08.21

Computer

SAFETY, SECURITY, AND RELIABILITY OF AUTONOMOUS VEHICLE SOFTWARE

IEEE

75 YEARS IEEE COMPUTER SOCIETY

vol. 54 no. 8

www.computer.org/computer

Volunteer Leadership Is Vital

Vote by Monday, 20 September at 12PM EDT

www.computer.org/election2021

IEEE

Digital Object Identifier 10.1109/MC.2021.3094698

Computer

20 **GUEST EDITORS' INTRODUCTION**
Safety, Security, and Reliability of Autonomous Vehicle Software
W. ERIC WONG AND ZIJIANG YANG

AUGUST 2021
FEATURES

22 Safety, Complexity, and Automated Driving: Holistic Perspectives on Safety Assurance
SIMON BURTON, JOHN MCDERMID, PHILIP GARNET, AND ROB WEAVER

33 Blockchain-Based Continuous Auditing for Dynamic Data Sharing in Autonomous Vehicular Networks
HAIYANG YU, SHUAI MA, QI HU, AND ZHEN YANG

46 Ethics in Autonomous Vehicle Software: The Dilemmas
SACHIN MOTWANI, TARUN SHARMA, AND ANUBHA GUPTA

Digital Object Identifier 10.1109/MC.2021.3055687

AUGUST 2021

CONTENTS

ABOUT THIS ISSUE
SAFETY, SECURITY, AND RELIABILITY OF AUTONOMOUS VEHICLE SOFTWARE

This issue looks at how to improve autonomous vehicle software.

FEATURES CONTINUED

56 An Online Multistep-Forward Voltage-Prediction Approach Based on an LSTM-TD Model and KF Algorithm
YE NI, ZHILONG XIA, CHUNRONG FANG, ZHENYU CHEN, AND FANGTONG ZHAO

66 Toward Improving Confidence in Autonomous Vehicle Software: A Study on Traffic Sign Recognition Systems
KOOROSH ASLANSEFAT, SOHAG KABIR, AMR ABDULLATIF, VINOD VASUDEVAN, AND YIANNIS PAPADOPOULOS

77 From Neuron Coverage to Steering Angle: Testing Autonomous Vehicles Effectively
JACK TOOHEY, M S RAUNAK, AND DAVE BINKLEY

COLUMNS

4 LETTERS
Terminate With Extreme Prejudice
TERENCE KELLY

8 50 & 25 YEARS AGO
ERICH NEUHOLD

9 COMPUTING THROUGH TIME
ERGUN AKLEMAN

11 EIC'S MESSAGE
Where's the Silicon?
NIR KSHETRI AND JEFFREY VOAS

13 THE IEEE COMPUTER SOCIETY'S 75TH ANNIVERSARY
Insights Into the Origins of the IEEE Computer Society and the Invention of Electronic Digital Computing
VLADIMIR GETOV

86 BODY OF KNOWLEDGE
For the Common Defense
DAVID ALAN GRIER

89 IT INNOVATION
Conversational Artificial Intelligence: Changing Tomorrow's Health Care Today
MARK CAMPBELL AND MLAĐAN JOVANOVIĆ

94 EDUCATION
Educate Back Better: A Perspective from Industry
DEJAN MILOJICIC

98 COMPUTING'S ECONOMICS
The Economics of Blockchain-Based Supply Chain Traceability in Developing Countries
NIR KSHETRI

104 OPEN SOURCE EXPANDED
Making Open Source Project Health Transparent
SEAN P. GOGGINS, MATT GERMONPREZ, AND KEVIN LUMBARD

112 CYBERTRUST
Cyber Resilience: by Design or by Intervention?
ALEXANDER KOTT, MAUREEN S. GOLAN, BENJAMIN D. TRUMP, AND IGOR LINKOV

118 COMPUTING ARCHITECTURES
Future Computers: Digital, Quantum, Biological
PHILIP TRELEAVEN

125 DATA
2021 State of the Practice in Data Privacy and Security
PREETI S. CHAUHAN AND NIR KSHETRI

Departments

5 Elsewhere in the CS

Membership News

Cover3 IEEE Computer Society Information

Circulation: *Computer* (ISSN 0018-9162) is published monthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720; voice +1 714 821 8380; fax +1 714 821 4010; IEEE Computer Society Headquarters, 2001 L Street NW, Suite 700, Washington, DC 20036. IEEE Computer Society membership includes a subscription to *Computer* magazine.

Postmaster: Send undelivered copies and address changes to *Computer*, IEEE Membership Processing Dept., 445 Hoes Lane, Piscataway, NJ 08855. Periodicals Postage Paid at New York, New York, and at additional mailing offices. Canadian GST #125634188. Canada Post Corporation (Canadian distribution) publications mail agreement number 40013885. Return undeliverable Canadian addresses to PO Box 122, Niagara Falls, ON L2E 6S8 Canada. Printed in USA.

Computer

IEEE COMPUTER SOCIETY http://computer.org // +1 714 821 8380
COMPUTER http://computer.org/computer // computer@computer.org

Digital Object Identifier 10.1109/MC.2021.3055688

LETTERS

TERMINATE WITH EXTREME PREJUDICE

To the Editor:

I am writing to raise awareness that a pervasive software paradigm is prone to a serious performance pitfall. At least one widespread instance of the problem has been remarkably adept at evading detection.

The paradigm in question is the *work queue* at the heart of myriad programs: software repeatedly dequeues a task and performs corresponding work, which may enqueue new tasks, until the queue is empty. The performance bug arises when output attains its final state long before the work queue drains; subsequent effort to empty the queue is wasted because it does not change the output.

The obvious solution—a "stop when done" termination test—is not always obvious to algorithm designers and developers coding in the work queue paradigm. More than once, I've seen production software that descends into a tragicomic frenzy of needless toil merely to drain a work queue, with no externally observable effect whatsoever beyond raising the CPU temperature. Work queues naturally incline toward self-inflicted busywork unless thoughtfully supervised.

Unfortunately, such vigilance is itself exceptionally difficult. We might hope that a modicum of peer review would suffice to exorcise gratuitous inefficiency from queue-centric designs. One of the most widely known elementary algorithms of all time, however, shows that extensive scrutiny is not proof against this problem.

Top textbooks, such as those by Cormen et al. (*Introduction to Algorithms*, third edition) and Sedgewick and Wayne (*Algorithms*, fourth edition), have withstood decades of intense critical attention from generations of academics and practitioners. The breadth-first search algorithm presented in these and similar texts terminates when its work queue drains, which may occur long after all output is finalized.

The paradigm in question is the work queue at the heart of myriad programs: software repeatedly dequeues a task and performs corresponding work, which may enqueue new tasks, until the queue is empty.

Compared to an "efficient BFS" that terminates when the output reaches quiescence, the classic textbook BFS is like a penny in a fuse box: never better and sometimes catastrophically worse—that is, sometimes slower by a factor proportional to the number of vertices in the input graph. For detailed evaluations, see https://queue.acm.org/detail.cfm?id=3424304.

The root cause of the "work queue run mad" antipattern is a confusion of ends versus means. Work queues are the latter. They provide reminders to be considered, not commands to be blindly obeyed, as computations unfold. Unfortunately, experience shows that work queues can distract attention from the simple fact that every program's purpose is to compute its output.

Terence Kelly
tpkelly@acm.org

Digital Object Identifier 10.1109/MC.2021.3055874
Date of current version: 30 July 2021

ELSEWHERE IN THE CS

Computer Highlights Society Magazines

The IEEE Computer Society's lineup of 12 peer-reviewed technical magazines covers cutting-edge topics ranging from software design and computer graphics to Internet computing and security, from scientific applications and machine intelligence to visualization and microchip design. Here are highlights from recent issues.

computing in SCIENCE & ENGINEERING

Using Jupyter for Reproducible Scientific Workflows

In this article from the March/April 2021 issue of *Computing in Science & Engineering*, the authors report two case studies—one in computational magnetism and another in computational mathematics—where domain-specific software was exposed to the Jupyter environment. This enables high-level control of simulations and computation, interactive exploration of computational results, batch processing on high-performance computing resources, and reproducible workflow documentation in Jupyter notebooks. In the first study, Ubermag drives existing computational micromagnetics software through a domain-specific language embedded in Python. In the second study, a dedicated Jupyter kernel interfaces with the GAP system for computational discrete algebra and its dedicated programming language.

IEEE Annals of the History of Computing

Recovering Software for the Whirlwind Computer

The late 1940s were seminal years in the development of electronic digital computing machines. One of these new generations of computing devices was the vacuum-tube Whirlwind computer, designed in the late 1940s at the Massachusetts Institute of Technology. Designed from the start for real-time control applications, Whirlwind evolved to be a key element in a proof of concept for national air defense. This article from the January–March 2021 issue of *IEEE Annals of the History of Computing* starts with a review of Whirlwind and its evolving software environment, goes on to describe the current effort to decode some of the Whirlwind software artifacts remaining in museum archives, and then describes some of the material the work has made accessible.

IEEE Computer Graphics AND APPLICATIONS

QuteVis: Visually Studying Transportation Patterns Using Multisketch Query of Joint Traffic Situations

QuteVis uses multisketch query and visualization to discover specific times and days in history with specified joint traffic patterns at different city locations. Users can use touch-input devices to define, edit, and modify multiple sketches on a city map. A set of visualizations and interactions is provided to help users browse and compare retrieved traffic situations and discover potential influential factors. QuteVis is built on a transport database that integrates heterogeneous data sources with an optimized spatial indexing and weighted similarity computation. An evaluation with real-world data and domain experts demonstrates that QuteVis is useful in urban transportation applications in modern cities. Read more in this article from the March/April 2021 issue of *IEEE Computer Graphics and Applications*.

Digital Object Identifier 10.1109/MC.2021.3055875
Date of current version: 30 July 2021

IEEE Intelligent Systems

Differentially Private Collaborative Coupling Learning for Recommender Systems

Coupling learning is designed to estimate, discover, and extract the interactions and relationships among learning components. It provides insights into complex interactive data and has been extensively incorporated into recommender systems to enhance the interpretability of sophisticated relationships between users and items. Coupling learning can be further fostered once the trending collaborative learning can be engaged to take advantage of the cross-platform data. To facilitate this, privacy-preserving solutions are in high demand. In this article from the January/February 2021 issue of *IEEE Intelligent Systems*, the authors develop a distributed collaborative coupling learning system, which enables differential privacy and defends against the adversary who has gained full knowledge of the training mechanism. It also addresses the privacy-utility tradeoff by a provable tight sensitivity bound.

IEEE Internet Computing

Cyberbullying Detection With Fairness Constraints

Cyberbullying is a widespread adverse phenomenon among online social interactions in today's digital society. While numerous computational studies focus on enhancing the cyberbullying detection performance of machine learning algorithms, proposed models tend to carry and reinforce unintended social biases. In this article from the January/February 2021 issue of *IEEE Internet Computing*, the authors try to answer the research question, "Can we mitigate the unintended bias of cyberbullying detection models by guiding the model training with fairness constraints?" They propose a model training scheme that can employ fairness constraints and validate the approach with different data sets. Various types of unintended biases can be successfully mitigated without impairing the model quality.

IT Professional

Simulation-Based Training via a "Readymade" Virtual World Platform: Teaching and Learning With Minecraft Education

Integrating information and communication technologies in teaching and learning is recognized as a relevant and promising practice. However, integration can be complex, tedious, and draining. The authors of this article from the March/April 2021 issue of *IT Professional* propose an innovative approach for enhancing teaching and learning in a simple, pleasant, and effective way: repurposing a virtual world platform to develop simulation-based training. Minecraft Education, a virtual world platform, was used to develop simulation-based training on agile project management. So far, 153 university students have applied the Scrum framework to execute a simulated project in Minecraft Education. Results show that students' perceived learning is very positive and that their learning experience was stimulating and challenging. Recommendations are formulated to guide practitioners and teachers adopting and adapting a "readymade" virtual world platform for teaching and learning.

IEEE micro

The Design Process for Google's Training Chips: TPUv2 and TPUv3

Google has been deploying computers for machine learning training since 2017, powering key Google services. These tensor processing units are composed of chips, systems, and software, all codesigned in house. In this article from the March/April 2021 issue of *IEEE Micro*, the authors detail the circumstances that led to this outcome, the challenges and opportunities observed, and the approach taken for the chips. They also present a quick review of performance and a retrospective on the results.

IEEE MultiMedia

AffectiveNet: Affective-Motion Feature Learning for Microexpression Recognition

The interpretation of microemotions from video clips is a challenging task. In this article from the January–March 2021 issue of *IEEE MultiMedia*, the authors propose an affective-motion imaging that cumulates rapid and short-lived variational information of microexpressions into a single response. An affective-motion feature learning network, called AffectiveNet, can perceive subtle changes and learns the most discriminative dynamic features to describe the emotion classes. AffectiveNet holds two blocks: MICRoFeat and MFL. The MICRoFeat block conserves the scale-invariant features, which allows the network to capture both coarse and tiny edge variations. The MFL block learns microlevel dynamic variations from two intermediate convolutional layers. The effectiveness of the proposed network is tested over four data sets using two experimental setups: person-independent and cross-data set validation.

IEEE pervasive COMPUTING

MOSAIK: A Formal Model for Self-Organizing Manufacturing Systems

In this article from the January–March 2021 issue of *IEEE Pervasive Computing*, the authors review past and current system architectures displaying self-organization in the domain of manufacturing. Based on a corpus of 84 reference papers, they find that multiagent systems (MAS) play a significant role in self-organization, especially MAS featuring bio-inspired algorithms for agent coordination. The emergence of new classes of cyberphysical systems further strengthens the prevalence of MAS on the subject. The authors devise the MOSAIK model, a generic model synthesizing all system architectures found in the corpus. The MOSAIK model can be used as a reference for formally comparing distinct architectures.

IEEE SECURITY & PRIVACY

The Fusion of Secure Function Evaluation and Logic Synthesis

Designing custom secure function evaluation compilers has been an active research area. However, the intelligent adaptation of the integrated circuit synthesis tools outperforms these compilers. It is time for the custom compilers to embrace this trend. Read more in this article from the March/April 2021 issue of *IEEE Security & Privacy*.

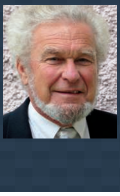
IEEE Software

Visualizing Change in Agile Safety-Critical Systems

High-dependability software systems must be developed and maintained using rigorous safety-assurance practices. By leveraging traceability, we can visualize and analyze changes as they occur, mitigate potential hazards, and support greater agility. Read more in this article from the May/June 2021 issue of *IEEE Software*.

Editorial: Unless otherwise stated, bylined articles, as well as product and service descriptions, reflect the author's or firm's opinion. Inclusion in *Computer* does not necessarily constitute endorsement by the IEEE or the IEEE Computer Society. All submissions are subject to editing for style, clarity, and space.

50 & 25 YEARS AGO



EDITOR ERICH NEUHOLD
University of Vienna
erich.neuhold@univie.ac.at

AUGUST 1971

In the early years, *Computer* was only published bimonthly. Therefore, we will have to skip our interesting and/or informative extractions for August. The next one will appear in the September 2021 issue of *Computer*, and we hope you will eagerly wait for our next publication of this column.

AUGUST 1996

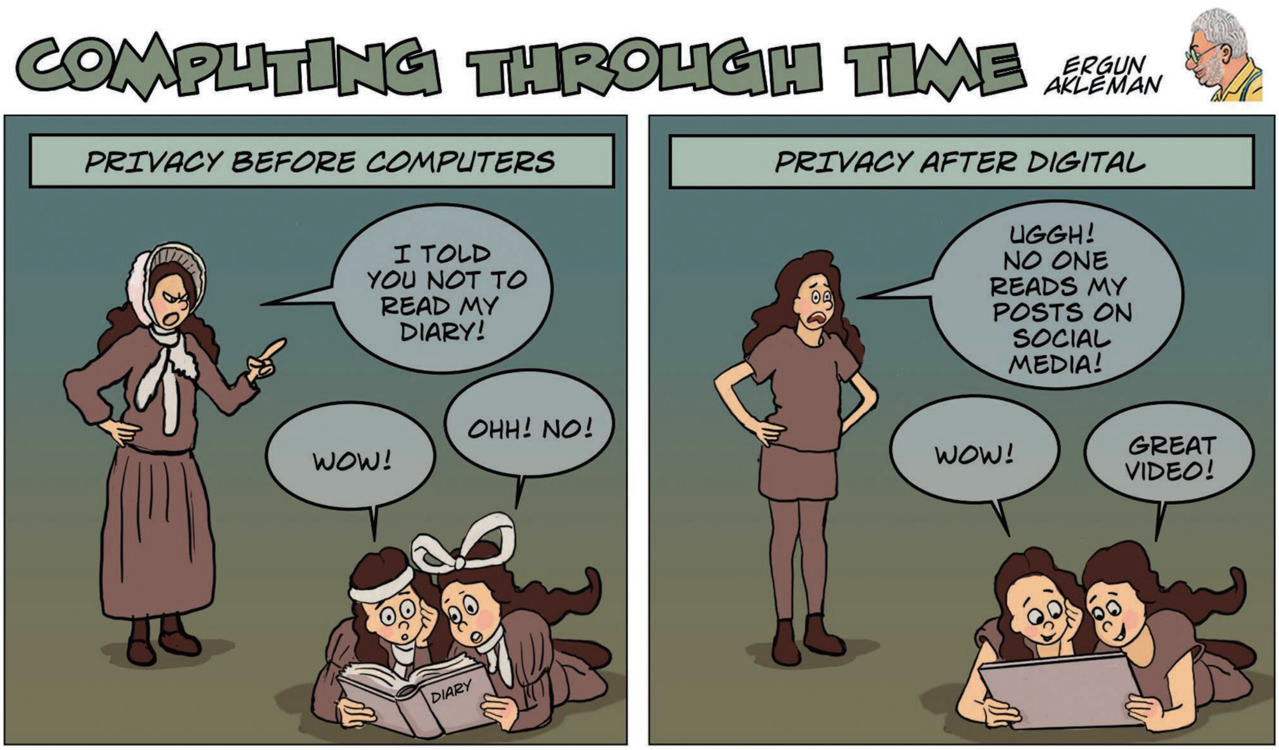
https://www.computer.org/csdl/magazine/co/1996/08

Developments Fuel Wireless Revolution; Thomas Kineshige (p. 16) "Analysts estimate there will be 40 million cellular users by the end of this year. … Developments in personal communication services (PCS), personal digital assistants (PDAs) and wireless notebooks technologies promise still more growth. … GSM, the European digital cellular phone standard is used by about 150 wireless providers in about 80 countries. … A survey conducted this year show that 24% of the US' network developers have committed to PCS, 26% have committed to AT&T's TDMA … 48% have committed to CDMA … By the year 2000 we will see all of the major TDMA and GSM networks evolve towards a spread-spectrum solution, a broadband, SDMA based interface, McClellan said." *[Editor's note: Already in 1996, despite this prediction, it was quite clear that GSM and its successor would conquer the world. Today's smartphones are based on this technology, which allows worldwide communication. Other solutions still exist but have serious limitations when used internationally.]*

Java, the Web, and Software Development; Edward Yourdon (p. 25) "What's the big deal about Java and the Web? The fact that they mark the death of fatware and the birth of dynamic computing built on rented components. … The recent explosion of interest in the Internet and the World Wide Web is just the latest example. The Internet is older than my college friend, but the WWW and programming languages like Java are quite new-and quite revolutionary. … Even if you ignore the competitive marketing pressures that led to the creation of 1,200 spreadsheet features, an important question remains: Why can't we provide users with only the features they want?" (p. 26) "But we do know that Java has created the opportunity to download the specific functionality a user wants at the moment the user requests that functionality. Obviously, this vision implies a user connected to the Internet and a functionality that can be invoked from within a Web browser. … Indeed, the notion of renting functionality via the Internet may further invigorate one segment of the soft-ware industry, the component developers. … Java and Internet-enabled applications will have the most significant effect on brand-new applications that reside on the Internet. But I also believe we'll see a significant transition from today's "traditional" client-server applications to this new model." (p. 28) "Security—a major topic of concern on the Internet—rarely played a significant role in the design of traditional client-server applications. … On the Internet, a casual attitude toward security is likely to have grave consequences." (p. 29) "Internet-based applications will be the subject of debate. In this regard, it's interesting to note that Java is currently about 30 times slower than C+ +." *[Editor's note: This article is really worthwhile to read. It is from the heyday of the Internet and languages like Java. The many predictions and suggestions are all very important. Some have become reality quickly, while others are still problems today, for example, the Internet and application security and privacy issues coming from neglected attitudes in the past.]*

Collaboratories: Doing Science on the Internet; Richard T. Kouzes et al. (p. 40) "The success of many complex scientific investigations hinges on bringing the capabilities of diverse individuals from multiple institutions together with state-of-the-art instrumentation. … Facilitating collaboration among a widely distributed scientific community is highly complex. Although a collaboratory is potentially nothing less than the village square of the Information Age, it is a synthetic place requiring social adaptation. … Collaboratory

Digital Object Identifier 10.1109/MC.2021.3055876
Date of current version: 30 July 2021



PRIVACY HAS HISTORICAL ROOTS IN ARISTOTLE'S TWO SPHERES OF LIFE: THE PUBLIC AND THE PRIVATE. HOWEVER, THE CONCEPT OF UNIVERSAL INDIVIDUAL PRIVACY IS A MODERN ONE. THE SYSTEMATIC TREATISES OF PRIVACY APPEARED AROUND THE 1890S, WITH THE DEVELOPMENT OF PRIVACY LAWS. THE WAY WE VIEW PRIVACY TODAY IS QUICKLY CHANGING AS TECHNOLOGY KEEPS ADVANCING.

Digital Object Identifier 10.1109/MC.2021.3083372
Date of current version: 30 July 2021

developers must consider psychosocial issues such as autonomy, trust, sense of place, and attention to ritual. … Technology solutions abound, but often fail to find a human problem to solve." (p. 41) "Groupware applications for a collaboratories will have to be selected and implemented with a clear understanding of the social and political concerns that characterize joint scientific work. Among these are issues of authorship, acknowledgment of contributions, esteem of peers, and recognition by professional role models. Without such characteristics, collaboratory systems will not find acceptance" (p. 45) "In the past 25 years collaboratories have sprung up and have been used extensively in the internet environment. But most of the systems offered had to be parameterized extensively to be accepted with regards of the mentioned concerns." *[Editor's note: Unfortunately, when looking at the forced home office, teleresearch, and teaching activities due to COVID-19, most of these concerns have been ignored, leading to a serious loss of productivity, teamwork, and learning, which are only now beginning to be recognized for their severe impacts.]*

Distributed Computing Using Autonomous Objects; Lubomir F. Bic et al. (p. 55) "Sensors that supply data to computer systems are inherently unreliable. … To improve sensor-system reliability, researchers have actively studied the practical problem of combining, or fusing, the data from many independent sensors into one reliable sensor reading. … The central question is, how can an automated system be certain to make the correct decision in the presence of faulty data? Much depends on the system's accuracy—the distance between its results and the desired results—and on the system's precision—the size of the value range it returns." (p. 57) "To satisfy the requirements of both the inexact-agreement problem and the sensor-fusion problem, we merged the optimal region algorithm with FCA to produce an algorithm that provides the best accuracy possible and increases the precision of distributed decision-making." *[Editor's note: This is a very interesting paper, especially now in the age of the Internet of Things. It discusses, in depth, a number of already-proposed algorithms that solve the problem and then offers a hybrid solution that eliminates some of the problems the others have.]*

Toward Intelligent Meeting Agents; Hsinchun Chen et al. (p. 62) "An experiment with an AI-based software agent shows that it can help users organize and consolidate ideas from electronic brainstorming. The agent recalled concepts as effectively as experienced human meeting facilitators and in a fifth of the time." (p. 63) "A major advantage of electronic meetings is that members can brainstorm in parallel. Indeed, electronic meeting systems are generally very effective during idea generation. A major disadvantage is that all the ideas from brainstorming—typically several hundred comments—must be organized. … must meet several challenges: Information overload—Lack of a collaborative vocabulary—Pressure to synthesize tasks—Sensitive topics and lack of trust." (p. 64) "Basically, a meeting facilitator can invoke the agent at any time to produce a category list of ideas. Each category is linked to specific comments, which users can browse." (p. 68) "Once again, the agent's lists were comparable to facilitators' lists in concept recall, but significantly inferior in concept precision." *[Editor's note: This work on supporting electronic brainstorming for finding interesting/executable ideas out of the many discussed during the meeting contains many interesting aspects. Some of them made it into the multitude of recommender systems that are used everywhere today.]*

3D User Interfaces for General-Purpose 3D Animation; Jean-Francis Balaguer et al. (p. 71) "Virtual Studio provides nonprofessional animators with inexpensive and easy-to-use 3D animation systems that have the functionality of complex and costly professional platforms." (p. 72) "You can use 3D devices to specify complex 3D motion and virtual tools to control application objects. An underlying constraint solver that automatically maintains inter-relationships among objects will tightly couple application and interface objects. … Device configurations: … The immersive configuration's goal is to convince users that they are part of the world they manipulate or, in other words, to create virtual reality. … Virtual Studio's desktop configuration, shown in Figure 3, uses a Spaceball and a mouse as input devices." (p. 75) "Defining animations: When users define animations in Virtual Studio, they first express the desire to record changes in various elements of the virtual world. A controller object that is connected to each element monitors animated objects' state changes." *[Editor's note: This article discusses the many problems related to creating and using 3D worlds. It reflects the state of the art of 25 years ago where it is interesting to see that, despite the impressive progress since then, some problems still remain, both on the interaction front and with true 3D world aspects.]*

WW II Colossus Computer Is Restored to Operation; Patricia J. Douglas et al. (p. 79) "On June 6, 1996, the 52nd anniversary of D-day, a rebuilt Colossus computer went into operation, bringing back memories of World War II for many attending the ceremony. Present were Tommy Flowers, original designer of the Colossus, and Bill Tutte, the cryptanalyst who first broke the Fish cipher. Several others who had worked at Bletchley Park in the UK, the home of Allied code-breaking operations during the war, also attended." *[Editor's note: The Colossus computer (10 of them were built during the war) was kept secret for 30 years, and most of the design and user documentation was deliberately destroyed. In Wikipedia, a very interesting account is given about the problem of its reconstruction as well as about the methods used for its code breaking.]*

Digital Library Task Force: Nabil R. Adam et al. (p. 89) "Until recently, electronic information sources served mainly specialized clients, but now these sources will be accessed by a wide range of users, ranging from computer specialists, discipline experts, engineers, and the general public, including novice computer users and students at all levels. These trends have created an emerging, important discipline: digital libraries. … A typical digital library uses a variety of database-management systems. Current DBMSs range from relational and extended relational systems to object-oriented database systems. Relational DBMSs are most often used for the storage of metadata and indexes with attributes that contain pointers to files in a file system." *[Editor's note: In 1996, the term* digital libraries *became popular in research environments. However, at the same time, the World Wide Web arose and eventually took over in the research community. Digital library technology, of course, still exists wherever controlled, indexed information is to be stored and retrieved.]*

Web Publishing: Speed Changes Everything; Steve Hitchcock (p. 91) "Because information is the lifeblood of professionals, any change will significantly affect readers as well as publishers. … Those who criticize the Web's content quality, such as T. Matthew Ciolek (*Computer*, January 1996, pp. 106–108), miss the point … Readers will demand quality and will also expect materials to be delivered at near-instantaneous speed, but that quality will be judged in the context of the new information medium that is the Web itself. … Nor is the author immune to the new demands of on-line publishing. Clear and unambiguous expression, good grammar and phraseology, and logical structuring of argument take time, but are all sequential. … Researchers must now find ways to adapt new, more transient information structures to meet quality expectations while recognizing valuable contributions and focusing debate." *[Editor's note: The changes envisioned in this article did occur much slower than predicted. The world of "quality" publications, on one side, has remain unchanged with regard to reviews and copyright enforcement. On the other side is the web, with its multitude of information-bearing sites. Quality has been sacrificed with fake and easily manipulated news, resulting in information and privacy violations everywhere. The web was built assuming that ethical and moral principles exist in all of its users, which, as it turns out, is a very wrong assumption.]*

EDITOR IN CHIEF JEFFREY VOAS
IEEE Fellow; j.voas@ieee.org

Where's the Silicon?

