Compatible, Scalable Cryptocurrencies," *IACR Cryptology ePrint Archive*, 2017, p. 300.
18. Y. Gilad et al., "Algorand: Scaling byzantine agreements for cryptocurrencies," *Proceedings of the 26th Symposium on Operating Systems Principles*, 2017, pp. 51–68.
19. V. Buterin, *A next-generation smart contract and decentralized application platform*, white paper, 2014.
20. *LiteCoin*, 2018; https://litecoin.com/.
21. *Ripple*, 2018; https://www.ripple.com/.

ABOUT THE AUTHORS

Bin Yu is a PhD student at Monash University/Data61 CSIRO. Contact her at bin.yu@monash.edu.

Jarod Wright is an industrial trainee at Data61 CSIRO. Contact him at Jarod.Wright@data61.csiro.au.

Surya Nepal is a principal research scientist at Data61 CSIRO. Contact him at Surya.Nepal@data61.csiro.au.

Liming Zhu is a research director at Data61 CSIRO. Contact him at Liming.Zhu@data61.csiro.au.

Joseph Liu is a senior lecturer at Monash University. Contact him at Joseph.Liu@monash.edu.

Rajiv Ranjan is a chair professor in Computing Science and Internet of Things at Newcastle University. Contact him at raj.ranjan@ncl.ac.uk.

Trust Erosion: Dealing with Unknown-Unknowns in Cloud Security

David A. Maluf
Cisco Systems

Raghuram S. Sudhaakar
Cisco Systems

Kim-Kwang Raymond Choo
University of Texas at San Antonio

Editor:
Kim-Kwang Raymond Choo
raymond.choo@
fulbrightmail.org

Although today's average cloud computing environment may incorporate security in most aspects of its design and infrastructure, the mere operation of the network exposes it to attacks. A typical attack starts with probing for weaknesses and/or vulnerabilities that can be exploited. And it is at this stage that the battle seems to be already lost, as the average network is insufficiently equipped—mostly for economic reasons—to even know that they are under probing, let alone thwart an attack. In many cases, cloud systems are caught unaware of situations where friends turn into foes, nullifying established security measures. Threats will always dwell on new (previously unknown) methods to compromise established security measures (i.e., a rat race between defenders and attackers, particularly well-resourced attackers). These methods largely fall outside the adapted models used by current security measures that protect cloud-based systems. After-the-fact analysis has driven security researchers to extend models to include assumptions about newly discovered threat(s). Solutions are then designed to deter these new threats. These models may also be generalized with additional measures mapping futuristic predictions—these are also referred to as known-unknowns.

In a world of finite resources and time, cloud security aspects of a network will follow the security aspects of any system from the point of view of the perceived *Risk* at the time of design or deployment. In any case, systems are only guaranteed up to a certain level. In other words, these guarantees are limited to the amount of accepted risk. As in most risk assessments, risk is decomposed into terms of likelihood and impact: $Risk = Likelihood \times Impact$.

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD

The Likelihood

One would come to the realization that cloud security (and cyber security in general) shortfalls can be categorized as either unaddressed *known* threats or as *unknown* threats (as part of the intriguing category of what is unpredictable). Indeed, the unpredictable class is the infamous *unknown-unknown* type of problems that cripple many decision aspects of the industrial world beyond cyber security.

One could argue that many security breaches have been of the *unknown-unknown* class of problems, or otherwise due to negligence or badly managed risks. The only approach, so far, to address or analyze these *unknown-unknowns* has been after the fact. Basically, one waits until something happens before responding, a cliffhanger episode for any IT department. After-the-fact analysis of the incidents would lead, on many occasions, to the identity of the relevant variables, depending on a range of factors such as the sophistication of the attackers. The formalism of subsequent digital forensic outcome, if any, is the basis of the *known-unknown* category.

For clarity, the likelihood is a product of the *known-unknown* category. For many, the likelihood (e.g., probability) is the science that identifies the *known*-part from the past to present, and the -*unknown* part is the art of predicting any component of the *known*-part into the future. In context of the *unknown-unknown* type problem, likelihood is foreign to that concept. U*nknown-unknown* has been a fascinating, often referred to, but rarely touched upon scientifically.

The Impact

Fundamentally, a threat to a system exists only if the system in question exists. The existence is then quantified with an associated value. This notion is known among economists and is typically defined as the utility value of the system. The utility value could be a national security index or the common agreed value such as US dollars in an industrial world. Both the national security index and common agreed value are two examples of contextualizing the impact. As a consequence, when the impact of the threat is fully materialized, it is quantified to be limited by the utility value of the system it threatens. Simply put, the maximum threat to a $1 dollar investment is $1 whereas the maximum threat to $1B is $1B.

There are two outcomes when analyzing the impact of cyber security: First, the impact of cyber security is relative. Second, the impact itself is governed by the *risks* and grows exponentially with the size of underlying system, its utility, complexity and most importantly time.

The Risk

Many standards govern the computation of risk. For example, in a popular approach, considering an economic angle, risk is accepted when it is below a certain threshold. In other words, assets can be sacrificed as long as the impact of the loss is below a threshold.

Cyber security is in a unique niche and has wedged itself to be an essential part in the world of economics. While economics is a mature field, in terms of risk management, the economics of cyber security is an ad hoc science, yet it plays a fundamental role in risk management.

The status quo of cloud and cyber security hinges on this fact. The economics of addressing the unknown unknowns has been perceived to be above the inflection point for utility.

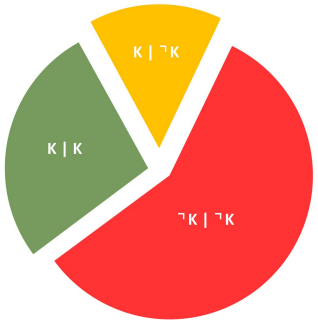


Figure 1. Conceptual delineation of the information segregation {k | k}, {k| ¬k}. Degradation in trust (increased risk) is proportional to the ratio of the known and unknown

TRUST EROSION

A distinction between digital assets (e.g., digital intellectual property) and physical assets (e.g., a rover on Mars) is the capitalization and cost of opportunity. While a rover would cost $1 billion USD in capitalization and many years to make, digital assets will far exceed their value proposition, risks and liabilities included, because of the premise in the economics of digitization. Cloud security and CYber INTelligence (CYBINT) differ at this level because CYBINT has a substantial upfront capitalization. The premise of the said maximization is solely dependent on the interpretation of *impact.* Obviously, the meaning of impact differs for both use cases. CYBINT impact is the nation's worth at risk, whereas the impact of a loss in the cloud service provider is its capitalization market value. The volatility in the digital asset context dictates the purpose of trust; which is to model such a ratio in the digital world that is less volatile and more physically defined in the analog world.

Security systems can be readily built once risk factors are understood. The fundamental assumption, however, is that the unknown-unknown class consists of risks that are not known or understood. Indeed, the nebulous nature of the problem limits the choices of solutions.

To avoid the pitfall of yet another new security system, it is essential to assert foundational parameters attributed to a system and its functioning. One can easily identify size (e.g., infrastructure), complexity, and utility as fundamental characteristics of the system. When analyzing the unknown-unknowns, these characteristics, in three dimensions, define only a sample or possibly a starting point. The time dimension creates the most significant impact to the risks posed to the system. It is imperative to note that the risk from unknown-unknowns monotonically increases over time. Further, the increase of potential risk is not linear with time because it would mean that the unknown-unknowns remain the same, contradicting the initial assumptions.

This means that given any complexity model, and regardless of whether it is growing or shrinking, the unknown risk will not go away but will follow an exponential growth curve over time if left unchecked.

Examination of Figure 1 determines that the total area grows exponentially on the premise of $\{\neg k \mid \neg k\}$. Both $\{k \mid k\}$ and $\{k \mid \neg k\}$ shrink at the same rate of the lack of new knowledge, $\{\neg k \mid \neg k\}$ grows. In layman's terms, the *cost of ignorance over time* is not forgiving, and can have significant impact on the organizations' bottom-line.

The growth of {¬k | ¬k} is non-linear, but the assumption made of its growth being exponential is fair.[8] It can also be easily argued that a linear or sublinear growth of the {¬k | ¬k} means that the system is of marginal utility or value. The choice of the exponential growth factor however, is determined from continuous attempts to learn from the data and the continuous deviation from what was learned. The growth of the unknowns in a system can be tracked by differentially measuring the foundational parameters of the system. Even if the underlying system function is not modelled, a measure or the rate of change of these parameters can be extremely informative. In fact, it will asymptotically follow the rate of growth because the growth naturally obeys the underpinning economics and the system complexity.

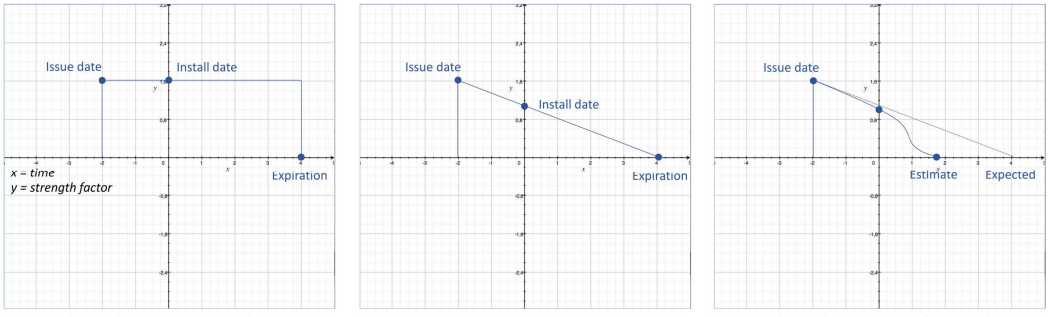
TRUST EROSION AS A SUM OF DEVIATIONS

The concept of {k | k}, {k | ¬k} and {¬k | ¬k} is very challenging to represent mathematically. One can model the { k | k} framework using set theory, and model {k | ¬k} using probability theory. On the other hand, {¬k | ¬k} is boundless and unknown by definition.

To this end, we suppose the underlying cloud security complexity growth follows the Solow economic model,[9] and therefore, the risk will follow a similarly exponential growth. In fact, the unknown-unknown growth factor can be readily assessed asymptotically from the actual data as a differential analysis of the expectation and deviation of the foundational parameters. The value proposition lies in how these differential analyses are made.

One can define a *Trust Erosion Factor* T_F as the exponent factor of all materialized and unmaterialized risks (i.e., deviations in the measured parameters). The Trust Erosion Factor (TEF) is a function over time and the growth (in both directions). It is trivial to understand that with a stale risk (i.e. a risk that has zero deviations), the trust in the system will erode over time.

Until there is a change to the underlying system, the unaccounted erosion is an increasing function over time (see Figure 2). It is notable that most systems do not rely on a utility function but some fixed assumptions made in time. One can further the assertion about the concept of a continuous TEF utility value, that when factored back into the system model, it reflects a degradation to the known assumptions and unknown modeled on a tendency of the growth.



Trust degradation as a step function | Trust degradation over time | Trust degradation over data and time

Figure 2. Erosion may occur at a different rate than assumed. Degradation is way more dynamic than a step function as used in the state-of-the-art.

AN INFORMATION THEORY BASED APPROACH TO UNKNOWNS

Trust Erosion is defined as the ratio of the impact, both perceived and hypothetical, of the unknown over what is known. While there is no clue whatsoever about the unknown, the impact is estimated to follow an exponential growth over time. Therefore, the known will decline in the same order of scale. The break-even is reached when learning is fast enough to catch up with the rate of the growth of the unknown. In truth, cloud systems are far from being adaptive and dynamic enough as they ought to be in today's total dependency on a digital world.

When these concepts are formalized into a Bayesian framework, they take the form of a time series likelihood optimization.[3] The framework is a continuous differential analysis comparing measurements to a *posterior* of a likelihood estimation. The resulting error or deviation is the most informative gain.[2] The deviations (unknown) are marginalized (unmodeled) for being highly non-linear.[7] It becomes imperative to solve for their magnitude, or $\|\{\neg k \mid \neg k\}\|$, as an information measure. It is not necessary to have models of the deviations to make decisions. In many cases, high entropy data—as it may get closer to random—are mathematically complex as it may turn to be highly nonlinear and close to noise in dimensionality. Measuring the magnitude (information content) is a simpler task and can quantify uncertainty. The measure of uncertainty is computed using the Shannon Information Theory.

Naturally, the proposal lends to finding the maximum bit encoding measures as information. This is readily available from the data without the need to decode the data at all time. Independence is achieved across platforms, and the information index measure spans all data, whether encrypted or not. The magnitude of the uncertainty is entropic and will be addressed with corresponding information theory mathematical measures.

Optimally sampling in the uncertainty space coincides with the Maximum Entropy Sampling theory. This lends itself to maximizing the information and minimizing the entropy for non-stationary processes,[4] which coincides with the nature of sampling the unknown. Sampling for deviations is the same problem as sampling points on a multidimensional function and is a stochastic process. The use of Maximum Likelihood Estimation (MLE) and Maximum Entropy Sampling (MES) approaches to maximize information follows the typical implementation of these theories.[1,5,6]

In a practical example, measuring the weight of a bucket is informative to a user who has developed a sensitivity to weights without having to sort out what is in the bucket. Comparing bucket weights sampled in different locations will become informative about underlying topology changes, or the deviations in a cyber security case. A more popular use case than "bucket comparison analysis" is the science behind Searching for Extraterrestrial Intelligence (SETI) computational algorithms, even though "intelligence" is not a well-defined concept.

The science behind building a TEF score, roughly speaking, is when applying Shannon's theorem to the information content carried by the deviation, one can deduce that any transformation scheme cannot, on average, have more than one bit of information per bit of deviation. This means that any value less than one bit of information per bit of deviation can be attained. The entropy of a deviation per bit multiplied by the length of that deviation is a measure of how much total information the deviation contains. Therefore, the maximum amount of information is calculated as the entropy of all the deviations.[10,11,12]

CYBER WARFARE

Warfare is the oldest school ever known to the human civilization and precedes literacy. Fundamentally, cyber warfare is not much different. An example is digital secret keys are changed periodically; which mimics an old strategy of moving the residency of a king every now and then. The question is when to move the king's residency precisely, versus a periodic strategy. Even the knowledge of the period, if leaked, may be detrimental on its own. Protecting the key and the king is very similar from a strategy perspective.

The analysis of the unknowns in cyber security is analogous to the analysis of the ingress and egress behaviors made at the gates of a city. Historically, travelers were denied passage when the travelers' ingress exceeded their egress. This made the generals evaluate the risk ratio, first, by the traveler count, and later, by the maximum volume and capacity of their transports (reflecting disguised materials), against an existing defense system. The formula was simple: assess the worst-case scenario given the ingress minus egress. Optimal reasoning mimicked the maximum entropy assessment of a threat if all travelers and all their transports constituted a threat. Inversely, reasoning with confidence intervals, or less than the maximum, is a limited system. An example for such a model is limiting the threat to men of an age range. Cyber security models today mimic such models, which have stated assumptions and set limits to deter threats. However, they fall short as attacks have been successful in finding digital concealments and workarounds. An interesting observation is that cities were built as hubs and worked to maximize the control over the transport mechanism. Today's cloud is not much different.

The maximum likelihood (paranoid approach) has both advantages and disadvantages. The maximum likelihood penalizes the optimization at some cost (impact). With the latter example, the impact could be represented as a trade□off analysis between trading economic loss against safety.

The objective in the analogy of the ingress and egress analysis over known physical quantities, such as volume and weight over time, serves the purpose of distinguishing the separation of the

quantitative analysis (the data) from its qualitative counterpart (the models). The qualitative analyses from the analogy are numerous, such as safety threats, illegal trafficking, etc. The physical measurements do not change.

Not surprisingly, this methodology is still an optimal approach used in modern warfare, social and economic behaviors. Military checkpoints will report the ingress and egress quantitative measures regardless of the qualitative assessment to Command and Control (also referred to as C2). In unperfected warfare methods, decisions are qualitatively made at checkpoints.

The digital world is not different from the physical one, digital I/O equates to the ingress and egress in physical dimensions.

The clear distinction, as in many analytical threat systems, is the premise that there is no knowledge about the threat, the enemy. This applies to every aspect underlying the {¬k |¬k} assumptions. However, the assumption that "the risk has a growth rate, which is measurable" is a fair deduction, without explicit knowledge of the risk.

TRUST EROSION EXAMPLE

If we could narrow down cyber security challenges to one thing, it might summarize into chasing the unknown gaps. One prominent gap is the way of handling digital secrets over time, or the secrets' lifecycle. If we think for a moment about the digital asset lifecycle, we realize that the dream of the immortal bits has taken over the reality. What that means, is that while some digital assets may live forever, their intended trust factor will degrade with time and usage.

The digital asset lifecycle establishes a framework to protect digital assets from their design and use intended to be applied in an automated and proactive manner. A comparison of the digital asset lifecycle to a physical system is insightful. A physical system pertains to wear and tear and deteriorates over time and usage. For example in the past, one would change the vehicle oil once a certain distance was travelled. Nowadays, modern vehicles are equipped with sensors that quantify more precisely the oil degradation level.

In general, systems may pertain to many types of digital secrets, such as private keys, symmetrical (session) keys, password databases, and hash keys. We can start with what is the most valued secret asset in a system—the private key. To date, private keys are statically deployed; and once deployed, many assumptions are made at a point in time.

In reality, it has been intimidating from a practical sense to replace decades of practice in key management. The intent is to track statistically the made assumptions augmenting the management with a continuous monitoring for individual asset levels. While secrets may remain untouched, their functions are assessed in a statistical sense over time and usage. That means the success of a secret does not solely depend on the strength of the key, but also how it is used and for how long it used.

Indeed, as for the vehicle analogy, advanced onboard sensors turn out to be very good indicators of the vehicle components' health. A digital monitor for a digital asset would do the same. In the scope of cyber security, research have shown for example that a watchdog for a key usage can determine in real□time the degradation of the secrecy of the keys exposing statistical leakage. The Digital Erosion Factor (DEF) and TEF are two proactive cyber□security strategies for the statistical leakage framework, a totally orthogonal field to the state of the art in cryptography.

Maintenance of the key carries the process of changing keys—this is deterministically defined when the maintenance follows a fixed time period. In parallel, security products have followed an evolutionary strategy of patching, and therefore, the system overall weakens over time. Failure modes become opportune times when many weaknesses of security systems align. For example, modern authentication may appear to increase the overall authentication but it may hide the growing weakness of user passwords. In other words, the keys may have already been compromised but because the complete authentication process is still effective, the compromise itself may remain dormant and thus unknown.

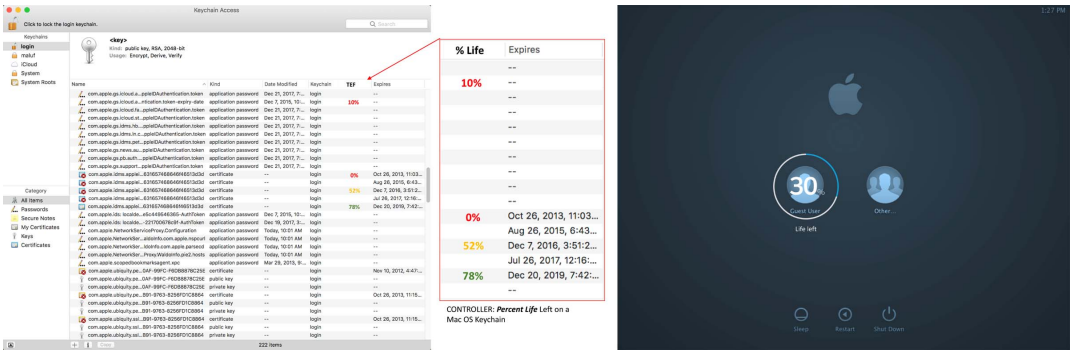


Figure 3. Example of TEF score for normalized trust of cloud and personal computer logins. A keychain may reflect the trust score of a secret. One secret with a low score is a backdoor to another secret with a higher score. Online cloud and personal computer logins become adaptive to usage versus a fixed period of time when passwords expire.

The convention of secrets being replaced or upgraded periodically, in a fairly arbitrary fashion, is a backdoor to the new layers of security solutions. Furthermore, digital asset protection in security revolves around designing very complex secrets; therefore, it stops short in deterring failure modality beyond the threat of deciphering the secrets.

Secrets are typically designed with a difficulty level that reflects an index of randomness and complexity. With the diverse types of configurations in security setups, it is challenging to quantify an index strength for original setups to start with. Typical usage and conditions on the difficulties in many cyber security instantiations are reflected in the secrets' expiration date. Expiration dates are derived analytically from the secrets' known states in certain usage conditions. To the extent of this paper, a user's passphrase is one type of private key where a typical user is familiar with the strength during the creation process.

CLOUD AS AN AGGREGATION POINT FOR TRUST

Trust degradation is a precursor index to failure. The use-cases of scoring the trust degradation in a system span to almost every aspect in networking, edge and cloud included. A well-devised TEF will cover many use cases such as: (i) better and adaptive private key management (e.g., rekeying); (ii) better and adaptive end-user experience password management and its fine grain monitoring in a data center; (iii) better and adaptive digital asset certifications; (iv) troubleshooting; and (v) real-time scalability and risk assessment for extremely large networks, for example in federated cloud environments.

The features of a digital trust scoring will start to reflect the likelihood of erosion of trust created on Day 0. Platform independency is achieved when the score is a degradation of the trust and not the trust value alone. A trust value may be erroneous at the beginning, but the rate of change should lead to continuous evaluation. What that means, the originating trust is set as *a priori.* Thus, erosion is a function of time against the assumed original trust. In the example of an expiration date or a combinatorial complexity erosion of a private key, the realization of a trust erosion is not a Boolean fail pass type, but a relative factor number. On a comprehensive integrated analytical dashboard, the trust factor produces the *percent life left* of given a digital secret (see Figure 3).

Either way, *scoring* can be segmented to atomic assets, whether per category or subcategory, and attributes of a system. A generalization can be reached as the information entropy addresses maximum information compressibility. The maximum information calculation covers encrypted and unencrypted data when a network is at play. It is understood that network topology change is also captured through time sensitivity.

The outcome with such a setup at large scale is a comprehensive analysis when coupled with monitoring tools. With a passive watchdog, continuously measuring trust is a noninvasive technique promoting advanced classification techniques. Figure 4 illustrates a mass scale of erosion

propagation. A C2 dash board plots the behavior in real time. From Figure 4, a threat may be expanded to diverse cyber security mongering (e.g., geo-political) beyond malware and advanced persistent threat.

An earlier column[13] posited the importance of ensuring systems are designed to be forensically-ready (i.e., forensic-by-design). The concepts proposed in this paper could be readily extended to build a secure and stand-alone digital black box (akin to black boxes in a plane), where artifacts or data of forensic interest are being collected automatically and stored securely in the black box to facilitate future investigation (e.g., after-the-fact analysis).

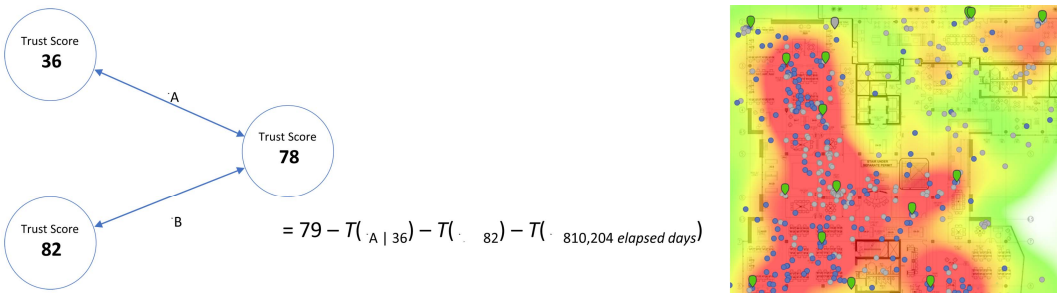


Figure 4. Example of TEF score propagation. Right side: A heat map illustrating trust erosion when a threat mongering is acting out behind a firewall on multiple Access Points degrading the trust of users. .

REFERENCES

1. H.P. Wynn, "Maximum Entropy Sampling and General Equivalence Theory," *mODa 7 — Advances in Model-Oriented Design and Analysis*, Di Bucchianico A., Läuter H., Wynn H.P., Contributions to Statistics, Physica, Heidelberg, 2004.
2. J. Sacks et al., "Design and Analysis of Computer Experiments," *Statistical Science*, vol. 4, no. 4, 1989, pp. 409–435.
3. N. Youssef, *Optimal Experimental Design for Computer Experiments*, dissertation, London School of Economics, 2011.
4. C.J. Paciorek and M.J. Schervish, "Spatial Modelling Using a New Class of Nonstationary Covariance Functions," *Environmetrics*, vol. 17, no. 5, 2006, pp. 483–506.
5. P. Sebastiani and H.P. Wynn, "Maximum Entropy Sampling and Optimal Bayesian Experimental Design," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 62, no. 1, 2000, pp. 145–157.
6. M.C. Shewry and H.P. Wynn, "Maximum Entropy Sampling," *Journal of Applied Statistics*, vol. 14, no. 2, 1987, pp. 165–170.
7. R.R. Zhou, N. Serban, and N. Gebraeel, "Degradation Modeling Applied to Residual Lifetime Prediction Using Functional Data Analysis," *Annals of Applied Statistics*, vol. 5, no. B, 2011, pp. 1586–1610.
8. R. Brincker, L. Zhang, and P. Andersen, "Modal Identification from Ambient Responses using Frequency Domain Decomposition," *Proceedings of the 18th International Modal Analysis Conference* (IMAC), 2000.
9. R.M. Solow, "A Contribution to the Theory of Economic Growth," *The Quarterly Journal of Economics*, vol. 70, no. 1, 1956, pp. 65–94.
10. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, D.L. Schilling, Wiley Series in Telecommunications, Wiley, 1991.
11. P. Sebastiani and H.P. Wynn, "Bayesian Experimental Design and Shannon Information," *Proc. of the Section on Bayesian Statistical Science*, 1997, pp. 176–181.
12. K. Chaloner and I. Verdinelli, "Bayesian Experimental Design: A Review," *Statistical Science*, vol. 10, no. 3, 1995, pp. 273–304.
13. N.H. Ab Rahaman et al., "Forensic-by-Design Framework for Cyber-Physical Cloud Systems," *IEEE Cloud Computing*, vol. 3, no. 1, 2016, pp. 50–59.