Nir Kshetri, University of North Carolina at Greensboro
Jeffrey Voas, IEEE Fellow

The world is facing a silicon shortage. We look at some of the key facts and figures to gain some insights.

A Wendy's hamburger commercial asked, "Where's the beef?" We're asking, "Where's the silicon?" Recent news clips concerning a global silicon shortage intrigued us. It's a supply chain case study that comes with an interesting twist: is this problem natural or manufactured? We can't answer that, but here are a few tidbits:

1. Silicon is a metalloid,[1] and China is the largest producer.[2]
2. Silicon demand has been growing for microprocessors to support 5G, self-driving vehicles, artificial intelligence, and other uses. Pandemic lockdowns accelerated the demand for work-from-home products, and China has been accumulating integrated circuits (see Figure 1).
3. In the late 1980s,[3] the prevalent business model among semiconductor designers changed to outsourced manufacturing. In the fabless model, a product vendor designs and sells hardware and semiconductor chips but relies on chip-making factories known as *foundries* to manufacture silicon wafers.
4. East Asia has emerged as the global epicenter of fabless. More specifically, Taiwan Semiconductor Manufacturing Company (TSMC) has been credited with pioneering the "foundry and fabless" model. According to TrendForce, TSMC and Samsung have foundry market shares of 55% and 18%, respectively.[4] About three-quarters of the global semiconductor manufacturing capacity, as well as key suppliers of essential materials, are in Asia (Figure 2).
5. East Asia's dominance is clear in the manufacturing of semiconductor devices. Currently, 100% of the world's highly advanced logic semiconductor (smaller than 10 nm) manufacturing capacity is in two Asian economies: Taiwan, 92%, and South Korea, 8%.[5] In 2020, Samsung and TSMC introduced 5-nm chips. A plan exists to produce 3-nm chips in 2022.[6, 7]
6. A single semiconductor fabrication plant costs US$10–20 billion.[8] Public support is almost certainly required. One important lesson from Taiwan and South Korea was the role that governments played.[5]
7. The United States recently took legislative measures; a 2020 bill, the CHIPS for America Act (H.R. 7178), provides incentives to enable R&D in the semiconductor industry and secure supply chains.[4]

Digital Object Identifier 10.1109/MC.2021.3055877
Date of current version: 30 July 2021

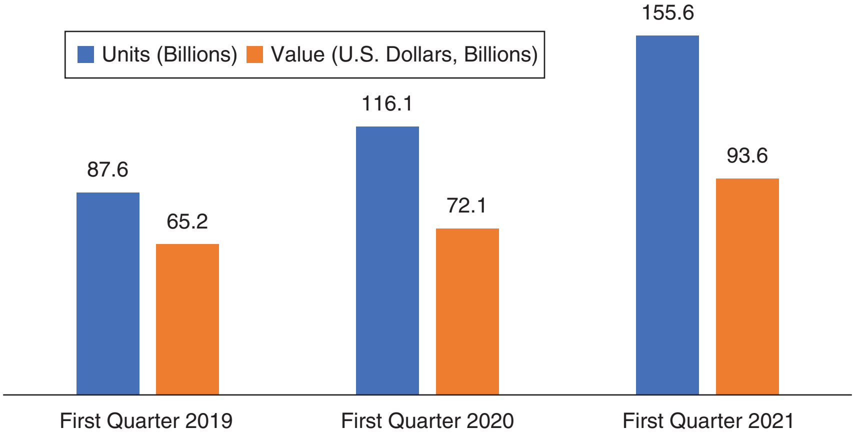


FIGURE 1. China's integrated circuit imports.

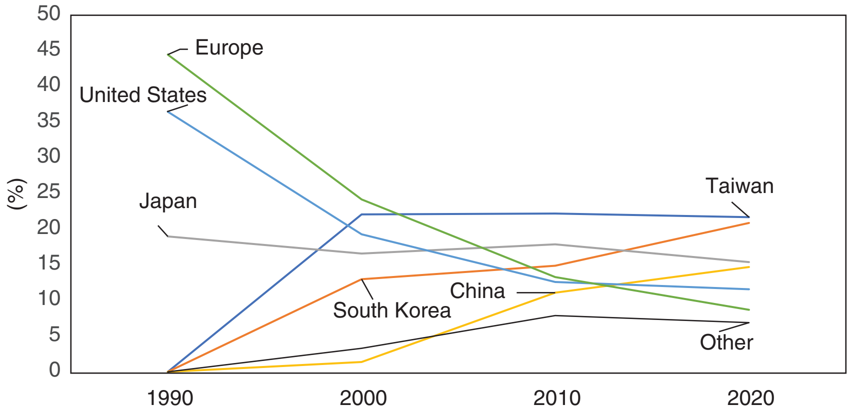


FIGURE 2. Major economies' share of the global semiconductor manufacturing capacity.

So that's some of what we found. We're still hoping to answer our original questions. Expect to read more about this in future *Computer* issues.

REFERENCES

1. S. Pappas. "Facts about silicon." LiveScience, Apr. 27, 2018, https://www.livescience.com/28893-silicon.html#:~:text=Silicon%20is%20neither%20metal%20nor,both%20metals%20and%20non%2Dmetals (accessed May 15, 2021).
2. "Major countries in silicon production from 2010 to 2020 (in 1,000 metric tons)." Statista. https://www.statista.com/statistics/268108/world-silicon-production-by-country/ (accessed on Apr. 26, 2021).
3. J. Ye, "China semiconductor imports surge to all-time high in March amid global chip shortage," South China Morning Post, Apr. 13, 2021. https://www.scmp.com/tech/big-tech/article/3129383/china-semiconductor-imports-surge-all-time-high-march-amid-global (accessed May 15, 2021).
4. A. Kharpal, "How Asia came to dominate chipmaking and what the U.S. wants to do about it," CNBC, Englewood Cliffs, NJ, Apr. 12, 2021. [Online]. Available: https://www.cnbc.com/2021/04/12/us-semiconductor-policy-looks-to-cut-out-china-secure-supply-chain.html
5. "SIA urges U.S. government action to strengthen America's semiconductor supply chain," Semiconductor Industry Association, Washington, D.C., Apr. 5, 2021. [Online]. Available: https://www.semiconductors.org/sia-urges-u-s-government-action-to-strengthen-americas-semiconductor-supply-chain/
6. R. Sharma, "Pound for pound, Taiwan is the most important place in the world." NY Times, Dec. 14, 2020. https://www.nytimes.com/2020/12/14/opinion/taiwan-computer-chips.html (accessed Feb. 15, 2021).
7. S. Kim, "South Korea and Taiwan's chip power rattles the U.S. and China," Bloomberg, Mar. 3, 2021. https://www.bloomberg.com/news/articles/2021-03-03/chip-shortage-taiwan-south-korea-s-manufacturing-lead-worries-u-s-china (accessed Mar. 15, 2021).
8. Y. To, "China chases semiconductor self-sufficiency," East Asia Forum, Feb. 22, 2021. https://www.eastasiaforum.org/2021/02/22/china-chases-semiconductor-self-sufficiency/ (accessed Mar. 15, 2021).

DISCLAIMER

The authors are completely responsible for the content in this message. The opinions expressed are their own.

NIR KSHETRI is a professor of management in the Bryan School of Business and Economics, University of North Carolina at Greensboro, Greensboro, North Carolina, 27412, USA, and the "Computing's Economics" column editor at *Computer*. Contact him at nbkshetr@uncg.edu.

JEFFREY VOAS, Gaithersburg, Maryland, USA, is the editor in chief of *Computer*. He is a Fellow of IEEE. Contact him at j.voas@ieee.org.

75 YEARS IEEE COMPUTER SOCIETY™

Insights Into the Origins of the IEEE Computer Society and the Invention of Electronic Digital Computing

Vladimir Getov, University of Westminster

Using some nearly forgotten facts, this article reviews the origins of the IEEE Computer Society and the first Technical Committee on Electronic Computers, established about 10 years after the revolutionary invention of electronic digital computing.

Digital Object Identifier 10.1109/MC.2021.3084067
Date of current version: 30 July 2021

For decades, the IEEE Computer Society (CS) has been, by far, the largest professional unit within IEEE. As early as 1968, its predecessor, the Computer Group, surpassed the 10,000-member milestone.[11] Since then, the IEEE CS has grown substantially, servicing more than 225,000 community members in 2021. This has been in line with the rapid global developments of computer science and engineering—the main driving force of the unprecedented digital revolution, which has transformed modern life.

The professional community activities in computer science and engineering began in the first year after the World War II. The CS origins can be traced back to 1946 when there were two independent and sometimes rival organizations: the American Institute of Electrical Engineers (AIEE) and the Institute of Radio Engineers (IRE). These two institutions eventually merged in 1963, creating IEEE. A lot has been written on this matter over the last 75 years, but there are still many questions that require answers and clarifications, particularly because the IRE and AIEE had initially created separate committees dedicated to the new field of computing.

Within the AIEE, the Subcommittee on Large-Scale Computing Devices was initially formed as part of the Basic Sciences Committee during May/June 1946 with Charles Concordia as the first chair.[20] In 1947, with the backing of AIEE leadership, the group started preparing for elevation to full committee status while dropping "Large-Scale" from its title. Eventually, the AIEE Committee on Computing Devices was formally approved by the Board of Directors on 29 January 1948 and achieved full official standing on 1 August 1948.

The uniqueness of the IRE lies in the fact that, during the first half of the 20th century, it was the most scientific of all of the American engineering societies. This spirit found its expression in the institute's high membership standards, a stress upon creativity, its democratic elections, and its dedication to international scientific collaboration. There was a strong convergence between the values implicit in the development of a highly scientific field and the professional atmosphere created by the founders of the IRE. In line with these principles, the IRE Technical Committee (TC) on Electronic Computers was initially formed in 1946 as a subcommittee, just like the AIEE Subcommittee on Large-Scale Computing Devices. However, it is still unclear why, in 1951, the IRE decided to establish the Professional Group on Electronic Computers (PGEC)[21] with no evidence showing continuity between the new group and the existing TC. The PGEC creation was a move in the right direction as PGEC turned out to be very successful in the 1950s and the early 1960s, before the merger of the IRE and the AIEE in 1963.

THE INCEPTION OF THE IRE TC ON ELECTRONIC COMPUTERS

The first IRE TC on Electronic Computers played an important historic role, and, therefore, we shall review some nearly forgotten facts. The earliest evidence of professional activities in this new area within the IRE dates back to 1946. The technical program of the highly successful 1947 IRE National Convention (3–6 March 1947) was published in January[6,7] and February 1947.[8,9] It included 122 papers, five of which were presented in the session, "Electronic Digital Computers," on Tuesday, 4 March 1947 in the West Ballroom, The Commodore Hotel, New York City.[10] These papers had to have been prepared and submitted in 1946 for inclusion on the technical program. The authors—J.W. Forrester [Massachusetts Institute of Technology (MIT)], S.N. Alexander [National Bureau of Standards (NBS)], H.H. Goldstine [Institute for Advanced Study (IAS)], J.A. Rajchman (RCA), and P. Crawford [U.S. Office of Naval Research (ONR)]—were officially listed as inaugural members of the IRE TC on Electronic Computers in 1948. Several interesting exhibits were presented in special demonstration rooms, including component units of the Electronic Discrete Variable Computer (EDVAC), which was declassified only one week before the show.

A total of 73 IRE TC and subcommittee meetings were held in 1947.[2] These included the Technical Subcommittee on Electronic Digital Computers, which must have been established in 1946. Later in 1947, Arthur Burks published in *Proceedings of the IRE* his highly cited ENIAC paper,[17] submitted in 1946. The 1947 annual report of the secretary, Haraden Pratt, as received by the IRE Board of Directors, includes the following statement:[2]

> *Three new important Technical Committees, … and* ***the Electronic Computers Committee****, were created, which indicates an increased trend in Technical Committee activities for the future.*

An extensive online search established that, after the successful start of the Subcommittee and the creation of the IRE TC on Electronic Computers in 1947, the IRE Executive Committee approved a bylaw amendment to include the TC on Electronic Computers in Section 80 and appointed James R. Weiner as the first chairman of the same TC at its meeting on 6 January 1948:[3]

> ***Bylaw Section 80.*** *Dr. Goldsmith moved that the Constitution and Laws Committee be instructed to prepare a Bylaw amending Bylaw Section 80 to include the new Technical Committee on "Electronic Computers." (Unanimously approved.)*
>
> ***Chairman, Technical Committee on Electronic Computers.*** *Dr. Goldsmith moved that the Executive Committee approve the appointment of J.R. Weiner as Chairman of the Technical*

> *Committee on Electronic Computers. (Unanimously approved.)*

These decisions not only confirmed the elevation to full committee status but also the drop of "Digital" from the title, which would leave room for the inclusion of analog (continuous) computers in the definition of scope, as approved by the IRE Executive Committee on 2 March 1948:[4]

> *Definition of Scope of the Electronic Computers Committee. Mr. S. L. Bailey moved that the Executive Committee approve the following definition of scope of the Electronic Computers Committee, submitted by the Committee Chairman, James R. Weiner, with the suggestion that the Technical Secretary investigate the use of the term "continuous" in this application:*
>
> *The Technical Committee on Electronic Computers is responsible for all work relating to digital and continuous computers. Included are applications to scientific computing, fire control, and industrial control problems. A primary duty of the Committee will include the compilation of a glossary of definitions designed to correct the many current ambiguities. Additional duties of the Committee include standardization of test methods, coordination with the Papers Procurement Committee, and computer session planning. (Unanimously approved.)*

This TC continued to be very active in 1948 with a committee meeting and nine papers presented in two regular sessions, "Computers I – Systems" and "Computers II – Components," at the 1948 IRE National Convention 22–25 March 1948.[5] Many of the authors of these papers were founding members of the IRE TC on Electronic Computers. Another fascinating fact is that among the invited speakers for the special session "Advances Significant to Electronics" were Norbert Wiener (MIT), "Cybernetics"; Claude Shannon (Bell Labs) , "Information Theory"; and John von Neumann (IAS), "Computer Theory."[5]

In his article from 1991,[1] Merlin Smith gives a list of the 21 founding members of the first IRE TC on Electronic Computers (1 May 1948). Regretfully, this article does not provide a reference to the original publication, which was eventually found in *Proceedings of the IRE*: [2]

> ***Technical Committee on Electronic Computers***
> *Chairman: J.R. Weiner (Raytheon Manufacturing Co.); Vice-Chairman: G.R. Stibitz (Bell Labs); Members: S.N. Alexander (NBS), J.V. Atanasoff (NOL), J.H. Bigelow (Princeton Univ.), Perry Crawford (ONR), C.S. Draper (MIT), J.P. Eckert Jr. (Eckert-Mauchly Computer Corp.), J.W. Forrester (MIT), H. Goldstine (IAS), E.L. Harder (Westinghouse Electric Co.), B.L. Havens (Columbia University), E. Lakatos, G.D. McCann (California State Polytechnic Inst.), C.H. Page, J.A. Rajchman (RCA), Nathaniel Rochester (Sylvania Electric Products Inc.), Robert Serrell (RCA), T.K. Sharpless (Univ. of Pennsylvania), R. Snyder (Univ. of Pennsylvania), and C.F. West (Raytheon Manufacturing Co.).*

The IRE TC on Electronic Computers inaugural roster (as listed above) is an impressive collection of early computer pioneers that includes the inventor of electronic digital computing: John Vincent Atanasoff.

THE INVENTION OF ELECTRONIC DIGITAL COMPUTING

The 1930s saw an unprecedented ebullience of scientific ideas and proposals looking for different solutions to the automatic calculations challenge. The variety of interests from academic institutions, commercial endeavors, and government administration further added to the richness of possible approaches. It was in these circumstances that Atanasoff's exceptional invention and development of electronic digital computing marked the beginning of the information revolution.[18] In 1937, after significant research and practical investigations, Atanasoff came up with the basic design principles of electronic digital computing.[13] These included the use of

- electronics technology for computational speed as opposed to mechanical or electromechanical technology
- binary arithmetic for simplicity of implementation as opposed to decimal arithmetic
- digital calculations for accuracy as opposed to analog calculations
- dynamically refreshed memory for low cost and reliability.

Based on these revolutionary concepts and after further practical investigation, a proof-of-concept prototype became operational and was demonstrated in October 1939. This was followed by the development of a full-scale computing machine [called the Atanasoff-Berry Computer (ABC) since the late 1960s] for solving systems of equations using digital electronics. It was demonstrated between 1939 and 1942 by Atanasoff and his graduate assistant, Clifford E. Berry.

A few years after Atanasoff's project, the British "Colossus" electronic digital computer was independently designed and built at the Post Office Research Labs at Dollis Hill in North London between March and December 1943.[22] The first prototype, developed for the top-secret code-breaking facility (known as "Station X") at Bletchley Park, was first operational on Christmas Day, 1943. After receiving highly favorable feedback from its initial operation, the project team, led by

Thomas H. (Tommy) Flowers, built another 10 improved Colossus versions, which were extensively used to break the German "fish" codes in the last two years of World War II.

Around the same time, another groundbreaking project, classified as confidential and led by J. Presper Eckert Jr. and John W. Mauchly, started in June 1943 at the University of Pennsylvania. This was the Electronic Numerical Integrator and Computer (ENIAC), which became operational in December 1945 and was announced to the media in February 1946, followed by a well-run publicity campaign. ENIAC's popularity remains widely acknowledged within the professional community to this day.

There are, however, some important facts which need to be emphasized again. In June 1941, Mauchly visited Atanasoff in Iowa State to learn in detail about his project. It is well documented that Atanasoff had submitted a patent application to Iowa State, but the university neglected to file it. Atanasoff also informed Mauchly about this matter and stated in his letter to him (dated 7 October 1941): "I have no qualms about having informed you about our device, but it does require that we refrain from making public any details for the time being."

Soon after that, ENIAC's design and construction were derived with no acknowledgment of Atanasoff's invention. As stated in Judge Earl R. Larson's Federal Court's historic decision (19 October 1973):

> *Between 1937 and 1942, Atanasoff, then a professor of physics and mathematics at Iowa State College, Ames, Iowa, developed and built an automatic electronic digital computer. The work of Atanasoff was known to Mauchly before any effort pertinent to the ENIAC machine or patent began. Eckert and Mauchly did not themselves first invent the automatic electronic digital computer, but instead derived that subject matter from Dr. John Vincent Atanasoff.*

Despite the long list of publications, discussions, interviews, and announcements, some of the details are still not well elucidated. For example, it is often believed that the computer's name, "ABC," was given during the development of the project. In fact, Atanasoff started using the acronym in the late 1960s to recognize Berry's contribution in his testimony for the court case on 15 June 1971.[13] Before that, he referred to it as the "computing machine"[14] while the only other book providing a two-page description of the "Atanasoff-Berry computer" was published in 1966.[12]

This again confirms Atanasoff as the sole inventor of electronic digital computing concepts and his innovative contributions to the creation of modern computers. While Berry had proven himself as a gifted young engineer at the time of his graduation in the summer of 1939, we have not seen published evidence about any contributions by him to the invention of electronic digital computing. In one of his letters (12 July 1963) to R.K. Richards,[23] Berry mentions September 1939 as the first month when he was fully occupied as a graduate assistant, building the frame without any real idea about what was going to go in the machine. At the same time, in October 1939, Atanasoff demonstrated the operation of his first partial prototype. Even the topic area of Berry's master's thesis, titled "Design of an Electrical Data Recording and Reading Mechanism" (1941), had been provided solely by his supervisor.[13] Atanasoff and Berry worked very closely together on the implementation, and existing publications are very clear about their productive partnership.

Berry quickly developed as one of the most forward-looking computer designers at that time. Applying his skills from the computing machine project with Atanasoff in his future work, he was initially the project manager of the electronic analog machine

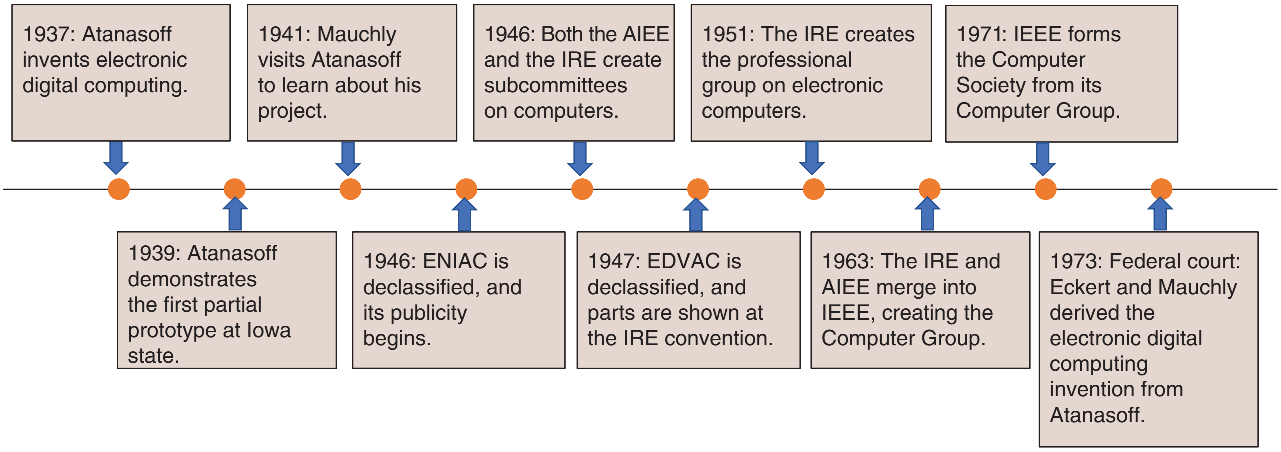

1937: Atanasoff invents electronic digital computing.
1939: Atanasoff demonstrates the first partial prototype at Iowa state.
1941: Mauchly visits Atanasoff to learn about his project.
1946: ENIAC is declassified, and its publicity begins.
1946: Both the AIEE and the IRE create subcommittees on computers.
1947: EDVAC is declassified, and parts are shown at the IRE convention.
1951: The IRE creates the professional group on electronic computers.
1963: The IRE and AIEE merge into IEEE, creating the Computer Group.
1971: IEEE forms the Computer Society from its Computer Group.
1973: Federal court: Eckert and Mauchly derived the electronic digital computing invention from Atanasoff.

FIGURE 1. The main milestones in the first 35 years of electronic digital computing.

CEC 30-103[15] at the ElectroData part of the Consolidated Engineering Corporation (CEC). After ElectroData's acquisition by Burroughs Corporation in 1956, Berry led Burroughs into electronic digital computing.

The list of early commercially available electronic computers includes (in chronological order): CEC 30–103 (1949), Universal Automatic Computer (UNIVAC) I (1951), and IBM 701 (1952). Both CEC 30-103 and the later completed CEC-30-201 (1954) were electronic analog computers and, therefore, relatively small systems while UNIVAC I and IBM 701 were electronic digital computers.

Of all of the brilliant scientists who have shaped the early years of electronic computing, Atanasoff was the first one to use digital electronics to implement arithmetic operations.[24] About 10 years after his revolutionary invention, organized technical activities in the field began (see Figure 1). The IRE TC on Electronic Computers built critical mass from the very beginning of its existence, initially as a subcommittee. It played a particularly influential role with its very strong membership, which included Atanasoff, Eckert, Stibitz, Weiner, Rochester, Goldstine, Alexander, and others. Atanasoff's design principles propagated via ENIAC and EDVAC (both declassified in 1946–1947) to most of the commercially available modern computers and remain at the core of electronic digital computing technologies to the present day.

A fully identical reconstruction of Atanasoff's original computing machine was completed[16] and demonstrated[25] in the 1990s. Since 2011, when it was moved to the Computer History Museum in Mountain View, this reconstruction has been attracting the wider public's attention.[19]

ACKNOWLEDGMENTS

This article presents the personal views of the author. Numerous discussions with colleagues and friends, including John V. Atanasoff II, Gordon Bell, Kiril Boyanov, Carl Chang, Dick Gowen, John Gustafson, Anne Marie Kelly, Blagovest Sendov, Mike Williams, and Steve Yau, have helped in gathering important facts, listening to stories, and developing a better understanding of the fascinating chain of events in the early years of modern computer history.

This article has supplementary downloadable material available at https://doi.org/10.1109/MC.2021.3084067, provided by the author.

REFERENCES

1. M. G. Smith, "IEEE Computer Society: Four decades of service," *Computer*, vol. 24, no. 9, pp. 6–12, Sept. 1991. doi: 10.1109/2.84894.
2. "Institute news and radio notes," *Proc. IRE*, vol. 36, no. 6, p. 761, June 1948. doi: 10.1109/JRPROC.1948.230545.
3. "Institute news and radio notes," *Proc. IRE*, vol. 36, no. 4, p. 505, Apr. 1948. doi: 10.1109/JRPROC.1948.229654.
4. "Institute news and radio notes," *Proc. IRE*, vol. 36, no. 5, p. 633, May 1948. doi: 10.1109/JRPROC.1948.226203.
5. "Institute news and radio notes," *Proc. IRE*, vol. 36, no. 3, pp. 365–380, Mar. 1948. doi: 10.1109/JRPROC.1948.233924.
6. "1947 IRE National Convention and Radio Engineering Show," *Proc. IRE: J. Commun. Electron. Eng.*, vol. 35, no. 1, p. 3, Jan. 1947. [Online]. Available: https://worldradiohistory.com/hd2/IDX-Site-Technical/Engineering-General/Archive-IRE-IDX/IDX/40s/IRE-1947-01-OCR-Page-0003.pdf
7. "Institute news and radio notes," *Proc. IRE*, vol. 35, no. 1, p. 49, Jan. 1947. doi: 10.1109/JRPROC.1947.231221.
8. "Program for the 1947 IRE National Convention," *Communications*, vol. 27, no. 2, pp. 18–40, Feb. 1947. [Online]. Available: https://www.rsp-italy.it/Electronics/Magazines/Communications/Communications1947 02.pdf
9. "Extensive plans set for 1947 IRE National Convention March 3, 4, 5, and 6 in New York," *Proc. IRE*, vol. 35, no. 2, pp. 172–184, Feb. 1947. doi: 10.1109/JRPROC.1947.231597.
10. 1947 IRE National Convention, "Institute news and radio notes," *Proc. IRE*, vol. 35, no. 5, pp. 499–503, May 1947. 10.1109/JRPROC.1947.226253.
11. J. D. Ryder and D. G. Fink, *Engineers & Electrons: A Century of Electrical Progress*. Piscataway, NJ: IEEE Press, 1984. [Online]. Available: https://ethw.org/w/images/c/cc/Engineers_&_Electrons.pdf
12. R. K. Richards, *Electronic Digital Systems*. Hoboken, NJ: Wiley, 1966.
13. J. V. Atanasoff, "Advent of electronic digital computing," *IEEE Ann. Hist. Comput.*, vol. 6, no. 3, pp. 229–282, July–Sept. 1984. doi: 10.1109/MAHC.1984.10028.
14. J. V. Atanasoff, *Computing Machine for the Solution of Large Systems of Linear Algebraic Equations*. Ames, IA: Iowa State College, Aug. 1940. Republished in B. Randell (Ed.), *The Origins of Digital Computers*, pp. 305–325. Springer-Verlag, 1973.
15. J. Strickland, "Stories," *Volunteer Inform. Exchange*, vol. 1, no. 7, p. 2, Computer History Museum, 2011. [Online]. Available: http://s3data.computerhistory.org/chmedu/VIE-007.pdf
16. J. Gustafson, "Reconstruction of the Atanasoff–Berry computer," in *The First Computers—History and Architectures*, R. Rojas and U. Hashagen, Eds. Cambridge, MA: MIT Press, 2000, pp. 91–106. [Online]. Available: https://www.researchgate.net/publication/262402944_Reconstruction_of_the_Atanasoff-Berry_computer
17. A. W. Burks, "Electronic computing circuits of the ENIAC," *Proc. IRE*, vol. 35, no. 8, pp. 756–767, Aug. 1947. doi: 10.1109/JRPROC.1947.234265.
18. A. R. Burks and A. W. Burks, *The First Electronic Computer: The Atanasoff Story*. Ann Arbor, MI: Univ. of Michigan Press, 1988.

19. C. Severance. *Interview with John C. Hollar–History of Computing, Computing Conversations, IEEE Computer Society* (Aug. 24, 2013). Accessed June 20, 2021. [Online Video]. Available: http://youtu.be/poh9ROv_LR8
20. C. Concordia, "In the beginning there was the AIEE Committee on computing devices (the first 25 years)," *Computer*, vol. 9, no. 12, pp. 42–44, Dec. 1976. doi: 10.1109/C-M.1976.218469.
21. M. M. Astrahan, "In the beginning there was the IRE professional group on electronic computers (the first 25 years)," *Computer*, vol. 9, no. 12, pp. 43–44, Dec. 1976. doi: 10.1109/C-M.1976.218469.
22. T. H. Flowers, "The design of Colossus," *IEEE Ann. Hist. Comput.*, vol. 5, no. 3, pp. 239–252, July 1983. doi: 10.1109/MAHC.1983.10079.
23. J. R. Berry, "Clifford Edward Berry, 1918–1963: His role in early computers," *IEEE Ann. Hist. Comput.*, vol. 8, no. 4, pp. 361–369, Oct. 1986. doi: 10.1109/MAHC.1986.10066.
24. J.V. Atanasoff. *The First Electronic Digital Computer, Computer History Museum* (Nov. 11, 1980). [Online Video]. Accessed June 20, 2021. Available: https://youtu.be/Yxrcp1QSPvw
25. J. Gustafson and C. Shorb. *The Atanasoff–Berry Computer in Operation*. Iowa State University (1999). [Online Video]. Accessed June 20, 2021. Available: https://youtu.be/YyxGIbtMS9E

VLADIMIR GETOV is with the University of Westminster, London, W1W 6UW, United Kingdom. Contact him at V.S.Getov@westminster.ac.uk.

Digital Object Identifier 10.1109/MC.2021.3094699

Digital Object Identifier 10.1109/MC.2021.3094700

Safety, Security, and Reliability of Autonomous Vehicle Software

W. Eric Wong, University of Texas at Dallas

Zijiang Yang, Xi'an Jiaotong University

Digital Object Identifier 10.1109/MC.2021.3084655
Date of current version: 30 July 2021

This special issue is focused on the safety, security, and reliability of autonomous vehicle software. Among all of the submissions, five articles were accepted covering different topics of the scope.

Autonomous vehicles are becoming ubiquitous and are having a greater impact on our everyday life. A dependable and reliable autonomous driving system that will not only reduce the number of accidents but also minimize driving-related human stress is in demand. However, recent fatal accidents due to the application of immature and unreliable autonomous vehicle software have undermined our trust in these systems. In response, this theme issue solicits original work that makes important contributions to ensure the safety, security, and reliability of autonomous vehicle software.

Below is a summary of the six articles accepted for this issue.

The first article is "Safety, Complexity, and Automated Driving: Holistic Perspectives on Safety Assurance" by Simon Burton, John McDermid, Philip Garnet, and Rob Weaver. The authors propose a framework to identify, analyze, and manage factors that impact the safety of complex automated driving systems.

The second article, "Blockchain-Based Continuous Auditing for Dynamic Data Sharing in Autonomous Vehicular Networks," emphasizes that cloud servers have made it possible to share massive data between autonomous vehicles to improve the driving experience and service quality. Data security is an important issue that needs to be addressed. The authors are Haiyang Yu, Shuai Ma, Qi Hu, and Zhen Yang.

The third article, by Sachin Motwani, Tarun Sharma, and Anubha Gupta, is "Ethics in Autonomous Vehicle Software: The Dilemmas." The ethical dilemma is a major challenge faced by the automobile industry while designing safe, secure, and reliable software for autonomous vehicles. The article presents some novel solutions to this problem.

The fourth article, "An Online Multistep-Forward Voltage-Prediction Approach Based on an LSTM-TD Model KF Algorithm," is authored by Ye Ni, Zhilong Xia, Chunrong Fang, and Zhenyu Chen, and Fangtong Zhao. Since the existing approaches often take too much time to predict the remaining voltage for battery management in electronic vehicles, a multistep-forward approach combining a long short-term memory time distributed model and a Kalman filter algorithm is recommended for time-efficient voltage prediction.

The title of the fifth article is "Toward Improving Confidence in Autonomous Vehicle Software: A Study on Traffic Sign Recognition Systems," with Koorosh Aslansefat, Sohag Kabir, Amr Abdullatif, Vinod Vasudevan Nair, and Yiannis Papadopoulos as the authors. It reviews the issue of distributional shift and its implications for the safety of learning-based classification tasks in autonomous vehicle software. SafeML II (an extension of SafeML) using a bootstrap-based *p*-value calculation is presented to improve the empirical cumulative distribution function-based statistical distance measure.

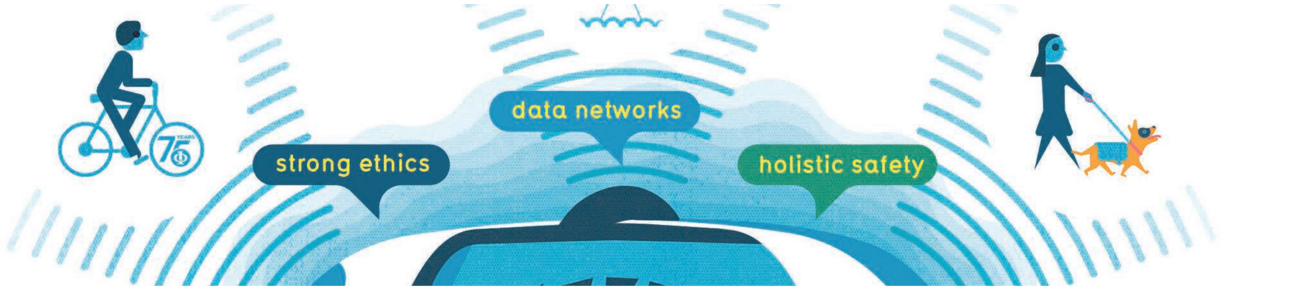
The last article, by Jack Toohey, M S Raunak, and Dave Binkley, is "From Neuron Coverage to Steering Angle: Testing Autonomous Vehicles Effectively." The authors explore the use of image transformation to create new test images and how these images impact the neuron coverage achieved by a deep neural network-based system for autonomous vehicle operations.

We would like to thank the authors of the six articles in this issue for sharing their knowledge and experiences on how to improve the safety, security, and reliability of autonomous vehicle software. We also thank all of the reviewers for helping us evaluate the articles and selecting those of high quality to be included in this theme issue. Special appreciation also goes to Dr. Jeffrey Voas, editor in chief of *Computer*, and IEEE staff members for their support during the preparation of this issue. C

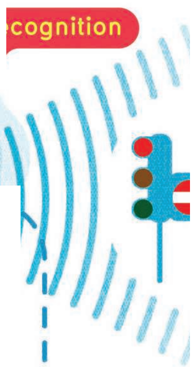
ABOUT THE AUTHORS

W. ERIC WONG is a professor and the founding director of the Advanced Research Center for Software Testing and Quality Assurance in Computer Science at The University of Texas at Dallas, Richardson, Texas, 75080, USA. His research interests include software testing, debugging, risk analysis/metrics, safety, and reliability. Wong received a Ph.D. in computer science from Purdue University. Contact him at ewong@utdallas.edu.

ZIJIANG YANG is a professor at Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China. His research focuses on developing formal method-based tools to support the debugging, analysis, and verification of complex systems. Yang received a Ph.D. from the University of Pennsylvania. Contact him at zijiang@gmail.com.



Safety, Complexity, and Automated Driving: Holistic Perspectives on Safety Assurance



Simon Burton, Fraunhofer Institute for Cognitive Systems and University of York
John McDermid and Philip Garnet, University of York
Rob Weaver, Independent Consultant

This article extends safety assurance approaches for automated driving by explicitly acknowledging the complexity of the emergent system behavior. We introduce a framework for reasoning about factors that contribute to this complexity as a means of structuring the interdisciplinary perspectives required to achieve an acceptable level of residual risk.

Human error is by far the greatest contributing factor to fatal incidents on roads in the United Kingom,[1] while environmental effects (8%) and vehicle defects (2%) play a relatively insignificant role in comparison. Automated driving systems (ADSs) have the potential to make roads significantly safer by optimizing traffic flow, recognizing and reacting to hazards on the route ahead, and limiting the impact of inattentive and unreliable human drivers. The first steps on the path to introducing highly automated driving technologies are already underway, with a recent public consultation regarding regulations for automated lane-keeping systems (ALKSs) in the United Kingdom[2] and Honda announcing limited availability of the technology in its Legend vehicles. These systems would control longitudinal and lateral movements of the vehicles at relatively low speeds, for example, in traffic jam situations on motorways. ALKSs differ from driver assistance systems

Digital Object Identifier 10.1109/MC.2021.3073430
Date of current version: 30 July 2021

currently on the market by not requiring continuous attention of the driver while operating in the automated driving mode.

The authors believe that the introduction of an ADS requires the consideration of safety beyond the traditional focus of the engineering community. We, therefore, do not restrict our analysis to technological aspects but consider the role of governance, management, operation, and assessment of human factors to establish a holistic view of safety. In doing so, we explicitly acknowledge the emergent complexity of the system and suggest that a "vehicle-centric" focus for engineering safe automated driving may be inadequate.

As part of a study[3] commissioned by Engineering X, an initiative coordinated by the Royal Academy of Engineering and supported by the Lloyd's Register Foundation, the authors were tasked with producing a framework (hereafter referred to as the safer complex systems framework) to provide conceptual clarity around the factors that lead to systemic failures in complex systems that have a safety impact.

This article builds on the results of the study. The following section discusses the relation of complexity to the safety of autonomous systems. We then introduce a framework for identifying factors that impact the safety of complex systems. The framework is illustrated by using it to model an incident involving a prototypical automated vehicle and then used to analyze the potential risks associated with interactions among systems when deploying ALKSs on public roads. Recommendations for addressing complexity in the safety assurance of automated driving are provided, as are directions for future work.

COMPLEXITY AND SAFETY

What is a complex system?

Complex systems theory defines a system as complex if some of the behaviors of the system are emergent properties of the interactions between the parts of the system, where the behaviors would not be predicted based on knowledge of the parts and their interactions alone. From the perspective of complexity science, there are a number of characteristics[4] that are shared by most, if not all, complex systems. These include the following, among others:

- *Semipermeable boundaries*: The boundaries between the system and environment are dependent on the scope of the system under consideration, known as the system of interest. This may vary depending on the objectives of the analysis. For example, if focusing on the functional performance of an automated vehicle, the system could be viewed as a set of electronic components that sense the environment, decide on control actions, and implement them via actuators. However, when considering a mobility service as a whole, the system includes other traffic participants, emergency services, and city or highway infrastructure as well as the impact on the use of public transport.[5]
- *Nonlinearity, mode transitions, and tipping points*: The system may respond in different ways to similar inputs depending on its state or context. Nonlinear behavior can also be caused by coupled feedback both within the system of interest and between the system and its environment. It is common to talk about complex systems going through critical transitions, widely referred to as tipping points. Tipping points can also be transitions into unsafe states, and these can be emergent properties of the system itself. The seemingly spontaneous occurrence of traffic jams and stop–start traffic on motorways are examples of such behavior within traffic systems.
- *Self-organization and ad hoc systems*: Systems can also emerge in an ad hoc manner through a convergence of parts, perhaps by a process of self-organization or self-assembly. Here the (semipermeable) boundary may change as the system evolves. The adaption in the behavior of human road users in response to automated vehicles is an example of self-organization, where the humans become part of a larger ad hoc system. The ability of traffic to spontaneously respond to approaching emergency vehicles, even at complex intersections, is an example of ad hoc self-organization.

Safety of complex systems

Traditional systems safety engineering focuses on component faults and their interactions with other system components and therefore requires some model of the system so that these interactions can be analyzed. In contrast, complex systems can give rise to systemic failures that do not necessarily arise from faults in individual system parts. This bears a strong and deliberate relationship to the definition of complex systems and the notion of emergence. Systemic failures originate from interactions between parts of the system and interactions with or dependencies on the environment rather than faulty components, buggy software functions, or wear and tear.

The difficulty of arguing the safety of an ADS lies in the inherent complexity and unpredictability of the ever-changing environment in which it operates. The system observes this complex, unpredictable environment using sensors that themselves have inherent inaccuracies due to the physical limitations of their sensing modalities. This uncertainty is countered by using multiple sensor types and algorithms that make use of heuristics or machine learning to interpret the sensing data. However, these algorithms are themselves inherently imprecise and introduce an additional level of uncertainty.[6]

The unpredictable nature of the impact of the vehicle's actions on its environment (for example, the reactions of other drivers and road users) "closes the loop" to the complex environment to be interpreted by the vehicle. Thus, implementing an ADS brings with it the potential for systemic failures due to the interactions between these uncertainties within the perception and control cycle. For example, could the introduction of an ALKS increase the propensity for stop–start traffic on motorways?

Risk reduction measures or controls to reduce the probability of safety-related failures can include engineering changes at design time or procedures and processes implemented during operation. Controls can be grouped in broad terms into those that enable robustness (that is, the ability of a system to cope with foreseen events), which we contrast with resilience (that is, the ability of a system to absorb the unforeseeable and remain unchanged).

Both resilience and robustness are tools for reducing risk, with resilience being more important in dealing with the uncertainties arising from complexity. Complexity science uses these terms rather differently. For example, resilience is used to mean that the system returns to its original state or maintains its original function. Here, resilience might mean that the system changes its behavior (or even purpose) but continues to operate safely in the presence of unforeseen events.

THE SAFER COMPLEX SYSTEMS FRAMEWORK

The framework we propose provides a structure for reasoning about factors that contribute to systemic failures due to complexity, and it contextualizes the measures and controls to manage risk.

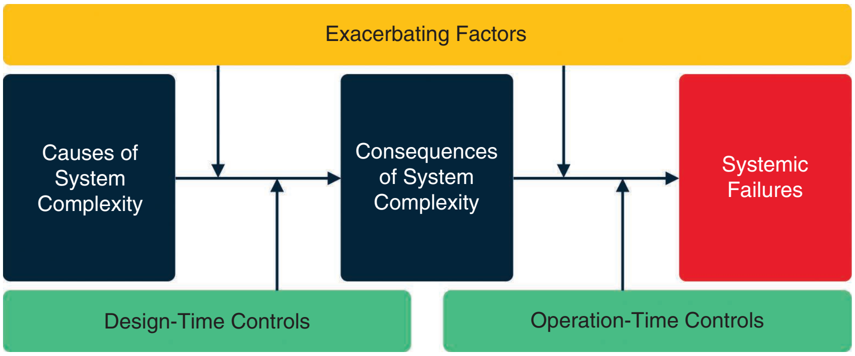


FIGURE 1. The safer complex systems framework.

As visualized in Figure 1, the central axis of the safer complex systems framework shows a flow from the causes of system complexity via their consequences to systemic failures. This is analogous to the causal relationships among faults, erroneous system states, and eventual system failures underlying traditional functional safety engineering as promoted by Avizienis et al.[7] However, systemic failures arise out of emergent properties of the system caused by complexity, not from faults in individual system elements. Also, the interdependencies between system elements as well as the causes and consequences of complexity are more subtle than a simple cause and effect relationship. Therefore, the flow in the diagram should not be interpreted as deterministic; instead, the consideration of factors and their relationships can lead to insights into how systemic failures occur.

The emergence of systemic failures can be tempered by controls at design time and during operation. These reduce the likelihood that systemic failures arise by either suppressing the causes of complexity or by reducing the likelihood that emergent complexity leads to the failure to maintain a system objective. The framework also recognizes the exacerbating factors that can make systemic failure more likely by either amplifying the consequences of system complexity or undermining control measures. The inherent uncertainty in our knowledge of the system and its boundaries as well as politicized decision making are examples of exacerbating factors that can increase the consequences of complexity or undermine the controls.

The relationships between the various elements of the model provide different perspectives on the analysis. For example, to identify measures for reducing the negative impact of emergent complexity within the system, it is important to both understand its potential causes as well as its impact on the system objectives.

In many cases, the causes of complexity and the controls for managing safety are regulatory, organizational, or financial instead of—or in addition to—technical. Furthermore, not all systems are explicitly engineered; they can also arise from ad hoc interactions between systems or components previously considered unrelated. This requires radically different viewpoints to previously applied safety engineering and management techniques. The framework is intended to address both designed and ad hoc systems and considers a system through the lens of the following strongly interacting viewpoints. These can be seen as layers within the overall model, akin to those found in Rasmussen's risk management framework.[8]

- *Governance*: This layer consists of the incentives and requirements for organizations to adhere to best practices through direct regulation, so-called soft-law approaches, or a consensus in the form of national and international standards. Through these means, governments and authorities represent societal expectations on the acceptable level of residual risk that is to be associated with the systems.
- *Management*: This layer coordinates the tasks involved in the design, operation, and maintenance of the systems, enabling risk management and informed design tradeoffs across corporate boundaries, management of supply chain dynamics, and long-term institutional knowledge of long-lived and evolving systems.
- *Task and technical*: This layer covers the technical design and safety analysis process that allows systems to be deployed at an acceptable level of risk and are then actively monitored to identify deviations between what was predicted and what is actually happening so that any gaps can be identified and rectified. This layer includes both the technological components and the tasks performed by the users, operators, and stakeholders within a sociotechnical context. In some cases, users may be unwilling or unknowing participants in the system who are nevertheless impacted by risk.

While developing the framework, 28 case studies from a number of domains, including aerospace, mobility, health care, and supply networks,[3] were analyzed to identify common categories of causes, consequences, systemic failures, exacerbating factors, and controls across these three layers. This resulted in a set of guide words that could be used as part of a structured analysis performed by interdisciplinary experts and can also be based on a specific investigation of previous system failures. The framework is not intended to replace existing safety analysis and management approaches. Instead, it provides an additional perspective to allow the perceived system boundaries, stakeholders, and influencing factors to be called into question, thereby providing a more robust basis for finding the gaps in current safety thinking and providing the context for more specific safety analyses.

> **THE ADAPTION IN THE BEHAVIOR OF HUMAN ROAD USERS IN RESPONSE TO AUTOMATED VEHICLES IS AN EXAMPLE OF SELF-ORGANIZATION, WHERE THE HUMANS BECOME PART OF A LARGER AD HOC SYSTEM.**

APPLICATION OF THE FRAMEWORK

We illustrate the framework by examining the issues surrounding the introduction of an ADS. For the purposes of this article, we understand an ADS as a system that takes over control of the vehicle while driving under a given set of conditions. During this time, drivers can direct their attention to other pursuits while the system takes control of the vehicle. Drivers must be available to take over control when the boundary of the operational design domain (ODD) is met.

An essential first step in the analysis is to determine an initial scope of the system of interest. Note that, as a result of the analysis, factors outside of the assumed system scope could be determined to be relevant, leading to a revision of that scope as part of an iterative process. For the purposes of our examples, the system scope shall be defined as follows: traffic infrastructure, traffic participants (both manually driven and automatically controlled vehicles), emergency services, regulations, and responsible authorities.

The next step in the application of the framework is to analyze the factors that lead to (intractable) complexity and, therefore, the potential for systemic failures. Our first example is a post hoc analysis of a collision involving a prototypical ADS-enabled vehicle. The description of the collision summarized in the following is based on a U.S. National Transportation Safety Board accident report and recommendations.[9]

In March 2018, an automated test vehicle operated by Uber Advanced Technologies Group (Uber ATG) was involved in a collision in Tempe, Arizona, that fatally injured a pedestrian who was crossing a dual carriageway while pushing a bicycle. The circumstances surrounding this incident highlight many of the risks involved in introducing automated driving technologies as well as the potential for systemic failures at the task and technical, management, and governance layers.

An analysis of the vehicle data demonstrated that the vehicle variously misclassified the pedestrian as a vehicle, an unknown object, and a bicycle. On each new classification, the object trajectory prediction algorithm reset and assigned a new classification-dependent trajectory prediction.

At 1.2 s before the impact, the system identified an unavoidable collision. However, to avoid the consequences of false-positive misclassifications, the system was designed to suppress any braking maneuvers in such a case due to the assumption that an attentive operator would take control. The safety driver was viewing content on her mobile phone and did not react to prevent the impact in time. Furthermore, the emergency braking systems preinstalled within the vehicle had been deactivated so as not to conflict with the prototypical functions under test.

The investigation[9] identified the inattentiveness of the operator as the most probable cause of the crash. However, it also identified a number of additional contributing factors, including inadequate safety risk assessment procedures at Uber ATG and ineffective oversight of the vehicle operators, including a lack of mechanisms for addressing operators' automation complacency. Additional factors were identified, included the ambiguous nature of the piece of ground separating the directions of the carriageway, which appeared to include pedestrian walkways, and ineffective oversight of automated vehicle testing by Arizona's Department of Transportation.

The safer complex systems framework is now used to identify the causes, consequences, and exacerbating factors leading to the collision. The results are summarized in Figure 2. The following manifestations of system complexity were identified:

- *Governance*: Rapid technological change, insufficient awareness of the associated risks, and the competing objectives of accommodating business needs versus regulatory responsibilities led to a loss of regulatory control at the state level and inappropriate deployment decisions. This ultimately led to an increased risk to other traffic participants and an avoidable collision.
- *Management and operation*: Inadequate engineering and release processes, coupled with market pressure, a prioritization of avoidance of false positives versus false negatives, and the transfer of responsibility to an inadequately trained and supervised operator led not only to a technically inadequate system but also to operational procedures that did not adequately account for unanticipated classes of risk.
- *Task and technical*: The complexity of the environment and the behavior of different agents within it was underestimated, and emergent behaviors related to the interaction of the system and (the attentiveness of) the safety driver as well as of pedestrians and their surroundings were not adequately considered. This led to a failure of the core driving function as well as the primary backup, which, in this case, was the safety driver.