About the Authors

David A. Maluf is a distinguished engineer at Cisco Systems. Formerly, he worked at NASA with different leadership capacities in R&D for over 12 years. Before joining Cisco, he was the founder and CTO for two high-tech companies, which have been both successfully acquired. Maluf received a PhD in electrical engineering from McGill University. Contact him at dmaluf@cisco.com.

Raghuram S. Sudhaakar is a technical leader at Cisco Systems and currently leads engineering for Cisco's connected car business. Previously, he held different engineering leadership and R&D positions in the IoT Business and CTO office at Cisco systems. He received a PhD in computer science from State University of New York at Buffalo. Contact him at rsudhaak@cisco.com.

Kim-Kwang Raymond Choo holds the Cloud Technology Endowed Professorship in the Department of Information Systems and Cyber Security at the University of Texas at San Antonio. His research interests include cyber and information security and digital forensics. He is a senior member of IEEE, a Fellow of the Australian Computer Society, and has a PhD in information security from Queensland University of Technology. Contact him at raymond.choo@fulbrightmail.org.

BIOMETRICS-AS-A-SERVICE: CLOUD-BASED TECHNOLOGY, SYSTEMS, AND APPLICATIONS

Silvio Barra
University of Cagliari

Kim-Kwang Raymond Choo
University of Texas at San Antonio

Michele Nappi
The University of Salerno

Arcangelo Castiglione
University of Salerno

Fabio Narducci
University of Naples "Parthenope"

Rajiv Ranjan
Newcastle University

The guest editors of the *IEEE Cloud Computing* special issue on Biometrics-as-a-Service discuss the benefits and challenges of using cloud computing with biometric authentication systems as well as the articles included in this issue. Three potential research topics are also discussed, including the use of machine/deep-learning techniques to circumvent existing biometric authentication solutions.

Interest and use of biometric authentication systems in cloud services continue to increase, partly because biometric credentials are harder to compromise in comparison to conventional password-based authentication. However, a number of challenges exist in deploying biometric authentication systems in cloud services (e.g. *function creep*-prone: gradual broadening of technology or system usage beyond the purpose for which it was originally intended). This special issue reports on state-of-the-art advances on this topic.

WHY THE NEED FOR BIOMETRICS IN THE CLOUD?

Cloud computing[1] is widely used in both scientific and business activities, as well as by individual users.[2] From the hardware infrastructure perspective, cloud computing helps to overcome

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

limitations in standalone computing. For example, organizations do not need to make significant investment in their computational processing and storage infrastructure, and can progressively scale up or down as needed. This drives down the cost in terms of hardware and ongoing maintenance supply, which is particularly crucial for small- and medium-sized organizations, and results in an entirely new ecosystem. For example, we now have organizations dedicated to the provision of infrastructure to manage cloud platforms and lease their resources to users based on their needs, and providers who pay for the rented resources from one or many infrastructure providers to serve other users.[3] Major cloud computing organizations include Google, Amazon, Microsoft, and Alibaba.

Challenges associated with the deployment and utilization of cloud services have been widely discussed in the literature. For example, ensuring the privacy of data while providing timely and secure access in a cloud computing environment, particularly in a federated or multi-cloud environment, can be extremely challenging.[4] This is partly due to differing privacy and related regulations and requirements on the management and storage of data between jurisdictions.[5,6] This necessitates the design of authentication system that ensures that data can only be accessed by authorized users.

Here, Biometrics-as-a-Service (BaaS) is a potentially attractive solution to providing ubiquitous authentication to cloud services. With BaaS, a service provider can offer a light way of accessing data, based on an individual's biometric traits (like fingerprint scanning or facial recognition); thus, mitigating potential fraudulent activities and streamlining customer service, without costly, time-consuming and resource-intensive software acquisition and integration processes. The potential for BaaS is also evidenced by recent services (i.e., biometric recognition to be used as a service on the cloud) offered by Fujitsu, BioID, ImageWare Systems, Animetrics, Aware and IriTech. Thus, this is the focus of this special issue.

Cloud-based biometric authentication (also referred to as biometrics-as-a-service) is a relatively new trend, replacing conventional password-based authentication system.[7] Using biometrics as a way of authentication on cloud computing architecture has potential benefits, such as scalability, cost-effectiveness, reliability, hardware agnostic, and allowing ubiquitous access to private data and services. In fact, biometric credentials have the advantage of not relying on the user's memory.

Existing biometric authentication literature generally focuses on how to acquire and/or process biometric traits for reliable recognition. Generally, biometric data (iris or face scan, fingerprint and so on) is captured during enrollment and converted into metadata (templates) for storage. User authentication takes place at a later stage by a matching process between the live acquired trait and previously stored template.

While biometric authentication systems offer a number of benefits over conventional password-based authentication systems, such systems are not perfect. For example, one's biometric traits cannot be replaced once they have been compromised. Hence, ensuring the secure storage of biometric traits is crucial but not sufficient, since biometric traits transmitted over public networks could be copied and exfiltrated by an eavesdropper. In addition, it has been demonstrated that using a fingerprint template rather than the original image does not guarantee the user's privacy.[8]

IN THIS SPECIAL ISSUE

The first contribution to this special issue, co-authored by Yang et al., is entitled "Tensor-based Big Biometric Data Reduction in Cloud," in which the authors proposed a biometric data tensor reduction solution for the cloud computing environment. Specifically, they model big biometric data using a tensor-based representation and use tensor decomposition techniques to achieve multidimensionality reduction of the big biometric data in the cloud.

The potential security and privacy challenges in cloud-connected mobile applications have been widely studied, and a number of solutions presented. Similarly, Fenu et al., in their article "Controlling User Access to Cloud-Connected Mobile Applications by Means of Biometrics," propose a continuous authentication approach, which integrates physical (face) and behavioral (touch and hand movements) biometrics to control user access to cloud-based mobile services.

De Marsico et al. present a smart peephole based on remote biometric services, in the article "House in the (biometric) cloud: a possible application." In their approach, minimal processing is carried out locally.

In the last article, entitled "Cognitive and Biometric Approaches to Secure Services Management in Cloud-Based Technologies," Ogiela et al. explain how different security procedures in data and service management can be applied in both the cloud and fog computing environments, as well as in distributed computing infrastructures. Service management in cloud computing has been presented in connection with secure cognitive management systems, supporting management tasks and securing important data using CAPTCHA solutions. All such protocols can use personal features and biometric patterns. Application of cognitive and biometric features allows the creation of personalized procedures, tailored for users or groups of participants who seek to gain access to particular data repositories or receive specific services. Furthermore, the use of these protocols allows new solutions in the area of user-oriented service management protocols to be developed.

CONCLUSIONS

While the articles in this special issue have contributed to the knowledge base on the topic, there are many more challenges that need to be addressed, such as those discussed by Popović and Hocenski.[9] For example,

1. Do we have solutions that assure us that when cloud service provider claims to have destroyed our biometric data, no copy of such data remains on their system?
2. Do we have solutions that allow us to know how and where our biometric data is stored at any point in time?
3. What about the secure storage of biometric traits?

Cryptosystems for biometrics can be broadly categorized into those that derive the key directly from the biometric trait acquired on-the-fly; and those that generate the key by binding the biometric trait and a random binary key. In both cases, the biometric trait does not need to be stored (except during enrolment when the acquired biometric data is used to generate the encrypted key). Once a user has been successfully enrolled, information from the acquired original biometric trait is no longer used or saved. However, when the biometric authentication takes place on cloud architectures, potential attacks to privacy may occur during the transmission of the acquired biometric trait through the network. For example, spoofing attacks can lead to identity theft,[10] which is particularly critical in biometrics due to the infeasibility of changing users' trait. A video recording or in some cases, a photograph of the authorized person, can be used to gain access to protected data. Thus, the interest in cancellable biometrics.[11,12] Cancellable biometrics refers to the systematic distortion applied intentionally on the original biometric image, with the aim of deriving a "new" trait used for authentication. In the event that the cancelable feature is compromised, a new distortion is applied on the original trait so that the same biometrics is mapped to a new template.

Potential research topic 1: To design a privacy-preserving computation framework, in order to handle robust and efficient biometrics fusion processing. This can facilitate situation-based identification and sharing in the cloud.

Potential research topic 2: To design solutions that can automatically scale or de-scale existing privacy preserving algorithms.

Potential research topic 3: To design machine/deep-learning based solutions that can be used to facilitate existing BaaS approaches.

ACKNOWLEDGEMENTS

We thank the authors for submitting their work to this special issue, the anonymous reviewers for providing constructive feedback, and Dr. Mazin Yousif, editor-in-chief of *IEEE Cloud Computing*, for his great support throughout the entire publication process.

REFERENCES

1. M. Armbrust et al., *Above the Clouds: A Berkeley View of Cloud Computing*, technical report Technical Report No. UCB/ EECS-2009-28, Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, 2009.
2. P. Mell and T. Grance, "The NIST definition of cloud computing," NIST Special Publication 800-145, National Institute of Standards and Technology, 2011.
3. Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: State-of-the-art and research challenges," *Journal of Internet Services and Applications* (AICT 14), vol. 1, no. 1, 2014, pp. 7–18.
4. A. Castiglione et al., "Biometrics in the cloud: Challenges and research opportunities," *IEEE Cloud Computing*, vol. 4, no. 4, 2017, pp. 12–17.
5. A.J. Brown et al., "Cloud Forecasting: Legal Visibility Issues in Saturated Environments," *Computer Law & Security Review*, vol. In press, 2018; doi.org/10.1016/j.clsr.2018.05.031.
6. C. Hooper, B. Martini, and K.-K. R. Choo, "Cloud computing and its implications for cybercrime investigations in Australia," *Computer Law & Security Review*, vol. 29, no. 2, 2013, pp. 152–163.
7. S.C. Eastwood et al., "Biometric-enabled authentication machines: A survey of open-set real-world applications," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 2, 2016, pp. 231–242.
8. A. Ross, J. Shah, and A.K. Jain, "From template to image: Reconstructing fingerprints from minutiae points," *IEEE Transactions on Human-Machine Systems*, vol. 29, no. 4, 2007, pp. 544–560.
9. K. Popović and Ž. Hocenski, "Cloud computing security issues and challenges," *Proceedings of the 36th International Convention on Information and Communication Technology, Electronics and Microelectronics* (MIPRO 10), 2010, pp. 344–349.
10. N. Evans et al., "Guest editorial: Special issue on biometric spoofing and countermeasures," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, 2015, pp. 699–702.
11. V.M. Patel, N.K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Processing Magazine*, vol. 32, no. 5, 2015, pp. 54–65.
12. K.M.S. Soyjaudah, G. Ramsawock, and M.Y. Khodabacchus, "Cloud computing authentication using cancellable biometrics," *IEEE AFRICON*, 2013, pp. 1–4.

ABOUT THE AUTHORS

Silvio Barra is an assistant professor at the University of Cagliari and research collaborator at the Biometric and Image Processing Lab of the University of Salerno. He received BS and MS degrees cum laude at the University of Salerno in 2009 and 2012. In 2016, he received a PhD at the University of Cagliari. His research interests include biometric recognition, machine intelligence, and pattern analysis in images, signals and video. Contact him at silvio.barra@unica.it.

Kim-Kwang Raymond Choo holds the Cloud Technology Endowed Professorship in the Department of Information Systems and Cyber Security at the University of Texas at San Antonio. Choo has a PhD in information security from Queensland University of Technology. His research interests include cyber and information security and digital forensics. He is a senior member of IEEE, a Fellow of the Australian Computer Society, an Honorary

Commander, 502nd Air Base Wing, Joint Base San Antonio-Fort Sam Houston. Contact him at raymond.choo@fulbrightmail.org.

Michele Nappi is an associate professor of computer science at the University of Salerno and team leader of the Biometric and Image Processing Lab. His research interests include pattern recognition, image processing, image compression and indexing, multimedia databases and biometrics, human computer interaction, VR\AR. Nappi is an author of more than 160 papers in peer-reviewed international journals, international conferences and book chapters, and co-editor of several international books. He is an IEEE Senior Member, and president of the Italian Chapter of the IEEE Biometrics Council. Contact him at mnappi@unisa.it.

Arcangelo Castiglione is a post-doctoral fellow with the Department of Computer Science and an adjunct professor with the Department of Industrial Engineering at the University of Salern. He received a BS, MS, and PhD in computer science from the University of Salerno. His research mainly focuses on cryptography, multimedia data protection and network security. In 2015 he was a visiting researcher at the Laboratory of Cryptography and Cognitive Informatics at AGH University of Science and Technology, and at the School of Mathematics and Computer Science at Fujian Normal University. Contact him at arcastiglione@unisa.it.

Fabio Narducci is an assistant professor at the University of Naples "Parthenope," adjunct professor at the University of Molise, and research collaborator at the Biometric and Image Processing Lab of the University of Salerno. He received a PhD in computer science at the Virtual Reality Lab of the University of Salerno. His research interests include biometrics, gesture recognition, augmented reality, virtual environments, mobile and wearable computing, human computer interaction, haptics. Contact him at fabio.narducci@uniparthenope.it.

Rajiv Ranjan is a reader in the School of Computing Science at Newcastle University; chair professor in the School of Computer, Chinese University of Geosciences; and a visiting scientist at Data61, CSIRO. He has a PhD in computer science and software engineering from the University of Melbourne. Ranjan's research interests include grid computing, peer-to-peer networks, cloud computing, Internet of Things, and big data analytics. Contact him at raj.ranjan@ncl.ac.uk or http://rajivranjan.net.

Tensor-Based Big Biometric Data Reduction in the Cloud

Jun Feng
Huazhong University of Science and Technology

Laurence T. Yang
Huazhong University of Science and Technology

Ronghao Zhang
Huazhong University of Science and Technology

When dealing with big biometric data, data reduction becomes a challenge. This article proposes a novel big biometric data reduction solution in the cloud environment, including a big biometric data reduction approach based on tensor decomposition, an incremental big biometric tensor reduction approach, and a secure big biometric data reduction approach, which can significantly reduce big biometric data.

Digital biometrics relates to the recognition of identity using measurable physiological, behavioral or cognitive characteristics.[1,2] The characteristics used as biometric data in biometric systems are from fingerprints, face, gait, iris, voice, signature, biological signals, and so on. Digital biometrics can be used for authentication or access control. Because biometric data cannot be lost and do not need to be remembered, digital biometrics has had a wide range of security applications where security is critical, such as in banks, markets, or the healthcare industry.[3]

With the recent growth in the complexity, number, and storage volume of computer smartphones, tablet devices, and wearable devices available for consumers, there has been a corresponding growth in demand for digital biometric analysis. The mobile biometrics market report of Biometrics Research Group shows that due to the security provided by mobile biometrics, 700 million users will pay $750 billion in annual worldwide mobile payment transactions by 2020.

The volume of data issue has been raised for years, and has been described as the greatest issue challenging digital biometric practitioners, resulting in major case backlogs across large and small agencies.[4]

Concerns about the increasing volume of data are regularly raised in academic publications, with various proposed solutions.[5] These include: data mining, content retrieval, visualization techniques, triage processes, distributed and parallel processing, grid computing, analytical algorithms, neural network techniques, digital biometrics-as-a-service, and artificial intelligence systems.[6]

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD

A process of data mining has successfully been applied to increasingly larger datasets of business and science for many years, and one of the first stages is data reduction relevant to a task.[7] The data reduction aspect of data mining has many significant applications to reduce processing and analysis time and storage demands.[8] Concerns have been raised relating to data reduction using traditional reduction solutions, which can result in crucial evidence being missed, such as statistical reduction methods resulting in calls for more research into data mining of digital forensic data. With this in mind, a data reduction approach applicable to digital forensic data was proposed, and research was undertaken exploring this methodology.[5]

However, existing methods are not designed for big biometric high-order data reduction. Some biometric data in practical biometric systems are usually high-order. Tensors[9] can naturally represent high-order data. Therefore, a tensor-based data reduction approach is required for biometric systems.

Recently, cloud computing has attracted a large number of users or companies outsourcing big data storage and computations to the cloud.[10,11] When outsourcing big biometric data reduction to the cloud, users privacy is an important concern.[12,13] Users hope that their biometric data, such as fingerprint data, face data, and iris data, are not exposed to cloud. It is, therefore, necessary for cloud-assisted biometric systems to have a secure big biometric data reduction approach.

To solve these problems, we propose a novel big biometric data reduction approach in the cloud for biometric systems with the following contributions.

- We propose using tensor to model big biometric data. A tensor reduction approach based on tensor decomposition is presented to reduce big biometric data. An incremental tensor reduction is also presented.
- We develop a privacy-preserving approach to address the privacy problem of biometric data reduction over encrypted biometric data in a federated cloud environment. The privacy-preserving big biometric data reduction approach can utilize the cloud and protect users' privacy.
- Some concise cases are presented to demonstrate the effectiveness of the proposed big biometric tensor reduction approaches.

In this article, we first provide the tensor reduction, the incremental tensor reduction, and the secure tensor reduction. We then discuss the proposed tensor decomposition approach as applied to digital biometric data subsets, and our performance evaluation results. The final section outlines conclusions and scope for future work.

BIG BIOMETRIC DATA REDUCTION APPROACH BASED ON TENSOR DECOMPOSITION IN CLOUD

This section proposes tensor decomposition based approach to reduce big biometric data in the cloud. We first utilize tensor to model the big biometric data. We then present a tensor reduction with incremental tensor reduction of big biometric data. Finally, we show the secure tensor reduction in cloud. For reference, Table 1 shows the notations used in this article.

Table 1. Table of symbols.

Symbol	Definition
T	tensor
M	matrix
$\times_h$	h-mode product
x	plaintext
pk/sk	public key / secret key
$E_{pk}(x)$	encryption of message x with pk
$D_{sk}(y)$	decryption of ciphertext y with sk
$[\![x]\!]$	ciphertext of message x, $[\![x]\!] = E_{pk}(x)$
$Ai_1\ i_2$	$(i_1\ i_2 \cdots)$ element of tensor/matrix A

TENSOR FORMULATION OF BIG BIOMETRIC DATA

Tensor[9] is an emerging and promising tool for big data analysis, big data mining, and biometrics. A tensor is a general form of a matrix or a vector. For example, a one-order tensor is a vector, and a two-order tensor is a matrix. Multidimensional big biometric data can be naturally represented by using tensors. Tensors can provide more effective big data processing for biometric systems.

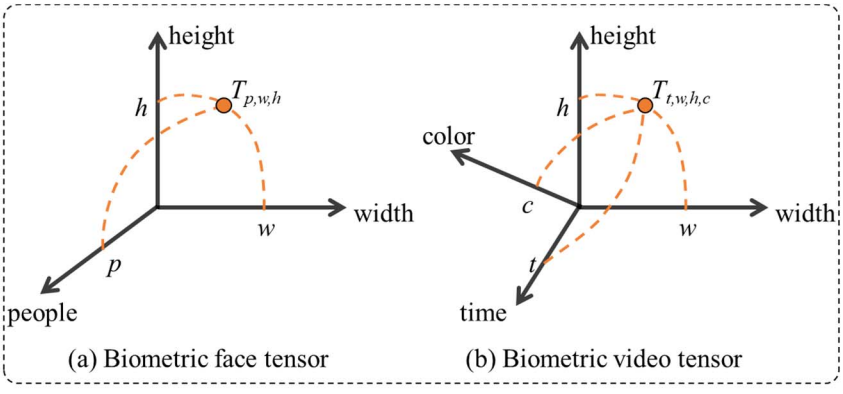


Figure 1. Tensor formulation examples of big biometric data.

In this article, we propose to employ tensors to model big biometric data. For instance, in biometric face recognition systems, we can create a three-order tensor $T \in R^{I_{people} \times I_{width} \times I_{height}}$ to model face data, where the tensor orders I_{people}, I_{width}, and I_{height} represent the number of facial images, the width the of images, and the height of the images, respectively. The biometric face tensor is illustrated in Figure 1(a). The element $t_{p,w,h} = 90$ represents the gray value of the w-th width and the h-th height for the p-th image is 90.

In a similar fashion, a biometric video can be naturally represented as four -order tensor $T \in R^{I_{time} \times I_{width} \times I_{height} \times I_{color}}$ with I_{time}, I_{width}, I_{height}, and I_{color} denoting time, width, height and color information (R, G and B). Figure 1(b) illustrates the biometric video tensor.

TENSOR REDUCTION OF BIG BIOMETRIC DATA

Like singular value decomposition and principal component analysis, tensor decomposition also can reduce big biometric data. However, singular value decomposition and principal component analysis can only reduce two-dimensional biometric data (i.e. biometric matrix). The biometric data are usually multidimensional, especially in the era of big data. Therefore, we propose using tensor decomposition to reduce multidimensional big biometric data (i.e. biometric tensors).

During tensor decomposition, the h-mode product of the tensor T by the matrix U is denoted as $T \times_h M$ where $T \in R^{J_1 \times J_2 \times \ldots \times J_H}$ denotes an H-th order tensor, $t_{j_1 j_2 \ldots j_{h-1} j_h j_{h+1} \ldots j_H}$ represents an element of the tensor T, $M \in R^{I_h \times J_h}$ denotes a matrix, and $m_{i_h j_h}$ is an element of the matrix M. The entry of $T \times_h M$ is written as follows:

$$(T \times_h M)_{j_1 j_2 \ldots j_{h-1} i_h j_{h+1} \ldots j_H} = \sum_{j_h=1}^{J_h} (t_{j_1 j_2 \ldots j_{h-1} j_h j_{h+1} \ldots j_H} \times m_{i_h j_h}) . \quad (1)$$

Tensor decomposition is a generalization of singular value decomposition. It can decompose a biometric tensor to a core tensor multiplied with a set of truncated orthogonal matrices. The tensor decomposition for an N-th order tensor T is defined as follows:

$$\begin{aligned} \mathcal{S} &= \mathcal{T} \times_1 U_1^{\mathrm{T}} \times_2 U_2^{\mathrm{T}} \ldots \times_H U_H^{\mathrm{T}}, \\ \hat{\mathcal{T}} &= \mathcal{S} \times_1 U_1 \times_2 U_2 \ldots \times_H U_H, \end{aligned} \quad (2)$$

where S, $\hat{\mathcal{T}}$ and U_i $(1 \le i \le H)$ denote the core tensor, the approximate tensor and the truncated orthogonal matrices, respectively.

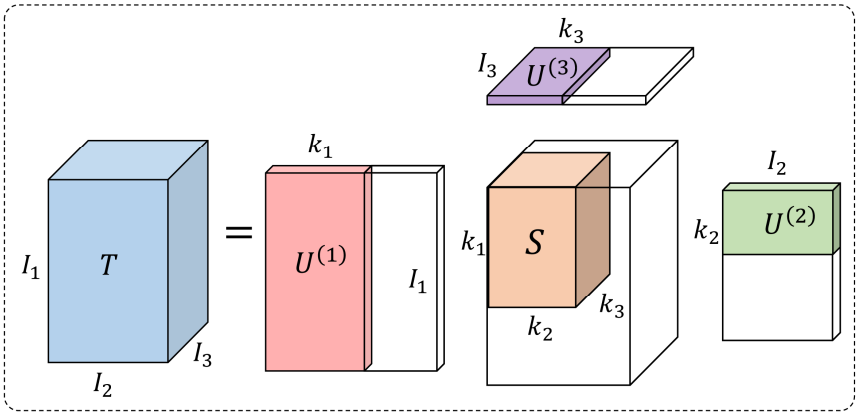
Since the matrices U_i $(1 \le i \le H)$ are truncated in tensor decomposition, the core tensor S and the truncated matrices U_i $(1 \le i \le H)$ are usually regarded as a reduced version of the original tensor T. Moreover, the reconstructed data in the approximate tensor $\hat{\mathcal{T}}$ are of higher quality than the original data in the tensor T. After tensor decomposition, we only need to store and process the core tensor S and the truncated matrices U_i $(1 \le i \le H)$ instead of the original tensor T. Figure 2(a) illustrates a three-order tensor decomposition. In the figure, the right of the equal sign denotes the unreduced result, while the green denotes the reduced result, which shows the tensor decomposition reduces the original data.

The biometric tensor reduction method is described as follows:

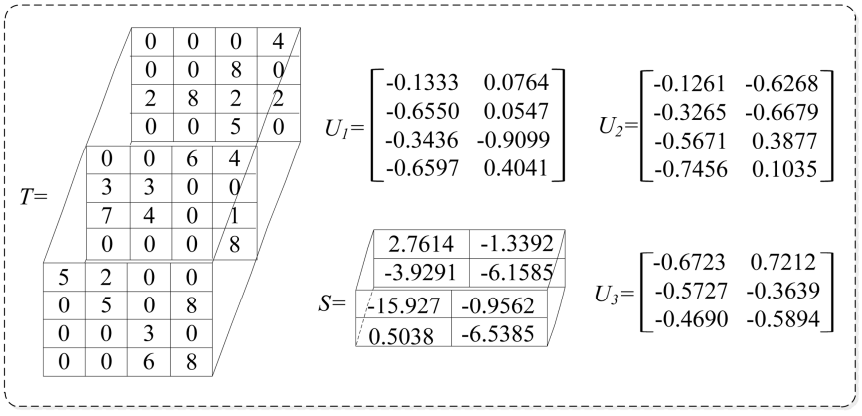
1. The original biometric tensor T is constructed from biometric data and biometric files, for example, pictures, documents, and Internet browsing history. For simplicity, suppose the order of biometric tensor T is 3.
2. The biometric tensor T is matricized on all modes, which generates three unfolded matrices.
3. To apply singular value decomposition on each unfolded matrix, three truncated orthogonal matrices are created. For a large-scale sparse matrix, the Lanczos method can be used to improve the efficiency of singular value decomposition.
4. The core biometric tensor can be computed by applying Equation 2.

We remark that for big biometric data, suppose the untruncated matrix is $m \times m$ matrix and the truncated matrix is $m \times n$ matrix, then the dimensionality m is far less n. Therefore, the big biometric data are greatly reduced.

For example, using the biometric tensor reduction method, we can decompose a $4 \times 4 \times 3$ biometric tensor T to generate a $2 \times 2 \times 2$ core tensor S, two 4×2 truncated matrices U_1, U_2 and one 3×2 truncated matrix U_3, which is shown in Figure 2(b). From the figure, it can be seen that the original biometric tensor is reduced.



(a) Visualization.



(b) Illustration.

Figure 2. Biometric tensor reduction.