From this perspective, functional insufficiencies of the system at the technical level as well as the behavior of the safety driver can be seen as emergent properties. They arose, in part, from the management and governance levels and the apparent failure of the duty holders to understand and manage the risks associated with operating such systems. There were insufficient measures in place to constrain the emergent risk of deploying the technical system in its environment. This may be, in part, due to a lack of understanding (competency gap) of the system scope to be considered as well as the potential for emergent behaviors within the system that included the vehicle, driver, pedestrian, and road layout as interacting constituent parts. The example also demonstrates the conflicting pressures to promote innovation in technologies such as automated driving that have the potential for improving overall road safety while, in parallel, managing the risk of integrating such technologies into existing traffic systems with an insufficient understanding of the emergent behaviors.

Consequences for the deployment of ALKSs on public roads

In this section, we apply the safer complex systems framework to a discussion of the risks associated with the deployment of ALKSs[2] onto public roads in the United Kingdom. While representing only a limited scope of ADS functionality, the analysis nevertheless highlights the potential risks if a holistic approach is not applied. The analysis was based on a review of regulations,[2] standards,[10] publicly available specifications,[11] and lessons learned from previous incidents.[9] The set of guide words identified when developing the framework was applied to identify the risk factors and allocate them within the framework to better understand their impact and interrelationships.

The analysis resulted in the identification of a number of themes where distinct relationships were found within the various components of the framework. These included the difficulty in defining a tolerable level of residual risk and liability for the systems, the issue of calibrated trust[12] and automation complacency, an appropriate definition of the ODD, risk transference between different stakeholders, and the interaction with existing traffic systems.

Figure 3 shows the causes of (unsafe) complexity arising from multiple jurisdictions, semipermeable system boundaries, and the heterogeneity and interconnectivity of the safety of ALKSs. At the governance level, multiple jurisdictions refer to the regulations for smart motorways and the ALKS itself. At its simplest, smart motorway rules mean that vehicles should not travel in lanes when a red X is shown and should move into other lanes; the ALKS regulations do not address such signs and do not require that vehicles are able to change lanes. A systemic failure that can be linked directly to this is the lack of a clear allocation of liability.

A further example of problems is the likely heterogeneity and interconnectivity

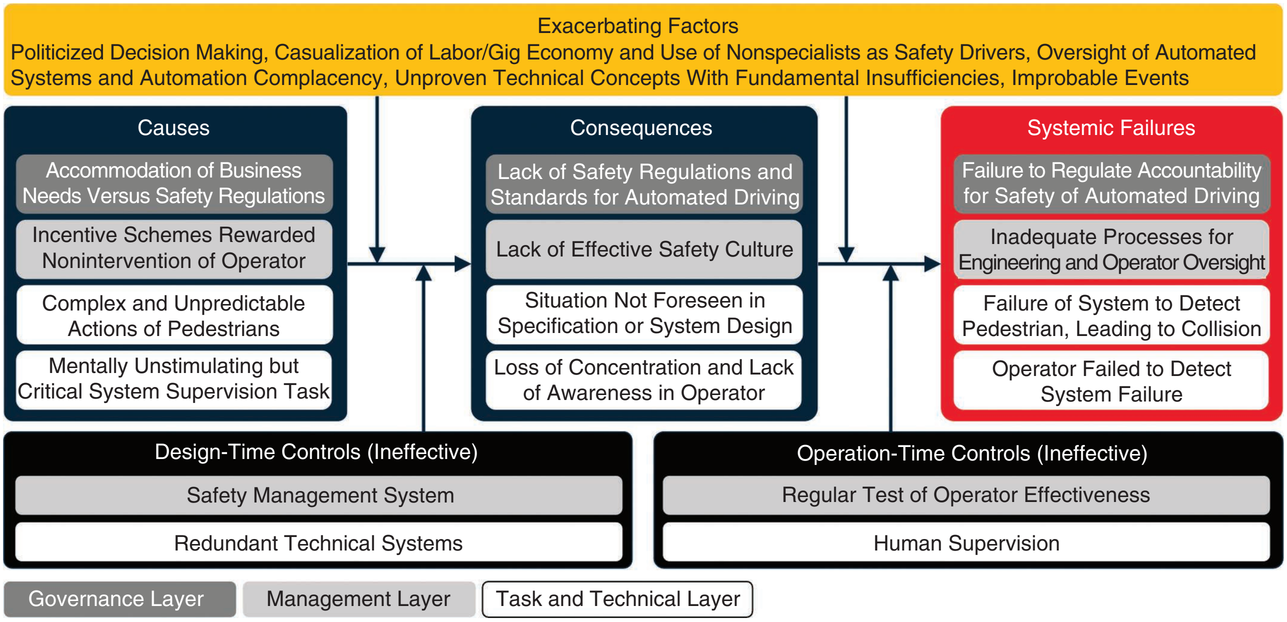

Exacerbating Factors
Politicized Decision Making, Casualization of Labor/Gig Economy and Use of Nonspecialists as Safety Drivers, Oversight of Automated Systems and Automation Complacency, Unproven Technical Concepts With Fundamental Insufficiencies, Improbable Events
Causes
Accommodation of Business Needs Versus Safety Regulations
Incentive Schemes Rewarded Nonintervention of Operator
Complex and Unpredictable Actions of Pedestrians
Mentally Unstimulating but Critical System Supervision Task
Consequences
Lack of Safety Regulations and Standards for Automated Driving
Lack of Effective Safety Culture
Situation Not Foreseen in Specification or System Design
Loss of Concentration and Lack of Awareness in Operator
Systemic Failures
Failure to Regulate Accountability for Safety of Automated Driving
Inadequate Processes for Engineering and Operator Oversight
Failure of System to Detect Pedestrian, Leading to Collision
Operator Failed to Detect System Failure
Design-Time Controls (Ineffective)
Safety Management System
Redundant Technical Systems
Operation-Time Controls (Ineffective)
Regular Test of Operator Effectiveness
Human Supervision
Governance Layer
Management Layer
Task and Technical Layer

FIGURE 2. An analysis of the factors contributing to the Uber Tempe fatality.

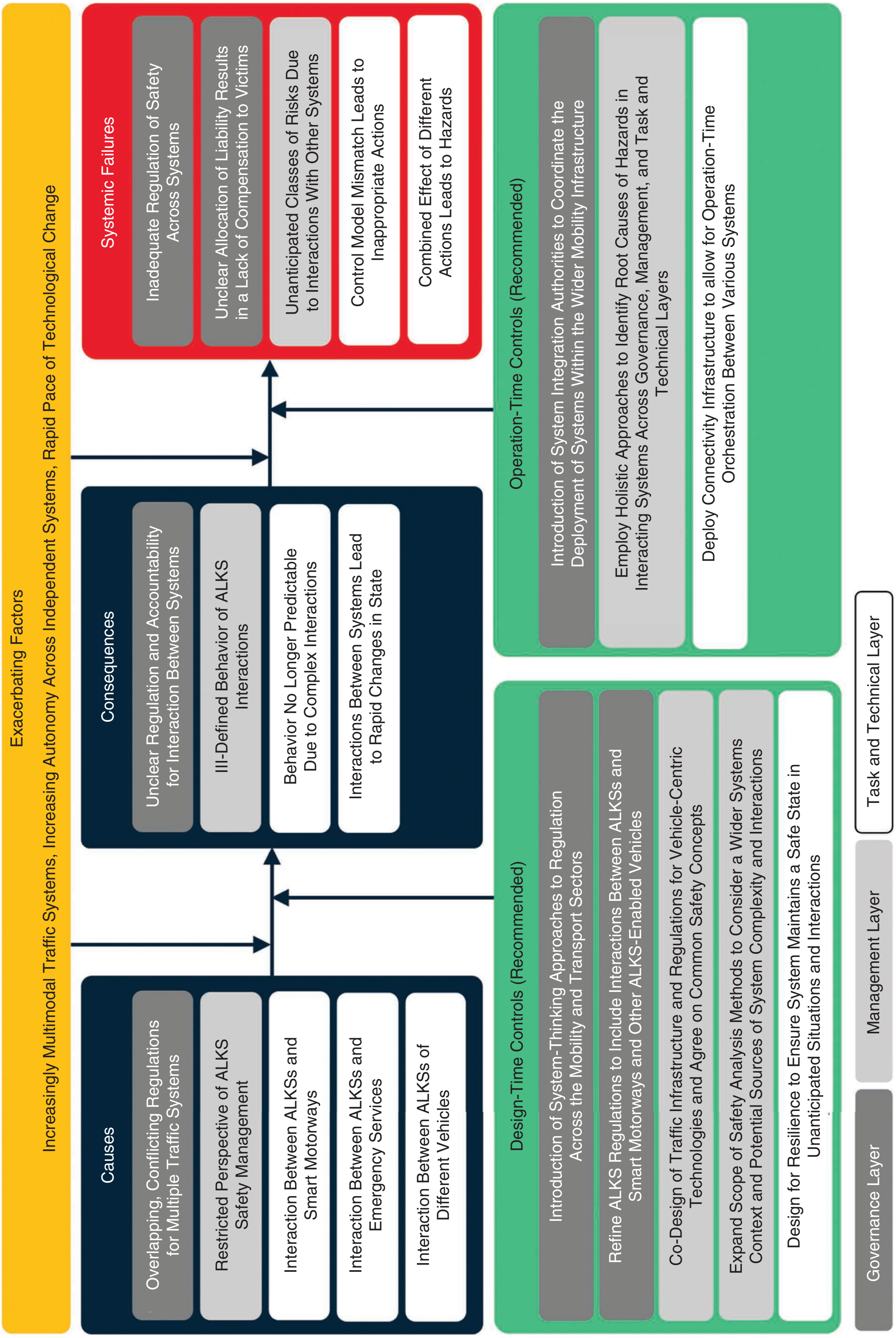

Exacerbating Factors
Increasingly Multimodal Traffic Systems, Increasing Autonomy Across Independent Systems, Rapid Pace of Technological Change
Causes
Overlapping, Conflicting Regulations for Multiple Traffic Systems
Restricted Perspective of ALKS Safety Management
Interaction Between ALKSs and Smart Motorways
Interaction Between ALKSs and Emergency Services
Interaction Between ALKSs of Different Vehicles
Consequences
Unclear Regulation and Accountability for Interaction Between Systems
Ill-Defined Behavior of ALKS Interactions
Behavior No Longer Predictable Due to Complex Interactions
Interactions Between Systems Lead to Rapid Changes in State
Systemic Failures
Inadequate Regulation of Safety Across Systems
Unclear Allocation of Liability Results in a Lack of Compensation to Victims
Unanticipated Classes of Risks Due to Interactions With Other Systems
Control Model Mismatch Leads to Inappropriate Actions
Combined Effect of Different Actions Leads to Hazards
Design-Time Controls (Recommended)
Introduction of System-Thinking Approaches to Regulation Across the Mobility and Transport Sectors
Refine ALKS Regulations to Include Interactions Between ALKSs and Smart Motorways and Other ALKS-Enabled Vehicles
Co-Design of Traffic Infrastructure and Regulations for Vehicle-Centric Technologies and Agree on Common Safety Concepts
Expand Scope of Safety Analysis Methods to Consider a Wider Systems Context and Potential Sources of System Complexity and Interactions
Design for Resilience to Ensure System Maintains a Safe State in Unanticipated Situations and Interactions
Operation-Time Controls (Recommended)
Introduction of System Integration Authorities to Coordinate the Deployment of Systems Within the Wider Mobility Infrastructure
Employ Holistic Approaches to Identify Root Causes of Hazards in Interacting Systems Across Governance, Management, and Task and Technical Layers
Deploy Connectivity Infrastructure to allow for Operation-Time Orchestration Between Various Systems
Governance Layer
Management Layer
Task and Technical Layer

FIGURE 3. An analysis of the interactions between the ALKSs and other traffic systems.

between vehicles fitted with ALKSs giving rise to emergent properties so that the behavior of the system-of-systems is no longer predictable—particularly if we consider how manually driven vehicles might behave if they see an ALKS-fitted vehicle proceed past an X. A design-time control to address this and the previously mentioned governance-level issue would involve refining the ALKS regulations to consider smart motorway infrastructure and to ensure consistency of behavior between different manufacturers' vehicles. This may well require additional detail to be added to the regulations and associated testing regimes.

A related operation-time control would be to enable the infrastructure to "orchestrate" the behavior of the ALKS-equipped vehicles, for example, to ensure that they all move in the same direction when approaching a lane with an X. Such orchestration could also deal with the interaction between ALKSs and emergency services by, for example, forcing the ALKS-equipped vehicles to create an "extra lane" by moving in opposing directions and thus allowing emergency vehicles through.

Thus, greater interconnectivity is likely to be needed to enable the safe introduction of ALKSs. However, this will inevitably lead to additional emergent properties, including issues of cybersecurity and a consequential interplay between safety and security. To address such changes necessitates further iterations and thinking through the impact of the changes, including the possibility of cybersecurity weaknesses introducing common-mode failures.

Due to the rapid technological changes driving the transformation of the mobility sector, it is not feasible to expect that traditional approaches to standards development will keep pace with the rate of change. Therefore, outcome-based safety regulations that stipulate requirements on what to argue instead of how to argue need to be developed; we should take a systems-oriented view, with additional focus on arguing the effectiveness of controls for reducing risk due to system complexity.

Published standards and regulations should be supported by publicly available specifications that provide more specific guidance and document the current industry consensus on topics such as assurance activities for machine learning in an automated driving context. These specifications can be developed in a more agile manner than full standards and can therefore be continuously updated to reflect the state of the art. ISO/TR 4804[11] and UL 4600[10] are two such recent examples. However, these publications do not reflect the complexity of the deployment challenges in considering complexities across the governance, management, and task and technical layers.

A consensus on safety targets for automated driving must be actively developed with a diverse range of stakeholder perspectives going beyond just manufacturers and technical approval authorities. This should consider both quantitative targets (for example, based on road safety statistics) as well as qualitative measures (based on engineering practices and operation-time controls) for achieving acceptable levels of residual risk. This will require cross-disciplinary dialog involving not only technical but also legal and ethics experts.[13] Wider engagement with the public in general is also required to consider the perspectives of those most impacted by risk and to gain an understanding of the expectations and assumptions made on the systems by the users. This is required to reach a level of trust and acceptance of the systems, without which the safety benefits of increased automation will not be realized.

Consequences for safety assurance of automated driving

The manifestations of complexity described in previous sections introduce uncertainty across the entire assurance process; models of the ODD used to design and validate the system are inevitably incomplete and imprecise due to the complexity of the environment; technology used within the systems both in terms of the sensors and actuators as well as the algorithms themselves are inherently imprecise (for example, based on the use of machine learning). Furthermore, due to a lack of clear definitions of safety targets and the lack of established best practices, the assurance case itself may suffer from assurance gaps, leading to inconclusive arguments. This section describes how complexity should be considered within all of the phases of the safety management processes and proposes the means for arguing that the residual risk of such systems is nevertheless tolerable.

Domain analysis helps us develop an understanding of the safety-relevant properties that must be maintained within the chosen operating environment. In addition to properties of the ODD itself, which are critical for ensuring the performance of perception and control functions, this phase must also ensure a sufficient understanding of the societal and legal expectations on the system. This analysis must take place within an assumed scope of the system under consideration and its interactions with its environment. This system scope shall be continuously validated and adjusted to ensure that critical interactions are considered in the safety assurance approach.

The system design refines the expectations on the system discovered during the domain analysis into the technical requirements on the system and identifies a system design capable of supporting the system's safety goals. The safer complex systems framework can provide indicators related to the risk associated with the complexity of the system and its context, operation, and interaction with human actors that should be considered during detailed safety analyses.

Causal approaches to safety analysis such as fault tree analysis[14] are based on a model[7] of how faults in individual system components cause an erroneous system state (error) that may subsequently lead to a failure of the system's service as perceived by its users. However, one of the consequences of complexity is the presence of unknown and unknowable faults and causes of systemic failures as well as a high level of interconnectivity and nonlinear interactions. There is also a need to be able to model the impact of uncertainties both in the environment as well as in the internal behavior of the system.[15]

System theoretic approaches such as those recommended by Rasmussen,[8] Leveson's system theoretic accident methods and processes (STAMP)/system theoretic process analysis,[16] and Hollnagel's functional resonance analysis method (FRAM)[17] take a more holistic view of risk factors within a system across the technical, management, and governance layers. Monkhouse et al.[18] recently proposed an enhanced vehicle control model that considers the shared cognitive nature of the driving task for driver assistance systems with limited autonomy. This approach allows for the subtle interactions between the task (human-controlled activities) and technical (system-controlled) activities within our framework to be more systematically analyzed, for example, by extending STAMP[16] or FRAM[17] type methods. We see these safety analysis methods as strongly complementary and believe our framework can provide important additional inputs to these analyses.

THE INCREASE IN THE USE OF SIMULATIONS DURING THE DESIGN AND VALIDATION OF THE SYSTEMS ALLOWS FOR MORE TARGETED TESTING OF CRITICAL AND RARE SITUATIONS.

An analysis of the complexity factors that could lead to systemic failures can provide information for verification and validation by determining the sets of assumptions that must be confirmed and the specific properties that must be validated during field tests and operation. Relating back to the Uber ATG collision described previously, this could include the validation of assumptions made regarding the performance of the safety driver or about the behavior and occurrence probabilities of pedestrians on certain types of roads.

Statistical arguments based on miles driven between incidents during field-based tests become both unfeasible and ineffective due to the effort required to collect the data and the difficulty in ensuring sufficient coverage of edge cases and critical situations. The increase in the use of simulations during the design and validation of the systems allows for more targeted testing of critical and rare situations. However, such approaches require additional arguments regarding accuracy and the ability to extrapolate the results of the simulation into the target domain.

We consider that the greatest benefit of explicitly considering complexity factors lies in the formulation of the assurance case[19] and, more specifically, the reduction of assurance gaps. Application of the safer complex systems framework can lead to an assurance case that better reflects the actual system context and risks associated with the system and its operating context. This could be achieved by integrating the consideration of complexity factors throughout the claims and evidence provided in the assurance or by formulating specific claims and targeted evidence focusing on the causes of systemic failures in the system.

The safety assurance of automotive systems currently places a strong focus on design-time controls and type approval. However, as the complexity and scope of the systems increase, and with it the sensitivity to an ever-evolving environment, it is unrealistic to believe that an adequate level of safety can both be achieved before the system is deployed and then maintained over the vehicle's lifetime—without ongoing controls and the potential for updates to the system. Operation-time measures are required for ensuring the safety of the systems, including the measurement of critical observation points within the system (the leading indicators of systemic failures) as well as whether assumptions made regarding the ODD (and, therefore, the validation approach) continue to hold. The assurance case for the system should be continuously evaluated and refined based on experiences in the field and changing expectations on the system. This holds true for automated driving applications but also for connected traffic infrastructure in general.

ABOUT THE AUTHORS

SIMON BURTON is the research division director at the Fraunhofer Institute for Cognitive Systems, Munich, 80686, Germany, and an honorary visiting professor at the University of York, York, YO10 5GH, U.K. His research interests include the safety assurance of complex, autonomous systems and the safety of machine learning. Burton received a Ph.D. in computer science from the University of York. Contact him at simon.burton@york.ac.uk.

JOHN McDERMID is a professor at the University of York, York, YO10 5GH, U.K. His research interests are in the safety and security of complex computer controlled systems, with a particular focus on safety of robotics and autonomous systems. McDermid recieved a Ph.D. from the University of Birmingham. Contact him at john.mcdermid@york.ac.uk.

PHILIP GARNET is a senior lecturer at the University of York, York, YO10 5GH, U.K. His research interests include the application of complex systems theory to organizations and how organizational culture, memory, and knowledge can be theorized as an emergent property of the system itself. Garnet received a Ph.D. from the University of York. Contact him at philip.garnett@york.ac.uk.

ROB WEAVER is an independent advisor and consultant in Canberra, ACT 2607, Australia. His research interests include aviation strategy and future concepts, risk and safety management, complex systems, training development and delivery, safety and regulatory compliance, and safety framework development. Weaver received a Ph.D. in computer science from the University of York. Contact him at rob@robweaveradvisory.com.

Assuring the safety of autonomous vehicles is a complex endeavor, and the deployment of automated vehicles within a public traffic infrastructure must be recognized as a complex system. By this, we do not only mean that it is technically difficult or involves many resource-intensive tasks that must somehow be managed within feasible economic constraints. Both are true. More than that, autonomous vehicles and their wider sociotechnical context demonstrate characteristics of complex systems in the stricter sense of the term. This has a huge impact on our ability to argue the safety of such systems.

This article proposes a framework by which the factors that impact the complexity of a system, thus leading to systemic safety failures, can be identified and used to inform a safety assurance process. We conclude from the analyses described in this article that ensuring and demonstrating the safety of ADSs requires a more comprehensive and holistic view of safety than for previous generations of vehicle electronic control systems. A systems-oriented approach that acknowledges complexity and includes coordinated measures across the governance, management, and task and technical layers is required to reach an adequate level of safety for ADSs. This will require closer collaboration between automotive manufacturers and suppliers, communications, and highway or city infrastructure as well as a better understanding of dependencies across the three layers and stakeholders within the framework, which includes the role of the general public.

The safer complex systems framework, at this stage in its development, seeks to provide an accessible overview of the factors that influence the safety of complex systems. As presented, the framework indicates only the highest-level dependencies between the elements of the framework. Further work will involve enriching the framework by integrating various safety analysis methods as well as domain-specific risk models to allow the framework to be integrated into safety analysis and management during system design. Work is also ongoing to validate the framework in other domains, including urban air mobility and health care.

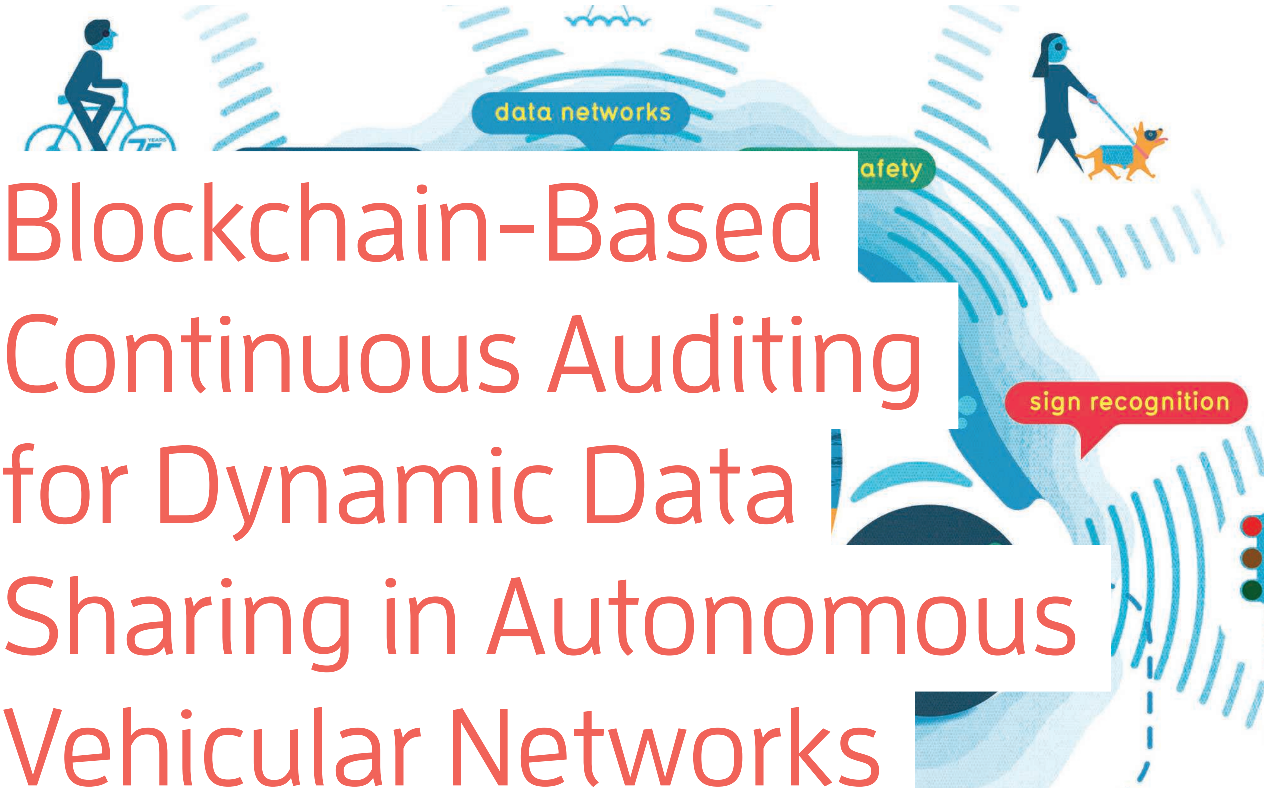
Most significantly, though, the authors see the strongest need in establishing a systems-thinking mindset that acknowledges complexity and uncertainty and provides an approach for handling both—particularly at the governance and management layers as it is ultimately here where the levers are most effective in ensuring that our traffic systems remain safe and become even safer through the introduction of autonomous technologies. The work of Buckle, which argues that for these approaches to become embedded in practice, we may need to apply maturity models for systems thinking, presents one possibly promising starting point.[20]

ACKNOWLEDGMENT

This project was supported by the Royal Academy of Engineering and the Lloyd's Register Foundation as part of the EngineeringX program. The authors would like to thank all of those who contributed to the project.

REFERENCES

1. "Contributory factors for reported road accidents," U.K. Dept. for Transport, London, U.K., RAS50, 2020. [Online]. Available: https://www.gov.uk/government/statistical-data-sets/ras50-contributory-factors#contributory-factors-for-reported-road-accidents-ras50–-excel-data-tables
2. "Safe use of automated lane keeping system on GB motorways: Call for evidence," U.K. Dept. for Transport and Centre for Connected Autonomous Vehicles, London. Accessed: Oct. 18, 2020. [Online]. Available: https://www.gov.uk/government/consultations/safe-use-of-automated-lane-keeping-system-on-gb-motorways-call-for-evidence
3. S. Burton, P. Garnett, J. McDermid, and R. Weaver, "Safer complex systems—An initial framework," Royal Society of Engineering, London, 2020. [Online]. Available: https: //www.raeng.org.uk/publications/reports/safer -complex-systems
4. P. Erdi, *Complexity Explained*. Berlin: Springer Science & Business Media, Nov. 2007.
5. C. Pakusch, G. Stevens, A. Boden, and P. Bossauer, "Unintended effects of autonomous driving: A study on mobility preferences in the future," *Sustainability*, vol. 10, no. 7, p. 2404, 2018. doi: 10.3390/su10072404.
6. S. Burton, L. Gauerhof, and C. Heinzemann, "Making the case for safety of machine learning in highly automated driving," in *Computer Safety, Reliability, and Security*, S. Tonetta, E. Schoitsch, and F. Bitsch, Eds. Cham: Springer International Publishing, 2017, pp. 5–16.
7. A. Avizienis, J-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Trans. Dependable Secure Comput.*, vol. 1, no. 1, pp. 11–33, 2004. doi: 10.1109/TDSC.2004.2.
8. J. Rasmussen, "Risk management in a dynamic society: A modelling problem," *Safety Sci.*, vol. 27, nos. 2–3, pp. 183–213, 1997. doi: 10.1016/S0925-7535(97)00052-0.
9. "Collision between vehicle controlled by developmental automated driving system and pedestrian," National Transportation Safety Board, Washington, D.C., 2019. [Online]. Available: https://www.ntsb.gov/news/events/Pages/2019-HWY18MH010-BMG.aspx
10. "Standard for safety for the evaluation of autonomous products," Underwriters Lab., Northbrook, IL, Tech. Rep. ANSI/UL 4600, 2019.
11. "Safety and cybersecurity for automated driving systems—Design, verification and validation," International Organization for Standardization, Geneva, Switzerland, Tech. Rep. ISO/PRF TR 4804:2020, 2020.
12. L. Dixon, "Autonowashing: The greenwashing of vehicle automation," *Transp. Res. Interdiscip. Perspect.*, vol. 5, p. 100113, May 2020. doi: 10.1016/j.trip.2020.100113.
13. S. Burton, I. Habli, T. Lawton, J. McDermid, P. Morgan, and Z. Porter, "Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective," *Artif. Intell.*, vol. 279, p. 103201, 2020. doi: 10.1016/j.artint.2019.103201.
14. E. Ruijters and M. Stoelinga, "Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools," *Comput. Sci. Rev.*, vols. 15–16, pp. 29–62, Feb. 2015. doi: 10.1016/j.cosrev.2015.03.001.
15. R. Gansch and A. Adee, "System theoretic view on uncertainties," in *Proc. Design, Automat. Test Europe Conf. Exhib. (DATE 2020)*, 2020, pp. 1345–1350.
16. N. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA: MIT Press, 2011.
17. E. Hollnagel, *FRAM, the Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*. Farnham, U.K.: Ashgate Publishing, 2012.
18. H. E. Monkhouse, I. Habli, and J. McDermid, "An enhanced vehicle control model for assessing highly automated driving safety," *Rel. Eng. Syst. Safety*, vol. 202, p. 107,061, 2020. doi: 10.1016/j.ress.2020.107061.
19. "Systems and software engineering—Systems and software assurance," International Organization for Standardization, Geneva, Switzerland, Tech. Rep. ISO/IEC/IEEE 15026:2019, 2019.
20. P. Buckle, "Maturity models for systems thinking," *Systems*, vol. 6, no. 2, p. 23, 2018. doi: 10.3390/systems6020023.



Blockchain-Based Continuous Auditing for Dynamic Data Sharing in Autonomous Vehicular Networks

Haiyang Yu, Shuai Ma, Qi Hu, and Zhen Yang, Beijing University of Technology

With the rapid development of intelligent transportation, massive data are generated by autonomous vehicle systems and shared among vehicles through cloud servers to improve the driving experience and service quality. However, cloud servers cannot be fully trusted and may lead to serious data security challenges.

The recent advances in artificial intelligence and automobile technologies are promoting the accelerated development of autonomous vehicles, which have great potential to enable fully automated intelligent transportation systems and play a vital role in our future society.[1] Autonomous vehicles are expected to bring a number of benefits, such as improved travel experience, reduced traffic congestion, increased safety, energy conservation, and pollution reduction.[2] To achieve these goals, a large amount of data (for example, traffic data, vehicle information, weather conditions, and so on) generated by autonomous driving systems should be shared and exploited to enhance self-driving performance. For instance, autonomous vehicles having the same destination can share traffic information to avoid traffic congestion or potential car accidents. As a consequence, data sharing in autonomous vehicular networks (AVNs) is becoming increasingly important.

In practice, an autonomous vehicle can generate up to 1 GB of data/s from cameras, radar, GPS, and so forth.[3] However, due to resource constraints, autonomous vehicles cannot support massive data storage and large-scale data sharing. To address this challenge, autonomous vehicles are allowed to outsource shared information to the cloud, as shown in Figure 1. Unfortunately, once data are outsourced, they will no longer be physically controlled by autonomous vehicles. Consequently, other

Digital Object Identifier 10.1109/MC.2021.3080332
Date of current version: 30 July 2021

autonomous vehicles cannot know whether the shared data that they frequently accessed from the cloud are intact or corrupted. On the other hand, cloud service providers (CSPs) are not always as secure and reliable as they claim to be. Incidents of outages and security breaches of noteworthy cloud services happen from time to time,[4] putting shared data as well as self-driving cars themselves at risk in the following ways. First, as all autonomous vehicles are connected to the cloud, an untrusted CSP can easily deceive thousands of autonomous vehicles and take control of them by tampering with the data shared with them. Second, when a CSP is hacked or controlled by malicious hackers, the data or programs can be forged maliciously, which gives rise to more serious consequences. For instance, hackers can easily make an autonomous vehicle deviate from its current route to cause a car accident by simply forging its navigational data in the cloud. They can even fully control a self-driving car by putting their own hijacked program into the system via over-the-air electronic communications. Therefore, how best to continuously check cloud data shared in the AVN to ensure shared data integrity over the whole storage period has become an increasingly important and challenging task.

There are a number of approaches[5–7] that deal with the data integrity problem in the cloud. However, the existing schemes need frequent communication and thus incur performance problems in AVNs. First, autonomous vehicles need to face all kinds of environments when operating outside. There is no strong guarantee that a signal can be found anywhere on the road. Thus, it is difficult for autonomous vehicles to stay connected with the cloud all of the time. Second, an AVN is connected with billions of smart vehicles, roadside units (RSUs), controllers, and so on. Frequent auditing requests will easily drain their communication and computation resources, or even power resources. Therefore, a lightweight approach has to be investigated in AVNs, which should reduce the communication load and CPU computation of autonomous vehicles for auditing other vehicles' data.

Another major concern in continuous auditing is data dynamics. As vehicular data (such as GPS, velocity, road congestion, and so forth) often change, they will be frequently updated by autonomous vehicle systems. As a result, the lightweight auditing approach designed in this article should also support data dynamics.

In this article, we use the sampling strategy and trapdoor delay function (TDF) to build a lightweight continuous auditing approach for shared data in AVNs and adopt chameleon hash functions to support data dynamics. Both the preprocessing and verification costs in the proposed scheme can be reduced. Furthermore, we consider leveraging the blockchain technology to optimize the construction. Blockchain is featured with tamper-proofing, collaborative maintenance, and smart contracts. The blockchain can be naturally embedded into AVN infrastructures and maintained by all of the RSU nodes. In this way, no autonomous vehicle needs to audit the shared data by itself. Instead, the smart contracts are responsible for automatically auditing the shared data. As a result, the autonomous vehicle can

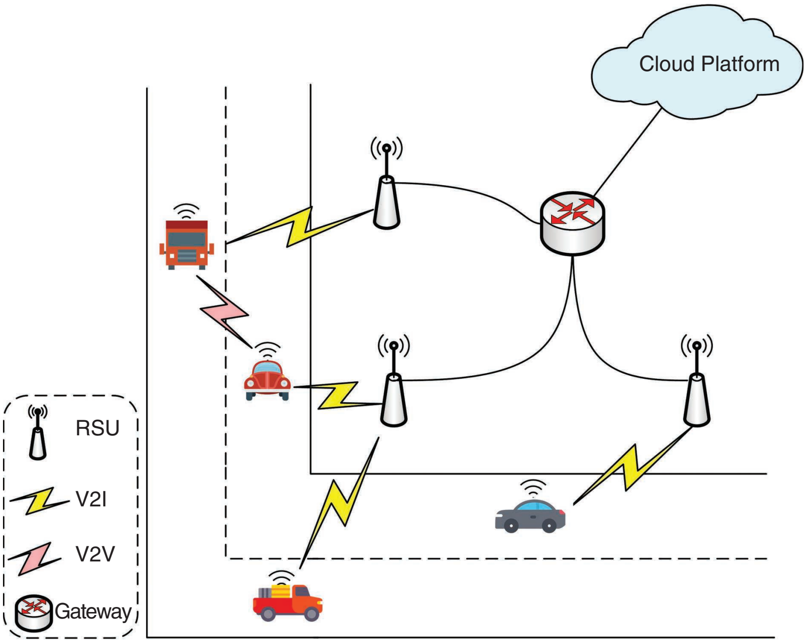


FIGURE 1. The framework of the cloud-based AVN. RSU: roadside unit; V2I: vehicle to infrastructure; V2V: vehicle to vehicle.

directly retrieve the auditing results calculated by the smart contracts from the blockchain.

Our contributions in this article can be summarized as follows.

- We exploit the sampling strategy together with the TDF and propose a blockchain-based continuous auditing scheme for shared data in AVNs. The proposed scheme is lightweight in terms of both its communication and computation performance.
- We further integrate our auditing approach with the chameleon hash function to support data dynamics, which allows the shared data to be updated during auditing.
- We provide a comprehensive performance analysis of our approach. Extensive experiments demonstrate that our scheme is more efficient and practical than the state-of-the-art ones.

RELATED WORK

Ateniese et al.[8] first put forward the concept of data integrity auditing for remote servers, which is named *provable data possession*. Juels and Kalisk[9] introduced a similar idea called *proof of retrievability* (*PoR*), which verifies cloud data integrity while retrieving corrupted files.

To improve the efficiency of data integrity auditing, Wang et al.[5] proposed an identity-based, public-provable data-possession scheme to eliminate complex certificate management from public key infrastructure. Abdallah et al.[10] presented a lightweight message authentication scheme for vehicle-to-grid connection. Similarly, Lin et al.[11] proposed an efficient message authentication scheme for vehicular ad hoc networks to ensure data integrity and authenticity, which reduced the authentication overhead on individual vehicles. However, these two schemes focused only on improving the performance of the authentication of transmitted data rather than cloud data.

As blockchain features tamper-proofing, researchers have attempted to integrate it into cloud-auditing schemes. Yue et al.[12] proposed a decentralized auditing scheme that stores the root of a Merkle hash tree in the blockchain. This scheme, however, was inefficient in the big data scenario as the tree root needed to be uploaded to the blockchain frequently when the data update occurred. Wang et al.[13] integrated blockchain to build a decentralized auditing framework in which blockchain nodes were responsible for auditing. This scheme solved the trust problem with the third-party auditor but did not consider the use of blockchain storage to further exploit blockchain. Xu et al.[14] and Wang et al.[15] considered storing metadata in the blockchain to protect the integrity of metadata and thus indirectly ensured the integrity of the corresponding data blocks. However, storing the metadata of all the data blocks made these solutions impractical in a big data scenario. Chen et al.[16] proposed a blockchain-based provable data-possession scheme for smart cities, which builds a decentralized auditing framework and supports dynamic auditing. Existing blockchain-based cloud-auditing schemes can fulfill the demand of validating cloud data integrity. However, all of these schemes inherently require interaction among data users and the cloud, which poses extra communication overhead for autonomous vehicles when applied to an AVN.

To frequently check the integrity and availability of data in remote storage, continuous auditing is proposed by Ateniese et al.[17] Their scheme is efficient in a static AVN due to low communication cost and simple verification. However, this scheme is not practical in a dynamic AVN that is full of shared data because of its inefficiency in supporting data dynamics. In AVNs, because data are frequently collected, updated, and shared with other autonomous vehicles, dynamic updating in continuous auditing should be efficient.

AS BLOCKCHAIN FEATURES TAMPER-PROOFING, RESEARCHERS HAVE ATTEMPTED TO INTEGRATE IT INTO CLOUD-AUDITING SCHEMES.

Motivation

Existing blockchain-based cloud-auditing schemes can fulfill the demand for validating cloud data integrity. However, all of these schemes inherently require interaction among data users and the cloud, which poses extra communication overhead for autonomous vehicles when applied to an AVN.

To frequently check the integrity and availability of data in remote storage, continuous auditing is proposed by Ateniese et al.[17] Their scheme is efficient in a static AVN due to low communication cost and simple verification. However, this scheme is not practical in a dynamic AVN that is full of shared data because of its heavy setup load and because it does not support data dynamics. In an AVN, data are frequently collected, updated, and shared with other autonomous vehicles, thus, both dynamic updating and preprocessing algorithms in continuous auditing should be efficient.

To deal with the aforementioned problems, in this article, we aim to investigate a lightweight approach as well as to support data dynamics in an AVN. The most challenging part is how to update shared data while not recalculating all the preprocessed tags stored in the cloud. This is because the tag computation process is sequentially executed, and thus, updating any file will affect the output of each round as well as the final output of the tag.

PROBLEM STATEMENT

System architecture

The system architecture depicted in Figure 2 consists of four types of entities: autonomous vehicles (data owners and data users), RSUs, the cloud, and blockchain.

- *Autonomous vehicles*
  - *Data owner*: A data owner is an autonomous vehicle that shares a large amount of local data with other autonomous vehicles via the cloud. It will frequently collect and upload dynamic shared data to the cloud.
  - *Data user*: A data user is a smart vehicle that wants to use shared cloud data for improving self-driving capability. Shared data should be checked first before the data user uses it.
- *RSU*: An RSU is responsible for communicating with autonomous vehicles and connecting with other RSUs to construct a blockchain network.
- *Cloud*: The cloud allows autonomous vehicles to store data in cloud servers and share data with other autonomous vehicles via cloud storage.
- *Blockchain*: The blockchain is maintained by RSUs and stores the storage proofs generated by the cloud. The smart contract in the blockchain can perform auditing verification on behalf of autonomous vehicles.

Algorithm formulation

The proposed scheme involves six algorithms, namely, Setup, Store, Challenge, Prove, Verify, and Update.

- *Setup*(λ, t, T) is run by a data owner, which takes a security parameter λ, auditing frequency t and storage time period T as input, and outputs a public-private key pair (pk, sk) and public parameters pm.
- *Store*(sk, $\{F\}$) is run by the data owner, which takes as input a secret key sk and a file set $\{F\}$, and computes a tag set Σ.

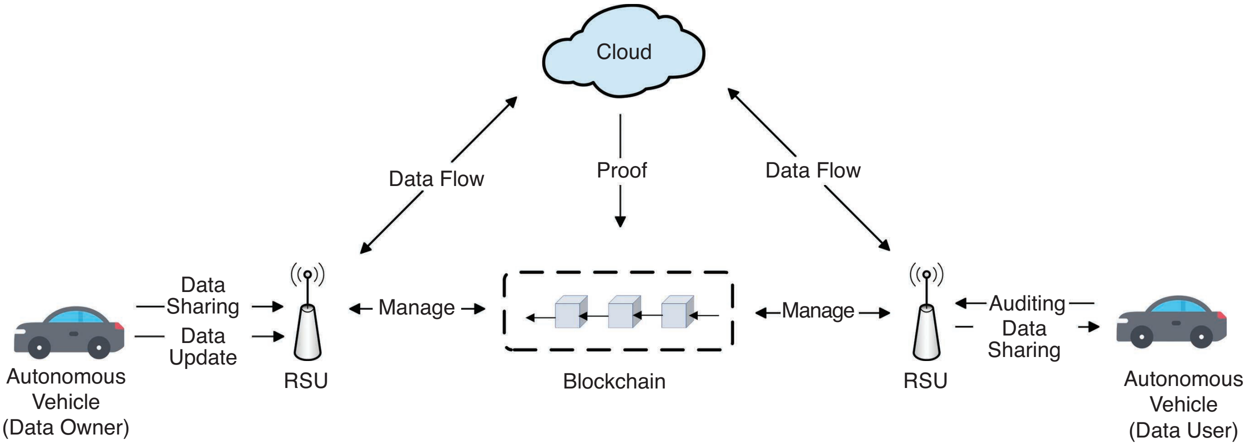


FIGURE 2. The system architecture of continuous auditing for shared data in the AVN.

- *Challenge(pm)* is run by the data owner. It takes as input the public parameters *pm*, generates a challenge *c* and a timer.
- *Prove*$(\Sigma, \{F\}, c)$ is run by the cloud storage provider, which takes the tag set Σ, the file set $\{F\}$ and the challenge *c* as input, and outputs the proof *pf* of the whole storage period and submits it to the blockchain immediately.
- *Verify*$(pk, \Sigma, pf, c, timer)$ is the verification algorithm, which takes as input the public key *pk*, the tag set Σ, the proof *pf*, the challenge *c* and the timer, outputs a bit $b \in (0, 1)$ as the verification result.
- *Update*(sk, F^*) is run by the data owner, which inputs a secret key *sk* and an updated file F^*, and outputs a new tag set Σ_u^*.

PRELIMINARIES

TDF

A TDF is a function that cannot be evaluated in less than a prescribed amount of time even when using multiple processors and parallelism. However, evaluating the function with a secret trapdoor can greatly reduce the time delay.

A TDF $F : X \rightarrow Y$ is a scheme consisting of the following three algorithms:

- *TDF.Setup*(λ, s) is a randomized algorithm that takes as input a security parameter λ and a delay parameter *s*, and outputs public parameters *pp* and a trapdoor *tr*. The delay parameter *s* is subexponential in λ.
- *TDF.Eval*(pp, x) takes as input $x \in X$ and outputs $y \in Y$.
- *TDF.TrapEval*(pp, tr, x) takes as input *x* and a trapdoor *tr*, outputs $y \in Y$.

The TDF can be easily instantiated via the Rivest–Shamir–Adleman trapdoor, as described by Wesolowski.[18]

Chameleon hash function

The chameleon hash function is a hash function in which the owner of the trapdoor keys can easily find collisions in the domain. Let *p*, *q* be prime such that $p = 2q + 1$, and let *g* be a generator for the subgroup of quadratic residues QR_p of Z_p^*. A chameleon hash function consists of three algorithms: *HGen*, *Hash*, and *HCol*.[19]

- $(y, x) \leftarrow HGen(1^k)$: The trapdoor key is a random value $x \in [1, q-1]$, and the hash key *hk* is equal to $y = g^x$.
- $h \leftarrow Hash(y, m, r, s)$: To hash a message $m \in 0, 1^*$, pick random $r, s \leftarrow Z_q$, and return $h = r - (y^{H(m\,||\,r)} \cdot g^s \bmod p) \bmod q$ where $H : \{0, 1\}^* \rightarrow Z_q$ is a standard-collision resistant hash function.
- $(r', s') \leftarrow HCol(x, (h, m, r, s), m')$: To compute a collision for message m', pick a random $k \in [1, q-1]$ and compute $r' = h + (g^k \bmod p) \bmod q$ and $s' = k - H(m'\,||\,r') \cdot x \bmod q$. Return (r', s').

PoR

PoR can provide the abilities of both data integrity checking and data recovery. In our scheme, we use only the PoR for data integrity checking and do not refer to data recovery. The typical construction of PoR can be adapted from Juels and Kaliski Jr.[9] Formally, a PoR scheme can be formulated as five algorithms:

- *PoR.KeyGen*(λ): Taking as input the security parameter λ, the KeyGen algorithm randomly chooses a public-private key pair (pk, sk).
- *PoR.Store*(sk, D): Taking as input a secret key *sk* and a file $D \in (0, 1)^*$, the Store algorithm computes the PoR tag *tg* for the file *D*.
- *PoR.Chal*$(sk, state)$: The Chal algorithm takes as input the private key *sk* and the *state*, and outputs a challenge *c*.
- *PoR.P*(c, pk, tg, D): Given the challenge *c*, the public key *pk*, the file tag *tg*, and the file D, the algorithm computes the response proof *p* for further verification.
- *PoR.V*(p, pk, c, tg): Given the response proof *p*, the public key *pk*, the challenge *c*, and the file tag *tg*, the algorithm outputs 1 if the proof *p* is valid, otherwise it outputs 0.

PROPOSED SCHEME

Overview

Here we present the concrete construction of dynamic continuous auditing schemes. By using the TDF, the cloud in our approach can continuously generate storage proofs without communicating with the auditor. Besides, verification only requires the calculation of one hash function. We also integrate the sampling strategy to relieve the computation burden on autonomous vehicles for preprocessing the shared data. Instead of computing all the files, only the randomly challenged file will be calculated in each time slot to generate the corresponding tag.

To support dynamic auditing, the chameleon hash function is adopted in our approach, which ensures that updating any challenged file will not result in the recalculation of the corresponding tag. The chameleon hash function is used to make sure its output (the challenge parameter) remains constant. Furthermore, we use the collision-generation algorithm of the chameleon hash function to update parameters in all the time slots for privacy concerns. In this way, the cloud cannot distinguish the difference among different time slots and thereby cannot know the information of the challenged set.

Detailed construction

The proposed approach includes six algorithms, namely, Setup, Store, Challenge, Prove, Verify, and Update.

Setup: T and Δt denote the storage period and the time slot, respectively. t is the auditing frequency so that $|\Delta t| = t$. The data owner generates a public-private key pair (pk, sk) and a pseudorandom permutation (PRP). It also generates the key pair (hpk, hsk) for the chameleon hash function by performing $HGen(1^k)$ The number of the checking round is $f = T/t$ and the TDF delay time is $dt < t - 2\delta T$, where δ is a public parameter in a TDF of δ–evaluation time and $0 < \delta \ll 1$. The data owner generates the TDF parameter $TDF.pp.$ as well as the trapdoor $TDF.tr.$ The data owner uses the symmetric encryption function $Enc.sk$ with secret key sk. The public information is the public keys (pk, hpk) and public parameters $pm = (T, t, dt, f, TDF.pp)$, and the secret information is the private keys (sk, hsk) and trapdoor $TDF.tr$.

ALGORITHM 1: CBDIC.STORE: THE FILE-STORING ALGORITHM.

```
Input: The file set $\{F\}$, the key pair $(pk, sk)$, checking round $f$, chameleon
       hash key pair $(hpk, hsk)$, and auditing times $l$
Output: The tag set $\Sigma$.
for $i = 1$ to $n$ do
    $\sigma'_i \leftarrow PoR.Store(sk, F_i)$;
end
for $i = 1$ to $l$ do
    $c_{i,1} \leftarrow PoR.Chal(sk, \sigma'_i)$;
    for $j = 1$ to $f$ do
        $id_{ij} \leftarrow PRP(c_{ij})$;
        $v_{ij} \leftarrow PoR.P(pk, c_{ij}, F_{id_{ij}}, \sigma'_{id_{ij}})$;
        if $j == f$ then
            break;
        end
        $u_{ij} = H(v_{ij})$;
        $q_{ij} \leftarrow TDF.TrapEval(pp, tr, u_{ij})$;
        $r_{ij}, s_{ij} \leftarrow Z_q^*$;
        $c_{i,j+1} \leftarrow CH(hpk, q_{ij}, r_{ij}, s_{ij})$;
    end
    $hid_i = H(id_{i,1}, \dots, id_{i,f})$;
    $c_i = H(c_{i,1}, \dots, c_{i,f})$;
    $q_i = H(q_{i,1}, \dots, q_{i,f})$;
    $r_i = H(r_{i,1}, \dots, r_{i,f})$;
    $s_i = H(s_{i,1}, \dots, s_{i,f})$;
    $\sigma_i = H(hid_i, c_i, q_i, r_i, s_i)$;
end
$C = Enc.sk(c_1, \dots, c_l)$;
$R = \{r_{ij}\}_{i \in [1,l], j \in [1,f]}$;
$S = \{s_{ij}\}_{i \in [1,l], j \in [1,f]}$;
$\Sigma = (C, R, S, \{\sigma'_i\}, \sigma_1, \dots, \sigma_l)$;
return $\Sigma$;
```

Store: The data owner randomly chooses the file to be checked in each time slot and generates the tag set for future auditing. It follows Algorithm 1 to generate the tag set Σ. Generally, the data owner should first compute PoR tags for each file and then sequentially perform the sampling process, *PoR.P* and *TDF.TrapEval* functions in each time slot to generate the verification tag.

Challenge: The data owner maintains a state $i \in [1, l]$ For each state $i < l$, it sends the challenge c_i to a nearby RSU. The RSU transmits the challenge c_i to the cloud and sets its timer to 0. It also submits the current time stamp to the blockchain as a part of the proof.

Prove: When receiving the challenge c_i, the cloud performs Algorithm 2 to generate the storage proof pf by using the challenge c_i as well as the file set $\{F\}$ and the tag set Σ. In short, the cloud sequentially executes the same sampling process, PoR and TDF function as in the Store algorithm, except for the TDF trapdoor. The cloud does not need communication and can continuously generate a part of the proof $(hid_{ij}, c_{ij}, q_{ij}, r_{ij}, s_{ij})$ for each time slot Δt_{ij}.

Verify: When receiving the storage proof *pf*, the RSU first checks the timer. If the time length is larger than $(1+\delta)T$ or less than T, the data owner will output 0. Otherwise, for the current state i, the RSU invokes the auditing contract and the smart contract checks whether $\sigma_i = H(pf)$. If the equation holds, the auditing contract outputs 1, otherwise it outputs 0. The data owner can access the auditing result in the auditing contract.

Update: The data owner updates a file F to F^* and then generates new tag set Σ^*_u by performing Algorithm 3. For the challenged file, it should first perform the sampling strategy, PoR and TDF functions in sequence, and then update r^*_{ij}, s^*_{ij} with the new q^*_{ij} For other files, the data owner need only update r^*_{ij}, s^*_{ij} with the original q_{ij}. In the end, the updated file F^* and tag set Σ^*_u will be transferred to the cloud and used to replace the original file and tag.