INCREMENTAL TENSOR REDUCTION OF BIG BIOMETRIC DATA

Because of the high-velocity feature of big biometric data, we propose using the incremental tensor reduction method to reduce big biometric data. The incremental tensor reduction method is summarized below, as shown in Figure 3.

Briefly, when the incremental biometric data arrive, first of all, the incremental matrices on all modes are obtained. Suppose a new unfolded matrix on mode i is $M_i = [M_{i1}; M_{i2}]$, where M_{i1} is the original unfolded matrix and M_{i2} is the incremental matrix. Let the singular value decomposition of the matrix M_{i1} be $M_{i1} = U_{i1} \Sigma_{i1} V_{i1}^T$.

Then, we apply incremental singular value decomposition on each unfolded matrix. The matrix M_i can be decomposed as follows:

$$[M_{i1}; M_{i2}] = [U_{i1}, J] \begin{bmatrix} \Sigma_{i1} & L \\ 0 & K \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & I \end{bmatrix}^T, \quad (3)$$

where the matrices J, L, and K are the unitary basis, projection coordinate, and projection coordinate, respectively. The truncated matrices of the new biometric tensor are obtained.

Finally, the core biometric tensor of the new biometric tensor is computed. More details can be found in the work of Castiglione and colleagues.[10]

Due to the incremental tensor reduction, we do not need to re-reduce the new biometric tensor. We only compute the incremental results, and then integrate the reduced results of the original biometric tensor and the incremental results to obtain the reduced results of the new biometric tensor. Obviously, the incremental tensor reduction improves the efficiency of big biometric data reduction.

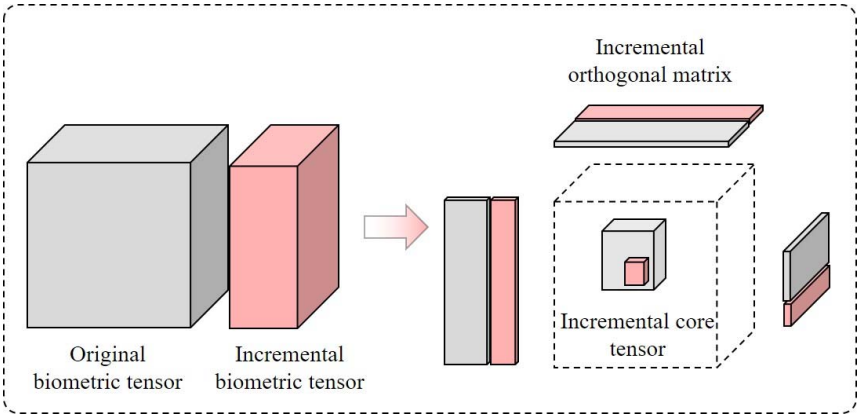


Figure 3. Incremental biometric tensor reduction.

PRIVACY-PRESERVING BIG BIOMETRIC DATA REDUCTION IN CLOUD

Building upon the above big biometric data reduction approach, we developed a privacy-preserving big biometric data reduction approach for protecting the users' privacy in cloud. The privacy-preserving big biometric data reduction approach is based on the Paillier cryptosystem with semantic security.

Suppose users would like to perform a big biometric data reduction, but they would not like to do it by themselves. The users intend to outsource the computing task of big biometric data reduction to public clouds and not allow clouds to see any biometric data. In our approach, the entities are users, and two public clouds are denoted by C_1 and C_2. The two clouds form a federated cloud. Assume that the entities are in a honest-but-curious model.

C_2 utilizes the Paillier cryptosystem to generate a public/private key pair (pk, sk) for the biometric system, and publishes the public key pk. To preserve users' privacy, biometric data x are respectively encrypted using the Paillier encryption with pk, $[\![x]\!] = E_{pk}(x)$, prior to outsourcing the encrypted biometric data $[\![x]\!]$ to C_1. The encrypted biometric data elements form the encrypted biometric tensor $[\![T]\!]$ on C_1.

C_1 and C_2 collectively carry out big biometric data reduction over the encrypted biometric tensor $[\![T]\!]$ in a privacy-preserving way. The Paillier cryptosystem exhibits the following two properties, namely homomorphic addition and homomorphic multiplication for any given ciphertexts $[\![m_1]\!]$, $[\![m_2]\!]$ and constant c:

$$\begin{aligned} &[\![m_1 + m_2]\!] = [\![m_1]\!] \times [\![m_2]\!] \bmod n^2, \\ &[\![cm_1]\!] = [\![m_1]\!]^c \bmod n^2. \end{aligned} \quad (4)$$

By using the properties, some secure sub-protocols (such as, the secure multiplication protocol, secure division protocol, and secure square root protocol) over encrypted data in a federated cloud environment are obtained. In the secure sub-protocols, the inputs and the outputs of C_1 are in encrypted form, the outputs are known only to C_1, and C_2 only can see the perturbation values

about users' biometric data. By using secure sub-protocols, the integrated secure protocol can be obtained to perform the big biometric data reduction in cloud.

In the privacy-preserving big biometric data reduction approach, once outsourcing the encrypted biometric data to the cloud, the users don't need to participate in any calculation of big biometric data reduction. This significantly reduces the computational burden on users.[14]

CASE STUDY AND SCENARIO ANALYSIS

Some concise cases are presented to demonstrate the effectiveness of the proposed big biometric tensor reduction approach. The approach has been applied to some real-world cases.

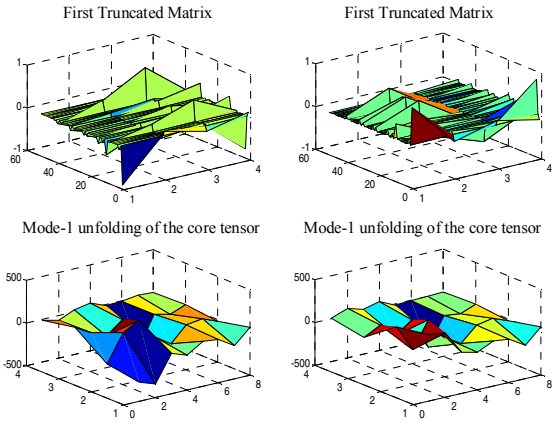
We took face biometrics as an example to illustrate the proposed big biometric tensor reduction approach. The biometric facial data were reduced by our proposed approaches. The reduction ratio and accuracy for the biometric face data are given in Figure 4(a).

The face data are from the facial recognition technology (FERET) dataset (https://www.nist.gov/itl/iad/image-group/color-feret-database), widely used in biometrics research. A subset of FERET dataset is used in our experiments. 60 pictures were selected and each picture was scaled to 30 × 30 pixels. The face data formed a third-order 60 × 30 × 30 tensor. By using the proposed reduction approach based on tensor decomposition, the biometric data were reduced to about 25% of the original tensor, which guaranteed 90.2% accuracy. By employing the proposed privacy-preserving approach, the biometric facial data were reduced to about 25% of the original tensor, which guaranteed 89.9% accuracy. In our experiments, when biometric facial data were updated, the incremental tensor reduction was utilized to update the core biometric facial data. For instance, the element $u_{2,11}$ of the truncated matrix U_1 changed from 0.0256 to 0.0055 when the biometric facial tensor elements $t_{20,20,1}$, $t_{20,20,2}$, $t_{20,20,3}$, $t_{20,20,4}$, and $t_{20,20,5}$ changed from 88, 11, 56, 25 and 150 to 171, 150, 132, 140 and 224. We used the incremental tensor reduction to speeds up the reduction process for the biometric facial tensor. In Figure 4(b and c), we plot the results of the tensor reduction and the results of the incremental tensor reduction for some biometric facial data. These results illustrate that both the approach based on tensor decomposition and the privacy-preserving approach can reduce the facial biometrics.

We took video biometrics as another example to illustrate the proposed big biometric tensor reduction approach. The video data are from the USF HumanID dataset (https://sites.google.com/site/tensormsl), which is widely used in biometrics research. The USF HumanID dataset consists of video clips of 71 subjects. The biometric video data were processed to generate thumbnail pictures with 32×22 pixels at certain intervals. The data formed a fourth-order 71×32×22×10 biometric video tensor where time mode is 10. Similarly, we applied the approach based on tensor decomposition and the privacy-preserving approach to the biometric video tensor. The reduction ratio and accuracy for the biometric video data are given in Figure 4(a). The biometric video tensor was reduced to 24% of the original data and the accuracy was 91.5% by using the approach based on tensor reduction, while the accuracy was 91.4% by means of the privacy-preserving approach.

Data	Reduction Ratio	Accuracy	
		Tensor Reduction	Privacy-Preserving Approach
Face Biometrics	25%	90.2%	89.9%
	20%	87.9%	87.8%
Video Biometrics	24%	91.5%	91.4%
	18%	90.6%	90.6%

(a) Reduction Ratio and Accuracy



(b) Tensor Reduction (c) Incremental Tensor Reduction

Figure 4. Experiment results of some biometric data reduction.

The above results demonstrate that the big biometric data reduction approach proposed in this article significantly reduces biometric data.

CONCLUSION

We proposed a novel big biometric tensor reduction approach to the cloud to improve the efficiency and effectiveness of biometric systems. A reduction approach using tensor decomposition with its incremental computation was developed to reduce multi-dimensional big biometric data, and a secure big biometric data reduction in clouds was presented to protect users' privacy. Finally, some biometrics examples were introduced to demonstrate the performance of our proposed big biometric tensor reduction approach. We have reduced the big biometric data to shed a light on the big biometric data research in computational biometrics. The proposed approach can be applied to tensor object recognitions in airport security checking systems or intelligent surveillance systems.

The incremental tensor reduction approach has been proposed in the article. Most of the operations (such as matrix-vector multiplication) in the proposed approach can be implemented in parallel. Significant efforts should be put into using the incremental tensor reduction approach and the distributed approach to further improve the efficiency of processing the big biometric data in practical apache-spark based computational biometrics.

ACKNOWLEDGMENTS

The research presented in this paper has been supported by the National Key Research & Development (R&D) Plan of China under Grant No. 2017YFB0801804, and the Shenzhen Fundamental Research Program under Grant No. JCYJ20170307172200714.

REFERENCES

1. A. Castiglione et al., "Biometrics in the Cloud: Challenges and Research Opportunities," *IEEE Cloud Computing*, vol. 4, no. 4, 2017, pp. 12–17.
2. M.R. Ogiela and L. Ogiela, "On Using Cognitive Models in Cryptography," *Proc. IEEE 30th Int'l Conf. Advanced Information Networking and Applications* (AINA 16), 2016, pp. 1055–1058.
3. M.R. Ogiela and L. Ogiela, "Application of Cognitive Cryptography in Fog and Cloud Computing," *Proc. Int'l Conf. Broadband and Wireless Computing, Communication and Applications* (BWCCA 17), 2017, pp. 293–298.
4. N. Memon, "How Biometric Authentication Poses New Challenges to Our Security and Privacy," *IEEE Signal Processing*, vol. 34, no. 4, 2017, pp. 196–194.
5. A. Castiglione et al., "Context Aware Ubiquitous Biometrics in Edge of Military Things," *IEEE Cloud Computing*, vol. 4, no. 6, 2017, pp. 16–20.
6. X. Wu et al., "Data Mining with Big Data," *IEEE Trans. Knowledge and Data Engineering*, vol. 26, no. 1, 2013, pp. 97–107.
7. J. Sun and C.K. Reddy, "Big Data Analytics for Healthcare," *Proc. of the 19th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD'13)*, 2013, p. 1525.
8. D. Zhang et al., "Real-Time Locating Systems Using Active RFID for Internet of Things," *IEEE Systems Journal*, vol. 10, no. 3, 2017, pp. 1226–1235.
9. X. Wang et al., "A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives," *IEEE Trans. Big Data*, 2017; doi.org/10.1109/TBDATA.2017.2757942.
10. A. Castiglione et al., "Cloud-based Adaptive Compression and Secure Management Services for 3D Healthcare Data," *Future Generation Computer Systems*, vol. 43, 2015, pp. 120–134.
11. M. Dong et al., "Multicloud-Based Evacuation Services for Emergency Management," *IEEE Cloud Computing*, vol. 1, no. 4, 2014, pp. 50–59.
12. D. He et al., "Security Analysis and Improvement of a Secure and Distributed Reprogramming Protocol for Wireless Sensor Networks," *IEEE Trans. Industrial Electronics*, vol. 60, no. 11, 2013, pp. 5348–5354.
13. H. Liu et al., "Role-Dependent Privacy Preservation for Secure V2G Networks in the Smart Grid," *IEEE Trans. Information Forensics and Security*, vol. 9, no. 2, 2017, pp. 208–220.
14. J. Feng et al., "Privacy-Preserving Tensor Decomposition over Encrypted Data in a Federated Cloud Environment," *IEEE Trans. Dependable and Secure Computing*, 2017.

ABOUT THE AUTHORS

Jun Feng is a PhD student in the School of Computer Science and Technology at Huazhong University of Science and Technology, Wuhan, China, and in the Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen, China. His research interests include cloud computing and big data security. Contact him at junfeng989@gmail.com.

Laurence T. Yang is a professor with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, with the Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen, China, and with the Department of Computer Science, St. Francis Xavier University, Antigonish, NS, Canada. His research interests include parallel and distributed computing, embedded and ubiquitous/pervasive computing, and big data. Contact him at ltyang@ieee.org.

Ronghao Zhang is currently working for the MS degree in School of Computer Science and Technology at Huazhong University of Science and Technology, Wuhan, China. His research interests include biometrics and cloud computing. Contact him at zrhfight@163.com.

Controlling User Access to Cloud-Connected Mobile Applications by Means of Biometrics

Gianni Fenu
University of Cagliari

Mirko Marras
University of Cagliari

Cloud-connected mobile applications are becoming a popular solution for ubiquitous access to online services, such as cloud data storage platforms. The adoption of such applications has security and privacy implications that are making individuals hesitant to migrate sensitive data to the cloud; thus, new secure authentication protocols are needed. In this article, we propose a continuous-authentication approach integrating physical (face) and behavioral (touch and hand movements) biometrics to control user access to cloud-based mobile services, going beyond one-time login. Experimental results show the security–usability tradeoff achieved by our approach.

Mobile cloud computing (MCC), one of the top next-generation computing disruptions, aims to enable mobile users to leverage infrastructure, platforms, and applications made available remotely by cloud providers (e.g., Google, IBM, and Microsoft). In 2016, the Cisco Global Cloud Index predicted that 59% (2.3 billion) of the Internet population will use personal cloud-connected storage by 2020.[1] In 2016, Statista forecasted that the total number of mobile phone users worldwide will reach 4.68 billion by 2019.[2]

Ubiquitous access to cloud-connected resources is associated with several security and privacy challenges. For instance, stolen accounts can be abused by impostor users to download sensitive data remotely stored in cloud-connected storage platforms. Existing access control methods in MCC mainly use passwords. Weak passwords are easy to remember but can be easily broken, whereas stronger passwords are more secure but difficult to remember. Providing a reasonable security–usability tradeoff becomes crucial.

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

MCC and mobile biometrics can share mutual benefits in order to move a step forward to secure the users' access to the cloud.[3] MCC can exploit the stronger verification capabilities of biometrics to enhance the security of access control for the cloud.[4] Acuity predicted that the use of mobile biometric technologies will increase and that the corresponding revenue will reach $50.6 billion annually by 2022.[5] At the same time, mobile biometrics can leverage cloud computational resources in order to be provided as a service, reducing management costs and improving performance.[6] In this direction, the Cisco Global Cloud Index predicted that, by 2020, 74% of cloud workloads will be software as a service.[1]

Leveraging mobile biometrics to protect access to users' data stored in the cloud promises to be a stronger solution than using passwords; biometrics cannot be easily stolen, forgotten, or guessed. Biometric authentication methods (i.e., one-to-one matching between the logged-in person's data and the account owner's data) analyze one or more physical and/or behavioral characteristics of users to verify their identities, such as fingerprints, irises, faces, typing, or touching. Existing methods tend to perform authentication only when the user's session starts. This makes it possible to hijack a session. Impostor users could access an account after the genuine user logs in. The same thing could happen if impostors succeed in fooling the system by passing the one-time login.

In response to this, continuous authentication is becoming increasingly popular. The goal is to regularly check the user's identity throughout the session.[7] Related methods can be good candidates to eliminate current user skepticism regarding security guarantees on cloud adoption.

In this article, we present a multibiometric approach that can be integrated in both web-based and native cloud-connected mobile applications in order to continuously and transparently check the user's identity throughout the session. To this end, the approach leverages the data collected from five sensors embedded in common mobile devices (see Figure 1). Front-facing cameras allow taking photos of the device users without their collaboration; touch sensors can detect contact between users' fingers and the screen and can track finger movement. Inertial measurement units (IMUs) provide data from accelerometers (measuring the acceleration force applied to a device on the three physical axes), gyroscopes (measuring the device rotation rate around each of the three physical axes) and magnetometers (measuring the geomagnetic field for the three physical axes) to track how users hold the device.

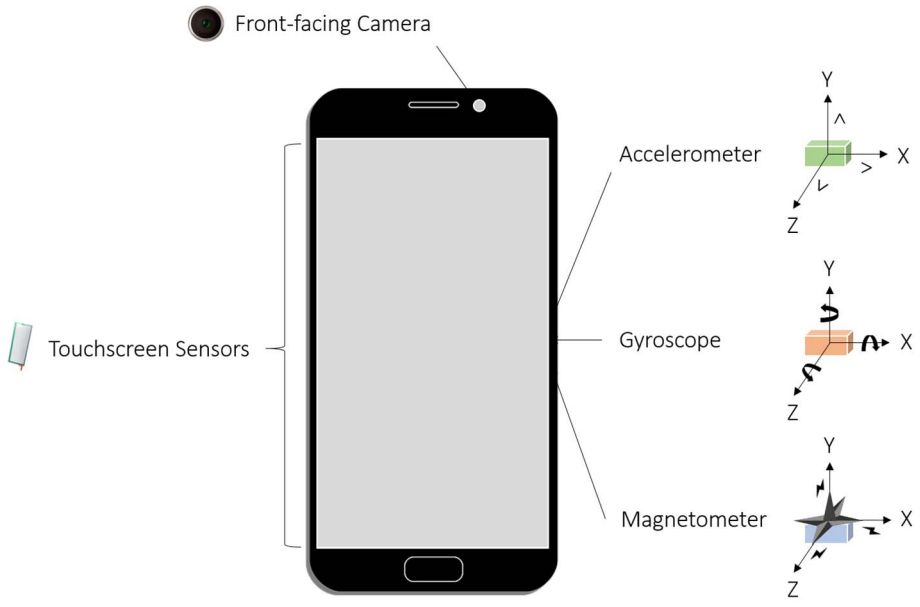


Figure 1. The set of mobile device sensors leveraged by the authentication approach.

On the basis of face images from front-facing cameras, touch data from touch sensors, and hand movement data from IMUs, three authentication subsystems individually compute the matching score between the probe (the data provided during authentication) and the template (the data stored in the database for the legitimate user). Then, the matching scores are fused. If the fusion score is over a given threshold, users can continue their normal activity during the session; otherwise, they are locked out. The proposed approach extends the research of "Leveraging Continuous Multi-modal Authentication for Access Control in Mobile Cloud Environments"[8] with

- a deeper contextualization in state-of-the-art solutions,
- an additional biometric authentication subsystem based on the data collected from multimotion sensors in IMUs, and
- an experimental evaluation of the new approach on chimeric users built from public datasets, investigating short- and mid-term enrolment-authentication scenarios.

Our approach promises to be more robust to impostors' attacks since both the user's face and the user's behavior must be continuously replicated during the session to fool the system.

EMERGING NEEDS FOR ACCESS CONTROL FOR CLOUD-CONNECTED MOBILE APPLICATIONS

User authentication in MCC refers to the process of validating mobile users' identities to ensure they can legitimately access requested cloud-based resources. Several studies have proposed authentication schemas tailored for controlling access to cloud-connected resources from laptops and desktop computers. Integrating authentication methods in MCC introduces challenges in relation to efficiency and efficacy owing to the more uncontrolled mobility context (e.g., limited computational resources and limited bandwidth). Existing MCC authentication methods are discussed in "Authentication in Mobile Cloud Computing: A Survey."[4]

Considering the deployment mode, existing approaches are subdivided into user-side or cloud-side approaches, according to where the authentication is performed. User-side approaches carry out both data acquisition and authentication processing entirely on the mobile device. Moving authentication within those devices makes user-side approaches less efficient. Moreover, a user's privacy can be compromised in the case of mobile-device robbery since the sensitive data (e.g., biometric data) is stored in the mobile device.

Cloud-side solutions compute the authentication processing on remote servers. Even though this presents benefits in performance and usability, security and privacy are negatively affected if they are not properly managed. Secure and privacy-aware authentication functionalities offered as a service over the cloud can be cost effective, scalable, reliable, and hardware agnostic, and provide security anytime and anywhere, in contrast with user-side approaches. We expect that such solutions could make it easier to manage the integration of authentication capabilities in the ever-increasing variety of cloud-connected applications.

Considering the leveraged data, existing methods are either identity-based or context-based. The first authenticates users through IDs and passwords, while the second authenticates users by analyzing passively collected data, such as the IP address, location, and application logs. Identity-based methods tend to suffer the common problems introduced by IDs and passwords (e.g., memorability), whereas context-based methods tend to use data specific to the application context, such as logs. However, logs vary among applications, so the related methods are dependent on the application in which they will be integrated and are not easily portable to other applications. Methods based on the location and IP address may limit users to accessing the cloud from only certain areas or networks. This is a limitation in current mobility scenarios.

In order to reduce such constraints, checking user access by means of biometrics can be a viable solution. The low use of biometric solutions so far is due to the fact that the related systems applied in mobile settings tend to suffer low matching accuracy. In fact, the acquisition conditions (e.g., illumination, poses, and face positions) are typically more uncontrolled in mobile scenarios,[9] and impostor users can take advantage of such drawbacks to improperly access applications.

Continuous authentication using multibiometric mechanisms has the potential to strengthen the overall security of existing systems while maintaining usability.

LEVERAGING BIOMETRICS FOR CONTINUOUS AUTHENTICATION ON MOBILE DEVICES

Continuous authentication is a promising area that has been actively studied over the years, but the related methods have been challenging to apply in real settings owing to their operational complexity. The goal is to regularly check the user's identity during the session. Related methods have been widely discussed in "Advances in User Authentication: Continuous Authentication."[7]

Existing applications include monitoring whether an unsuspected impostor has hijacked a genuine user's session on a device or on an online website, as well as identifying whether an authorized user has shared his or her credentials with others in e-learning exams.[10] Existing techniques for continuous authentication have extracted physiological or behavioral biometrics by leveraging built-in sensors. Physical biometrics, such as face biometrics, have been captured using front-facing cameras and analyzed to get users' distinctive features. The mouse, keyboard, microphone, touchscreen, and IMU have been used to measure behavioral biometrics (e.g., gait, voice, typing, touching, mouse movements, and hand movements). Such methods continuously check whether the probe comes from the genuine user; if this is true, the user can continue the session; otherwise, the user is locked out.

One of the common unimodal continuous approaches is based on face recognition. Many algorithms work well on images and videos that are collected in controlled settings. However, their performance degrades significantly on images that present variations in face conditions (e.g., pose, illumination, occlusion, and image quality). Recently, "Unconstrained Still/Video-Based Face Verification with Deep Convolutional Neural Networks" presented a deep-learning system for face recognition and verification tailored for acquisition in uncontrolled settings.[11] In our approach, we integrated a deep neural network to detect faces and extract features from them.

With the proliferation of mobile devices, other unimodal approaches employing touch-pattern-based authentication have become a new way to validate the identity of legitimate users. In "A Survey on Touch Dynamics Authentication in Mobile Devices," the authors discussed existing schemas that leverage touch sensor data to get users' distinctive features for authentication purposes.[12] In contrast, "Performance Analysis of Multi-motion Sensor Behavior for Active Smartphone Authentication" investigated the reliability and applicability of motion sensors for active authentication.[13] Extensive experiments have showed that sensor behavior exhibits sufficient discriminability and stability for active authentication. Therefore, we decided to use motion sensors as a data source for providing authentication. In "HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users," the hand's movement, orientation, and grasp are considered as a biometric modality to continuously authenticate users during typing activities.[14] The authors analyzed motion sensor behavior close to the time the touch action happens. In our approach, we modelled motion behavior before, during, and after touch interaction.

Unimodal continuous authentication approaches have had to deal with noisy data, intraclass variation, spoofing attacks (i.e., fooling a biometric system by copying or imitating the biometric identifying the legitimate user). More recent multimodal systems exploit evidence from multiple sources to mitigate such issues. For instance, "A Continuous User Authentication Scheme for Mobile Devices" presented a framework combining face and touch modalities.[15] Instead of a score-level fusion of the scores from face and touch authentication subsystems, the authors built a set of classifiers for each modality such that each sample results in two score sets. Then, these sets are combined following a late-integration approach with a stacked classifier. That classifier learns the nuances of the scores it obtains from each base classifier.

COMBINING FACE, TOUCH, AND HAND-MOTION PATTERNS FOR CONTINUOUS AUTHENTICATION

This section describes the proposed continuous-authentication approach. First, we briefly present how the overall authentication process works. Then, we go in depth into the functionalities of each module of the underlying architecture, including the face, touch, and hand-motion authentication steps.

The Proposed Approach

The reference architecture underlying the proposed approach is depicted in Figure 2. The implementation of each biometric module was properly carried out to ensure both the efficacy and efficiency of the overall solution. This approach aimed to provide a good tradeoff between security and usability.

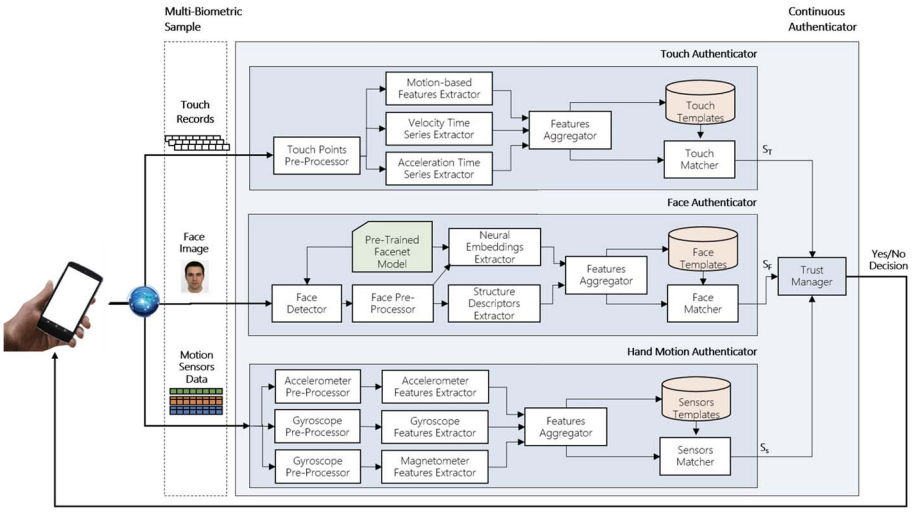


Figure 2. The reference architecture for continuous authentication based on physical (face) and behavioral (touch and hand movement) biometrics under uncontrolled settings.

The approach works as follows. Starting from the login step, for each touch stroke, a tracking module gets all the data about the touch stroke together with the image captured from the camera and the data from IMU sensors close to the time the user touches the screen. The sample is sent to the continuous authenticator. Inside it, each biometric authenticator individually performs authentication by comparing the features extracted from the face image, touch data, or motion sensor data with the corresponding templates and returns a matching score. Then, the matching scores are fused by the trust manager, which decides whether the user can continue the session.

The Face Authenticator

This module receives an image as input. It implements the FaceNet algorithm[16] for the localization of the relevant subregion of the image containing the face. FaceNet has been proven to work well when face samples are collected in the wild, including illumination changes, occlusion, and wide variation in poses and facial expressions. An image correction routine is used to normalize illumination on the face image.

Then, the module detects the key points of interest of the face (i.e., landmarks): the left eyebrow, the right eyebrow, the left eye, the right eye, the nose, and the mouth. The implemented landmark detector is based on an ensemble of regression trees trained to estimate facial-landmark positions from pixel intensity. The face region is cropped and aligned in preprocessing on the basis

of the eye regions to obtain a normalized rotation, translation, and scale representation of the face. Landmarks are also cropped and manipulated separately.

From the extracted image parts, the module extracts these types of features:

- the preprocessed face converted to grayscale and rescaled to 64 × 64 pixels;
- local binary pattern (LBP) features extracted for a cell size of 8 × 8 from the 64 × 64 rescaled grayscale face (LBP is a texture operator that labels pixels of an image by thresholding the neighborhood of each pixel. It then considers the result as a binary number.);
- eye-, nose-, and mouth-based bounding boxes extracted from the 64 × 64 rescaled grayscale face and resized to 16 × 20, 28 × 17, and 22 × 46 pixels, respectively;
- LBP features obtained for a cell size of 12 × 12 pixels from each of the resized landmark bounding boxes computed during the previous step; and
- FaceNet neural embeddings resulting from the face and the individual landmarks.

Features are vectorized as unidimensional vectors, and the feature vectors extracted from multiple training samples are averaged, then normalized in the range [0, 1] by using z-score normalization. A feature selection method is applied to reduce the dimensionality of each feature vector. The resulting feature vector constitutes the user's template. The matching score between the template and the probe is computed using cosine similarity and returned by the authenticator.

The Touch Authenticator

This module receives a set of data records related to an individual touch stroke as input. Each set of records is composed of one finger-down record, n consecutive move records, and one finger-up record. Every record r_i is encoded as a 4D list: $r_i = (x_i, y_i, c_i, t_i)$ for $i \in 1, \ldots, N_p$, where (x_i, y_i) are the location coordinates, c_i is the contact size applied at time t_i, and N_p is the number of records captured during the current touch stroke.

The module is set to process only strokes containing more than three data records. From each set of records, different types of features are extracted in relation to the contact during motion, the velocity, and the acceleration during the touch. For contact, velocity, and acceleration sequences, nine statistical features (e.g., mean, standard deviation, maximum, and minimum) are computed, as described in "HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users."[14] In addition, a 6D frequency-based feature vector (spectral centroid, spectral energy, spectral spread, spectral skewness, spectral kurtosis, and spectral flatness) is extracted from the time series representing the contact, velocity, and acceleration along the touch.

The features are concatenated, and the feature vectors extracted from multiple training samples are averaged feature-wise, then normalized in the range [0, 1] by using z-score normalization. Each feature is scaled per user using the related standard deviation. The cosine similarity between the feature vectors extracted from the probe and the template is returned as a matching score by the authenticator.

The Hand-Motion Authenticator

This module receives three sequences of observations—one sequence for each motion sensor. Inside each sequence, each observation represents the values from the three axes of that sensor at a given time, and it is in the form <*Timestamp*, *X*, *Y*, *Z*>. In order to get rotation invariance, to each observation is added a fourth value: the magnitude *M* of the combined vector that corresponds to the physical definition, being computed as the square root of the sum of the squares of the homologous values in the three sensor axes. Consequently, each observation is in the form <*Timestamp*, *X*, *Y*, *Z*, *M*>.

For each motion sensor, the module separately analyzes the *X*, *Y*, *Z*, and *M* values as time series and extracts from each time series a 16D time-based feature vector composed of 10 time-based

features (mean, standard deviation, average deviation, skewness, kurtosis, root-mean-square energy, minimum, maximum, nonnegative count, and zero-crossing rate) and six frequency-based features (spectral centroid, spectral energy, spectral spread, spectral skewness, spectral kurtosis, and spectral flatness). Therefore, from each sensor sequence, a 64-dimensional feature vector is extracted (4 time series × 16 features per time series).

Feature vectors extracted from multiple training samples are averaged. Feature vectors coming from the three motion sensors' data are vectorized and concatenated to form a unique feature vector. The feature vector is normalized in the range [0, 1] by using z-score normalization, and each feature is scaled per user using the related standard deviation. A feature selection method is applied to reduce the feature vector dimensionality. During verification, the module uses cosine similarity to calculate how similar the feature vectors from the probe and the template are.

The Trust Manager

This module is responsible for the fusion of the matching scores and of the computation of the global trust level of the user's genuineness, which is used to decide whether the current user is authenticated. For each biometric authenticator, an individual trust level of user genuineness is kept updated, taking into account past values in the session, by using formulas derived from those reported in "Performance Evaluation of Continuous Authentication Systems. IET Biometrics."[17] The parameters have been optimized for each authenticator during the preliminary stages.

For each biometric authenticator, the manager rewards or penalizes the trust in the user's genuineness on the basis of the amount of variation between the genuine user template and the current user probe. If there is only a small deviation for that biometric module, then the trust of the system in the user's genuineness for that biometric module tends to increase. For a large deviation, the trust tends to decrease. How much the trust increases or decreases dynamically depends on the amount of variation.

Then, the trust levels are fused together with a weighted sum. All the levels range between 0 and 1. The values assigned to the weights are based on the *equal-error rate* (EER) of the corresponding authentication module. The EER is the value when the *false-acceptance rate* (FAR) and *false-rejection rate* (FRR) are the same. If the global trust is over a given threshold, the user continues the session; otherwise, he or she is locked out.

EXPERIMENTAL EVALUATION

This section presents the experimental evaluation carried out on the proposed approach. First, we describe the chimeric dataset; then, we list the evaluation metrics and the experiments.

The Chimeric Dataset

To evaluate the applicability of our approach, we used the HMOG (hand movement, orientation, and grasp)[14] and MOBIO (Mobile Biometry)[18] multimodal datasets. HMOG includes behavioral data recorded by mobile device sensors for 100 users under three scenarios (reading, writing, and map navigating). For coherence with respect to MOBIO acquisition scenarios, we selected only HMOG sessions where users were sitting. Each volunteer performed 12 sessions, each lasting about 5 to 15 minutes. MOBIO consists of audio and image data taken from 150 people. For each user, 12 sessions were captured: six sessions included 21 videos; the other six sessions consisted of 11 videos.

Assuming the physical (i.e., face) biometrics and the behavioral (i.e., touch and sensor) biometrics were statistically independent, we built 100 chimeric users from these two datasets. It is accepted that results obtained in this way are worthy of full reliability. Given a user u_i from MOBIO and a user u_j from HMOG, a chimeric user c_{ij} is (u_i, u_j). Since HMOG contains 100 users, 100 of the 150 MOBIO subjects were randomly chosen. Then, 100 chimeras were built by randomly associating users.

Since the number of sessions per user is the same in both MOBIO and HMOG, we chronologically sorted sessions in the datasets, and we paired them using their position in the sorted lists. For each user's session, we created the biometric sample by counting and sampling the touch strokes the user had done during the session. Then, we associated each touch stroke with a single biometric sample. Each biometric sample was enriched with the data from each motion sensor, from 100 ms before the touch stroke until 100 ms after the touch stroke.

Finally, we merged all the videos of the current user in the current session, and we counted the total number of frames of the merged video. In order to simulate the acquisition setting defined by our approach, we randomly selected as many video frames as there were touch strokes in the current session of the current user. Then, we chronologically paired the extracted video frames with the touch stroke samples, creating a set of (video frame, touch stroke sample) pairs. Each experiment was repeated 10 times, and the results were averaged.

The Evaluation Metrics

The performance of our approach was measured with the following evaluation metrics:

- The FAR is the measure of the chance that the biometric system incorrectly accepts an access attempt by an impostor user. The FAR results from dividing the number of false acceptances by the number of impostor attempts. The lower the FAR is, the better.
- The FRR is the measure of the chance that the biometric system incorrectly fails to authenticate a legitimate user. The FRR is calculated as the ratio between the number of false rejects and the number of genuine attempts. The lower the FRR is, the better.
- The EER is the value obtained at the threshold level used by the biometric system where the FAR and FRR are equal. The lower the EER is, the better.

Short-Term Enrolment-Authentication Evaluation

The first experimental setup aimed to evaluate the performance of our approach when authentication is performed close in time to enrolment. In this setting, the influence of the interaction context variation and the aging effect should be low.

To simulate this scenario, for each user session, we trained the system with data recorded on the first 70% of the session and tested it on data tracked on the remaining 30%. Going over the scope of this experimental setup, the simulated scenario could also serve as continuous authentication within one session of device utilization. Assuming that the genuine user logs in to the application by entry-point authentication and that enrolment starts with the session, after the device observes a few samples, it changes to authentication mode. In this way, it can detect whether another person is using the account in cases in which the genuine user interacts with the device and, after a moment, leaves the device unattended. In this case, we should assume that the login to the account and the first part of the session (enrolment) are entirely carried out with the genuine user.

The EER of the individual biometric subsystems (the face, touch, and motion sensor matchers) as well as of the bimodal and trimodal fusion matchers obtained during this experimental setup are reported in Table 1. From the results with unimodal verification, we observed that the motion sensor subsystem achieved the lowest performance (EER = 12.45%) compared to the other combinations. Since the data tracked by motion sensors was exposed to large variations during the sessions, it is reasonable that the resulting EER was higher. The touch subsystem achieved promising but not exceptional results. The authentication capability depends mainly on the duration of the touch stroke and the amount of data the strokes consequently provide. For instance, swipe gestures enable getting more data in terms of touch records, compared to tap gestures. The face subsystem exhibited good performance when it was used alone.

Table 1. Experimental short- and mid-term enrolment-authentication results in terms of the equal-error rate (EER). Italics indicate the lowest EER for each pairing of a biometric-system type and an authentication setting.

Biometric-system type	Authenticators used	Authentication setting	
		Short term (%)	Mid term (%)
Unimodal	Face	*3.21*	*5.95*
	Touch	7.80	11.23
	Hand motion	12.45	17.56
Bimodal	Face + touch	*1.53*	*3.92*
	Face + hand motion	2.81	5.69
	Touch + hand motion	6.78	10.52
Trimodal	Face + touch + hand motion	*0.84*	*3.56*

Meanwhile, the EER for the multimodal subsystems was lower than for the unimodal subsystems, except for the bimodal subsystem combining touch and hand motion sensors, which performed worse than the individual face matcher (+3.57% EER). When the output of the face and touch subsystems was fused, the performance improved by –1.68% compared to the EER of the individual face matcher. Integrating all the modalities improved the EER by –0.69% compared to the best bimodal subsystem (face plus touch). The achieved EER is reasonable.

Mid-term Enrolment-Authentication Evaluation

The second experimental setup aimed to evaluate the performance of our approach when the authentication is performed some weeks after the enrolment. Thus, the aging effect and the different interaction context could influence recognition capabilities, especially for touch and hand motion.

This scenario assumes that the genuine user trains the approach during enrolment and then the template stays the same over many sessions. During the session, the system authenticates the user by comparing the probe with the template built during old sessions.

The scenario is more realistic than the scenario we depicted in the short-term-evaluation setup. In the latter, we assumed that the login and the first part of interaction during the session were performed by the genuine user. Even though that scenario includes a lot of real situations, it does not work in the case of the mobile device having been stolen in advance. The improper access is detected only if the impostor enters the session after the real user. To simulate the mid-term-evaluation setup, we trained the system with data from the first 50% of the sessions and tested it on data from last 50% for each type of task (e.g., reading and writing).

The EER obtained in the second experimental setup is reported in Table 1. The EER varied between 5.95% and 17.56% for the single-modality approaches. The motion sensor subsystem reported the largest decline in performance compared to the corresponding single-modality results in the short-term evaluation, while the performance was more stable for face and touch.

Integrating multimodal solutions improved the authentication performance, but the combination of touch plus motion sensors performed worse than face-only recognition in this experimental setup. Bimodal recognition resulted in higher EER than in the short-term evaluation. Moreover, the touch and motion sensor subsystems did not improve the face recognition capabilities as

much in this scenario. This is reasonable since behavioral biometrics usually tend to exhibit large variations among different sessions.

Furthermore, some variables of the interaction context could vary between enrolment sessions and authentication sessions (e.g., the emotional state, body posture, and hand posture). The aging effect could also influence the overall performance. Bimodal fusion enabled reaching reasonable but higher EER values compared to the short-term evaluation. The total number of scores being tested was big, and the number of scores correctly matched was relevant. The EER is reasonable even in this case.

CONCLUSION

In this article, we presented an approach for integrating face authentication with behavior authentication to secure user access to cloud-connected mobile applications, going beyond one-time login. The approach is promising and leaves a large space for improvement.

At the infrastructure level, we plan to employ big data architectures for large-scale fast computation, leverage the attractive properties of the cloud, and investigate communication and storage protocols to preserve privacy. At the algorithmic level, we will investigate

- other methods to improve the recognition scores returned by the biometric subsystems,
- the adoption of additional biometrics,
- the way the system trusts the user's genuineness, and
- the capability of working well when one motion signal is missing.

At the application level, we will deploy the algorithms as a service and validate them in real scenarios, such as cloud-based e-learning or storage platforms.

ACKNOWLEDGMENTS

Mirko Marras gratefully acknowledges the Sardinia Regional Government for the financial support of his PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020—Axis III "Education and Training," Thematic Goal 10, Priority of Investment 10ii, Specific Goal 10.5).

REFERENCES

1. *Cisco Global Cloud Index 2015–2020*, Cisco, 2016; https://www.cisco.com/c/dam/m/en_us/service-provider/ciscoknowledgenetwork/files/622_11_15-16-Cisco_GCI_CKN_2015-2020_AMER_EMEAR_NOV2016.pdf.
2. "Number of Mobile Phone Users Worldwide from 2015 to 2020 (in Billions)," *Statista*; https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide.
3. A. Albahdal and T.E. Boult, "Problems and Promises of Using the Cloud and Biometrics," *Proceedings of the 11th International IEEE Conference on Information Technology: Generations* (ITNG 14), 2014, pp. 293–300.
4. M. Alizadeh et al., "Authentication in Mobile Cloud Computing: A Survey," *Journal of Net-work and Computer Applications*, vol. 61, 2016, pp. 59–80.
5. "Mobile Biometric Market Forecast to Exceed $50.6 Billion in Annual Revenue in 2022 as Installed Base Grows to 5.5 Billion Biometric Smart Mobile Devices," *Acuity Market Intelligence*, PR Newswire, 14 September 2017; https://www.prnewswire.com/news-releases/mobile-biometric-market-forecast-to-

exceed-506-billion-in-annual-revenue-in-2022-as-installed-base-grows-to-55-billion-biometric-smart-mobile-devices-300519359.html.
6. A. Castiglione et al., "Biometrics in the Cloud: Challenges and Research Opportunities," *IEEE Cloud Computing*, vol. 4, no. 4, 2017, pp. 12–17.
7. D. Dasgupta, A. Roy, and A. Nag, "Continuous Authentication," *Advances in User Authentication*, Springer, 2017.
8. G. Fenu and M. Marras, "Leveraging Continuous Multi-modal Authentication for Access Control in Mobile Cloud Environments," *Proceedings of the International Conference on Image Analysis and Processing (ICIAIP 17)*, LNCS 10590, Springer, 2017.
9. V.M. Patel et al., "Continuous User Authentication on Mobile Devices: Recent Progress and Remaining Challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, 2016, pp. 49–61.
10. G. Fenu, M. Marras, and L. Boratto, "A Multi-biometric System for Continuous Student Authentication in E-learning Platforms," *Pattern Recognition Letters*, 2 April 2017; doi.org/10.1016/j.patrec.2017.03.027.
11. J.C. Chen et al., "Unconstrained Still/Video-Based Face Verification with Deep Convolutional Neural Networks," *International Journal of Computer Vision*, vol. 126, no. 2-4, 2016, pp. 272–291.
12. P.S. Teh et al., "A Survey on Touch Dynamics Authentication in Mobile Devices," *Computers & Security*, vol. 59, no. C, 2016, pp. 210–235.
13. C. Shen et al., "Performance Analysis of Multi-Motion Sensor Behavior for Active Smartphone Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, 2018, pp. 48–62.
14. Z. Sitová et al., "HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, 2016, pp. 877–892.
15. M. Smith-Creasey and M.A. Rajarajan, "A Continuous User Authentication Scheme for Mobile Devices," *Proceedings of the 14th Annual IEEE Conference on Privacy, Security and Trust* (PST 16), 2016, pp. 104–113.
16. F. Schroff, F. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 15), 2015.
17. P. Bours and S. Mondal, "Performance Evaluation of Continuous Authentication Systems," *IET Biometrics*, vol. 4, no. 4, 2015, pp. 220–226.
18. C. McCool et al., "Bi-modal Person Recognition on a Mobile Phone: Using Mobile Phone Data," *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops*, vol. ICMEW 12, 2012, pp. 635–640.

ABOUT THE AUTHORS

Gianni Fenu is an associate professor of computer science at the University of Cagliari. His research interests include biometrics, cloud computing, and recommender systems. Fenu received a master's in engineering from the University of Cagliari. Contact him at fenu@unica.it.

Mirko Marras is a PhD student in computer science at the University of Cagliari. His research interests include multibiometrics, cloud computing, cognitive computing, and technology-enhanced learning. Fenu received a master's in computer science from the University of Cagliari. Contact him at mirko.marras@unica.it.

House in the (Biometric) Cloud: A Possible Application

Maria De Marsico
Sapienza University of Rome

Eugenio Nemmi
Sapienza University of Rome

Bardh Prenkaj
Sapienza University of Rome

Gabriele Saturni
Sapienza University of Rome

This article presents a novel approach to extend cloud computing from company services to consumer biometrics. The proposed system recognizes the person at the door, allowing entrance or denying it according to the recognition result. Very little processing is required locally, and biometrics is implemented as a service.

This article does not present a novel technique for either cloud computing or biometric recognition. Rather, it deals with a novel approach to using cloud computing for consumer biometrics in everyday applications, even beyond mobile computing. The wide range of possible tasks includes, of course, security, yet in a wider context than access to company services and spaces.

As an example of this new way of exploiting cloud computing, a smart peephole implemented through remote biometric services can recognize the person at the door and automatically allow or deny entrance according to rules decided by the homeowner. Very little processing is carried out locally, and biometrics is implemented as a service.

Smart Peephole relies on Microsoft Cognitive Services (MCS), included in the Microsoft Azure platform. The user has to install nothing but a camera with a sound capture facility in correspondence with the peephole, and lightweight software able to communicate by (nowadays almost ubiquitous) a home connection with the remote server for biometric recognition. The capture-and-recognition activity is triggered by a movement detection module. The latter carries out a continuous activity, which is implemented locally to avoid a traffic jam on the network. The captured video and audio samples are sent in cascade to the corresponding remote services for recognition. The final result is sent back to possibly trigger a response action. The presented prototype includes face, speech, and emotion recognition. Its aim is to demonstrate the feasibility of the approach.

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

BIOMETRICS-AS-A-SERVICE: FROM COMPANY SERVICES TO HOME APPLICATIONS

Modern powerful and sophisticated devices allow a wide range of applications that have rapidly evolved from company and work environments to the everyday life of "normal" consumers. Mobile computing, especially with smartphones and tablets, is an example of the new frontier reached by the human–computer interaction paradigm. Cloud computing and the possibility of building efficient software through remote suites of functions open even newer lines of application development.

Biometric authentication or recognition has relatively recently joined the group of services that can exploit this technology. Its popularity has increased owing to the possible weaknesses of password-based and token-based approaches.[1] Any human trait—e.g., faces, palm prints, irises, or fingerprints—that is sufficiently discriminative and permanent over time, and universally available, can be used as a key for personal identification.[2] Any biometric system carries out the same workflow, by collecting biometric samples, processing them to extract relevant features, and recording the obtained templates (enrollment) in order to verify a person's identity or identify a person at a later time (recognition).

When the size of the data or the number of possible users increases, both processing and storage resources may become an issue. Moreover, the cost of writing a biometric application in-house or continually updating it can be too high for most users. This is the connection point between biometrics and cloud computing. Thanks to the services offered through the latter, it is possible to rely on smart remote distribution and deployment of storage, resources, and processing tools.

Biometrics-as-a-Service (BaaS) has been around quite stealthily for a while. This is mainly because its use is generally limited to company applications. The HYPR company (www.hypr.com) advertises its BaaS services as a way for companies to save millions in hardware and development. Unlike software as a service (SaaS), BaaS usually is not a single deliverable application but a set of tools that can be integrated through APIs into a new or existing application.

Is it possible to go farther? A recent paper by Jeremy Rose has an interesting question as its title: "Biometrics as a Service: The Next Giant Leap?"[3] The expression "giant leap" usually refers to epochal achievements of science and technology. Is it possible to foresee that this is the case with BaaS? Even normal consumers might be interested in remotely delegating some processing tasks that would require more expensive software as well as hardware equipment. This might spur the introduction of consumer biometric services in the cloud world, thanks to faster, wider, and more affordable communication facilities, and to more and more accurate sensors that are nowadays also embedded in personal devices.

Actually, a first step toward biometrics in consumer settings has been already made. For instance, the announced introduction of advanced face recognition on the latest smartphone models represents in itself a significant change of target with respect to purely security-critical applications, traditionally deployed in critical areas such as airports or restricted spaces. Here we hypothesize that cloud computing can further facilitate the spread of biometrics in smart homes. Microsoft Azure is an example of a cloud-computing platform offering a collection of services that include cognitive services, which in turn include face and voice recognition. Here, we use MCS to demonstrate the feasibility of BaaS in a domestic setting.

Smart Peephole automatically "observes" incoming "objects" (the MCS term for humans, animals, etc.) at a house's doorstep. It can therefore be deployed as a home security system implementing the basic function of a classic door peephole. However, the decision whether to open the door, thus accepting the person "object" in front of the door, or keep it closed, is taken by an automatic system rather than by the homeowner looking through the hole. Figure 1 illustrates the idea.

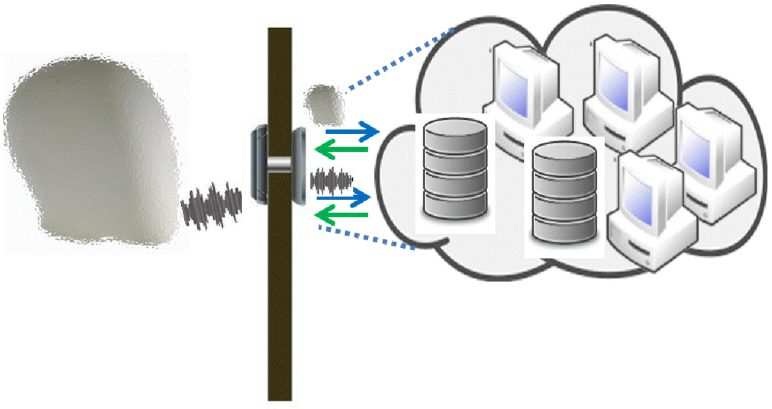


Figure 1. The idea behind Smart Peephole. The decision whether to open the door to a person is made by an automatic system rather than by the homeowner looking through a peephole.

The system is composed of four main modules, implementing different biometric recognition tasks to discriminate between members of the allowed group of people (*members* for short) registered by an administrator, and intruders.

The *Peephole Module* detects movement and recognizes whether, at the moment a movement is detected, there is a face in front of the peephole. Although face detection is available in MCS, this module works locally, in order to avoid a continuous burst of network requests and a server jam.

The *Face Identification Module* includes functions to both register allowed subjects in the system by uploading a number of face photos and identify a person later by matching the incoming face image with the gallery set (the enrolled faces).

The *Voice Verification Module* records a person's voice and uploads it to the system. When someone asks to enter the house, he or she has to speak into an appropriate microphone to be recognized as a member of the allowed group.

The *Emotion Detection Module* aims to catch the mood of the person in front of the peephole, by extracting and analyzing features from that person's facial expression.

SMART PEEPHOLE: DESIGN AND ARCHITECTURE OF BAAS AT HOME

The goal of Smart Peephole to grant secure and easy access to members is achieved by a combination of biometric measurements and modalities, entailing face and voice recognition in a multibiometric cascade. In the *enrollment* phase, the system registers a person by storing an arbitrary number (typically 20) of photos and recording his or her voice pronouncing a chosen passphrase. Both the actual presence of a face in each photo and the quality of the sound are remotely verified through suitable MCS APIs. The *identification* phase is triggered by a sequence of motion detection first and face detection afterward.

Nowadays it is possible to build new applications by the massive use of library functions made available in many contexts. A popular example is represented by OpenCV,[4] which is continually updated with (possibly basic versions of) the most popular algorithms in the literature. The functions provided cover a complete set of preprocessing, feature extraction, and classification techniques, and can be integrated seamlessly into one's own code. The next step is to rely on

frameworks following a black-box model, where only interfaces or APIs are exposed. Finally, the detached module can be a remote one, so that it is possible to consider this as an example of programming by services or cloud computing (SaaS).