Discussion

As TDFs are time consuming and chained together, the cloud cannot precompute all of the intermediate parameters $\{q_{ij}\}$ in advance. Moreover, if one file is corrupted, the final hash result of the storage proof *pf* must be different from the tag σ_i. Therefore, by providing a valid storage proof *pf*, the cloud can prove that it has sequentially executed the algorithm and continuously stored the shared data. An autonomous vehicle (a data owner or a data user) can make sure that the data shared via the cloud are continuously intact and available and that the data have never been replaced or modified by any adversaries for malicious purposes.

In the dynamic setting, to audit updated files, the corresponding tag set Σ should also be updated. However, recalculating each tag σ_i involves a lot of TDF computations. Although trapdoor can speed up TDF computations, it still incurs large computation costs for the data owner when considering frequent updating. Therefore, in our approach, we design an efficient Update algorithm that is different from the Store algorithm in calculating tags. Figure 3 illustrates the major difference between the Update and Store algorithms. Compared with the Store algorithm, the Update algorithm recalculates only the TDF for the updated file and regenerates r_{ij} and s_{ij} to make the $c_{i,j+1}$ output of the chameleon hash function the same as before. For the rest of challenged files, the Update algorithm need only update r_{ij} and s_{ij} and does not recompute the PRP, PoR, and TDF functions.

PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed scheme by comparing it with the state-of-the-art cPoSt.[17] Both our scheme and cPoSt adopt the PoR instantiation by Juels and Kaliski Jr.[9] and the TDF by Wesolowski.[18] The chameleon hash function is instantiated, as in the work of Ateniese and de Medeiros.[19]

To evaluate the performance of our scheme, we implement a prototype and run experiments on a Windows 10 system with a 3.70-GHz Intel Xeon CPU and 32GB DDR4 memory. We use SHA-3 to implement the hash functions in our scheme. The size of each file is set to 64 M. The length of each storage period T is set to 30 days. The sizes of the hash value and modulus in the chameleon hash are both 256 bits.

ALGORITHM 2: CBDIC.PROVE: THE STORAGE PROOF-GENERATION ALGORITHM.

```
Input: The initial challenge $c_i$, the file set $\{F\}$, checking round $f$, chameleon
       hash public key $hpk$, and the tag set $\Sigma$.
Output: The storage proof $pf$.
for $j = 1$ to $f$ do
    $id_{ij} \leftarrow PRP(c_{ij})$;
    $v_{ij} \leftarrow PoR.P(pk, c_{ij}, F_{id_{ij}}, \sigma'_{id_{ij}})$;
    if $j == f$ then
        break;
    end
    $u_{ij} = H(v_{ij})$;
    $q_{ij} \leftarrow TDF.TrapEval(pp, tr, u_{ij})$;
    $c_{i,j+1} \leftarrow CH(hpk, q_{ij}, r_{ij}, s_{ij})$;
end
$hid_i = H(id_{i,1}, \ldots, id_{i,f})$;
$c_i = H(c_{i,1}, \ldots, c_{i,f})$;
$q_i = H(q_{i,1}, \ldots, q_{i,f})$;
$r_i = H(r_{i,1}, \ldots, r_{i,f})$;
$s_i = H(s_{i,1}, \ldots, s_{i,f})$;
$pf = (hid_i, c_i, q_i, r_i, s_i)$;
return the storage proof $pf$;
```

Probability analysis

To evaluate the performance of the sampling strategy, we analyze the probability model of sampling.

As in Figure 4, we can see that the verification probability grows with the increase of checking rounds and challenged files. Suppose that the time length of the auditing frequency is 1 h. For the storage period of one month, our scheme can achieve 99.9% verification probability with $\rho = 0.1\%$, which can satisfy the needs of many practical applications.

Communication cost

In this section, we compare the proposed scheme with cPoSt on the communication overhead. Suppose that the size of each file is $|F|$, the size of the hash value is $|H|$, the number of updated files is z, and the current state is i. $|q|$ denotes the modulus in the chameleon hash.

As the two schemes have to submit all of the files in the Store algorithm and submit the updated file in the Update algorithm, we ignore the communication overhead of the transmitting files. Table 1 demonstrates the theoretical comparison of the communication overhead of the Store, Challenge, Prove, and Update algorithms. It is easy to see that our approach is much more efficient than cPoSt in both the Challenge and Prove algorithms. The communication overhead of our schemes is independent of the number of stored files n while the overhead of cPoSt is linear with n. This is because cPoSt has to check all the files within each time slot while our schemes check only one file in each time slot by using the sampling strategy. For the Store and Update algorithms, it is not easy to straightly compare two schemes from the table. The main difference with the Store algorithm between the two schemes is in $2ln|H|$ and $l((2f+1)|q|+|H|)$. Similarly, in the Update algorithm, the major difference lies in the $2(l-i)z|H$ of cPoSt and the $(l-i)(2f|q|+|H|)$ of our approach.

For a better illustration, we compare the performance of the Store and Update algorithms of the two schemes in Figure 5(a) and (b). The $|H|$ and $|q|$ are both 256 bits. The auditing frequency t is 1 h. Thus, the checking rounds f is 720. The auditing times l is set to 10. For the Update algorithm, the number of updated files z is 500.

As seen in Figure 5(a), the communication cost in the proposed scheme increases more slowly than does cPoSt in the Store algorithm, which is because our scheme generates and

ALGORITHM 3: CBDIC.UPDATE: THE FILE-STORING ALGORITHM.

Input: The updated file F^*, the key pair (pk, sk), checking rounds f, chameleon hash public-private key pair (hpk, hsk), the *state*, the initial challenge c_i, and auditing times l.

Output: The updated tag set Σ_u.

```
$\sigma_F^* \leftarrow PoR.Store(sk, F^*)$;
for $i = state + 1$ to $l$ do
    for $j = 1$ to $f$ do
        $id_{ij} \leftarrow PRP(c_{ij})$;
        if $F_{id_{ij}} == F^*$ then
            $v_{ij}^* \leftarrow PoR.P(pk, c_{ij}, F^*, \sigma^*)$;
            if $j == f$ then
                break;
            end
            $u_{ij}^* = H(v_{ij}^*)$;
            $q_{ij}^* \leftarrow TDF.TrapEval(pp, tr, u_{ij}^*)$;

            $r_{ij}^*, s_{ij}^* \leftarrow HCol(hsk, (c_{i,j+1}, q_{ij}, r_{ij}, s_{ij}), q_{ij}^*)$;
        end
        else
            $r_{ij}^*, s_{ij}^* \leftarrow HCol(hsk, (c_{i,j+1}, q_{ij}, r_{ij}, s_{ij}), q_{ij})$;
        end
    end
    $hid_i = H(id_{i,1}, \ldots, id_{i,f})$;
    $c_i = H(c_{i,1}, \ldots, c_{i,f})$;
    $q_i^* = H(q_{i,1}, \ldots, q_{i,f})$;
    $r_i^* = H(r_{i,1}^*, \ldots, r_{i,f}^*)$;
    $s_i^* = H(s_{i,1}^*, \ldots, s_{i,f}^*)$;
    $\sigma_i^* = H(hid_i, c_i, q_i^*, r_i^*, s_i^*)$;
end
$R^* = \{r_{ij}^*\}_{i \in (state, l], j \in [1, f]}$;
$S^* = \{s_{ij}^*\}_{i \in (state, l], j \in [1, f]}$;
$\Sigma_u^* = (R^*, S^*, \sigma_F^*, \sigma_{state+1}^*, \ldots, \sigma_l^*)$;
return $\Sigma_u^*$;
```

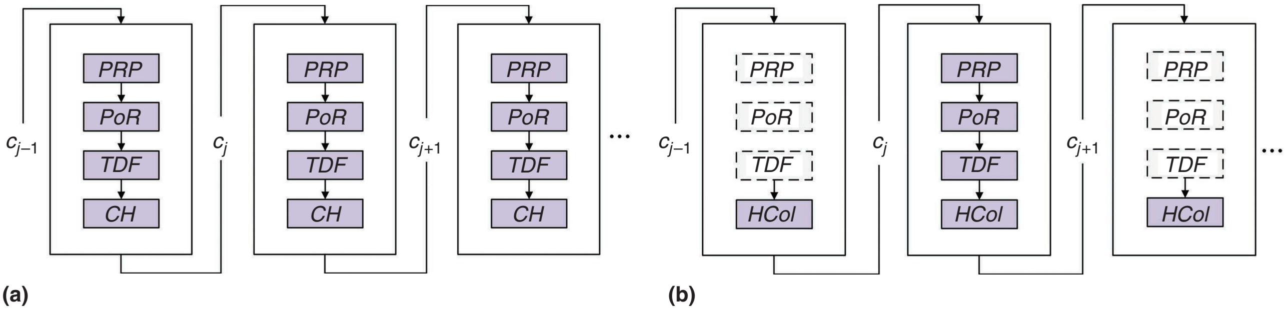


FIGURE 3. A comparison of the (a) Store and (b) Update algorithms. PRP: pseudorandom permutation; PoR: proof of retrievability; TDF: trapdoor delay function; CH: chameleon hash.

transmits only one tuple $(C, R, S, \sigma_1, \dots \sigma_l$ in the tag set for all the files, while cPoSt has to transfer a tag set for each file. Figure 5(b) demonstrates that our scheme incurs slightly more costs than does cPoSt in the Update algorithm. The reason for this is that the Update algorithm in our scheme eliminates the transmission of the updated challenge set C, but it needs to transfer extra R^* and S^* to the cloud to support dynamic auditing. The advantage of this design is that it greatly reduces the computation cost in the Update algorithm, which will be explained in the next section.

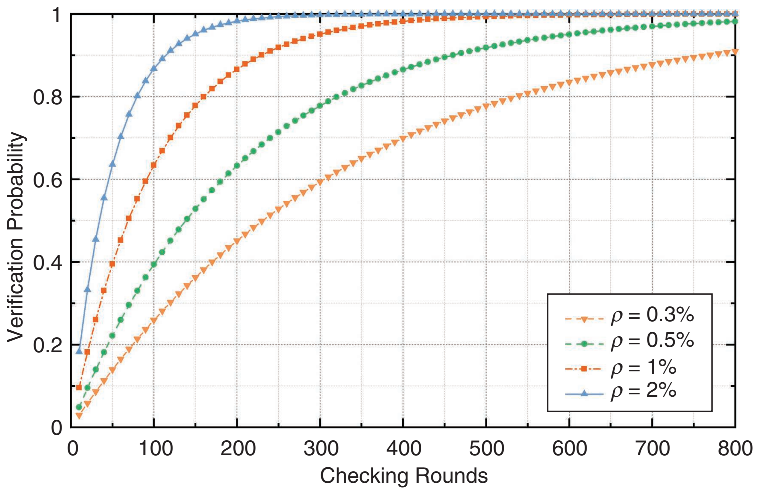


FIGURE 4. The verification probability versus different checking rounds f and corruption probability ρ in each time slot.

Computation cost

Store algorithm. In Figure 5, (c) and (d) illustrate the computation time of the Store algorithm of our approach and cPoSt. As presented in Figure 5(c), the computation time of both schemes grows linearly; however, our approach incurs much less computation cost than does cPoSt. The reason why is because we adopt the sampling strategy to build a lightweight probabilistic auditing scheme. The computation overhead can be greatly reduced because the data owner in our scheme

TABLE 1. A comparison of communication costs between our scheme and cPoSt.

Algorithm	cPoSt	Proposed scheme
Store	$2ln\lvert H\rvert + n\lvert H\rvert$	$l((2f+1)\lvert q\rvert + \lvert H\rvert + n\lvert H\rvert$
Challenge	$ln\lvert H\rvert$	$l\lvert q\rvert$
Prove	$2nl\lvert H\rvert$	$5l\lvert H\rvert$
Update	$2(l-i)z\lvert H\rvert + z\lvert H\rvert$	$(l-i)(2f\lvert q\rvert + \lvert H\rvert) + z\lvert H\rvert$

needs to only compute one tag set for all the files while computing one tag set for each file in cPoSt. This is also why more files incur more overhead in cPoSt while the cost in our approach remains constant.

Figure 5(d) demonstrates the comparison of the Store algorithm of two schemes versus the different auditing

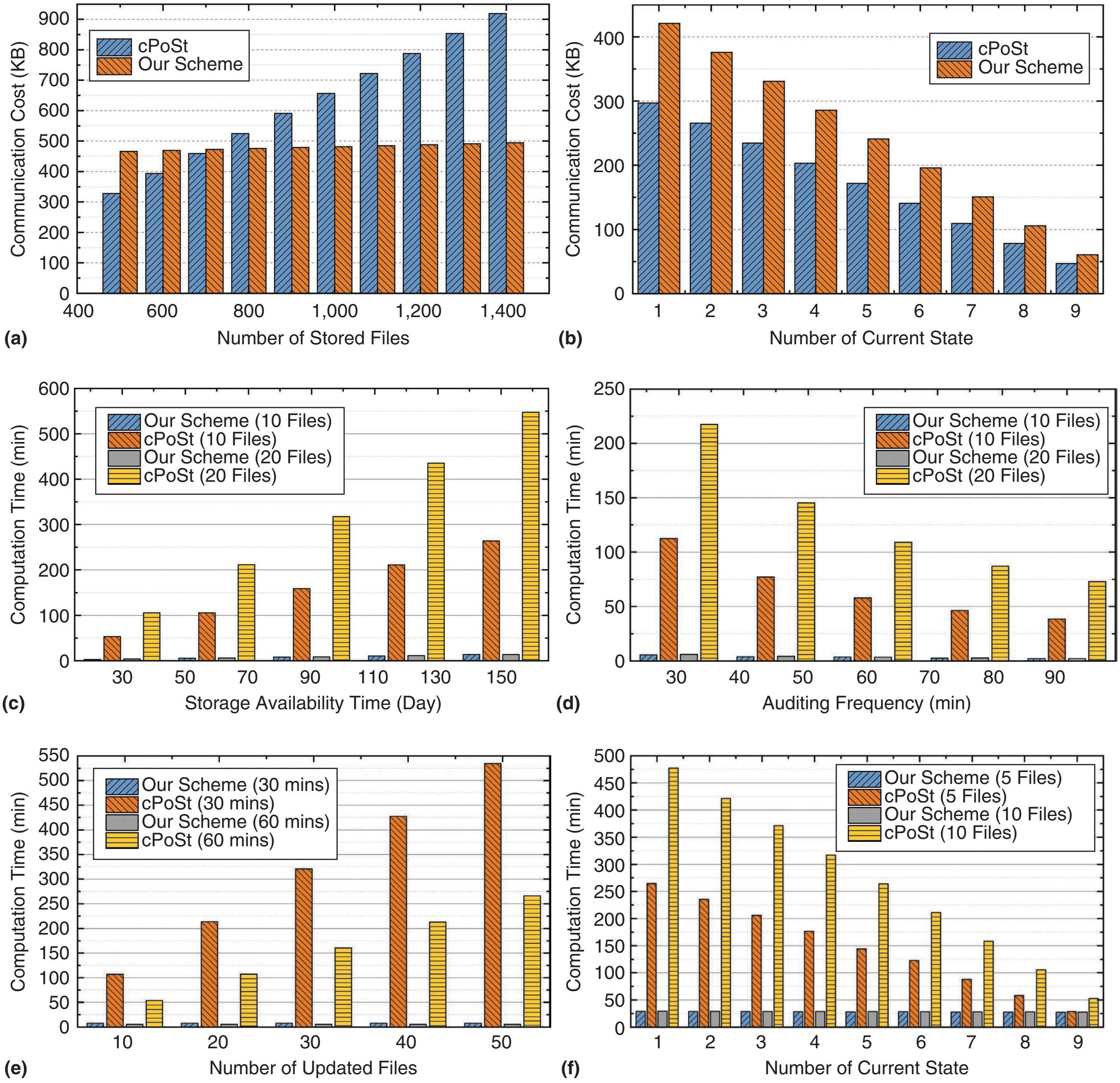


FIGURE 5. A comparison of the communication and computation costs. (a) A communication comparison of the Store algorithm, (b) a communication comparison of the Update algorithm, (c) the computation time of the Store algorithm versus different storage availability times, (d) the computation time of the Store algorithm versus different auditing frequencies, (e) the computation time of the Update algorithm versus different updated files, and (f) the computation time of the Update algorithm versus different states.

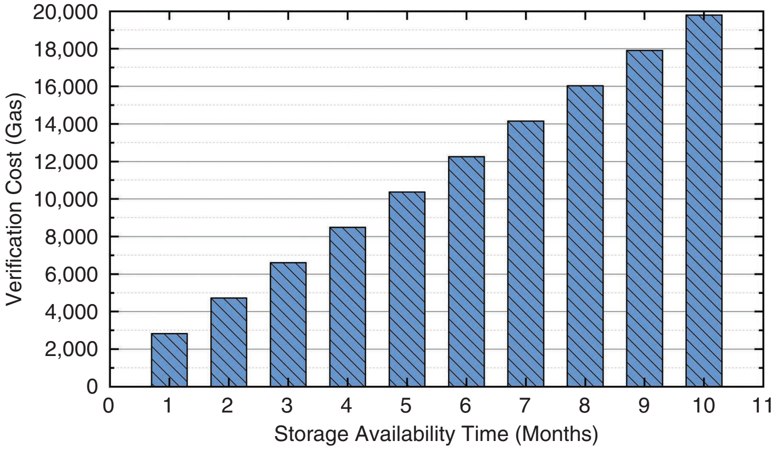


FIGURE 6. The verification cost of the auditing contract versus different storage availability times.

frequency and the number of files. As the auditing frequency t increases, the checking round f will decrease. As a result, the computation complexity and computation time of both schemes decrease. It is clear that our approach reduces much more computation overhead compared with cPoSt due to the similar reason expressed in Figure 5(c).

Prove algorithm. As computing TDF is inherently time consuming, both schemes require the same time as that of the storage period to run the Prove algorithm for one file. However, for proving the integrity of multiple files, our scheme will be much more efficient due to the same reason as in the Store algorithm.

Update algorithm. We show the performance of our Update algorithm in Figure 5(e) and (f). Updating one file in cPoSt requires the data owner to perform the Store algorithm again. As a result, the update phase is the same as in the Store algorithm. In Figure 5(e), the updating costs of both schemes grow linearly with the increase of the updated files in the challenge set. In addition, our Update algorithm can greatly reduce the computation costs compared with those of cPoSt. This is because by adopting the chameleon hash function, our Update algorithm can skip most time-consuming TDFs in the Store algorithm.

In Figure 5(f), we can see that the updating costs of both approaches decrease with the increase of the current state, because there will be fewer tag sets to be updated as the number of the current state increases. However, our approach maintains low costs against different current states.

Verify algorithm. The auditing contract based on the smart contract is implemented on Ethereum platform using Solidity language of version 0.6.4. The auditing frequency t is set as 1 h.

Figure 6 demonstrates the efficiency of the auditing contract with a different storage availability time, which is also related to different auditing times l. We can see that the gas cost of executing the auditing contract increases linearly with the storage availability time. In reality, the auditing contract is efficient and practical for many applications as it takes approximately only 19,000 Ethereum gas for auditing data over 10 months, which is worth less than US$0.26[20] (according to the gas and ETH prices on 1 November 2020). The price is much cheaper than most of the auditing services.

A data explosion in AVNs has brought a great challenge for protecting data integrity in cloud storage from internal or external threats. In this article, we proposed a lightweight cloud-auditing scheme that can continuously check the shared data in AVNs. By adopting a sampling strategy as well as a TDF, we built a lightweight scheme that greatly reduces the calculation overhead of tag generation. Furthermore, we extended the continuous auditing scheme to efficiently support data dynamics by integrating the chameleon hash function. Data updating in the proposed scheme requires only recalculating the TDF for the updated file instead of all the TDFs in an auditing tag. We provided an analysis of the validation probability of our scheme and then gave a comprehensive evaluation among the proposed approach and the state-of-the-art schemes. The extensive experiments show that our scheme has superior computation performance compared with the existing schemes.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2020YFB2009501), the National Natural Science Foundation of

ABOUT THE AUTHORS

HAIYANG YU is an assistant professor of computer science at Beijing University of Technology, Beijing, 100124, China. His research interests include decentralized storage, blockchain, and data security. Yu received a Ph.D. in computer science and technology from Beijing University of Technology. Contact him at yuhaiyang@bjut.edu.cn.

SHUAI MA is currently pursuing a master's degree in computer science at Beijing University of Technology, Beijing, 100124, China. His research interests include cloud storage auditing and adversarial examples. Ma received a B.S. in computer science from Beijing University of Technology. Contact him at usermashuai@emails.bjut.edu.cn.

QI HU is currently pursuing a master's degree in cyberspace security at Beijing University of Technology, Beijing, 100124, China. Her research interests include cloud storage auditing and data integrity checking. Hu received a B.S. in internet of things from Hebei Normal University. Contact her at gsky.lucky@gmail.com.

ZHEN YANG is a full professor of computer science and engineering at Beijing University of Technology, Beijing, 100124, China. His research interests include data mining, machine learning, trusted computing, and content security. Yang received a Ph.D. in signal processing from the Beijing University of Posts and Telecommunications. He is a senior member of the Chinese Institute of Electronics and a Member of IEEE. Contact him at yangzhen@bjut.edu.cn.

China (61671030), the Industrial Internet Innovation Development Project, the China Postdoctoral Science Foundation (2019M660377), and the Great Wall Scholar Program. It was also supported by the Engineering Research Center of Intelligent Perception and Autonomous Control, Ministry of Education. Zhen Yang is the corresponding author.

REFERENCES

1. H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, "Uav-enabled intelligent transportation systems for the smart city: Applications and challenges," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 22–28, 2017. doi: 10.1109/MCOM.2017.1600238CM.
2. W. Brenner and A. Herrmann, "An overview of technology, benefits and impact of automated and autonomous driving on the automotive industry," in *Digital Marketplaces Unleashed*, C. Linnhoff-Popien, R. Schneider, and M. Zaddach, Eds. Berlin: Springer-Verlag, 2018, pp. 427–442.
3. W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, "Internet of vehicles in big data era," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 1, pp. 19–35, 2017. doi: 10.1109/JAS.2017.7510736.
4. C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage," *IEEE Trans. Comput.*, vol. 62, no. 2, pp. 362–375, 2013. doi: 10.1109/TC.2011.245.
5. H. Wang, D. He, J. Yu, and Z. Wang, "Incentive and unconditionally anonymous identity-based public provable data possession," *IEEE Trans. Services Comput.*, vol. 12, no. 5, pp. 824–835, Sept. 2019. doi: 10.1109/TSC.2016.2633260.
6. J. Shen, D. Liu, D. He, X. Huang, and Y. Xiang, "Algebraic signatures-based data integrity auditing for efficient data dynamics in cloud computing," *IEEE Trans. Sustain. Comput.*, vol. 5, no. 2, pp. 161–173, 2020. doi: 10.1109/TSUSC.2017.2781232.
7. H. Wang, H. Qin, M. Zhao, X. Wei, H. Shen, and W. Susilo, "Blockchain-based fair payment smart contract for public cloud storage auditing," *Inf. Sci.*, vol. 519, pp. 348–362, May 2020. doi: 10.1016/j.ins.2020.01.051.
8. G. Ateniese et al., "Provable data possession at untrusted stores," in *Proc. 14th ACM Conf. Comput. Commun. Security*, Oct. 2007, pp. 598–609. doi: 10.1145/1315245.1315318.
9. A. Juels and B. S. Kaliski Jr., "Pors: Proofs of retrievability for large files," in *Proc. 14th ACM Conf. Comput. Commun. Security*, Oct. 2007, pp. 584–597. doi: 10.1145/1315245.1315317.
10. A. Abdallah and X. Shen, "Lightweight authentication and privacy-preserving scheme for v2g connections," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2615–2629, Jan. 2016. doi: 10.1109/TVT.2016.2577018.
11. X. Lin and X. Li, "Achieving efficient cooperative message authentication in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 62, pp. 3339–3348, Sept. 2013. doi: 10.1109/TVT.2013.2257188.

12. D. Yue, R. Li, Y. Zhang, W. Tian, and C. Peng, “Blockchain based data integrity verification in p2p cloud storage,” in *Proc. 2018 IEEE 24th Int. Conf. Parallel Distributed Syst. (ICPADS)*, pp. 561–568. doi: 10.1109/PADSW.2018.8644863.
13. C. Wang, S. Chen, Z. Feng, Y. Jiang, and X. Xue, “Blockchain-based data audit and access control mechanism in service collaboration,” in *Proc. 2019 IEEE Int. Conf. Web Services (ICWS)*, pp. 214–218. doi: 10.1109/ICWS.2019.00044.
14. Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, “Blockchain empowered arbitrable data auditing scheme for network storage as a service,” *IEEE Trans. Services Comput.*, vol. 13, no. 2, pp. 289–300, Mar.–Apr. 2019. doi: 10.1109/TSC.2019.2953033.
15. H. Wang, Q. Wang, and D. He, “Blockchain-based private provable data possession,” *IEEE Trans. Dependable Secure Comput.*, early access, Oct. 2019. doi: 10.1109/TDSC.2019.2949809.
16. R. Chen, Y. Li, Y. Yu, H. Li, X. Chen, and W. Susilo, “Blockchain-based dynamic provable data possession for smart cities,” *IEEE Internet Things J.*, vol. 7, no. 5, pp. 1–1, Jan. 2020. doi: 10.1109/JIOT.2019.2963789.
17. G. Ateniese, L. Chen, M. Etemad, and Q. Tang, “Proof of storage-time: Efficiently checking continuous data availability,” in *Proc. 27th Annu. Netw. Distributed Syst. Security Symp., NDSS 2020*, San Diego, CA, Jan. 2020, pp. 1–15.
18. B. Wesolowski, “Efficient verifiable delay functions,” in *Proc. 38th Annu. Int. Conf. Theory Appl. Cryptographic Techn. (EUROCRYPT 2019)*, Darmstadt, Germany, May 19–23, 2019, pp. 379–407. doi: 10.1007/978-3-030-17659-4_13.
19. G. Ateniese and B. de Medeiros, “On the key exposure problem in chameleon hashes,” in *Proc. Int. Conf. Security Commun. Netw.*, Springer-Verlag, 2004, pp. 165–179.
20. ETH Gas Station. Nov. 1, 2020. https://ethgasstation.info/ (accessed Nov. 1, 2020).