A rich set of precoded and verified functions is readily available. Given that the APIs stay unchanged, the algorithms and their implementations may be possibly updated and optimized in a way transparent to the programmer of a final application, even without affecting the existing software.

Microsoft Cognitive Services in Smart Peephole

Along the preceding lines, the Smart Peephole implementation relies on the choice of suitable API procedure calls from MCS. Such services are part of Microsoft Azure (http://azure.microsoft.com),[5] a collection of cloud-computing integrated services that can be used by developers to create, deploy, and manage different kinds of applications across a network. It was born as a solution for enterprise settings;[6] however, we demonstrate here that it can also be used on a smaller scale for consumer applications, like Smart Peephole. During preliminary testing, API method calls were really efficient and demonstrated high responsiveness, thus leading to the decision to exploit them.

The APIs used for the implementation of Smart Peephole are the *Face API*, *Speaker Verification API*, and *Emotion API*.

The Face API provides both face detection with attributes (up to 64 human faces in the same image with high-precision location) and face recognition. The rectangular region of interest indicating the face location in the image is returned along with each detected face. Optionally, the face detection function can extract a series of face-related characteristics such as pose, gender, age, head position, and the presence of facial hair and glasses. The provided face recognition functions are face verification, finding similar faces, face grouping, and person identification.

Smart Peephole exploits the face identification function. It can be used to match a probe subject against the member group created in advance and possibly updated over time. Each group may contain up to 1,000 person objects, each with one or more faces registered. If the probe face is identified as a person object in the group, the person ID will be returned.

Voice has unique characteristics that can be used to identify a person, and can be used to increase the level of security. MCS offers speaker recognition in two modalities: speaker verification and speaker identification. Smart Peephole exploits only the Speaker Verification API. The implemented approach is text dependent; i.e., the user is asked to pronounce a specific passphrase chosen and used during enrollment. The extracted voice features for the chosen phrase form a unique signature. The system finds the list of possible phrases by making another API call. The service requires at least three enrollment captures for each speaker before the profile can be used in verification scenarios. In Smart Peephole, passphrase pronunciation is used in verification modality in cascade after face identification, so that only the template of the person identified by face is further verified by speech.

The Emotion API returns the confidence across a set of emotions for each face in the image, as well as bounding boxes for such faces, using the Face API. If the application has already called the Face API, it can submit the identified region of interest as an optional input. The emotions detected are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. The Emotion Detection Module (an optional one, not strictly related to authentication) allows interacting with the person standing on the doorstep in a more human-like way, by ascertaining his or her mood, and taking a predefined appropriate action.

Most of the core biometric functions planned for the application can be reduced to API function calls that execute on highly efficient machines requiring a minimal temporal cost.

The other side of the coin for this service is the cost for full functionality and storage capacity. However, this is often the case for commercial cloud-distributed resources, and the increased use

of such services may also lead to a decrease in the required expense or to a finer tuning for real user needs. Given the commercial nature of the services, the algorithmic details are not available. In any case, they are beyond the scope of this work.

The Local vs. Remote Choice

In specific cases, according to the application setting, it may be worth partitioning the system operations between local and remote processing, in order to avoid remote jams and much higher service costs. As an example of this strategy, Smart Peephole preliminary detection operations—i.e., motion detection possibly triggering face detection—are carried out locally, although face detection is also available in MCS. In fact, these operations entail a continuous check of the space in front of the peephole, which would cause too much traffic on the network and too much load on the remote server.

The Peephole Module is the core of the proposed application and interacts with all the other modules. Among its functions, it is responsible for movement detection. The module uses dense optical flow (DOPTFlow) based on Gunner Farnebäck's algorithm.[7] It computes the optical flow for all points in a frame and exploits an empirically chosen threshold to infer when a person is in front of the peephole. In more detail, optical flow is the pattern of apparent motion of image objects, caused by the movement of objects or the camera, which can be estimated by matching two consecutive frames. Movement is represented as a field of 2D displacement vectors, each computed according to the difference of a point position from the first frame to the second one. Figure 2 shows an example of the movement caught by the vectors during the execution of the DOPTFlow algorithm.

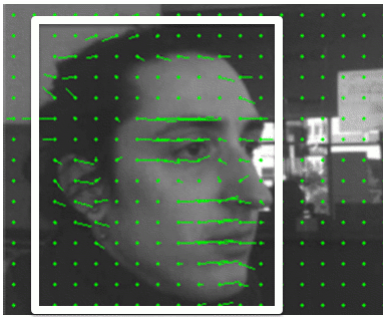


Figure 2. An example of result from the DOPTFlow (dense optical flow) algorithm. This image shows the movement caught by 2D displacement vectors during the algorithm's execution.

If the Peephole Module recognizes some movement, face detection is triggered, using the popular Viola-Jones algorithm[8] implemented in OpenCV. To improve the overall accuracy, if the system detects a face, it also searches for the eyes in order to decrease false positives. The system executes a call to the remote API only if it is confident that it has retrieved a face-like contour.

The Multibiometric, Cascade Modality Choices

Each biometric trait can be affected by distortions due to either intrinsic (e.g., facial expression or hoarseness) or extrinsic (e.g., illumination or noise) factors. To overcome the resulting limitations, a possible effective approach entails using more traits.[9] Since MCS provides both face and video biometric services, Smart Peephole exploits both of them.

In general, when more biometric traits are used, each of them is recognized according to the same modality, either identification or verification. Results are fused at possibly different levels:[10]

- the *feature level* (captured signals are fused, if possible),

- the *score level* (scores achieved separately are combined, possibly adopting some weighting policy), or
- the *decision level* (only final outcomes are combined—e.g., accepted or refused in verification mode).

However, in order to lower both the requested time and the possible cost of the overall operation, Smart Peephole entails different, asymmetric modalities. Face recognition is carried out in (open set) identification mode, since the subject might not be included among members. Afterward, and only if the subject face is identified as one of the enrolled ones, speaker recognition is carried out in verification mode, using the identity returned by the face API as an implicit claim.

Face identification is triggered by face detection and requires a person to stand in front of the camera. The Face Identification Module exploits the face identify function in MCS. This function identifies unknown faces from a person group. For each face in the set of detected ones, it computes similarities between the query face and all the faces in the enrolled group of members, and returns candidate identities ranked by similarity confidence for each query face. Of course, for each query face, the first candidate identity is taken into account, but having a longer list can support, if needed, further inspection of the results. The algorithm allows up to 10 faces to be identified independently in the same request.

This phase of the recognition is more demanding because it entails a 1:*N* matching for each query face. According to experiments carried out, identification works well for frontal and near-frontal faces, but, given the application context, this is not a real limitation. As a matter of fact, people who do not maintain such a pose in front of a peephole are likely trying to avoid recognition; therefore, they could be automatically marked as intruders.

If a valid identity is returned by the Face Identification Module, then the user is prompted to speak into an appropriate microphone for a further recognition operation. This second biometric step is carried out in verification mode, by a 1:1 matching with the identity returned by the face module. The system checks whether the audio recorded corresponds to the template stored in the gallery for the supposed user—i.e., to the pronunciation of the enrollment passphrase with the enrollment voice. In practice, the Voice Verification Module also checks at the same time the correctness of the pronounced passphrase.

A confidence level is returned, with the values Low, Normal, or High, which is associated with the verification result and can be optionally exploited to enforce the verification result. A Low confidence level could mean that the person is not speaking with the same timbre of voice or that the device being used for authentication is not working properly.

Both face identification and speaker verification must meet specific thresholds, which are chosen at system configuration time. If both of these recognition steps lead to a positive result, then the person is allowed to access the house, having been classified as a member with an associated identity. If face recognition fails, the person is rejected without further checking. If face recognition succeeds but speaker verification fails, it is possible to either refuse access or ask the person to provide a secret password, memorized in advance as a further identification element at enrollment time. If the password provided is wrong, then the person is definitively classified as an intruder.

It is worth stressing that the proposed architecture (see Figure 3) is completely modular, and the actual inclusion of the different modules can be adapted during system configuration. For instance, it would be possible to exclude the Voice Verification Module for some members, if this hinders accessibility (e.g., for deaf or mute members). The Password Checker Module can have the goal to reduce the number of false rejects due to speaker recognition, if the homeowner is confident enough, since in this case it would override the presence of the Voice Verification Module. It may further enforce security if, differently from Figure 3, the Password Checker Module is used in addition to the others (three positive responses are requested to grant access), not as an alternative. In our tests, this was omitted. In summary, the architecture is extremely flexible and adaptable.

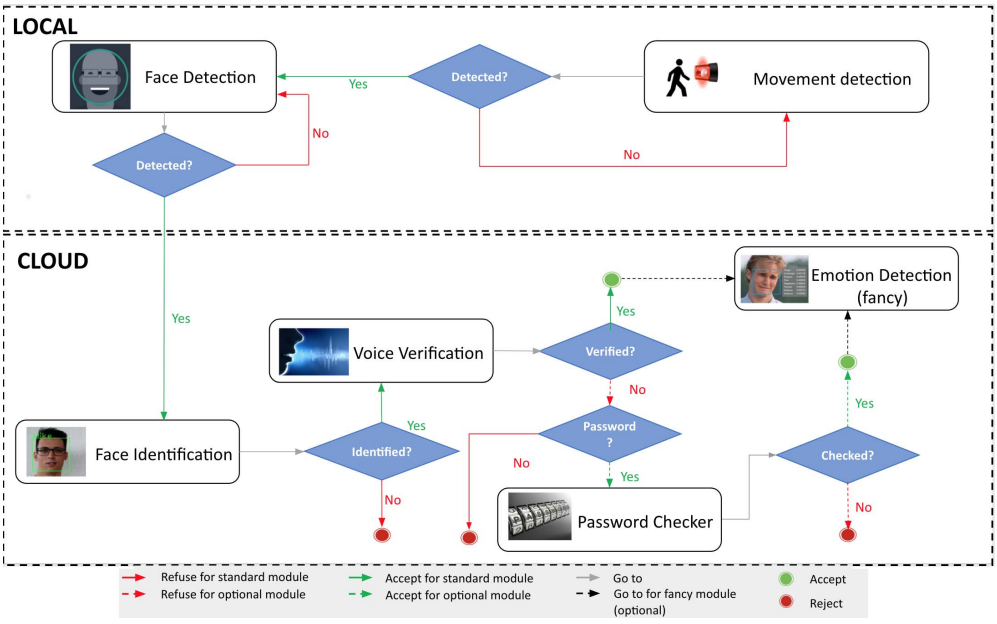


Figure 3. The multibiometric, cloud-based architecture of Smart Peephole.

It is worth mentioning that the performance of open set identification is measured by the same parameters as verification—e.g., the false-acceptance rate (FAR), false-rejection rate (FRR), and equal-error rate (EER, given by FAR = FRR), although it's computed in a different way.[11]

Figure 4 shows the possible correct or incorrect recognition paths without a password module, which at present is beyond the scope of this work.

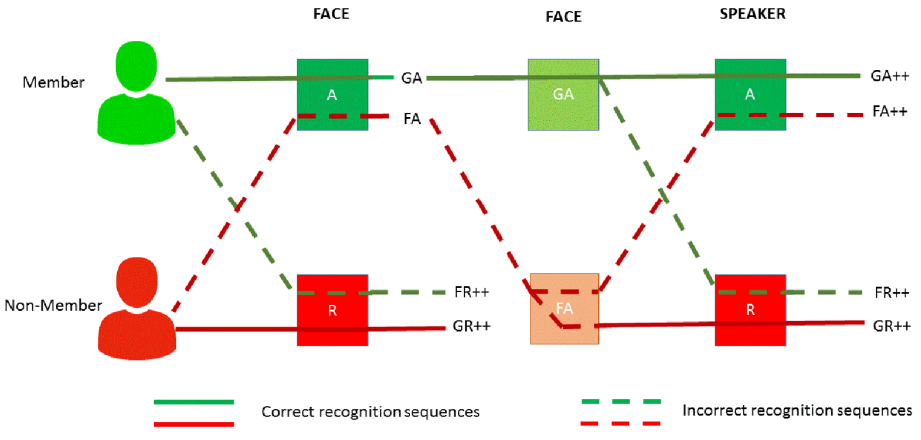


Figure 4. Correct and incorrect recognition paths in the multibiometric cascade. A = acceptance, GA = genuine acceptance, FA = false acceptance, R = rejection, GR = genuine rejection, and FR = false rejection. GA++, FA++, FR++, and GR++ indicate the single increment of the number of results corresponding to the acronyms.

It is reasonable to assume complete independence between face and voice features; therefore, it is possible to also assume independence between the respective recognition results. Accordingly, a coarse evaluation of the combined performance can be given by the following (where GAR and GRR stand for the genuine acceptance rate and genuine rejection rate):

- FAR(Total) = FAR(Face) × FAR(Speaker)
- FRR(Total) = FRR(Face) + (GAR(Face) × FRR(Speaker))
- GAR(Total) = GAR(Face) × GAR(Speaker)
- GRR(Total) = GRR(Face) + (FAR(Face) × GRR(Speaker))

After positive recognition, the optional Emotion Detection Module processes the face images resulting from the previous detection step. According to the response and to the preliminary system setting, the execution flow could vary, producing a different action for each different response code. At present, the system associates each emotion with the playing of a different song as the recognized person goes through the doorway.

The main problem encountered during the implementation is the ambiguous way humans show their emotions, according to the level of expression and possibly influenced by culture and personal character. For example, fear can be wrongly classified as sadness, happiness as neutrality, anger as disgust, and so on.

Figure 3 shows the final architecture and workflow of Smart Peephole.

SOME RESULTS

Of course, the aim of some tests carried out with Smart Peephole was not to evaluate the performances of MCS but to check the actual viability of a domestic solution based on MCS. Therefore, given the kind of application taken into account, large-scale testing was beyond the scope of the experiments carried out. In any case, the interested reader can find some results using a subset of Labeled Faces in the Wild (LWF)[12] in previous work.[13] We say only that, on a subset of LWF consisting of 55 persons (where five persons were genuine and the rest were considered impostors), the face recognition achieved EER = 0.02.

In the same way, since emotion recognition is nothing but an optional addition to the system, no further experiments were carried out after those presented in previous work.[13] Rather, the system was tested in a setting including a low number of members, in contrast with a number of intruders compatible with the possible average frequency by which an either unknown or unregistered person is in front of a private door.

Performance of open set identification was measured by the same parameters as verification: FAR, FRR, and EER. We report here the results from face identification and speaker verification separately and then in combination, according to the recognition protocol implemented by Smart Peephole (see Figure 3 and Figure 4). Moreover, it was interesting to measure the average response times for the different services, in order to assess their suitability for the intended application.

The experiments entailed four registered users (members); face recognition, being the most critical step, was put to the test in a more challenging condition with 30 nonmembers. As for speaker recognition, the same four registered users were each considered as members once, and then three times as nonmembers when their voice was matched against another returned identity, for a total of nine "impostor" attempts. Actually, the impostor attempts were much more than this because they also entailed pronouncing a different passphrase. However, as we had hoped (but which was not completely obvious), the Voice Verification Module always returned a Reject in this case. In order to provide more stringent results, those reported here consider only attempts with exactly the same passphrase. It is worth stressing again that, in any case, our aim was not to measure MCS performance.

The matching protocol elements are summarized in Figure 5, taking into account that not all pairwise matching results are available because registration to MCS was limited to members and no password input was configured as an alternative to speaker recognition. The top section of Figure 5 defines the labels for the ground truth data used for the experiments. The middle section specifies the possible inputs and functions applied to them with the corresponding return conditions. The bottom section describes the classification of the different outcomes.

Labels for ground truth data

Face_ID = real id of input face (either member or nonmember, always available in ground truth)

Speaker_ID = real id of input speaker (either member or nonmember, always available in ground truth)

ISMEMBER (ID) = true if an ID corresponds to a member

If Face_ID = Speaker_ID AND ISMEMBER(Face_ID) = true then genuine

if Face_ID = Speaker_ID and ISMEMBER(Face_ID) = false then impostor

if Face_ID != Speaker_ID then impostor (anyway!)

Inputs, functions and possible return values

PICTURE = input face corresponding to either to a member Face_ID or to an impostor

Face(PICTURE) calls Face identification module for the input face and either returns NULL or an ID

ID(PICTURE) = Face_ID

VOICE = input voice corresponding to either a member Speaker_ID or to an impostor

ID(VOICE) = Speaker_ID

Speaker(ID, VOICE) calls Speaker verification module for the input voice claiming ID identity

If Face(PICTURE) =NULL	Face module returns a Reject and the procedure ends
If Face(PICTURE) = ID_F	the Face module returns the valid id of a member ID_F, that is passed for verification to Speaker module
If Speaker(ID_F, VOICE) = YES	the Speaker module produces a final Accept
If Speaker(ID_F, VOICE) = NO	the Speaker module produces a final Reject

Some relevant examples of test results

If Face(PICTURE) = NULL AND ISMEMBER(ID(PICTURE)) = true	→ FR
If Face(PICTURE) = NULL AND ISMEMBER(ID(PICTURE)) = false	→ GR
If Face(PICTURE) = ID_F AND ID_F = ID(PICTURE) AND ID(VOICE) = Speaker_ID AND ID_F = Speaker_ID AND Speaker(ID_F, VOICE) = YES	→ GA
If Face(PICTURE) = ID_F AND ID_F = ID(PICTURE) AND ID(VOICE) = Speaker_ID AND ID_F = Speaker_ID AND Speaker(ID_F, VOICE) = NO	→ FR
If Face(PICTURE) = ID_F AND ISMEMBER(ID(PICTURE)) = false AND Speaker(ID_F, VOICE) = YES	→ FA
If Face(PICTURE) = ID_F AND ISMEMBER(ID(PICTURE)) = false AND Speaker(ID_F, VOICE) = NO	→ GR

The following might be examples of spoofing attacks with the face of a member but a different voice, therefore the FA can be critical

If Face(PICTURE) = ID_F AND ID_F = ID(PICTURE) AND ID(VOICE)= Speaker_ID AND ID_F != Speaker_ID AND Speaker(ID_F, VOICE) = YES	→ FA
If Face(PICTURE) = ID_F AND ID_F = ID(PICTURE) AND ID(VOICE)= Speaker_ID AND ID_F != Speaker_ID AND Speaker(ID_F, VOICE) = NO	→ GR

The following might be an example caused by family members whose faces resemble each other, therefore the FA is not critical

If Face(PICTURE) = ID_F AND ID_F != ID(PICTURE) AND ISMEMBER(ID(PICTURE)) = true AND Speaker(ID_F, VOICE) = YES	→ FA

Figure 5. The upper section shows the labels for the ground truth data used for the experiments. The center section shows the possible inputs and the functions applied to them, with the corresponding return conditions. The bottom section shows the classification of the different outcomes. GA = genuine acceptance, GR = genuine rejection, FA = false acceptance, and FR = false rejection.

The Face Identification Module alone achieved an optimal EER = 0 (FAR = FRR = 0); therefore, according to Figure 4, the possible system errors are due to the Voice Verification Module. This module alone achieved FAR = 0.07 and FRR = 0. We want to stress that, while the Face API allows choosing an acceptance threshold, this is not possible for the Speaker Verification API, which simply returns an Accept or a Reject answer with a confidence value. We considered all the Accept responses, notwithstanding the associated confidence. Given this and the independence of the two biometric traits, and thanks to the cascade, the final FAR of the system was 0.

It is interesting to consider two special cases that might be missed in classic error analysis. In the first case, if two member faces are very similar, face recognition might erroneously recognize

one as the other. Also, speaker verification might incorrectly succeed as well, since the voices may be very similar. However, this does not cause trouble in the context described here.

The second case is more critical. In a face-spoofing attempt, although not so frequent in the average domestic context, if no antispoofing procedure is added to the modules, face recognition may succeed when the impostor shows a photo of a member. This would be a "false living" rather than a false accept, in the sense that the error would not be the recognition result, but rather the missed detection of a nonliving artifact. The voice of the nonmember attacker might be erroneously recognized as the one associated in the gallery with the member returned by the face module. In this case, the actual FAR is equal to the error by the Voice Verification Module—i.e., 0.07. For a larger dataset, given the EER = 0.02 mentioned above, and still assuming FAR = 0.07 for speaker recognition, we would obtain a very good FAR = 0.0014, which significantly improves the results from both single traits. Moreover, Microsoft is continually improving and updating MCS.

Another interesting aspect to consider regards the response times, given that it is desirable that the answer to the access attempt is as timely as possible. Tables 1 and 2 show the results for the investigated modules, including the minimum, maximum, average, and median time. The response times are converted into seconds and computed over a total of 16 tries.

Table 1. The response time of the Face Identification Module.

	Response time(s)			
	1 person, 6 pictures	**1 person, 12 pictures**	**2 persons, 24 pictures**	**4 persons, 24 pictures**
Max.	1.48	1.37	2.50	5.01
Min.	0.94	0.78	0.87	1.54
Mean	1.08	1.07	1.17	2.10
Median	1.01	1.04	1.06	1.78

Table 2. The response time of the Voice Verification Module.

	Response time(s)		
	1st person (male)	**2nd person (female)**	**3rd person (female)**
Max.	4.4	4.87	4.65
Min.	3.86	3.17	3.23
Mean	4.10	4.14	4.01
Median	4.16	4.24	4.03

In Table 1, response times refer to an increasing number of members, each with a number of stored images ranging from six (the first and fourth columns) to 12 (the second and third columns). Of course, a higher number of images per person increases the possibility of correctly recognizing the person but slightly increases the response time.

Table 2 reports times for voice verification, for different voices. As we mentioned before, in this case, the system always works in verification mode (1:1 matching), and there is a fixed number of samples per user in the system gallery. In any case, the total time includes the transmission time, which is not possible to completely isolate and is not higher than the time required for a real homeowner to open the door (a maximum of less than 10 seconds for four registered members).

It is worth mentioning that some further local processing can be added to obtain the best samples to send to the cloud, in order to decrease the possibility of error. For instance, regarding the face, the best frame from the video can be selected according to some simple methods to evaluate the frontal pose and the illumination,[14] while some antispoofing technique can be applied to avoid simple presentation attacks.[15] In any case, the adopted techniques should be lightweight enough to still assure a timely total response.

CONCLUSION

Cloud computing can become the new frontier of software design, well beyond the enterprise context. In the same way, biometric recognition is spreading out of its traditional use for security applications. It is becoming an interesting service to exploit in smart ambient design too, and its use can be further spurred by the possibility of relying on robust precoded functions made available in the cloud (at possibly reasonable costs).

The aim of this work has been to evaluate the viability of a "domestic-centered" approach to biometrics in the cloud. Of course, when relying on third-party software, a developer inherits both strengths and limitations. Adapting to specific needs may result in a nontrivial task, and missing functionalities may still require expense and effort. For instance, at present no antispoofing function is available in MCS. Its local implementation may require extended equipment, and its design is related to a presently hot research topic. However, the results achieved so far encourage continued investigation along this line.

REFERENCES

1. R. Clarke, "Human identification in information systems: Management challenges and public policy issues," *Information Technology & People*, vol. 7, no. 4, 1994, pp. 6–37; https://www.emeraldinsight.com/doi/abs/10.1108/09593849410076799.
2. S. Prabhakar, A.K. Jain, and A. Ross, "An introduction to biometric recognition," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 14, no. 1, 2004, pp. 4–20; https://ieeexplore.ieee.org/abstract/document/1262027/.
3. J. Rose, "Biometrics as a service: the next giant leap?," *Biometric Technology Today*, vol. 2016, no. 3, 2016, pp. 7–9; https://www.sciencedirect.com/science/article/pii/S0969476516300509.
4. A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*, O'Reilly Media, Inc., 2016.
5. *Truly consistent hybrid cloud with Microsoft Azure*, white paper, Microsoft, 2017; https://azure.microsoft.com/en-us/resources/truly-consistent-hybrid-cloud-with-microsoft-azure/.
6. J. Nickel, *Mastering Identity and Access Management with Microsoft Azure*, Packt Publishing Ltd., 2016.
7. G. Farnebäck, "Two-frame motion estimation based on polynomial expansion.," *Scandinavian conference on Image analysis* (SCIA), Bigun J., Gustavsson T., 2003, pp. 363–370; https://link.springer.com/chapter/10.1007%2F3-540-45103-X_50.
8. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer*

Vision and Pattern Recognition (CVPR 01), 2001, pp. I–I; https://ieeexplore.ieee.org/abstract/document/990517/.
9. A.K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM*, vol. 47, no. 1, 2004, pp. 34–40; https://dl.acm.org/citation.cfm?id=962102.
10. A. Ross and A.K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, 2003, pp. 2115–2125; https://www.sciencedirect.com/science/article/pii/S0167865503000795.
11. P. Grother and P.J. Phillips, "Models of large population recognition performance," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Volume 2)* (CVPR 04), 2004, pp. II–II; https://ieeexplore.ieee.org/abstract/document/1315146/.
12. G.B. Huang et al., *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, technical report Technical Report 07-49, University of Massachusetts, Amherst, 2007; http://cs.brown.edu/courses/cs143/2011/proj4/papers/lfw.pdf.
13. M. De Marsico et al., "A Smart Peephole on the Cloud," *New Trends in Image Analysis and Processing – Proc. of First International Workshop on Biometrics-as-a-Service: Cloud-based Technology, Systems and Applications* (IW-BAAS 17), 2017, pp. 364–374; https://link.springer.com/chapter/10.1007%2F978-3-319-70742-6_34.
14. M. De Marsico, M. Nappi, and D. Riccio, "Measuring measures for face sample quality," *Proceedings of the 3rd international ACM Workshop on Multimedia in Forensics and Intelligence* (MIFOR 11), 2011, pp. 7–12; https://dl.acm.org/citation.cfm?id=2072524.
15. M. De Marsico et al., "Moving face spoofing detection via 3D projective invariants," *5th IAPR International Conference on Biometrics* (ICB 12), 2012, pp. 73–78; https://ieeexplore.ieee.org/abstract/document/6199761/.

ABOUT THE AUTHORS

Maria De Marsico is an associate professor at Sapienza University of Rome. Her main research interests include image processing, biometric systems, and multimodal interaction. She's an area editor of the *IEEE Biometrics Compendium* and the *Biometrics Newsletter*. She's a member of IEEE, ACM, IAPR, EAB, and INSTICC. Contact her at demarsico@di.uniroma1.it.

Eugenio Nemmi is completing his MSc in computer science at Sapienza University of Rome. His research interests include system security and cloud computing. Contact him at eugenio.nemmi@uniroma1.it.