Digital Object Identifier 10.1109/MC.2021.3094703

Ethics in Autonomous Vehicle Software: The Dilemmas

Sachin Motwani, Indraprastha Institute of Information Technology-Delhi

Tarun Sharma, Birla Institute of Technology and Science Pilani-Hyderabad Campus

Anubha Gupta, Indraprastha Institute of Information Technology-Delhi

Automation is everywhere in the automobile industry. Ethical dilemmas pose a major challenge while designing such systems. It becomes imperative to discuss these dilemmas, thereby helping the software architects to design safe, secure, and reliable software for autonomous vehicles.

Autonomous vehicles (AVs) are the automobiles trained to transport goods and people from one place to another with little or no human interference and with the ability to make unprecedented driving decisions in real time. A boost to AVs was provided by the advent of artificial intelligence to the research world. In the 1950s, scientists and engineers made attempts to create technology that could think like humans. The first autonomous car, which ran on the roads of France back in 1986, was built by German engineers with the help of a scientist named Ernst Dickmanns.[1] Over time, the automobile industry tried and tested various autonomous features with the goal of increasing driver/rider comfort. The recent key players in this domain are Google's Waymo LLC., Amazon-backed Zoox Inc., and Tesla Inc.

SAE International, the global organization that develops standards for various engineering industries, divided the levels of autonomy for a vehicle from 0 to 5.[2] Here, zero means *not autonomous*, that is, traditional cars

Digital Object Identifier 10.1109/MC.2021.3077576
Date of current version: 30 July 2021

driven completely by human drivers, followed by *driver assisted* and *partially automated vehicles*. These contain some simple autonomous features that are used under the complete supervision of the driver (for example, assisting the driver to park the car).

The third level of autonomy is when the driver is present but does not continuously monitor the vehicle's decisions. However, a driver needs to be ready for any potential threat or difficult situation. This feature is currently provided by some high-end cars (for example, the highway pilot mode or the mountain mode). The fourth level is that of high automation, wherein under certain controlled scenarios, the vehicle can work completely autonomously. The automobile industry is currently in a race to establish this level before they roll out AVs with level 5—fully AVs. In this level, the AV will commute securely and reliably anywhere in any conditions without human interventions and without the potential to damage someone or something inside or outside itself. Henceforth, an AV stands for level 5 automation in this article.

Finally, one key aspect while developing AVs and making them acceptable in the society is the ethical decisions that these vehicles may require to make. Regular discussions in both science and philosophy are being held regarding machine morals.[3,4] It will make all the difference in thve transition from no autonomy to complete autonomy. Some commonly debated questions include "Between a cat and a human, what will the AV choose to save from being run over, when saving both is not an option and not swerving means hitting both?" or "Will it run over an elderly woman or a toddler [Figure 1(a)] or crash itself, hurting the rider inside?"[5] Devising algorithms that will help AVs make these moral judgments is a formidable challenge. Further, the AV software architects are expected to improve the explainability of such user-centric algorithms.[6]

This article attempts to encapsulate the available technical literature and

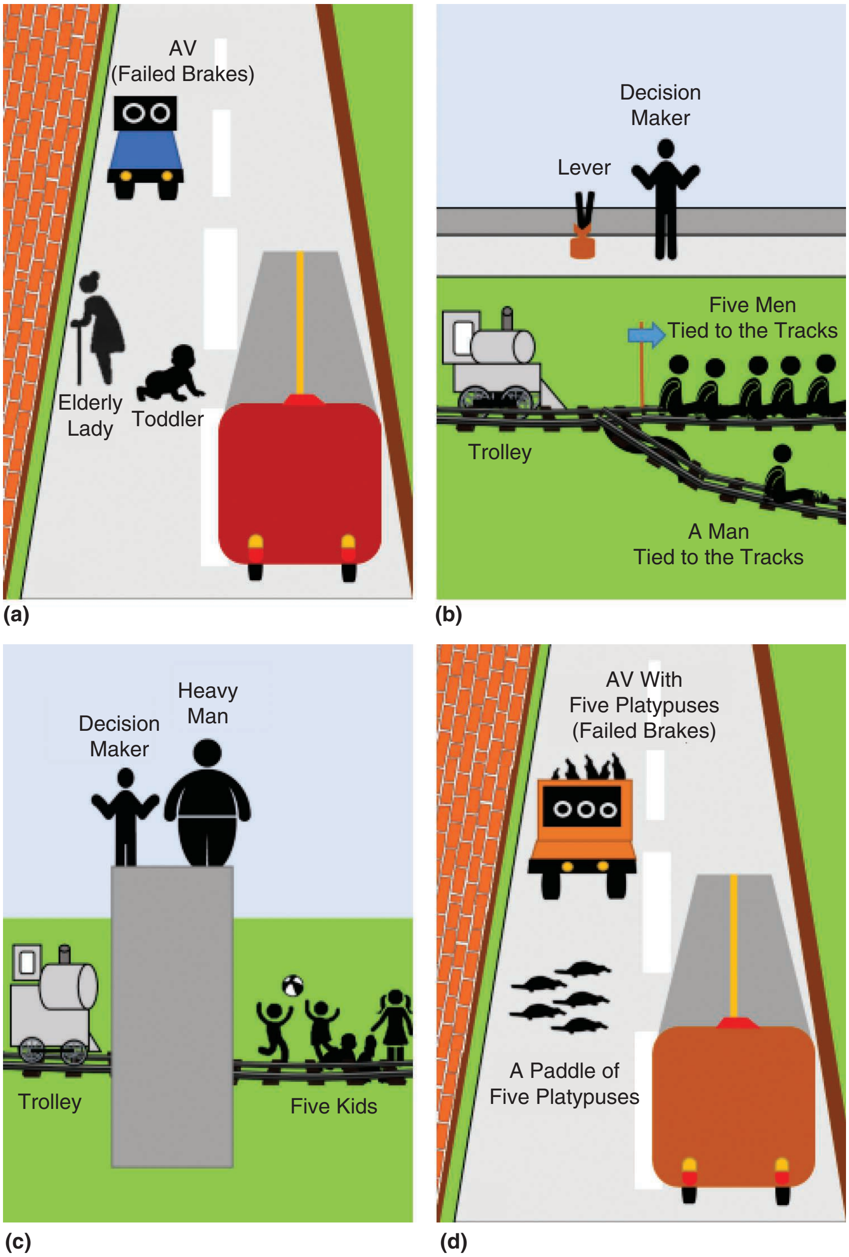


FIGURE 1. The different scenarios. A (a) dilemma to save an elderly lady or a toddler, (b) trolley dilemma, (c) footbridge dilemma, and (d) platypus dilemma.

philosophical opinions around the AV dilemmas. In the next section, "Possible Dilemma Situations," we present a concise and coherent write up on binary AV dilemmas under two separate categories: involving and not involving riders. We propose a unique classification framework of these dilemmas as a flowchart, which is unprecedented. This flowchart eases the approach to understand the dilemmas, allowing us to propose some of the most preferred/rational approaches as solutions that have the potential of high acceptability. These are presented in the "Toward the Solutions" section, where we discuss some guidelines and ground rules that have been set up by institutions worldwide, including philosophical surveys and experiments undertaken around AV dilemmas, along with what could be some rational approaches to handle these dilemmas.

THE THIRD LEVEL OF AUTONOMY IS WHEN THE DRIVER IS PRESENT BUT DOES NOT CONTINUOUSLY MONITOR THE VEHICLE'S DECISIONS.

For some of the dilemmas whose solutions are not yet clear, we present some rarely discussed perspectives. This discussion can, perhaps, help the reader make better informed decisions in their respective social, geographical, and cultural contexts. The "Summarizing Solutions" and "Future Scope" sections offer some explanations not found in the existing literature, which can potentially help architects build safe, secure, and reliable software for AVs.

POSSIBLE DILEMMA SITUATIONS

Machines cannot be allowed to make random decisions in situations of potential accidents. In fact, one of the major arguments for AVs is that, at least theoretically, they attempt to reduce road accidents. Hence, every case of a driving dilemma should be fed by the engineer with utmost care postdiscussion on the certainty of its consequences. In the words of American-German psychologist Kurt Lewin, this is a classic case of an approach-approach conflict,[7] but in machines instead of humans.

For example, the famous trolley dilemma[5] is the textbook example of AVs' ethics dilemma. It has been discussed for roughly a century now, earlier as a moral question, followed by a philosophical case study, and finally in its current context. As shown in Figure 1(b), the problem starts by giving you access to the lever of a runaway trolley, which you may use to redirect the trolley toward a single person tied to the track, leading to his/her untimely demise. Or you may let the trolley go on its way, killing five people tied to the track it is moving toward. What do you do? The *trolley dilemma* is an umbrella term used to describe a series of scenarios that involve such predicaments [for example, the footbridge dilemma presented in Figure 1(c)].[8]

Much literature is available that discusses such situations. The Moral Machine,[9] for instance, provides nine different contrasts of opinions: Should an AV place the lives of humans above those of animals, riders above those of pedestrians, women above those of men, younger over old, healthy over unhealthy, people of a higher social strata over a lower one, or people abiding by the law over lawbreakers? Should more people be saved rather than fewer? And lastly, should AVs actually take an action or stay on their dedicated task (inaction)? Universally accepted answers to these scenarios do not exist.

Interestingly, *motion safety*, that is, the ability to avoid collisions, is apparently one of the key advantages of adopting AVs.[10] According to the series of six studies conducted by Bonnefon et al.[4] in 2015, participants (U.S. residents only) approved of utilitarian[11] AVs (that is, AVs that will sacrifice their riders' life for the greater good) for others, but disagreed to accept the same fate for themselves.

Dilemmas involving riders

The most common scenario is when the rider would be in danger, and AVs have to make a choice between the rider and the one on the opposite side. The choices opposite to the rider could be one of the following: human/humans, animal/animals, property (any inanimate obstruction, for example, walls, boundaries, barricades, dividers, and so on), other vehicle/vehicles, reaching a destination by a certain deadline, or the driving laws that are being violated while saving the rider (see Figure 2). In fact, the rider could also be classified as a male, female, animal, adolescent, toddler, elderly lady, and so forth. But, for the sake of simplicity, let us define a rider as *any living being*.

Consider the example of a sudden brake-failure condition [Figure 1(d)]. Here the software is put in a situation where an AV experiences a sudden brake failure while traveling with a paddle of five platypuses and also encounters a paddle of five other platypuses crossing the road. The other side of the road is blocked by a concrete barricade. Hence, the AV has to decide to either keep moving and kill the paddle crossing the road or swerve and crash into the barricade, killing the paddle in the AV. The existing software programs are designed to be efficient for the day-to-day operations of vehicles and may not account for such exceptional cases. It is, however, imperative for the software to have built-in features for these scenarios.

Dilemmas not involving riders

This section deals with situations where the AV's rider is not in danger (see Figure 2). This includes the cases where the vehicle deals with the transportation of goods or when a vehicle is being called from one place to another, without any rider inside. In these situations, the vehicle could certainly make the simple choice of damaging itself over the living beings in its path. However, a possible scenario could be that the vehicle is instructed that damaging itself is not a choice. An example could be an AV carrying fuel, where a fire triggered by an accident will result in far more damage.

Therefore, a secondary choice is required to be made by selecting between two objects in its path. The cases that may arise here are numerous combinations of humans, animals, deadlines, law, property, and other vehicles. This secondary choice is even more complex because a number of possible combinations of choices exist, and the question goes beyond self-sacrifice. For instance, consider an AV that has to choose between two bicyclists traveling in adjacent lanes, where one is wearing a helmet but the other is not.[5] Other scenarios could be whether to break the law or damage some property or any other AV to reach the destination in time. Additionally, it also

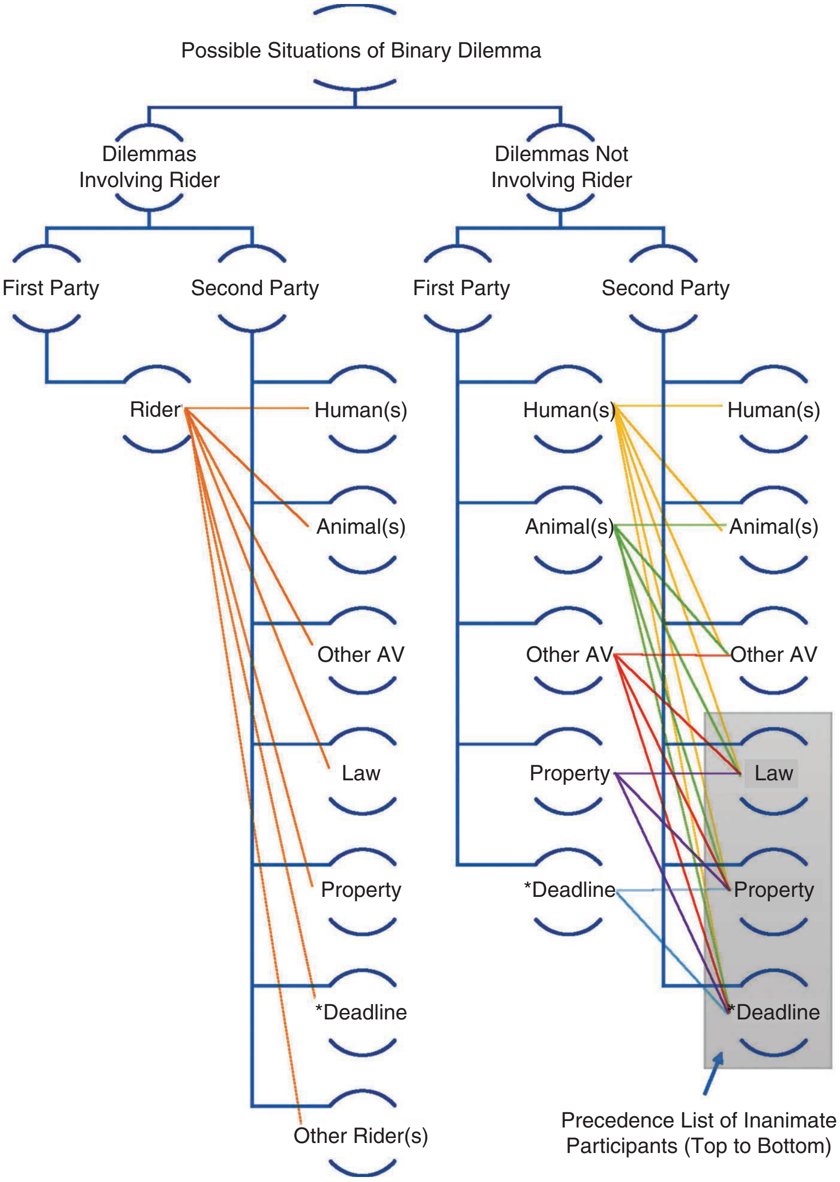


FIGURE 2. This flowchart depicts most of the possible situations of binary dilemmas that may occur for AV software. The colored lines match a pair of participants, forming a combination that can create a binary dilemma within the ethics of an AV controlling software. *The deadline implies reaching the destination in time.

includes the scenarios that require deciding between two other AVs.

TOWARD THE SOLUTIONS

The previous sections have briefly explained and summarized most of the possible binary dilemmas under specific assumptions. The rest of the article focuses on trying to find solutions to these problems or at least some of them. We discuss ideas and opinions from the literature as well as talk about our observations and findings. The architects of the software, and the ones who shall be proposing solutions to these problems, must ensure that the solutions are consistent and justified yet not outrageous (for general public) or discouraging to buyers.

The reader might argue that the described dilemmas will never occur with an AV, and hence, there is no need to discuss the solutions. Nevertheless, dismissing the argument might not be the best solution to the problem. The United Kingdom's Centre for Connected and Autonomous Vehicles points out an AV road-testing U.K. norm that mandates to "have a driver (in or out of the vehicle) who is ready, able, and willing to resume control of the vehicle [being tested]."[12]

In the future, simply pulling over and handing control back to the human operator might not be enough.[5] These are the best remedies that we have today. But, if the AV needs to be brought into reality, there must be more response options. Another strong argument for discussing these problems and solutions is that, for instance, "Wasn't the driver's seat supposed to be empty?" or "What if there are only kids in the AV?" More than that, it is a forceful U-turn from level 5 back to level 3 of autonomous driving control, which is certainly not a technically progressive solution. Also, the risk-to-time factor may not allow enough time for a driver to take immediate control, followed by taking a sane, quick, and critical decision.[13] This, in turn, may worsen the situation.

Most of the related institutions are focusing on building software that somehow avoids the dilemmas. For instance, the policy outlined by Germany suggests to "prevent accidents wherever this is practically possible."[14] Certainly, the AV should have good control over its speed and trajectory to avoid these situations. But the entire argument is based on the cases of unavoidable scenarios. Also, choosing one option over the other, or in other words, crash optimization, is seen as targeting[5] or diverting harm from one area to another.

As a starting point to a level 5 AV, a common suggestion is to mark a dedicated lane for these vehicles. However, not only is this an expensive solution both in terms of time and resources, the software ethics required to be built for this will be different from that needed in a level 5 AV. For instance, there is a lesser probability that some cars will drive inappropriately within this lane than when the AV can use the entire width of the road. As a result, restricting an AV to one lane is more of a solution than introducing a level 5 AV to a level 4 environment. Hence, this scenario is beneficial for a period of transition to a fully functional AV and is definitely not good in the long run.

A better suggestion is to have AVs programmed to be able to face such dilemmas. The software should have a provision of an emergency mode. In this mode, the system can take additional precautions to save the rider, and possibly any pedestrian, from any harm. An example of this could be partially inflating the airbags before an anticipated collision. Other preparations could include alarms and sirens that alert people around an AV. Communication could be established with the traffic-managing authorities immediately. With increased usage of Internet of Things devices, this communication is plausible. An alert about the threatening situation could also be sent to other AVs near that location, preventing it from worsening or becoming disastrous.

The other case considers the legality of decisions made by AVs. Recent recommendations by the European Commission suggests using current collision statistics to select vulnerable road users and design the AVs' driving algorithm accordingly.[6] It further recommends the organic selection of the dilemma outcomes initially and to use the data of the same to improve and develop traffic rules accordingly. Relative to the rules and regulations that will define the motion of any vehicle on the road with level 5 AVs, breaking the law to save someone might be counterproductive; doing so may lead to the *butterfly effect*, a term used to denote the sensitive dependence on initial conditions in the chaos theory. Here, a small change in the initial conditions of a nonlinear system results in large differences in the later states.[15] If an AV enters a no-entry lane to save a pedestrian, for instance, it may run over a number of laborers working on the street who were, perhaps, outside the field of view of the AV initially.

Thinking of an AV as a real-time operating system (RTOS), it is expected to be accurate in reaching a destination on time, that is, meeting a deadline. It is therefore expected to have a hard RTOS because hard RTOSs are designed to value time over any other parameter. The door-to-door travel time would be calculated by the vehicle in advance with accuracy. However, considering the harm it may cause while trying to meet the deadline, we argue to position the deadlines to the lowest rank in the precedence list of inanimate options under consideration (Figure 2). This list of precedence in Figure 2 has been suggested to avoid or minimize the damage to life. The priority list may hold the law at the top to avoid any mishappenings or the generation of new dangers, followed by the avoidance of damage to property that may otherwise unintentionally cause damage to life, and finally, the deadline. Conclusively, AVs ought to be soft RTOSs. Note that in this list of precedence among inanimate objects, the options of other vehicles has been omitted as they may also carry living beings.

Thus far in this section, we have discussed at length the generic situations where dangers to riders (chosen as the basis of categorization) did not really make a difference in the actions or consequences (for example, a list of precedence among inanimate objects or breaking the law). Let us now discuss where the presence or absence of a rider makes a difference.

DEVELOPERS SHOULD IMPLEMENT SECURE INSTALLATION TECHNIQUES FOR THE DRIVING SOFTWARE SO THAT ANY TAMPERING CAN BE PREVENTED OR IDENTIFIED IMMEDIATELY.

Dilemmas involving riders

The idea of utilitarian vehicles has been around for a while now. Thinking of the readiness to sacrifice one's own life intentionally when other options are (apparently) available, the idea is too brave to be handled by the general public. As bizarre as it sounds to a person thinking practically, hearing of it in articles such as "The Robot Car of Tomorrow May Just Be Programmed to Hit You"[16] is even more terrifying.

One of the major psychological potholes in the human mindset is that change is never accepted at first mention. The personal vehicles of today have complete control over the owner. The perception toward utilitarian AVs may firmly transform as people start to view situations from the perspective of both the pedestrian and the rider simultaneously. Martin et al.[8] also raised this thought by pointing out a flaw in the questions that were presented in the study by Bonnefon et al.[4] The former authors argue that the conflicting responses to the two questions about utilitarian AVs (Which option is moral: to kill a pedestrian or sacrifice the rider? And will they buy cars with such algorithms embedded?) were due to the presentation of only one side of the situation, namely, not highlighting the fact that an individual who buys an AV with a nonutilitarian algorithm would also, inevitably, be a pedestrian at some point in time.

Another argument about opposing the change is that, currently, people are more attached to their personal vehicles, akin to their loved ones.[17] A self-sacrificing vehicle could be viewed as a betrayal by someone beloved. However, thinking about the business model of AVs, which would require regular software updates and routine hardware examinations, there is a high probability of future AVs as interchangeable units under a transit system, similar to the one where you book a shared cab today. Skeete[2] supports this argument with some more concrete reasoning. As people start to shift their point of view from personal vehicles to hiring systems (mobility as a service), they may arrive at a consensus on deploying an algorithm of maximum overall safety into AVs.

Logically speaking, an AV could be installed with world-class safety features, including all kinds of modern technologies that help reduce damage to the rider. This is not likely for a living being walking on the road. Hence, the tradeoff in the case of a rider versus a pedestrian (any living being not in a vehicle) tends to weigh more toward saving the pedestrian. Also, this will encourage a sense of alertness in the rider to strictly abide by the safely precautions they are obligated to abide by inside an AV. In a nutshell, there is no unique solution for this case. It is like an iterative process that will be successful when both the owners and developers come to an agreeable point. The process can be overseen by the government through law, policies, and treaties. It is important to mention that, although governments may sign treaties to establish utilitarian AV driving software, it is likely that some private vehicle owners may indulge in unlawful tampering of the software to make it more rider friendly. As a result, developers should implement secure installation techniques for the driving software so that any tampering can be prevented or identified immediately.

Dilemmas not involving riders

The most critical and the debated case among the dilemmas is how to choose between an elderly lady and a toddler, or a disabled man and a blind girl, or a local and a tourist. Choosing between one of the two on some customary preferences (for example, preferring to not harm a female over a male) may violate the basic professional codes of ethics. For instance, according to the IEEE Code of Ethics, every member under its umbrella agrees to treat everyone impartially and to not engage in “acts of discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression.”[18] The Association for Computing Machinery, along with the IEEE Computer Society, has set similar rules for the software developers, namely, that they “shall serve the best interests of their client and employer consistent with the public interest.”[19]

> ONE OF THE BEST SOLUTIONS THAT APPEARS BOTH ETHICAL AND LOGICAL IN MANY SITUATIONS IS TO AGREE ON THE ALGORITHM OF LEAST HARM.

It is often suggested by many to leave things as they are, that is, refuse to change an AV’s course of action (namely, inaction) in case of a dilemma and let both be hit. However, it seems worse than hitting just one of the two. Another solution could be one that allows a random algorithm to choose one of the two. This too is a debated scenario from an ethics point of view. One cannot leave someone’s life to random decisions when there is a chance to use a possible algorithm of reason, however unpleasant and uncomfortable that reason may be.

Someone may emphasize strict pedestrian and driving laws to avoid getting into such dilemmas. However, one cannot give unrestricted rights to an AV to run over the ones who are bending the laws. In the classic two-bikers example discussed previously, if the AV chooses to select the biker not wearing the helmet because the biker isn’t following the traffic rules, it will encourage what is called *street justice*, which is not commonly considered a healthy practice (and is even punishable under many judicial systems). Most probably, the careless biker may not survive the collision and may even face severe injuries. On the contrary, if the law-abiding biker is chosen, it would encourage the chaos of not following the traffic rules to avoid being targeted by AVs.

By these contrasting arguments, the decision lies with a country’s jurisprudence on whether the law favors punishments and penalties or rewards and recognition. With respect to jurisprudence and linking them with the studies that show diverse opinions on the ethics of AVs based on customary laws,[9] the possibility of country-tailored autonomous driving software is very high. This will be akin to the current driving codes where the basic concepts remain uniform throughout the world, but there may be slight variations from country to country.

Another important argument in this debate is whether there should be a precedence in the choice to save among the living beings (if they are at risk). The life of a laborer, for instance, must not be put at risk over the life of a puppy. As prescribed by the German laws on the ethics of AVs, the algorithms must be programmed to accept a danger to animals or inanimate objects, in the case of conflict against preventing personal injury.[14] It further supports the argument that a minimum number of injuries should be caused in unavoidable circumstances. However, it also endorses that, in a bid to save the lives of the involved individuals, compromising the safety of noninvolved individuals could not be justified. This leads us to the argument that, according to the German guidelines, the lever in the trolley dilemma of Figure 1(b) should not be pulled, which otherwise involves an uninvolved individual.

In 2014, the Massachusetts Institute of Technology Media Lab launched a gamified experiment called *The Moral Machine*,[9] in which players made decisions on such dilemmas—variations of the trolley dilemma—on behalf of a self-driving car. This information was used to generate insights on collective ethical priorities of different cultures and derive conclusions on the ethics that would be expected from an AV in the respective cultural settings. The study concluded that, overall, people chose to save a human over an animal, save more lives over fewer, and prioritize young people over old. The study suggests that the majority of people would rather swerve the car in Figure 1(a) toward the elderly lady, saving a toddler. Similarly, the heavy man would be pushed by a majority, given a choice, as in Figure 1(c).

If we consider the scenario of saving the lives of pedestrians by risking those of riders, one may argue that pedestrians will be careless and not be held accountable for breaking laws. The counterargument could be that although the humans breaking the rules can be held accountable in a court of law, if some animals do so, it is the concerned authority who would be held accountable. For instance, the paddle riding in an AV shall be given priority over the one blocking the road [Figure 1(d)]. The accountability shall rest with the local bodies (say, the forest department) responsible for taking care of them. Heavy penalties and punishments could prevent many from breaking the rules. If property is misplaced, there is someone doing the mischief, and if the other vehicle is getting in the way wrongfully, it is already breaking the drivers' code of conduct. Regardless, there ought to be someone causing such troubles on the road who could be held responsible.