Bardh Prenkaj is completing his MSc in computer science at Sapienza University of Rome. His research interests include AI and biometric systems. Contact him at prenkaj.1602894@studenti.uniroma1.it.

Gabriele Saturni is working on his PhD in computer science at Sapienza University of Rome. His research interests include underwater sensor networks and cloud computing. Contact him at saturni@di.uniroma1.it.

Cognitive and Biometric Approaches to Secure Service Management in Cloud-Based Technologies

Marek R. Ogiela
AGH University of Science and Technology

Lidia Ogiela
AGH University of Science and Technology

This article describes new ideas for applying security procedures to data and service management in cloud and fog computing. Management in cloud computing is presented in connection with cognitive systems supporting management tasks and securing important data. The application of cognitive and biometric features allows creation of personalized procedures oriented at particular users or a group of protocol participants.

The application of perceptual, cognitive, and behavioral features can play an important role in advanced protocols dedicated to secure data management and remote-service provision. In many such applications, to guarantee the highest level of security, it is necessary to use some procedures that can be dedicated to a particular user or a group of participants. To define such human-oriented and secure-management protocols, unique or very characteristic personal features, including behavioral patterns as well as cognitive or visual-perception abilities, can be considered.[1–5]

This article describes selected example applications of personal and perceptual features to develop new secure-management protocols dedicated to data and service management performed in distributed computer infrastructure and in the cloud.[6,7]

Such unique personal or behavioral parameters can be extracted or evaluated by cognitive systems that support evaluating personal patterns and such biometric characteristics as palm or finger movements, body motions, or specific gestures.[4] Apart from such personal characteristics

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

evaluated by cognitive information systems, it is possible to create management procedures based on the perception abilities or the specific knowledge of a particular person. Such protocols can be linked to the application of individual perception thresholds associated with the recognition and interpretation of specific visual data. Once the above personal features are available, we can create secure procedures for data encryption, concealment, and transmission, as well as distributed-service management in fog and cloud environments.[8–10] Such personally oriented technologies can also play an important role in future pervasive-computing technologies and even the Internet of Things.[11,12]

The goal of this article is to present selected examples of the uses of personal features and perceptual abilities for secure-management purposes. The examples presented are based on extracting individual patterns and using them in secure data and service management processes.[10] Such solutions seem very promising for developing future security technologies, especially those based on visual patterns, that can also evaluate specific human habits, motion features, and other human perception capabilities. Such new procedures should expand existing cryptographic methodologies toward a new branch of cognitive cryptography.[4,5]

PERSONAL CHARACTERISTICS IN MANAGEMENT PROTOCOLS FOR CLOUD AND FOG COMPUTING

Different personal characteristics or motion patterns can be used in management protocols. The most important activity with regard to these types of personal features is the registration or extraction of unique or nonstandard personal characteristics, including motion and behavioral patterns. Cognitive-vision systems in connection with multimedia devices such as Leap Motion, Kinect, or motion capture sensors can be used for this extraction or evaluation. In particular, to create a personalized management procedure, we can consider the following types of personal patterns:

- *Palm or finger motion patterns*. Such movements can be recorded using Leap Motion or Kinect sensors. The features of such movements can then be evaluated using cognitive information systems that allow informative motion characteristics to be extracted.
- *Movement patterns of human body parts*. Such patterns can be recorded using motion capture devices connected to cognitive-vision systems.
- *Specific and very complex motion patterns, having the form of extreme exercises, sport or acrobatic techniques, etc*. The analysis of such patterns seems to be most difficult and requires the application of professional vision equipment as well as more sophisticated visual analyses allowing specific features to be extracted.

Palm or hand motion patterns can consist of simple gestures performed using a single finger, multiple fingers or even the whole hand, or two-hand gestures. In this group we can define fixed patterns, user-defined ones (available for several individuals only), natural gestures, or patterns imitating the handwriting of signatures.

Among body movements, it is particularly worthwhile to consider the very specific natural gait, which can be characteristic for particular individuals, or simple exercises that can be performed by most people in different, personalized ways.

Among the last group of complex motion patterns, it is possible to consider only advanced movements observed in sports, gymnastics, or dances and possible for only a small group of persons with great physical skills. Here, it is possible to analyze longer motion sequences presenting special movements learned over a longer time that other performers would have difficulty quickly repeating.

The personal features extracted should be stored in a personal-feature vector that can contain as much information as possible and describes biometric features or motion parameters, etc. For each application, it will be possible to select a small subset of such features that can be applied in a particular protocol.

The analysis of the above motion patterns supports extracting personal parameters that can be applied in secure-management procedures. The possibilities of analyzing such complex patterns are also strongly dependent on personal habits, physical abilities, age, etc.

One of the most important examples of personal-feature use in security protocols is associated with management protocols in a fog–cloud infrastructure. Such management solutions are very promising for the development of modern security protocols dedicated to different distributed infrastructures.

Many resources of strategic data or core information may be stored in the cloud with possible access by authorized persons, working at different hierarchical levels or nodes of this distributed infrastructure.[13–16] Linguistic or biometric threshold procedures executed by trusted instances at the fog level can be used in such a data distribution protocol.[17,18] This allows sharing any important resources using cognitive-cryptography approaches[19,20] and gives a group of users who have been granted the same security access the ability to reconstruct these resources. The shared information can be distributed between different groups of users, depending on the distribution keys established for them. The encrypted data can be decrypted once a group of trusted persons combine their secret parts into the one original piece of information.

Such a distribution procedure is also applicable to combining different parts of information stored in several places and encrypted with different keys associated with users or peers. The final information encrypted and stored in the cloud can be revealed by trusted instances, working at the fog level, but only when all the independent encryption keys are available. The idea of this protocol is presented in Figure 1.

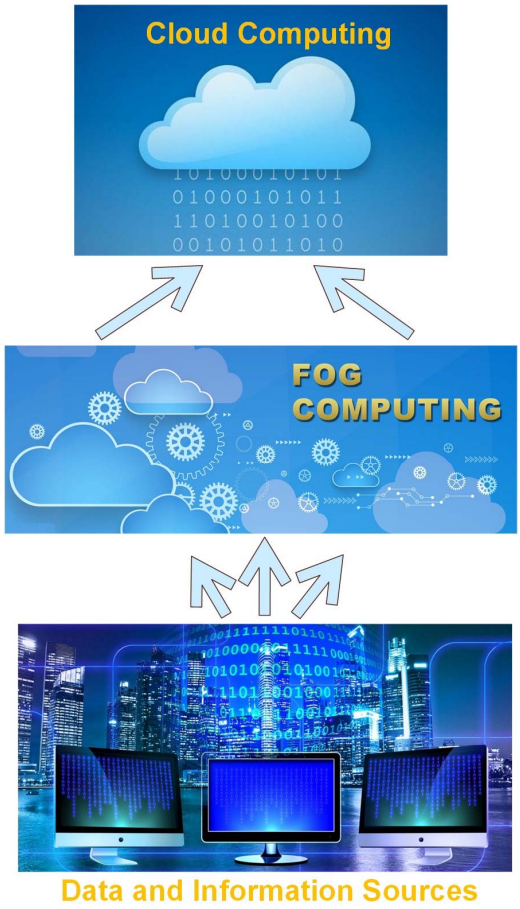


Figure 1. The flow of information in data-sharing procedures in fog and cloud architectures.

Thus, the whole procedure works as follows. At the bottom level, all nodes or peers simply register signals or acquire important information. Some of them—representing users—can also acquire personal features that form the keys for accessing encoded information or distributed data. If such data represents a big repository and should be preprocessed, this can also be done at the second level—i.e., the fog level. At this level, more sophisticated computing can be performed, and data can be fused. At the highest level, the protected information can be stored, shared, or made accessible to other users. In addition, the highest level (the cloud) is responsible for providing the services required by the lower layers and final users.

CAPTCHAS IN MANAGEMENT PROTOCOLS

A captcha is a special type of user verification procedure that confirms that a task was executed by a human, not a computer. Such verification codes can also be used for information sharing and providing services in cloud or fog infrastructures. The verification procedures usually performed with captchas consist of several levels, at which different verification techniques may be executed. It is possible to verify all users by confirming their understanding of all the commands executed.

The same situation can take place when we wish to restore any divided secret information or gain access to remote services or resources. Visual codes allow access to data or services to be granted to all participants who have made the correct decision or answered a question correctly. Visual codes are used in cryptographic protocols dedicated to data splitting and sharing. Captcha techniques allow creation of a lock securing access to secret information. In this solution, it is necessary to correctly select any combination of elements that fulfil particular conditions or have a particular meaning. For example, we can select all or only a small number of visual parts representing specific information (see Figure 2).

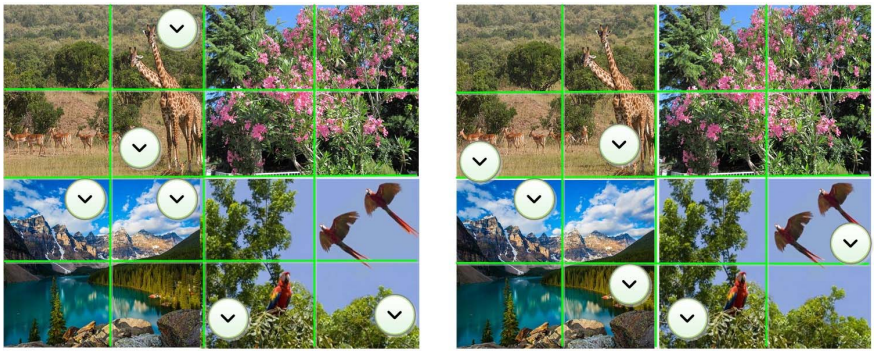


Figure 2. Examples of a visual captcha for a verification procedure with the question “Where is the animal?” or “Where do you see the mountain?”

Figure 2 presents two possible solutions in which users are verified with any two combinations of correct captchas and are allowed to access data or service resources. In service management procedures, it is possible to specify the required information such as the type of proposed services, the significance and quality of the services, and the cost and availability of the services provided.

PERCEPTION ABILITIES IN MANAGEMENT PROTOCOLS

Perception abilities can also be used in visual cryptography.[4] For such protocols, we can establish individual perceptual thresholds for each person or participant of a data or service management procedure. Very often, people notice different details when inspecting an image or picture,

or even recognize objects in a different way. In such cases of considering or adding extra information or secret parts, it is also possible to receive enough data to properly recognize the source image.[5]

Hence, in this procedure, it is possible to establish personal perception thresholds of particular users. Exceeding such an individual threshold established for a particular user allows him or her to properly recognize the original pattern, but other individuals cannot do this in the same way. Such a personal threshold depends on the personal expectation or knowledge of the content of the original images.

Figure 3 presents an example of such a threshold. Users without any knowledge or expectation of the content of this image of a man's head have to collect all the secret parts to recognize it. However, those who have some additional knowledge (e.g., individuals who know the man) can recognize him at an early stage.

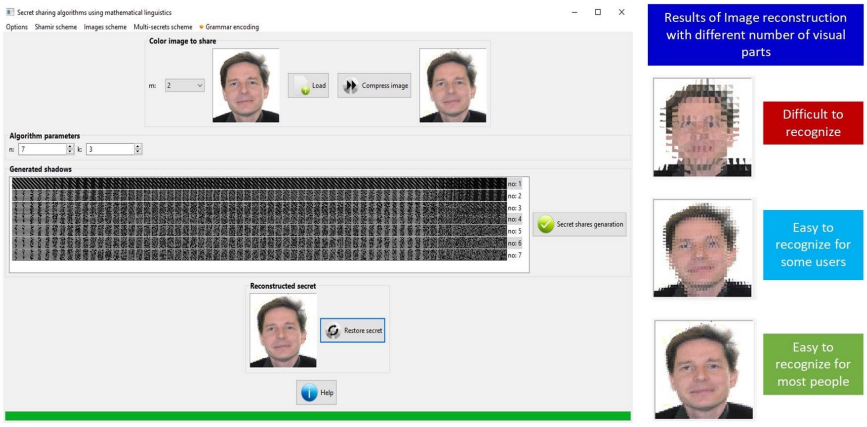


Figure 3. An example of image reconstruction based on a different number of secret parts (using a photo of one of the authors of this article).

SECURITY FEATURES OF MANAGEMENT PROTOCOLS

The use of perceptional and behavioral features in secure-management protocols supports the development of efficient and universal procedures dedicated specifically to groups of trusted users or single authorized individuals. During the development of personal secure-management protocols, it is possible to use several types of individual features and biometric patterns connected with morphometric parameters of human body parts, as well as behavioral features and perception abilities.

The protocols presented have several security features, which can be summarized as follows:

- Data and service management protocols are secure from the cryptographic point of view owing to the application of basic secure cryptographic algorithms that may be expanded to include the use of personal, behavioral, and perceptual features.
- Management protocols are fast and efficient. The application of personal or behavioral features requires evaluating specific parameters only once in the whole procedure. A personal-feature record can then be used many times for different management purposes.
- User-specific behavioral or perceptual characteristics should be stored in a safe place from which they may be taken when a personalized management protocol is to be performed.
- Personal characteristics can be evaluated in real time, and the protocol should prevent the use of counterfeit records that do not represent a live person in real time.
- The complexity of such protocols remains at the same polynomial level as that of the cognitive information systems that should be used to evaluate personal features.

CONCLUSION

This article describes some possibilities of using behavioral and perceptional features in secure information management protocols executed in cloud and fog infrastructures. Personal characteristics can be extracted from motion sequences presenting unique movements, and can then be used in management procedures aimed at information sharing, encryption, and distribution. Besides simple body movements, security protocols can make use of visual abilities and perceptive skills. Perceptual or behavioral features can be extracted by cognitive-vision systems, which allow personal unique parameters to be evaluated in specific human actions. The main idea of applying the above features in cloud and fog service management processes stems from the secure authentication and verification procedures.

The universality of such protocols allows them to be used at different management levels, depending on the infrastructure and the available cloud resources. All simple processes can be realized at a lower level on personal workstations or at the middle level in the fog. They can be also performed at the highest level—in the cloud—using high-performance infrastructure. The low-level analysis is usually to secure procedures and the authorization aspects tied to particular services. At the high level, it is also possible to use personal, visual, or behavioral analyses dedicated to personal cryptographic solutions and verification procedures.

The approach presented may also be expanded to allow its use by individuals with minor cognitive deficits. In such applications, it is necessary to source mainly personal features that do not change over time and are linked mainly to biometric patterns for security purposes. It is allowed to use some independent features that can be acquired without specific physical skills and can change over time.

ACKNOWLEDGMENTS

This work has been supported by the National Science Centre, Poland, under project DEC-2016/23/B/HS4/00616.

REFERENCES

1. M Gomez-Barrero et al., “Multi-biometric template protection based on Homomorphic Encryption,” *Pattern Recognition*, vol. 67, 2017, pp. 149–163.
2. C. Hahn and J. Hur, “Efficient and privacy-preserving biometric identification in cloud,” *ICT Express*, vol. 2, no. 3, 2016, pp. 135–139.
3. L Han, Q Xie, and W Liu, “An improved biometric based authentication scheme with user anonymity using elliptic curve cryptosystem,” *International Journal of Network Security*, vol. 19, no. 3, 2017, pp. 469–478.
4. M.R. Ogiela and L. Ogiela, “On Using Cognitive Models in Cryptography,” *The IEEE 30th International Conference on Advanced Information Networking and Applications* (AINA 16), 2016, pp. 1055–1058.
5. M.R. Ogiela and L. Ogiela, “Cognitive Keys in Personalized Cryptography,” *The 31st IEEE International Conference on Advanced Information Networking and Applications* (AINA 17), 2017, pp. 1050–1054.
6. M. Gregg and B. Schneier, *Security Practitioner and Cryptography Handbook and Study Guide Set*, Wiley, 2014.
7. L. Ogiela, *Cognitive information systems in management sciences*, Elsevier, Academic Press, 2017.
8. S. Costache et al., “Resource management in cloud platform as a service systems: Analysis and opportunities,” *Journal of Systems and Software*, vol. 132, 2017, pp. 98–118.
9. P. Hu et al., “Survey on fog computing: architecture, key technologies, applications and open issues,” *Journal of Network and Computer Applications*, vol. 98, 2017, pp. 27–42.

10. L. Ogiela and M.R. Ogiela, "Management Information Systems," *Lecture Notes in Electrical Engineering*, LNEE 331, Springer, 2015; doi.org/10.1007/978-94-017-9618-7_44.
11. L. Ogiela, "Towards cognitive economy," *Soft Computing*, vol. 18, no. 9, 2014, pp. 1675–1683.
12. L. Ogiela and M.R. Ogiela, "Data mining and semantic inference in cognitive systems," *2014 IEEE International Conference on Intelligent Networking and Collaborative Systems* (INCoS 14), 2014, pp. 257–261.
13. A. Castiglione, A. De Santis, and B. Masucci, "Hierarchical and Shared Key Assignment," *17th International Conference on Network-Based Information Systems*, 2014, pp. 263–270.
14. A Castiglione et al., "Supporting dynamic updates in storage clouds with the Akl–Taylor scheme," *Information Sciences*, vol. 387, no. C, 2017, pp. 56–74.
15. A. Castiglione et al., "Hierarchical and Shared Access Control," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, 2016, pp. 850–865.
16. A. Castiglione et al., "Cryptographic Hierarchical Access Control for Dynamic Structures," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, 2016, pp. 2349–2364.
17. M.R. Ogiela and U. Ogiela, "Linguistic Approach to Cryptographic Data Sharing," *The 2nd International Conference on Future Generation Communication and Networking* (FGCN 08), 2008, pp. 377–380.
18. M.R. Ogiela and U. Ogiela, "Grammar Encoding in DNA-Like Secret Sharing Infrastructure," *Proceedings of the 2010 International Conference on Advances in Computer Science and Information Technology* (AST/UCMA/ISA/ACN 10), 2010, pp. 175–182.
19. M.R. Ogiela and U. Ogiela, "Secure Information Management in Hierarchical Structures," *3rd International Conference on Advanced Science and Technology* (AST 11), 2011, pp. 1–8; 10.1109/ICITST.2009.5402551.
20. M.R. Ogiela and U. Ogiela, *Secure Information Management using Linguistic Threshold Approach*, Springer, 2014.

ABOUT THE AUTHORS

Marek R. Ogiela is a professor of computer science, a cognitive scientist, a cryptographer, and the head of the Cryptography and Cognitive Informatics Laboratory at the AGH University of Science and Technology. His research interests include pattern recognition, cognitive image analysis and semantic understanding, and cryptographic threshold schemes. Ogiela received a habilitation in computer science from the AGH University of Science and Technology. He's an SPIE Fellow, an IEEE Senior Member, and a member of the Interdisciplinary Scientific Committee of the Polish Academy of Arts and Sciences. Contact him at mogiela@agh.edu.pl.

Lidia Ogiela is a computer scientist, a mathematician, an economist, and a member of the Cryptography and Cognitive Informatics Laboratory at the AGH University of Science and Technology. Her research interests include cognitive analysis techniques and their application in intelligent information systems. Ogiela received a habilitation from the Faculty of Electrical Engineering and Computer Science at VŠB—Technical University of Ostrava. She's a member of IEEE, SIAM, SPIE, and the Cognitive Science Society. Contact her at logiela@agh.edu.pl.

Secure Data Collection, Storage, and Access in Cloud-Assisted IoT

Wei Wang
Huazhong University of Science and Technology

Peng Xu
Huazhong University of Science and Technology, Shenzhen Huazhong University of Science and Technology Research Institute

Laurence Tianruo Yang
Huazhong University of Science and Technology and St. Francis Xavier University

The cloud-assisted Internet of Things (IoT) provides a promising solution to data booming problems for the ability constraints of individual objects. However, with the leverage of cloud, IoT faces new security challenges for data mutuality between two parties, which is introduced for the first time in this paper and not currently addressed by traditional approaches. We investigate a secure cloud-assisted IoT data managing method to protect data confidentiality when collecting, storing, and accessing IoT data while limiting to effects of IoT scalability. We further present numerical results to show that the method is practical.

Recently, versatile IoT systems have been widely deployed in daily life, for example, in health care and traffic monitoring, which generate giga-level high-definition images and videos every minute. Massive IoT data require impractically large storage and high-performance computation that a normal user or smart object within IoT hardly supports. Cloud-assisted IoT is popularly applied to leverage the computation and storage capability of a cloud for massive IoT data.[1] A cloud is a powerful platform that can provide additional conveniences as a data distribution delegate. When an IoT user has legal requests for certain data being collected, stored, and accessed, he can directly delegate the requests to the cloud at any time with greater convenience.

However, the convenience that cloud brings to IoT comes at the cost of potentially new security risks, which have never been considered in a traditional IoT system. In both theory and practice, a cloud is widely recognized as an honest-but-curious party.[2] This means that a cloud will handle user-delegated tasks but hardly guarantee confidentiality of user data.

This disadvantage is a critical obstacle when building any cloud-assisted IoT system. Moreover, overcoming these security challenges is a big problem due to the versatile functions of cloud-

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

assisted IoT systems and the versatile security requirements of users. Unlike the security of traditional IoT,[3] this type of problem cannot be perfectly solved in a short time period. Normally, a trust-based system is applied as a solution to those risks. However, trust-based systems cannot provide provable security, which lowers the IoT security level. Similarly, in IoT scenarios, it is not practical to apply data anonymization and obfuscation[4] to guarantee security for dynamic operations (insertion or deletion,) especially when merging with the cloud, since they are applied for privacy preservation without provable security. We therefore propose a framework based on cryptographic methods to support data security in cloud-assisted IoT. In this article, traditional IoT refers to IoT without the assistance of a cloud. Except for confidentiality, other types of security issues such as integrity and authentication are not considered.

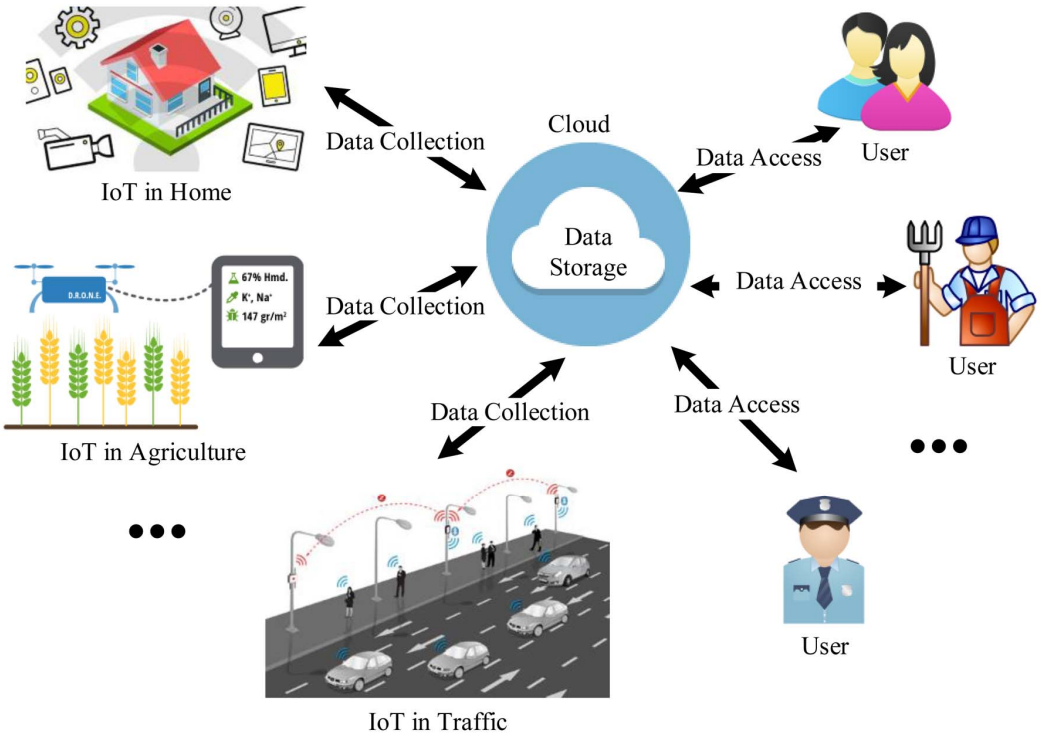


Figure 1. Some fundamental functions of a cloud-assisted IoT

In this article, we abstract some fundamental models of data transfer in a cloud-assisted IoT system (as shown in Figure 1), and discuss promising cryptographic methods to keep data confidentiality in these models. In a cloud-assisted IoT system, users can delegate their data collection tasks to the cloud, store their IoT data in the cloud and access the expected data on the cloud. Correspondingly, the first challenge is that IoT security models should be modified to define potential attackers that could appear when bridging IoT data in the cloud. The second challenge is the impacts of security to IoT scalability. Encrypted cipher will result in incremental burdens to the system as the number of users increases.

We investigate the proxy re-encryption scheme called Conditional Identity-based Broadcast Proxy re-encryption (CIBPRE)[5] and find it applicable to our security models. Hence, we respectively propose protocols to the data transfer models based on CIBPRE. The new method appears promising with provable security when re-encrypting data in the data collection step and extending keys to users by the key generation center (KGC). Our method could resist attacks from both IoT insiders and outsiders seeking to breach data confidentiality without increasing communication cost as the number of IoT users increases.

In the following sections, we mainly focus on our promising solutions against risks in the fundamental data transfer models. We respectively discuss our choice on the proper encryption methods in cloud-assisted IoT, review our CIBPRE schemes and introduce the secure data transfer protocols based on the CIBPRE scheme and encryption methods to achieve confidentiality when collecting, storing, and accessing IoT data with the assistance of a cloud. Finally, we analyze the performances of our proposed protocols and comparisons with traditional identity-based encryption (IBE) and public-key encryption (PKE) schemes.

SECURITY RISKS AND CHALLENGES

Over the past decade, research on the confidentiality of the traditional IoT has attracted a lot of attention. As a result, many cryptographic methods were proposed to guarantee IoT data confidentiality, while saving as much of the computation and communication costs as possible due to the limited IoT capabilities.[6–7] These previous works provide a fundamental background for considering the risks and challenges in cloud-assisted IoT when a cloud is leveraged.

Accordingly, we divide all smart objects of IoT into two categories: IoT-inside objects and IoT-edge objects. IoT-inside objects only communicate with other objects in an IoT, whereas IoT-edge objects are smart objects that also communicate with cloud servers. In this article, we only consider the new problems of IoT-edge confidentiality, which are caused by mutuality between IoT-edge objects and cloud servers. We omit IoT-inside confidentiality, or the confidentiality of data communicated among smart objects, which discussed elsewhere.[8-10]

In this section, we discuss IoT-edge confidentiality security models. These models establish rules, including who could be attackers seeking to break the confidentiality of cloud-assisted IoTs, and some of the challenges with adopting encryption methods to resist these attackers. Referring to Figure 1, the possible attackers are as follows:

- **IoT-edge objects.** In practice, such adversarial objects could be caused by people illegally controlling existing objects or forging new objects to join IoT systems. These adversarial objects are inside attackers that steal sensitive data from other legal objects.
- **The honest-but-curious cloud** (as mentioned in the first section). Some well-known and traditional protection strategies are ineffective for achieving our desired objectives.
- **Users that access IoT data from the cloud.** The cloud is a public platform that provides data collection, storage, and access services for multiple users. In practice, different users obviously have different rights to access different IoT data. Some users may be curious about other users' IoT data.
- **Eavesdroppers.** An eavesdropper can obtain all the transferred data, such as the data transferred between an IoT-edge object and the cloud, and the data between the cloud and users. We do not consider an eavesdropper that would like to eavesdrop on the inside communication of an IoT system because such an eavesdropper has been widely considered in previous works.[6-7]

Singh and colleagues explored some of the possible risks of cloud-assisted IoT, listing the data confidentiality of IoT as essential, but did not present effective solutions.[11] Similar works also looked at preventing these risks via traditional PKI schemes that we list in the following section.[12] Most previous works are based on traditional cryptographic methods.

Our work consists of three phases: security guarantees on data collection, data storage, and data access. Different phases consider different attackers. In the data collection phase, the main attacks are usually caused by adversarial IoT-edge objects and eavesdroppers. The main attackers in the data storage phase are users that access IoT data from the cloud. In the data access phase, the adversarial users and eavesdroppers are the main attackers.

According to the different characteristics of attackers, the following challenges must be addressed:

- To resist eavesdroppers, all communications should be made via a secure channel or encrypted. Moreover, no eavesdropper should know the decryption keys.

- To resist adversarial IoT-edge objects, all IoT-edge objects must have different keys to encrypt their data in the data collection phase. In other words, no object can decrypt another objects' ciphertexts.
- To resist the risk caused by the cloud, all IoT data are stored as ciphertexts in the cloud. Moreover, the cloud cannot decrypt any ciphertext.
- To resist adversarial users, traditional methods of access control are ineffective. Traditional access control allows a server to respond to a user's data request if the user has the corresponding right. All data are usually stored as plaintexts in the server. Hence, the traditional access control can resist the adversarial users if the server is fully trusted. Otherwise, the server will bypass the traditional access control and directly send sensitive data to the adversarial users. Because the cloud is honest-but-curious, it is clear that traditional access control cannot be used to achieve our objectives. In the following section, we suggest that encryption-based access control is a promising solution.

In summary, the above discussions demonstrate the security challenges in our work and suggest that encryption is a promising method to address them. However, the use of encryption alone does not address all of our objectives. In cryptography, there are many different encryption methods that have distinct properties. Our next task is choosing a specific encryption method.

ENCRYPTION SCHEMES IN A CLOUD-ASSISTED IOT

The first step in establishing a secure cloud-assisted IoT is to choose perfect encryption methods between two categorized schemes, public-key encryption (PKE) and symmetric-key encryption (SKE).

The primary difference between PKE and SKE is whether to apply asymmetric key or symmetric keys. When adopting PKE in cloud-assisted IoT, users and IoT-edge objects do not need to be online simultaneously. In contrast, with SKE, users and IoT-edge objects must be online simultaneously.

PKE usually requires much more time than SKE to generate a ciphertext. However, the execution time is not a significant weakness of PKE since the time cost of PKE is not directly applied to the file but to its private key. When encrypting a file with a slightly larger size, the time cost of PKE will not be the main factor affecting the performance of the cloud-assisted IoT system.

In summary, PKE (especially identity-based encryption [IBE]), is a better choice than SKE. Recall that cloud is helpful for IoT because more and more IoT systems generate massive data that users typically do not have the capacity to handle. For massive IoT data, the time cost of PKE will not be a significant factor affecting the performance of the cloud-assisted IoT system.

When employing PKE in the cloud-assisted IoT system of our work, all IoT-edge objects and users have individual public and private keys, and private keys are used to decrypt the corresponding PKE ciphertexts. Normally, a public-key management system such as public-key infrastructure (PKI)[13] is required; otherwise, the intended receiver's public key may not be obtained. However, the management system could be too complex for a cloud-assisted IoT system to be practical. Hence, we introduce IBE[14] to avoid the requirement of a management system. In IBE, anyone can assume a public identity as a public key. Hence, it is very easy to obtain others' identities. Conclusively, IBE is a promising choice to our secure cloud-assisted IoT. Owing to its unique properties, we apply a specific IBE scheme proposed in our previous work[5] in this article.

A SPECIFIC IBE SCHEME

We previously proposed a special IBE scheme called conditional identity-based broadcast proxy re-encryption (CIBPRE).[5] In a CIBPRE system, a trusted KGC initializes the system parameters of CIBPRE and generates private keys for users. To confidentially share data with multiple receivers, a sender can encrypt the data with the intended receivers' identities under a data-sharing condition.

When receiving the encrypted data, these receivers can independently decrypt the data using their private keys. If the sender would later also like to share the data associated with the same condition with other receivers, the sender can delegate a re-encryption key labeled with the condition to the proxy. Then, the proxy can re-encrypt the initial ciphertexts matching the condition to the resulting receiver set. When receiving the re-encrypted ciphertexts, these new receivers can independently decrypt the data using their private keys. The initial ciphertexts may be stored remotely while being kept secret. The sender does not need to download and re-encrypt repetitively but can rather delegate a single key matching the condition to the proxy. These features make CIBPRE a versatile tool for securing remotely stored data, especially when there are different receivers to share the data.

Let $N \in \mathbb{N}$ be the maximal size of the receiver set for one CIBPRE encryption or re-encryption. Let $(X, \mathbf{SE}_x, \mathbf{SD}_x)$ be a SKE scheme such as AES (the popular choice in practice), where X is the symmetric-key space and $\mathbf{SE}_x$ and $\mathbf{SD}_x$ respectively denote the encryption and decryption algorithms, both with a symmetric key $x \in X$. CIBPRE consists of following algorithms:

- **Setup**(λ, N): Given a security parameter $\lambda \in N$ and value N, this algorithm outputs the master public parameters **PK** and the master secret parameters **MK**, where $(X, \mathbf{SE}_x, \mathbf{SD}_x) \subset \mathbf{PK}$.
- **Extract**(**MK**, *ID*): Given **MK** and an identity *ID*, this algorithm outputs the private key SK_{ID}.
- **Enc**(**PK**, $\mathcal{S}$, F, α): Given **PK**, a set $\mathcal{S}$ of some identities (where $|\mathcal{S}| \leq N$), data F and a condition α, this algorithm randomly chooses a secret key $k \subseteq X$, generates an initial CIBPRE ciphertext C_1 of k and a SKE ciphertext $C_2 = \mathbf{SE}_k(F)$ of F and outputs an initial ciphertext $C = (C_1, C_2)$.
- **RKExtract**(**PK**, *ID*, SK_{ID}, $\mathcal{S}'$, α): Given **PK**, an identity *ID* and its private key SK_{ID}, a set $\mathcal{S}'$ of some identities (where $|\mathcal{S}'| \leq N$) and a condition α, this algorithm outputs a re-encryption key $d_{ID \to \mathcal{S}'|\alpha}$.
- **ReEnc**(**PK**, $d_{ID \to \mathcal{S}'|\alpha}$, C, $\mathcal{S}$): Given **PK**, a re-encryption key $d_{ID \to \mathcal{S}'|\alpha}$, an initial ciphertext $C = (C_1, C_2)$ and a set $\mathcal{S}$ of some identities (where $|\mathcal{S}| \leq N$), this algorithm generates a re-encrypted CIBPRE ciphertext $\tilde{C}_1$ of C_2 and outputs a re-encrypted ciphertext $\tilde{C} = (\tilde{C}_1, C_2)$.
- **Dec-1**(**PK**, *ID*, SK_{ID}, C, $\mathcal{S}$): Given **PK**, an identity *ID* and its private key SK_{ID}, an initial ciphertext $C = (C_1, C_2)$, and a set $\mathcal{S}$ of some identities (where $|\mathcal{S}| \leq N$), if $ID \in \mathcal{S}$, this algorithm decrypts the initial CIBPRE ciphertext C_1 to get a secret key k and outputs data $F = \mathbf{SD}_k(C_2)$.
- **Dec-2**(**PK**, ID', SK_{ID}, $\tilde{C}$, $\mathcal{S}'$): Given **PK**, an identity *ID* and its private key SK_{ID}, a re-encrypted ciphertext $\tilde{C} = (\tilde{C}_1, C_2)$ and a set $\mathcal{S}'$ of some identities (where $|\mathcal{S}'| \leq N$), if $ID' \in \mathcal{S}$, this algorithm decrypts the re-encrypted CIBPRE ciphertext $\tilde{C}_1$ to get a secret key k and outputs data $F = \mathbf{SD}_k(C_2)$.

Additional mathematical details about CIBPRE are left out of this paper due to the limited space.

CONFIDENTIAL DATA COLLECTION

In this section, we describe how to apply CIBPRE to achieve confidential data collection in a cloud-assisted IoT system. In the data collection phase, a user (note that an IoT-edge object can also be a user) can delegate a data collection task to the cloud, and all related IoT-edge objects then upload their data to the cloud. As we have mentioned previously, all IoT-edge objects must have different keys to encrypt their data before sending their data to the cloud. This requirement

is easy with CIBPRE because the real key for encrypting data in algorithm **Enc** is randomly chosen. Suppose that KGC has published the generated master public parameters **PK** and has generated private keys for all IoT-edge objects and users and that KGC never generates any private key for the cloud (this assumption is valid because no one would like to generate such critical information for a potential attacker). These assumptions are also valid in the other phases of our work. Figure 2 shows the main steps of the data collection phase.

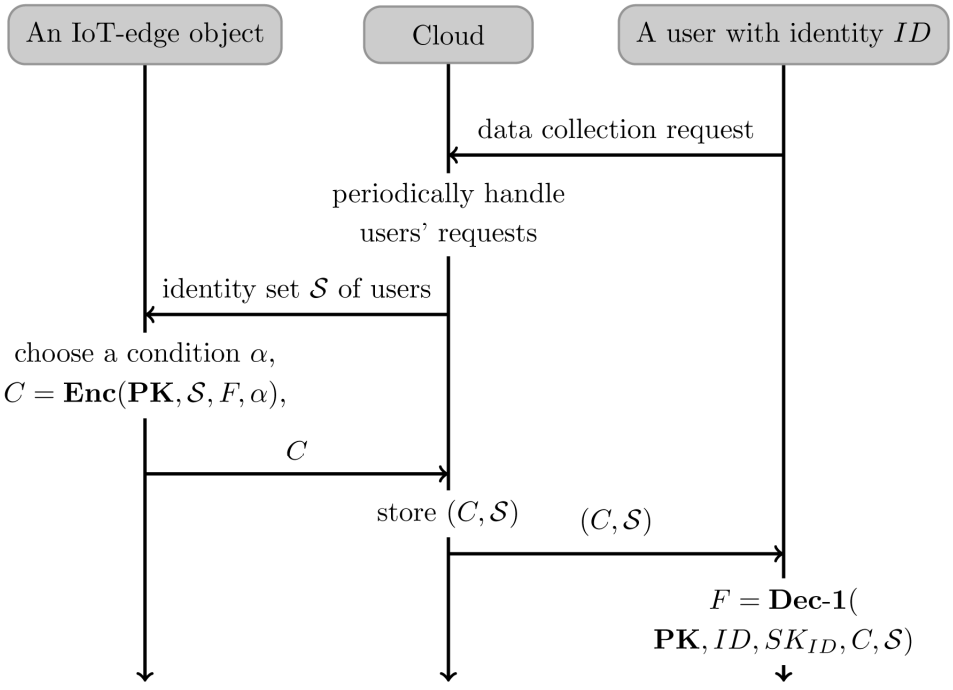


Figure 2. The main steps of the data collection phase.

The details are described as follows:

1. Users send their data collection request and identities to the cloud.
2. The cloud periodically handles users' requests. Suppose that the users in identity set $\mathcal{S}$ want to collect the data F from an IoT-edge object. The cloud sends the set $\mathcal{S}$ and data collection request to the IoT-edge object.
3. The IoT-edge object verifies the users' identities in the set $\mathcal{S}$ and eliminates the invalid identities. Suppose that all users in the set $\mathcal{S}$ are valid. The IoT-edge object chooses a data-sharing condition α (which will be useful in the data access phase), encrypts the data F by algorithm C = **Enc**(**PK**, $\mathcal{S}$, F, α), and sends the ciphertext C to the cloud.
4. The cloud stores the ciphertext C and the set $\mathcal{S}$ and forwards (C, $\mathcal{S}$) to the users in $\mathcal{S}$.
5. All users in $\mathcal{S}$ independently decrypt the data F using algorithm **Dec-1**(**PK**, ID, SK_{ID}, C, $\mathcal{S}$).

In step 2 above, we allow the cloud to periodically handle users' requests. This method can save us the communication cost associated with ciphertexts because CIBPRE allows the IoT-edge object to generate a constant-size ciphertext for multiple users. However, if the cloud handles user requests one by one, it is obvious that the size of the generated ciphertexts is linearly related to the number of users. In addition, a user can start a data collection request, as is done in the above data collection phase, and an IoT-edge object can also actively start a data collection task by itself. To achieve this, an IoT-edge object encrypts its data by its own identity using algorithm **Enc** and uploads the generated initial ciphertext to the cloud. Because this step is very easy, we omit it in the above data collection phase.

With the respect of confidentiality, all data are transferred as ciphertexts. According to the confidentiality of CIBPRE, only the users in the set $\mathcal{S}$ can decrypt the ciphertext C. In other words, none of the eavesdroppers, cloud and non-intended users can learn anything about the data F from the ciphertext C.

CONFIDENTIAL DATA STORAGE

According to the above data collection phase, it is easily determined that the data storage phase is confidential. Without loss of generality, suppose that the cloud wants to learn something encrypted in the ciphertext C, which was generated in the above data collection phase. The confidentiality of CIBPRE guarantees that except the users in the set $\mathcal{S}$, no one can learn anything about the data F from the ciphertext C. Hence, the only possible method for the cloud to break the ciphertext C in the data storage phase is colluding with one of the users in the set $\mathcal{S}$. However, it is practical to assume none of the users in the set $\mathcal{S}$ collude with the cloud, as no one would like to actively leak his sensitive data to an attacker.

In the data access phase, the cloud can break the ciphertext C using another possible method, which is discussed in the next section.

CONFIDENTIAL DATA ACCESS

In this section, we show how to apply CIBPRE to confidential data access in a cloud-assisted IoT system. In the data access phase, a user can share his collected IoT data with other users with the assistance of the cloud. At the same time, the cloud cannot disobey the user's request to share the non-expected data with other users or share the expected data with non-intended users. Otherwise, the cloud can possibly know the users' data. Suppose that a user with identity ID' wants to share another user's data F, where the latter user has identity ID and the data F were stored as an initial ciphertext C in the cloud (this step was achieved in the above data-collection phase). Figure 3 shows the main steps of the data access phase.

The details are described as follows:

1. The user ID' sends his identity and data-sharing request to the cloud.
2. The cloud periodically handles users' data-sharing requests. Suppose that all users in identity set $\mathcal{S}'$ want to share the same data of user ID. The cloud sends the data-sharing request and the set St to the user ID.
3. The user ID verifies the validity of the identities in the set $\mathcal{S}'$ and eliminates the invalid identities. Invalid identities mean that the corresponding users have no right to share the requested data. Suppose that all identities in the set $\mathcal{S}'$ are valid. Then, user ID chooses the same condition α that was used to generate the initial ciphertext C and generates and sends a re-encryption key $d_{ID\rightarrow\mathcal{S}'|\alpha}$ = **RKExtract**(**PK**, ID, SK_{ID}, $\mathcal{S}'$, α) to the cloud.
4. The cloud re-encrypts the initial ciphertext C to generate a re-encrypted ciphertext $\tilde{C}$ = **ReEnc**(**PK**, $d_{ID\rightarrow\mathcal{S}'|\alpha}$, C, $\mathcal{S}$) sends ($\tilde{C}$, $\mathcal{S}'$) to the users in the set $\mathcal{S}'$.
5. All users in the set $\mathcal{S}'$ independently decrypt the re-encrypted CIBPRE ciphertext $\tilde{C}$ to get the data F = **Dec-2**(**PK**, ID', SK_{ID}, $\tilde{C}$, $\mathcal{S}'$).

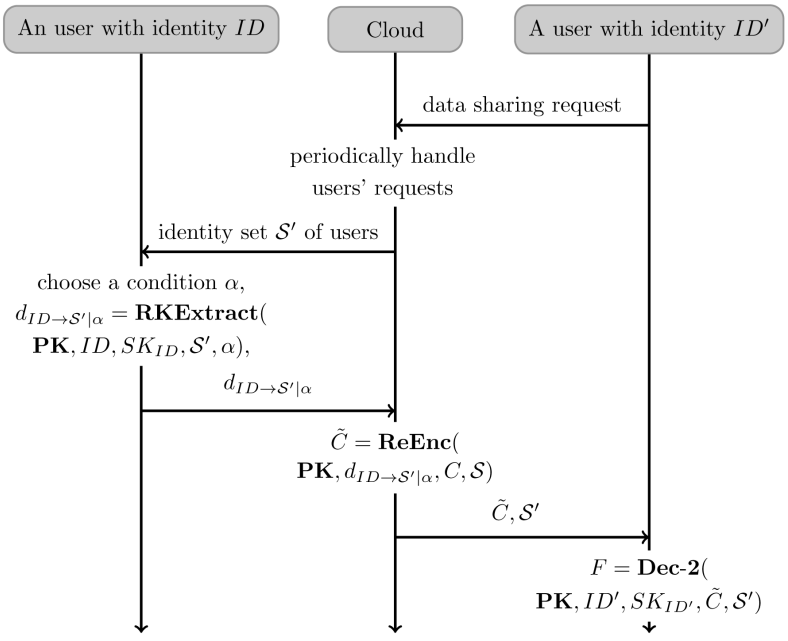


Figure 3. The main steps of the data access phase.

Note that we also allow the cloud to periodically handle users' data-sharing requests in step 2 above. This method provides the same advantage as the similar treatment does in the data collection phase.

In terms of confidentiality, CIBPRE guarantees that (1) only the users in the set $\mathcal{S}'$ can decrypt the re-encrypted ciphertext $\tilde{C}$ and (2) the cloud cannot share data with different sharing conditions with any user. In other words, any initial ciphertext with different sharing conditions cannot be correctly re-encrypted by the cloud.

Compared with SKE, CIBPRE makes the data access phase much more convenient. Suppose that we only adopt SKE for the data access phase and that all of a user's data are encrypted using the user's secret key. For example, Alice encrypted her data with her secret key s and stored the generated ciphertext in the cloud; when Bob wants to share Alice's data, Alice must confidentially send the secret key s to Bob. Without the help of PKE, the confidential transfer of s is difficult to achieve in practice.

PERFORMANCE

We coded the main CIBPRE algorithms to demonstrate the performance of the above phases. Because the time cost of SKE is linearly related to the size of the data, we exclude this cost in the main algorithms. The related system parameters are listed in Table 1, including the hardware, the operation system, and the cryptographic program library PBC and the chosen elliptic curve to code the above mentioned CIBPRE scheme. Note we apply scalable users in our experiments unlike our previous work did,[4] since the PBC-based code supports large-scaled simulation of IoT better than the Miracl-based code.[4] The experimental results are as follows:

- Figure 4 shows the time costs of algorithms **Enc** and **Dec-1** for different numbers of users who want to collect an IoT-edge object's data. And the time costs of that two algorithms are linear with the number of users. For example, suppose that there are 49 users who want to collect an IoT-edge object's data. The object takes approximately 267ms to generate an initial ciphertext, and each user takes approximately 73ms to decrypt the initial ciphertext.

- Figure 5 shows the time costs of algorithms **ReEnc** and **Dec-2** for different numbers of users who want to share a user's data. And the time costs of that two algorithms are linear with the number of users. For example, suppose that there are 49 users who want to share another user's data. The cloud takes approximately 76ms to re-encrypt an initial ciphertext and generate a re-encrypted ciphertext, and for sharing data, each user takes approximately 66ms to decrypt the re-encrypted ciphertext. We do not test the time cost of algorithm **RKExtract** because this algorithm has a time cost similar to that of algorithm **Enc**.

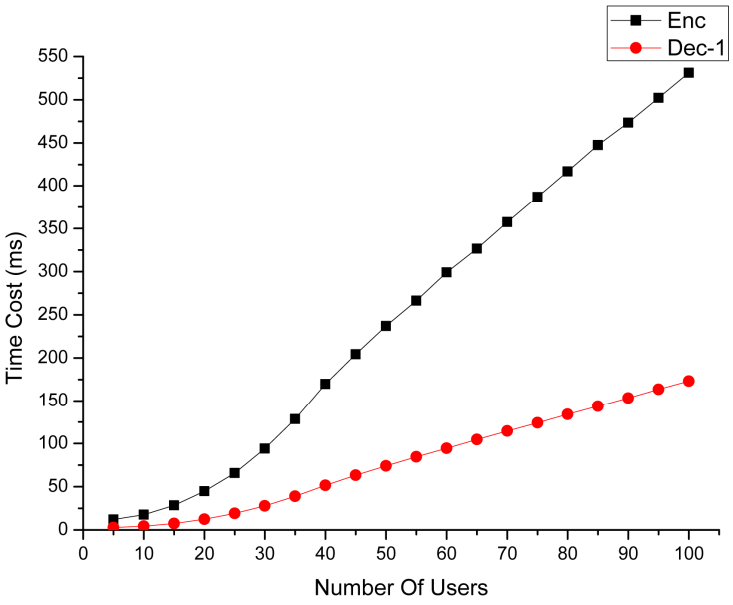


Figure 4. The time costs in the data collection phase for different number of users.

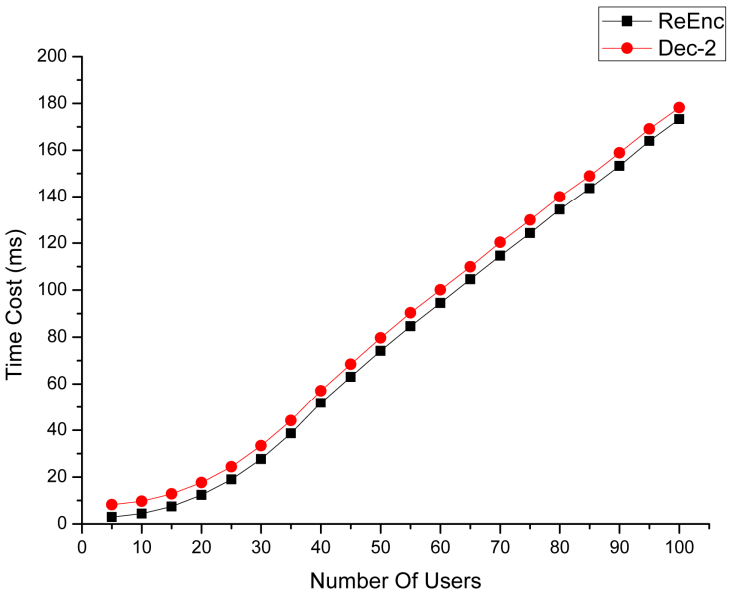


Figure 5. The time cost in the data access phase for different number of users.

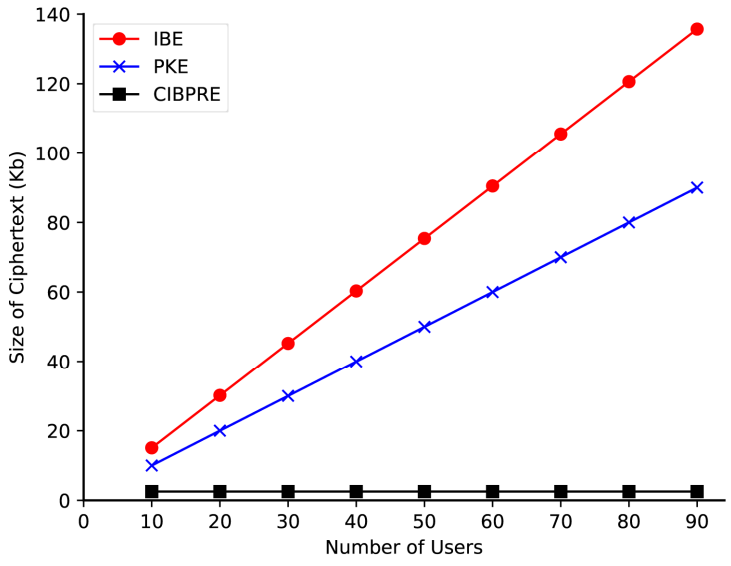


Figure 6. The communication cost in the data access phase for different number of users when comparing with several traditional approaches.

Table 1. Configuration and System Parameters

Hardware	Intel(R) Xeon(R) CPU E5-2420 v2 @ 2.20GHz
OS	CentOS release 6.6 (Final)
Program Library	PBC[15]
Mathematical Parameters	
Elliptical Curve	$y^2=x^3+x$
Base Field	87807107996633125224377819847540498158068831994142082110286533992664756308802229C57078625179422662221423155858769582317459277713367317481324925129998224791
Order	730750818665451621361119245571504901405976559617
The default unit is decimal.	

SYSTEM ANALYSES

Compared with the traditional IBE and PKE schemes, the biggest advantages of our CIBPRE-based system are as follows: (1) saving the communication cost, and (2) efficiently sharing the collected (or old) data with the other new users. More details are as follows.

Suppose that there are N users who want to collect an IoT-edge object's data. When employing the traditional IBE or PKE scheme to achieve the cloud-assisted IoT system, the object must randomly choose N secret keys, and encrypt these secret keys respectively by the identities or public

keys of the users. The size of the generated ciphertext is linear with the number of users. For example, suppose that the traditional IBE scheme is the one proposed in Boneh and Frankling[14] with the system parameters in Table 1, then the generated ciphertext has size about O(1544 *N*) bits; suppose the traditional PKE scheme is the RSA scheme, then the generated ciphertext has size about O(1024 *N*) bits. In contrast, the generated ciphertext has the constant size (about 2584 bits) in our CIBPRE-based cloud-assisted IoT system with the system parameters in Table 1, as shown in Figure 6. In other words, our system achieves a constant communication cost that is independent of the number of users. Note that we do not take into account the binary length of the IoT-edge object's data when evaluating the size of the generated ciphertext, since this cost is the same for the above three cryptographic schemes. When sharing the collected (or old) data with other new users, the traditional IBE or PKE scheme requires the IoT-edge object doing the same work as collecting new data does. This means that the object will require communication and time costs that are linear to the number of the new users. And the time cost also relies on the binary length of the collected (or old) data. In contrast, our system saves the communication and time costs, since the IoT-edge object only sends a re-encryption key with the constant size of about 2080 bits to the cloud, and the time cost to generate the re-encryption key is independent of the binary length of the collected (or old) data.

SUMMARY

Cloud-assisted IoT is a popular and useful framework for handling massive IoT data. This paper focuses on the data confidentiality when collecting, storing and accessing IoT data with the assistance of a cloud and introduces a promising method called CIBPRE for this purpose. CIBPRE allows users to confidentially collect and store an IoT-edge object's data with the assistance of the cloud and share these data with others. In addition to confidentiality, CIBPRE is advantageous in terms of performance. Its "broadcast" property allows the data collection and data access functions to be achieved in a batch manner. Its "conditional" property allows the data to be accessed in a fine-grained manner. Its "identity-based" property avoids the complex public-key management of the traditional PKE. This paper also provides numerical results demonstrating the feasibility of our work.

ACKNOWLEDGEMENT

This work is partly supported by the National Natural Science Foundation of China under grant no. 61472156 and no. 61702206, the Shenzhen Fundamental Research Program under grant no. JCYJ20170413114215614, and the Fundamental Research Funds for the Central Universities under grant no. 2017KFYXJJ062.

REFERENCES

1. J. Gubbi et al., "Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions," *Future Generation Computer Systems*, vol. 29, no. 7, 2013, pp. 1645–1660.
2. M. Kaufman, "Data Security in the World of Cloud Computing," *IEEE Security & Privacy*, vol. 7, no. 4, 2009, pp. 61–64.
3. Y. Liu et al., "IOT secure transmission based on integration of IBE and PKI/CA," *International Journal of Control & Automation*, vol. 6, no. 2, 2013, pp. 245–254.
4. D.E. Bakken et al., "Data obfuscation: Anonymity and desensitization of usable data sets," *IEEE Security & Privacy*, vol. 2, no. 6, 2004, pp. 34–41.
5. P. Xu et al., "Conditional Identity-based Broadcast Proxy Re-Encryption and Its Application to Cloud Email," *IEEE Transactions on Computers*, vol. 65, no. 1, 2016, pp. 66–79.
6. R. Roman, P. Najera, and J. Lopez, "Securing the Internet of Things," *Computer*, vol. 44, no. 9, 2011, pp. 51–58.

7. K.T. Nguyen, M. Laurent, and N. Oualha, "Survey on Secure Communication Protocols for the Internet of Things," *Ad Hoc Networks*, vol. 32, 2015, pp. 17–31.
8. R.H. Weber, "Internet of Things -- New Security and Privacy Challenges," *Computer Law & Security Review*, vol. 26, no. 1, 2010, pp. 23–30.
9. A. Mukherjee, "Physical-Layer Security in the Internet of Things: Sensing and Coomunication Confidentiality under Resource Constraints," *Proceedings of the IEEE*, vol. 103, no. 10, 2015, pp. 1747–1761.
10. I.E. Bagci et al., "Fusion: Coalesced Confidential Storage and Communication Framework for the IoT," *Security and Communication Networks*, vol. 9, no. 15, 2016, pp. 2656–2673.
11. J. Singh et al., "Twenty Security Considerations for Cloud-supported Internet of Things," *IEEE Internet of Things Journal*, vol. 3, no. 3, 2016, pp. 269–284.
12. V. Bhuse, "Security and Privacy Challenges for Healthcare Records and Wearable Sensors in Cloud," *Transactions on IoT and Cloud Computing*, vol. 2, no. 3, 2014, pp. 11–17.
13. A. Boldyreva et al., "A Closer Look at PKI: Security and Efficiency," *Public Key Cryptography*, T. Okamoto, X. Wang, Lecture Notes in Computer Science, vol. 4450, Springer, 2007; doi.org/10.1007/978-3-540-71677-8_30.
14. D. Boneh and M. Franklin, "Identity-Based Encryption from the Weil Pairing," *Advances in Cryptology — CRYPTO 2001*, Lecture Notes in Computer Science, vol. 2139, Springer, 2001.
15. B. Lynn, *The Pairing-Based Cryptography Library (PBC) Ver. 0.5.14*; https://crypto.stanford.edu/pbc/download.html.

ABOUT THE AUTHORS

Wei Wang is a lecturer with the Cyber-Physical-Social Systems Lab, School of Computer Science and Technology, Huazhong University of Science and Technology. Her research interests include cloud security, network coding and multimedia transmission. She has a PhD in electronic and communication engineering from Huazhong University of Science and Technology. Contact her at viviawangwei@mail.hust.edu.cn.

Peng Xu is an associate professor at Huazhong University of Science and Technology. His research interests include big data, network security, and cryptography. He has a PhD in computer science from Huazhong University of Science and Technology. Contact him at xupeng@mail.hust.edu.cn.

Laurence T. Yang is the director of the Cyber-Physical-Social Systems Lab, School of Computer Science and Technology, Huazhong University of Science and Technology. He also works in the Department of Computer Science at St. Francis Xavier University. His research interests include parallel and distributed computing, embedded and ubiquitous computing, and security. He has a PhD in computer science from the University of Victoria, Canada. Yang is a Fellow of the Canadian Academy of Engineering. Contact him at ltyang@gmail.com.

Revenue Growth is the Primary Benefit of the Cloud

Brad Power
MAXOS

Joe Weinman

Cloud computing is too often seen as a tactical way to reduce costs, when its most important benefit is as a strategic way to grow revenues. Such revenue growth can come about in a variety of ways, such as through faster innovation of new products, processes, and customer interactions; identifying more customers and closing more purchases; and improving customer relationships through more targeted offers and better service and experiences. Companies that clearly understand the relative magnitude of cost savings and revenue growth and orient themselves toward the latter will better exploit the cloud and related technologies such as big data, artificial intelligence (AI), the Internet of Things, and blockchain, and thus strengthen their competitive advantage and customer value.

A typical, but simplistic view says that the economies of scale that large cloud service providers achieve drive down the unit cost of compute.[1] This logic does have some truth to it, but is flawed on several levels. First, cloud providers enjoy cost advantages not only through economies of scale, but also through better utilization through aggregation of statistically uncorrelated workloads, as well as selling spare cycles through mechanisms such as "spot instances." They also can enjoy economies through greenfield siting, where power and land are cheap, and tax benefits may be conferred by local authorities. However, large diversified enterprises with sufficient technical competence, operations capabilities, purchasing power, etc., can enjoy similar cost advantages. Moreover, enterprises doing it themselves don't have to pay the additional cost structure penalty comprising a cloud service provider's profit margin and general, sales, and administrative expenses.[2] As a result, real-world experience is mixed, with some saving money by moving to the cloud and others saving money by moving out of public clouds.[3] Many others do best with a hybrid strategy that can offer an optimal balance of lower costs through fixed resources for baseline demand and elastic pay-per-use resources for variable demand beyond the baseline.[4] Still others use a multicloud approach, either to cobble together digital support for an end-to-end workflow,[5] for reliability, or to arbitrage price differentials among cloud service pro-

Copublished by the IEEE CS and IEEE ComSoc
2325-6095/18/$33.00 USD ©2018 IEEE

viders.[6] Many use a combination of all of the above, leading to a mix of private and public, multiple public clouds, and centralized facilities and dispersed ones—the hybrid multicloud fog.[7]

Moreover, even if the *cost savings* are dramatic, their impact barely moves the needle on *overall* corporate financials. Since IT costs typically average roughly 4% of revenue, a compelling 25% cost reduction in IT only represents a 1% impact on the company. This is good for a CFO, CIO, or procurement executive's performance review, but not enough to guarantee business success in the world of global hyper-competition and disruption that most companies find themselves in.

No, the greatest impact of the cloud—broadly defined not just as Infrastructure-as-a-Service (IaaS) but new architectures such as the hybrid multi-cloud fog and layers through the application layer—is in growing revenues, not in cutting costs. Several mechanisms are at work, because the cloud helps to:

- **accelerate innovation**, by providing easy-to-use enabling tools and reducing friction;
- **find and reach more customers**, in more places, more precisely, with better experiences and more compelling content; and
- **enhance customer relationships**, and thus, stickiness and loyalty.

These form a virtuous cycle, enabled by the cloud. Faster, cheaper, and better innovation in products, services, and processes thanks to the cloud leads to differentiated offers in the marketplace. Together with precision marketing and global reach, this enables companies to find more customers, close sales, and deliver products and services to those customers, globally. Relationships with those customers move from mere anonymous transactions to a higher degree of intimacy, and advanced big data algorithms such as recommendation engines enable "collective intimacy."

The cloud is also associated with other benefits, such as accelerated time to market and time to volume, and business agility and elasticity. However, these also directly correlate to revenue growth. Reducing time to market implies generating first-mover advantage, which in today's world can imply signing up sticky customers and ecosystem partners before someone else does. It also can imply gaining a larger share of the profit pool before it dissipates later in the product lifecycle across a sea of competitors. Resource elasticity helps revenue-generating services scale, ensuring revenue is maximized to the extent the market will allow. Finally, the concept of "business agility" can be a bit amorphous, but it generally refers to the ability to rapidly respond to shifts in market dynamics, customer needs, or competitor moves. While a portion of these benefits surely accrues to cost reduction, a majority generally is realized as enhanced revenue.

All of these benefits together can mean more compelling products and customers, more frequent purchases, and/or larger purchase sizes, all leading to higher revenues: the true impact of the cloud.

ACCELERATE INNOVATION

The cloud accelerates innovation in many different ways. By eschewing ownership in favor of on-demand access, it eliminates much of the friction and risk involved in experimentation, a necessary component of innovation.

For example, in the early 1950s, when GE wanted to experiment with state-of-the-art technology—a "stored-program electronic computer"—it reportedly needed to plunk down millions of dollars for a UNIVAC. It also needed what was, in effect, a ruse, to spend that money—using a mundane justification of automating payroll, manufacturing operations, ordering and billing, and accounting. As Paul Ceruzzi remarks, "GE needed to assure its stockholders that it was not embarking on a wild scheme of purchasing exotic, fragile, and expensive equipment,"[8] even as it had the foresight to envision that this technology could be revolutionary. One can only imagine how much time it must have taken to formulate a business case, socialize the approach with various leadership communities, none who would get credit but all who would shoulder the blame if anything went wrong, gain the necessary approvals, place the order, build or repurpose a site, accept the delivery of the equipment, etc.

Today, early adopters no longer need to take crazy risks to invest heavily and "bet the farm" on promising but unproven new technologies. The fundamental profile of the technology adoption lifecycle is being restructured through on-demand cloud access models with pay-per-use pricing rather than fixed cost or up-front capital investment; microservices, cloud functions, and APIs; "try-before-you-buy" free tiers, freemium, or introductory trial periods; elastic capacity; and so forth, which all enable customers to experiment with bleeding-edge technologies for free or at little cost or risk.

Free trials have been a means of incenting customer adoption for decades, if not centuries. It is a particularly appropriate strategy for cloud-based software, because the marginal cost of a customer trial is essentially zero, and the product does not wear out with a trial, unlike, say, a car used for test drives. In many cases, services are offered entirely free to customers/users, either as a loss-leader in a broader strategy, or based on 3rd party or multi-sided monetization.

To pick some examples to illustrate the variety of approaches, Microsoft offers free 30-day trials of Office 365 that are intended to convert to a monthly subscription. Google offers some services, such as Gmail and Google Docs, for free, with a paid version with enhanced features for business (with a free 14-day trial); others, such as Google search, are advertiser-supported. AWS offers free tiers for various offers, such as AWS Lambda functions (i.e., serverless computing). With Lambda, for example, the first million transactions each month are free, although there are some surprisingly complex economics behind this.[9]

What is particularly of interest in today's cloud offers is that it's not only mainstream or commodity functions that have free trials of freemium models, or products that are still in beta, but highly advanced technologies. For example, IBM offers free trials for dozens of products, such as SPSS statistical analysis and MaaS360 mobile device security. But it also offers free trials for its most advanced capabilities.

Customers can experiment with the IBM Watson API or IBM Q quantum computing environment for free, and then expand into paid use. This approach not only has benefits for customers, it also means that service providers can rapidly learn from early adopters and identify valid industry use cases with which to tune and iterate their offers. For example, Staples was able to rapidly prototype a new version of its iconic "Easy" button that tied to an IBM Watson chatbot.

Moreover, cloud users have a range of options to accelerate the testing and deployment of new products, services, and processes. Rather than relying on standard packages, which don't provide competitive advantage; or having to wait interminably for in-house development efforts to build complex features from scratch, CIOs and their teams can rely on a mix of standard tools and environments used as is, which might be ERP functions or collaboration tools such as email or shared documents; open source software *and* hardware and networks; custom mixes of multicloud SaaS tools that enable workflows unique to the industry or offering strategic advantage to the customer,[10] or environments that are built in-house but leverage APIs or serverless computing/cloud functions to rapidly assemble Lego blocks of software to build unique, proprietary functionality.

The digital capabilities and implementations that result from this approach offers the best of all worlds: alignment with business strategy and processes; low risk; fast time to market; low up-front and operating costs; high scalability and elasticity. This software can be used to power differentiating business processes or digitalize products and services.

The cloud also democratizes innovation. Rather than innovation being restricted to deep-pocketed corporate labs, anyone, anywhere, at any age can be an innovator. A great example is Tanmay Bakshi, who released his first iPhone app at age nine (thanks to app dev SDKs and Apple's cloud-based App Store) and as an eleven-year-old heard about Watson. Within a week, he had built a Watson-based app.[11] Indeed, virtually anyone can gain access from the comfort of their living room, often at no charge, to the most advanced technologies in the world. They can then rapidly move from awareness, to experimentation, to trials and proofs of concepts, to production readiness and global scale.

Many business model innovations, such as pay-as-you-drive insurance, are ultimately based on a cloud-enabled architecture entailing real-time data collection and processing. In this approach, an automobile insurance company underwrites policies based not just on demographic information

such as customer age and zip code, or driving history and recent violations, but can price premiums in real time and by the mile based on live data streams such as roadway congestion, weather conditions, vehicle speed and acceleration, lane changes, and so forth, aggregated and processed in the cloud.

The cloud can accelerate innovation in many other ways.[12] This includes cloud-mediated idea markets, innovation networks, contests and challenges such as GE FlightQuest, the Netflix prize and Fold.it, hackathons, and ultimately, machine innovation as exemplified by Google DeepMind AlphaGo's move 37 in Game 2 against one of the world's top Go players, Lee Sedol[13]; IBM Chef Watson, which creates new recipes; Melvin, the algorithm that designs quantum physics experiments[14]; and the final frontier, machine creativity in formulating scientific theories through automated hypothesis generation.[15]

FIND AND REACH MORE CUSTOMERS

To find likely prospects has always been a challenge. As John Wanamaker famously said a century ago, "Half the money I spend on advertising is wasted; the trouble is I don't know which half." But today, cloud Software-as-a-Service (SaaS) tools can tie in to other cloud elements such as social media and email to optimize targeting. In other words, one cloud provides the AI processing; other clouds provide the raw data; still others provide the means to message customers. For example, a Harley-Davidson dealership in New York was able to increase sales leads by almost 3000%, using "Albert," an AI-based SaaS package that supports autonomous media buying across multiple marketing channels, testing and optimization of those channels and creative content, and various other activities such as identifying "lookalikes," i.e., prospects that match existing customer segments.[16]

Reaching more customers has always meant utilizing new channels for marketing and distribution. At one time the new channel was the mail-order catalog. Then TV, with sponsored shows ("soap" operas) and advertising. Then it was tele-sales, and then ecommerce.

Today, basic ecommerce has evolved, and, marketing and distribution includes owned, earned, and sponsored media. It includes paid search, banner and pop-up ads. But it also includes entirely new mechanisms. Alibaba created a virtual reality Macy's store for Singles Day, and shipped Google-cardboard-like smartphone holders turning every phone into VR goggles. At the other extreme, Amazon Dash buttons make every washing machine, kitchen cabinet, or refrigerator a storefront. Smart speakers such as the Amazon Echo family can be used to order anything that a browser pointed to Amazon.com can. All of these are enabled by global wireless and wireline infrastructure linking these points of presence to the cloud.

In the physical world, scaling distribution means building more branches and retail stores and the supply chain feeding them including manufacturing or service operations infrastructure. In the virtual world, as long as the software architecture is designed to scale, the cloud—both in terms of elastic web and app servers but also content delivery networks—provides the essentially limitless resource infrastructure that enables companies to grow without needing to site and build out their own data centers. Moreover, the latest cloud architectures utilize a global fabric of hyperscale datacenters distributed across global regions, tied through global networks and interconnection facilities to each other and a highly dispersed fog/edge.[17] This reduces latency as well as backhaul traffic and thus bandwidth requirements.[18]

ENHANCE CUSTOMER RELATIONSHIPS

Customer relationships are also enhanced by cloud-centered approaches. For example, consumer packaged goods manufacturers used to manufacture products and sell them through intermediaries such as retailers and VARs. They often had no idea who the customer was or how they were using the product. Now, products are increasingly smart, digital, and connected. Smart home devices like Wi-Fi lightbulbs and door locks are one example. Modern vehicles such as Teslas are another, with over-the-air upgrades and data collection for global optimization. Netflix is another example. Netflix collects trillions of data points from each viewer, such as search intent,

navigation behavior on the "home page" displayed in browsers, phones, and smart TVs; viewing behavior such as watching, pausing, and rewinding; contexts such as mobile device or TV, time of day, and location; social graphs; and convolves this with external data such as director, filming location, actors, and subjective evaluations of content metrics, such as whether the content is romantic or funny. All this data is aggregated and maintained in the cloud, and used to generate recommendations intended to increase customer satisfaction and reduce churn, thereby increasing customer lifetime value. This also enhances referral marketing, whereby happy Netflix customers act as a virtual salesforce helping to convince prospects to subscribe.[19–21]

TOP LINE TRUMPS BOTTOM LINE

Cost savings can help the cloud provide value to IT organizations, but the true value of the cloud is in growing revenues for the corporation.

REFERENCES

1. D. Sholler and D. Scott, "Economies of Scale Are the Key to Cloud Computing Benefits," *Gartner.com*, 30 June 2008; https://www.gartner.com/doc/710610/economies-scale-key-cloud-computing.
2. J. Weinman, *Cloudonomics: The Business Value of Cloud Computing*, Wiley, 2012.
3. J. Weinman, "Migrating to—or away from—the Public Cloud," *IEEE Cloud Computing*, vol. 3, no. 2, IEEE Cloud Computing, 2016, pp. 6–10.
4. J. Weinman, "Hybrid Cloud Economics," *IEEE Cloud Computing*, vol. 3, no. 1, 2016, pp. 18–22.
5. B. Power, "Digital Transformation Through SaaS Multiclouds," *IEEE Cloud Computing*, vol. 5, no. 3, 2018, pp. 27–30.
6. J. Weinman, "Cloud Pricing and Markets," *IEEE Cloud Computing*, vol. 2, no. 1, 2015, pp. 10–13.
7. J. Weinman, "The Economics of the Hybrid Multicloud Fog," *IEEE Cloud Computing*, 2017, pp. 16–21.
8. P.E. Cerruzi, *A History of Modern Computing*, MIT Press, 2003.
9. A. Eivy, "Be Wary of the Economics of 'Serverless' Cloud Computing," *IEEE Cloud Computing*, vol. 4, no. 2, 2017, pp. 6–12.
10. B. Power, "Digital Transformation Through SaaS Multiclouds," *IEEE Cloud Computing*, vol. 5, no. 3, 2018, pp. 27–30.
11. R. Umoh, "How this self-taught 14-year-old kid became an AI expert for IBM," *CNBC.com*, 25 January 2018; https://www.cnbc.com/2018/01/25/how-self-taught-14-year-old-tanmay-bakshi-became-an-ai-expert-for-ibm.html.
12. J. Weinman, *Digital Disciplines: Attaining Market Leadership via the Cloud, Big Data, Social, Mobile, and the Internet of Things*, Wiley CIO, 2015.
13. C. Metz, "In Two Moves, AlphaGo and Lee Sedol Redefined the Future," *Wired*, 16 March 2016; https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/.
14. C.Q. Choi, "Physicists Unleash AI to Devise Unthinkable Experiments," *Scientific American*, 22 March 2016; https://www.scientificamerican.com/article/physicists-unleash-ai-to-devise-unthinkable-experiments.
15. "Computer Says 'Try This'," *The Economist*, 4 October 2014; https://www.economist.com/science-and-technology/2014/10/04/computer-says-try-this.
16. B. Power, "How Harley-Davidson Used Artificial Intelligence to Increase New York Sales Leads by 2,930%," *Harvard Business Review*, 30 May 2017; https://hbr.org/2017/05/how-harley-davidson-used-predictive-analytics-to-increase-new-york-sales-leads-by-2930.
17. J. Weinman, "The Economics of the Hybrid Multicloud Fog," *IEEE Cloud Computing*, vol. 4, no. 1, 2017, pp. 16–21.
18. J. Weinman, "The 10 Laws of Fogonomics," *IEEE Cloud Computing*, vol. 4, no. 6, 2017, pp. 8–14.

19. J. Weinman, *Digital Disciplines: Attaining Market Leadership via the Cloud, Big Data, Social, Mobile, and the Internet of Things*, Wiley CIO, 2015.
20. X. Amatriain, "Big & personal: data and models behind Netflix recommendations," *Proceedings of the 2nd international workshop on big data, streams and heterogeneous source Mining: Algorithms, systems, programming models and applications* (BigMine 13), 2013, pp. 1–6.
21. C.A. Gomez-Uribe and N. Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems*, vol. 6, no. 4, 2016, p. 13.

ABOUT THE AUTHORS

Brad Power is a consultant who helps organizations that must make faster changes to their products, services, and systems to compete with start-ups and leading software companies. He is a principal at MAXOS, a partner at FCB Partners, a Questrom Digital Fellow at Boston University, an advisor to the Innovation Scout, and a frequent contributor to the *Harvard Business Review*. He received a BS in mathematical sciences from Stanford University and an MBA from UCLA. Contact him at bradfordpower@gmail.com.

Joe Weinman is a frequent global keynoter and author of *Cloudonomics* and *Digital Disciplines*, both available in Chinese editions. He also serves on the advisory boards of several technology companies. Weinman has a BS in computer science from Cornell University and an MS in computer science from the University of Wisconsin-Madison. He has completed executive education at the International Institute for Management Development in Lausanne. Weinman has been awarded 24 patents in areas such as cloud computing, distributed storage, data networking, mobile telephony, consumer products, and encryption. Contact him at joeweinman@gmail.com.

IEEE computer society

PURPOSE: The IEEE Computer Society is the world's largest association of computing professionals and is the leading provider of technical information in the field.
MEMBERSHIP: Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.
COMPUTER SOCIETY WEBSITE: www.computer.org
OMBUDSMAN: Direct unresolved complaints to ombudsman@computer.org.
CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.
AVAILABLE INFORMATION: To check membership status, report an address change, or obtain more information on any of the following, email Customer Service at help@computer.org or call +1 714 821 8380 (international) or our toll-free number, +1 800 272 6657 (US):

- Membership applications
- Publications catalog
- Draft standards and order forms
- Technical committee list
- Technical committee application
- Chapter start-up procedures
- Student scholarship information
- Volunteer leaders/staff directory
- IEEE senior member grade application (requires 10 years practice and significant performance in five of those 10)

PUBLICATIONS AND ACTIVITIES
***Computer*:** The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.
Periodicals: The society publishes 13 magazines, 19 transactions, and one letters. Refer to membership application or request information as noted above.
Conference Proceedings & Books: Conference Publishing Services publishes more than 275 titles every year.
Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.
Technical Committees: TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.
Conferences/Education: The society holds about 200 conferences each year and sponsors many educational activities, including computing science accreditation.
Certifications: The society offers two software developer credentials. For more information, visit www.computer.org/certification.

COMPUTER SOCIETY OFFICES

Washington, D.C.: 2001 L St., Ste. 700, Washington, D.C. 20036-4928
Phone: +1 202 371 0101 • **Fax:** +1 202 728 9614
Email: hq.ofc@computer.org
Los Alamitos: 10662 Los Vaqueros Circle, Los Alamitos, CA 90720 **Phone:** +1 714 821 8380
Email: help@computer.org

MEMBERSHIP & PUBLICATION ORDERS
Phone: +1 800 272 6657 • **Fax:** +1 714 821 4641 • Email: help@computer.org
Asia/Pacific: Watanabe Building, 1-4-2 Minami-Aoyama, Minato-ku, Tokyo 107-0062, Japan
Phone: +81 3 3408 3118 • **Fax:** +81 3 3408 3553
Email: tokyo.ofc@computer.org

IEEE

revised 03 July 2018

IEEE computer society

SHARE THE GIFT OF KNOWLEDGE:

Give Your Favorite Student a Membership to the IEEE Computer Society!

With an **IEEE Computer Society Membership**, your student will be able to build their network, learn new skills, and access the best minds in computer science before they're even out of school. Your gift includes thousands of key resources that will quickly transition them from classroom to conference room, such as:

- **A subscription to *Computer* magazine** (12 issues per year)
- **A subscription to *ComputingEdge*** (12 issues per year)
- **Local chapter membership**
- **Full access to the Computer Society Digital Library**
- **Eligible for 3 student scholarships where we give away US$40,000 yearly**
- **Skillsoft:** Learn new skills anytime with access to 3,000 online courses, 11,000 training videos, and 6,500 technical books.
- **Books24x7:** On-demand access to 15,000 technical and business resources.
- **Unlimited access to computer.org and myCS**
- **Conference discounts**
- **Members-only webinars**
- **Deep member discounts** on programs, products, and services

Give Your Gift at: www.computer.org/2018gift