One of the best solutions that appears both ethical and logical in many situations is to agree on the algorithm of least harm. It should be designed to calculate the cost function of damage[5] that is caused from making a decision. Here weights can be given to multiple dilemmatic factors. For example, if an AV hits a sport utility vehicle versus hitting a hatchback, the cost should be less for a rider who agrees to utilitarian laws because this will ensure lesser damage to the party opposite to the rider. Likewise, the cost would be less if one collides with a concrete wall compared to if one collides with a kangaroo that has jumped on the road because a kangaroo may not survive the collision. But such weights need to be designed with caution. With the approach of the aforementioned examples, one would try to give a higher cost to the biker without a helmet compared to the biker with a helmet because the biker without a helmet has a higher chance of fatality/damage in the case of a collision. However, this is undesirable because this would imply saving a biker without a helmet over a biker with a helmet, thus promoting lawlessness.

SUMMARIZING SOLUTIONS

The plausible solutions for ethical dilemmas under the specific assumptions in the previous discussion can be summarized as follows.

1. The algorithms should be designed to try and avoid unanswered dilemmas as much as possible. In case of an emergency, there can be sirens to caution nearby people and an automated system that alerts the traffic authorities as well as the other proximate AVs immediately. Similarly, the software can be provisioned for an emergency mode to ensure additional safety for both the rider(s) and pedestrian(s).
2. In cases involving only inanimate objects, breaking the law is counterproductive and may result in chaotic situations. To save lives and avoid property damage, adherence to the law is given first priority. This is followed by the avoidance of damage to property, and the third priority would be given to reaching the destination in time (meeting the deadline).
3. In cases where the rider is in danger, AVs may choose to save the pedestrian over the rider. Although the AV could be designed to ensure world-class safety measures, pedestrians do not have any safety arrangements. Hence, if an AV chooses to save the pedestrian, it will ensure that the AV owner is using the best safety-related AV accessories. The law-breaking pedestrian could be punished and penalized under the law, as applicable.
4. To make the AV driving software secure, arrangements should be made to immediately catch any unlawful tampering of the software, its memory element, or data.
5. In cases where the rider is out of danger, the pedestrian would not be selected on the basis of some discriminatory characteristic, such as age, gender, nationality, and so on. Nevertheless, if such a dilemma arises, it is anticipated that the decision would be made based on the cultural practices and/or the jurisprudence of the land. Therefore, AV driving software having variations across political boundaries is a highly likely scenario.
6. As suggested in the policy reports of some countries, the international community may soon outline AV driving standards and frame policies

ABOUT THE AUTHORS

SACHIN MOTWANI is a Ph.D. candidate at Indraprastha Institute of Information Technology-Delhi, 110020, India. His primary research interests include signal processing and artificial intelligence and their applications in biomedical areas. Motwani received a B.Tech. in electronics and communication from Guru Tegh Bahadur Institute of Technology. He is a Student Member of IEEE. Contact him at sachinmotwani20@computer.org.

TARUN SHARMA is currently pursuing his M.E. in microelectronics at the Birla Institute of Technology and Science Pilani-Hyderabad Campus, Hyderabad, Telangana, 500078, India. His research interests include machine ethics computer architecture, and VLSI design. Sharma received his bachelor's in electrical and electronics engineering from Guru Tegh Bahadur Institute of Technology. He is a Graduate Student Member of IEEE. Contact him at tarun96@ieee.org.

ANUBHA GUPTA is a professor in the Department of Electronics and Communication Engineering, Indraprastha Institute of Information Technology-Delhi, (IIIT-Delhi), Delhi, 110020, India, where she also leads the signal processing and biomedical imaging lab. Her current research interests include application of machine learning and deep learning in cancer imaging, cancer genomics, and biomedical areas; fMRI/EEG/MRI/DTI signal and image processing; and issues in higher education including pedagogy, assessment, access and retention. Gupta received a Ph.D. in electrical engineering from IIT-Delhi. She is an associate editor of *IEEE Access*. She is a Senior Member of IEEE. Contact her at anubha@iiitd.ac.in.

via a consensus on solutions to dilemmas. It will make the software development as well as the adaptation of AVs much simpler.

7. One of the best algorithms that can be justified both ethically and logically is that of least overall harm, although this solution fails in situations that may promote lawlessness.
8. Some of these dilemmas are more complex and require further analysis and discussion. There is a tradeoff on the inherent ethics versus greater good of a decision in many such dilemmas. The philosophical discourses of human decision-making processes can help in solving such critical issues.

This article discussed the binary ethical dilemmas, under some assumptions, that an AV's driving software architect may have to solve and justify to improve the safety, security and reliability of an AV. These were categorized under two broad scenarios and various subcategories to make it easy to unlock the solutions to some of these problems. In particular, an elaborate discussion on the available and newly suggested solutions was undertaken, followed by a summary of the most reasonable ones.

While categorizing the scenarios, we considered cases with respect to who is standing in front of the AV. However, we did not consider who is sitting inside the AV. In the rarest of cases, the AV may have to decide the safety of some of its riders while in danger. Such dilemmas can be taken as a part of the extension of this current work.

Some key challenges of AV design remain, such as data generation and storage. These include information-storage facilities within the AV as well as the processing that might happen onsite or through cloud computing. The security of these data and their communication are also important. Moreover, the software should be reliable in time-sensitive scenarios, that is, it should respond quickly in situations requiring near-instantaneous response times.

In the article, we discussed only the AVs that are used for transportation and delivery purposes. However, a broader definition of AVs includes any self-controlled machine capable of locomotion, such as autonomous, unmanned aerial vehicles (also known as *drones*). These drones might be used for logistics or combat purposes. Similarly, in the future, we can expect the introduction of autonomous, weaponized military vehicles. Such vehicles may also experience ethical dilemmas that may not be in line with the aforementioned discussions. Instead, a more rigorous and rough code of conduct needs to be established by AV software developers after thorough discussions with the military personal desiring to deploy such vehicles in their armed forces.

Similarly, emergency vehicles such as ambulances need to take a different approach when it comes to ethics. Reaching their destinations on time (meeting the deadline) should be the utmost priority for such vehicles, unlike the cases presented in this article. In fact, the AV discussed previously must give way to these emergency AVs. For emergency vehicles, therefore, a firm RTOS that balances the time and safety constraint will have to be created. Hence, a rigorous study and a deeper understanding of such machine ethics will become a new branch of ethical software development.

REFERENCES

1. J. Delcker, "The man who invented the self-driving car (in 1986)," Politico, July 2018. https://www.politico.eu/article/delf-driving-car-born-1986-ernst-dickmanns-mercedes/ (accessed Mar. 20, 2021).
2. J.-P. Skeete, "Level 5 autonomy: The new face of disruption in road transport," *Technol. Forecasting Soc. Change*, vol. 134, pp. 22–34, Sept. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0040162517314737, doi: 10.1016/j.techfore.2018.05.003.
3. W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*. London: Oxford Univ. Press, 2008.
4. J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science*, vol. 352, no. 6293, pp. 1573–1576, 2016. doi: 10.1126/science.aaf2654.
5. P. Lin, "Why ethics matters for autonomous cars," in *Autonomous driving*, M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, Eds. Berlin, Heidelberg: Springer-Verlag, 2016, pp. 69–85. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2F978-3-662-48847-8.pdf
6. J.-F. Bonnefon et al., "Ethics of connected and automated vehicles: Recommendations on road safety, privacy, fairness, explainability and responsibility," European Commission, Brussels, Belgium, 2020. [Online]. Available: https://ec.europa.eu/cip/contact/index_en.htm
7. C. I. Hovland and R. R. Sears, "Experiments on motor conflict. I. Types of conflict and their modes of resolution," *J. Experimental Psychol.*, vol. 23, no. 5, p. 477, 1938. doi: 10.1037/h0054758.
8. R. Martin, I. Kusev, A. J. Cooke, V. Baranova, P. Van Schaik, and P. Kusev, "Commentary: The social dilemma of autonomous vehicles," *Front. Psychol.*, vol. 8, p. 808, 2017. doi: 10.3389/fpsyg.2017.00808.
9. E. Awad et al., "The moral machine experiment," *Nature*, vol. 563, no. 7729, pp. 59–64, 2018. doi: 10.1038/s41586-018-0637-6.
10. T. Fraichard, "Will the driver seat ever be empty?" INRIA, Le Chesnay Cedex, France, Research Rep. RR-8493, Mar. 2014. [Online]. Available: https://hal.inria.fr/hal-00965176
11. F. Rosen, *Classical Utilitarianism from Hume to Mill*. Evanston, IL: Routledge, 2005.
12. Centre for Connected and Autonomous Vehicles, U.K., *Ensuring Safety and Security; Innovation is great: Connected and automated vehicles in the* U.K.: *2020 Information Booklet*. Nuneaton, U.K.: HM Government, Oct. 2020. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/929352/innovation-is-great-connected-and-automated-vehicles-booklet.pdf
13. G. Meyer and S. Beiker, *Road Vehicle automation, Human Factors and Challenges*, 3rd ed. Berlin: Springer-Verlag, 2019, vol. 201955.
14. C. Luetge, "The German ethics code for automated and connected driving," *Philosophy Technol.*, vol. 30, no. 4, pp. 547–558, 2017. doi: 10.1007/s13347-017-0284-0.
15. E. Lorenz, "The butterfly effect," *World Sci. Ser. Nonlinear Sci. Ser. A*, vol. 39, pp. 91–94, 2000. [Online]. Available: https://books.google.co.in/books?hl=en&lr=&id=olJqDQAAQBAJ&oi=fnd&pg=PA91&dq=E.+Lorenz,+%E2%80%9CThe+butterfly+effect,%E2%80%9D+World+Sci.+Ser.+Nonlinear+Sci.+Ser.+A,+vol.+39,+pp.+91%E2%80%9394,+2000&ots=_3zRhAX8ag&sig=uID4Sdcl0sp2SoiPhbwAV-QCW_A&redir_esc=y#v=onepage&q&f=false
16. P. Lin, "The robot car of tomorrow may just be programmed to hit you," Wired, June 2014. [Online]. Available: https://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/
17. D. Neil, "Could self-driving cars spell the end of ownership," *Wall Street J.*, 2015. [Online]. Available: https://www.wsj.com/articles/could-self-driving-cars-spell-the-end-of-ownership-1448986572
18. IEEE Policies, "Section 7—Professional Activities (Part A—IEEE Policies), 7.8 IEEE Code of Ethics. [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html
19. "Code of ethics," IEEE-CS/ACM Joint Task Force on Software Engineering Ethics and Professional Practices, Washington, D.C., 1999. [Online]. Available: https://www.computer.org/education/code-of-ethics

An Online Multistep-Forward Voltage-Prediction Approach Based on an LSTM-TD Model and KF Algorithm

Ye Ni, Zhilong Xia, Chunrong Fang, and Zhenyu Chen, Nanjing University

Fangtong Zhao, The University of Akron

We propose a multistep-forward voltage-prediction approach combining a long short-term memory time-distributed model and the Kalman filter algorithm to improve prediction efficiency and reduce the demand for computing capability.

Electric vehicles (EVs) based on lithium-ion batteries (LIBs) are gradually expanding their market share due to the benefits they provide in reducing oil consumption and gas emissions. Among the complex EV systems, the battery management system (BMS)[1] is the most critical component, achieving core functions such as state monitoring[2] and analysis, safety protection, and energy control.

The prediction of battery-state parameters is an important part of quality assurance in BMSs. Previous studies have shown that the precise prediction or estimation of battery-state parameters can help prevent battery defects and reduce driving risks.

The industry generally uses traditional coulomb counting[3] or voltage state of charge lookup tables[4] to realize battery-state prediction. However, the mathematical integration method easily accumulates arithmetic errors, and the method based on the lookup table is often limited by the charging environment. The approach based[5] on the equivalent circuit model (ECM) can accurately simulate the structure of the battery so it avoids the influence of environmental factors. Moreover, a recent study[6] has proposed an effective method of estimating the state of battery health

Digital Object Identifier 10.1109/MC.2021.3070314
Date of current version: 30 July 2021

based on the improved, unscented Kalman filter (KF) and ECM. The ECM is very useful in verifying the accuracy of a voltage-prediction approach. However, model-based approaches also bring the problem of being too idealistic, which ignores onboard environmental factors.

Data-driven prediction approaches, such as estimation models based on support vector machines[7] and neural networks (NNs),[8] have been proposed. These approaches can achieve high prediction performance by historical state data without the physical and chemical properties of the LIBs, such as charging and discharging characteristics. In recent years, the hybrid approach of NN and fusion algorithms[9] have shown high precision. However, due to the insufficient computing power of onboard chips, many approaches, especially NN-based prediction, have challenges at deployment time. The time and space efficiency of artificial intelligence models restricts their application.[10]

NN-based approaches usually introduce the long short-term memory (LSTM) model to predict voltage. However, the conventional LSTM model only outputs one voltage at a time, so it has to use the current prediction as the next input and run multiple times to achieve the multistep-forward prediction features. Not only can a BMS increase the efficiency of the prediction, but it can also predict the state of the LIBs for a period when supporting multistep prediction.

In this article, we propose an LSTM time-distributed (LSTM-TD)-based approach to predict the terminal voltages of LIBs in the charging process and improve the efficiency of the predictions. The LSTM-TD model replaces the last output layer of the conventional LSTM model with a TD layer and can output multiple voltages each time. We further use the KF algorithm to smooth the output voltages of the LSTM-TD model.

Briefly, the contributions of our work can be summarized as follows.

1. We realize multipoint outputs in voltage prediction for the first time by constructing an LSTM-TD model, which improves the efficiency of the prediction process.
2. We introduce the KF algorithm to smooth the predicted voltages of the LSTM-TD model and optimize the worst voltage in a set of predicted values of each prediction round. Through the KF algorithm, we improve the prediction accuracy.
3. We design a concrete approach that combines the LSTM-TD model and KF algorithm. Then, we conduct a series of experiments to verify it. The results show that our approach improves performance in computational resource consumption.

METHODOLOGY

Voltage prediction is closely related to the historical process of LIBs, so it is a time-step type of prediction. The LSTM-TD model is based on the LSTM layer, which has very good prediction performance based on time steps. The LSTM-TD model can be seen as a complex nonlinear function, represented by Θ. The input domain of the LSTM-TD model is composed of the feature vector of LIBs, and the output domain is the trend of voltage during the charging process of LIBs

$$\mathbf{X} \rightarrow \Theta \rightarrow \{H_1, H_2, \ldots, H_5\} \tag{1}$$

where $\mathbf{X}$ represents the input feature of LIBs, and $H_{\{1\ldots5\}}$ represents the future trend of voltage in the charging process.

The LSTM-TD model can output multivoltages in one prediction round; thus, the LSTM-TD model outputs the voltage trend of the future. We call this feature where one model can predict future trends the *multistep-forward prediction*. Through the multistep-forward feature, the efficiency of voltage prediction can be greatly improved. To further increase the accuracy of the LSTM-TD model, we introduce the KF algorithm. The architecture of our approach is illustrated in Figure 1(a).

LSTM-TD model

The LSTM-TD model consists of two LSTM layers and a TD layer. The architecture of the LSTM-TD model is presented in Figure 1(c). We use two LSTM layers with 100 and 50 nodes, referencing Yang et al.,[11] which uses three LSTM layers, each with 50 nodes, to learn the nonlinear relationship between the LIBs' charging voltage and previous states. Through the TD layer, the LSTM-TD model can output intermediate results of each time step and realize the multistep-forward feature.

The LSTM layer is displayed in Figure 1(b), and it consists of three gates: a forget, input, and output gate. When the data continuously enter, the LSTM layer maintains a cell state to store historical information, represented by $\mathbf{C_t}$. The LSTM model has two inputs. The first input is the present LIBs' state (for example, current) represented by $\mathbf{X_t}$, and the second input is the output of the last series, indicated by $\mathbf{H_{t-1}}$. The LSTM model combines the previous time step $\mathbf{H_{t-1}}$ with present input X_t as a real input vector $[\mathbf{H_{t-1}}, \mathbf{X_t}]$ at time t. Then, the combined input vector goes through three gates to calculate output $\mathbf{H_t}$ and update cell state to $\mathbf{C_t}$.

The essential formulas are expressed as follows:

$$\mathbf{f_t} = \sigma(\mathbf{W_f} \cdot [\mathbf{H_{t-1}}, \mathbf{X_t}] + \mathbf{b_f}), \tag{2}$$

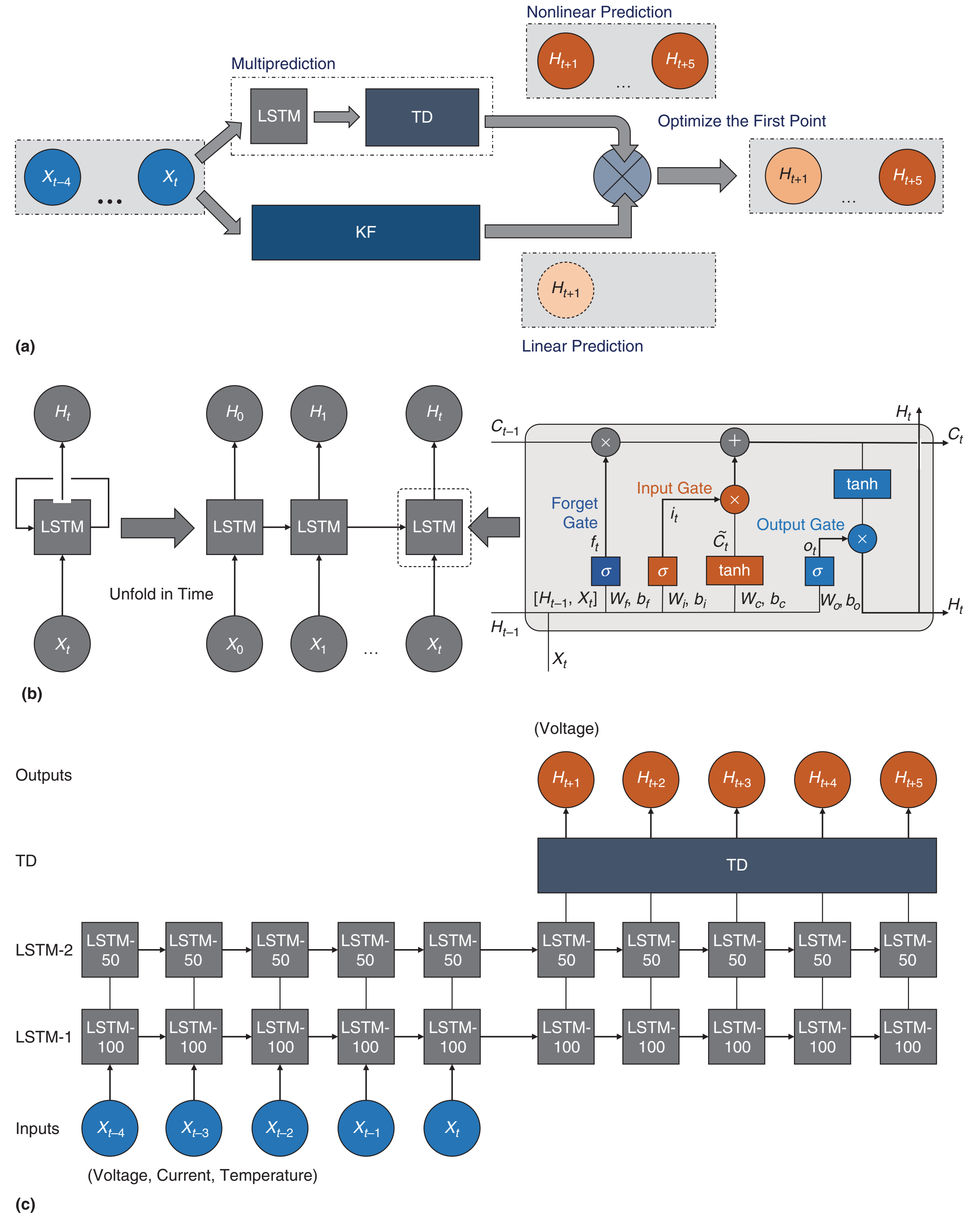


FIGURE 1. The architectures of (a) the multistep-forward prediction approach, (b) the LSTM layer, and (c) the LSTM-TD mode. (*continued*)

$$\mathbf{i_t} = \sigma(\mathbf{W_i} \cdot [\mathbf{H_{t-1}}, \mathbf{X_t}] + \mathbf{b_i}), \quad (3)$$

$$\tilde{\mathbf{C}}_\mathbf{t} = \tan h(\mathbf{W_c} \cdot [\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_c}), \quad (4)$$

$$\mathbf{C_t} = \mathbf{f_t} * \mathbf{C_{t-1}} + \mathbf{i_t} * \tilde{\mathbf{C}}_\mathbf{t}, \quad (5)$$

$$\mathbf{o_t} = \sigma(\mathbf{W_o} \cdot [\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_o}), \quad (6)$$

$$\mathbf{h_t} = \mathbf{o_t} * \tan h(\mathbf{C_t}), \quad (7)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (8)$$

$$\tan h(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}, \quad (9)$$

where $\mathbf{W}_{\{f,i,c,o\}}$ and $\mathbf{b}_{\{f,i,c,o\}}$ are parameters of the LSTM layer, representing the weight and bias matrix, respectively. $\mathbf{f_t}$ represents the forget gate and determines how much information in the cell state needs to be forgotten, and $\mathbf{i_t}$ represents the input gate and consists of two parts: the first part decides which information to update, and the second part goes through a tan$h$ function to get the candidate cell state $\tilde{\mathbf{C}}_\mathbf{t}$. The output gate o_t combines the updated cell state and input information to calculate output results. As displayed in Figure 1, we also implemented a comparative voltage-prediction approach based on a conventional LSTM model [Figure 1(d)].

KF

The KF algorithm can be split into two parts: prediction and update. The prediction equation is

$$\hat{\mathbf{X}}_k = \mathbf{F}\hat{\mathbf{X}}'_{k-1}, \quad (10)$$

$$\mathbf{P_k} = \mathbf{F}\,\hat{\mathbf{P}}'_{k-1}\mathbf{F}^\mathbf{T} + \mathbf{Q}, \quad (11)$$

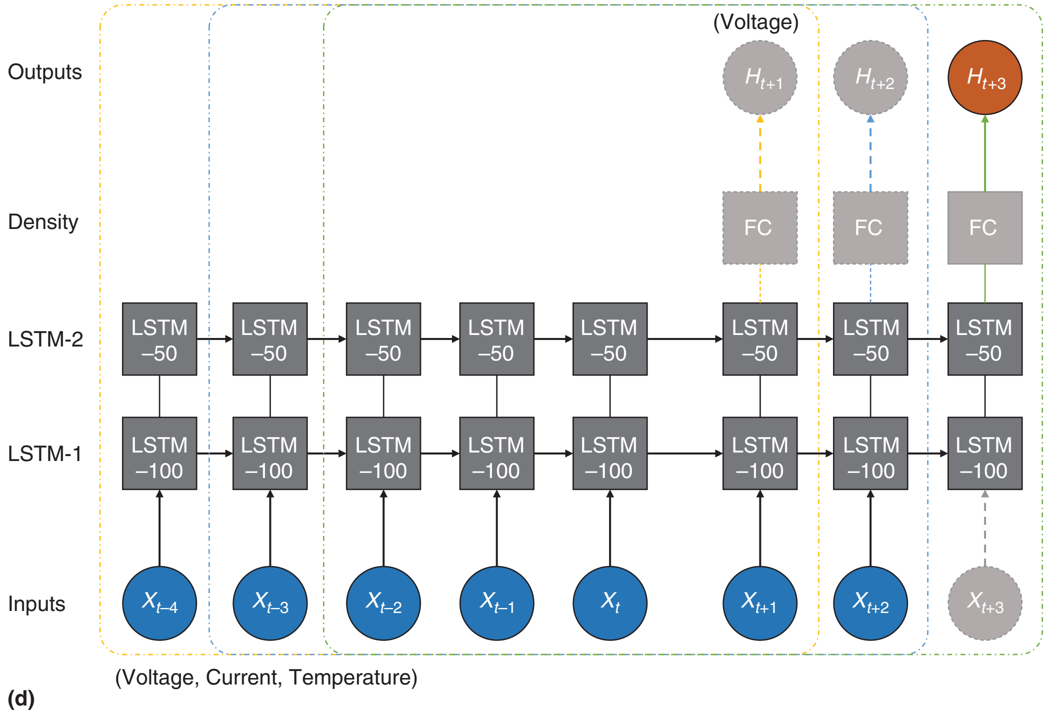
where $\hat{\mathbf{X}}_k$ and $\hat{\mathbf{X}}_{k-1}$ are the prediction state of the system at time k and the optimized state at time $k-1$, respectively, $\mathbf{F}$ is the state-transition matrix that predicts the current state based on the previous state, and $\mathbf{P}$ and $\mathbf{Q}$ are covariance matrices that present prediction error and system noise, respectively.

$$\mathbf{K} = \mathbf{P_k}\,\mathbf{H^T}(\mathbf{HP_k}\,\mathbf{H^T} + \mathbf{R})^{-1}, \quad (12)$$

$$\mathbf{P}'_k = \mathbf{P_k} - \mathbf{KHP_k}, \quad (13)$$

$$\hat{\mathbf{X}}'_k = \hat{\mathbf{X}}_k + \mathbf{K}(\mathbf{Z_k} - \mathbf{H}\hat{\mathbf{X}}_k), \quad (14)$$

where $\mathbf{H}$ indicates the mapping matrix whose mission is mapping the state vector to the observation values, $\mathbf{R}$ indicates observation noise and is the variance of a standard normal distribution, $\mathbf{K}$ is the Kalman gain, $\mathbf{Z_k}$ is the observation vector at time k, and $\hat{\mathbf{X}}'_k$ and $\hat{\mathbf{P}}'_k$ are the optimized state vector and final prediction error covariance matrix at time k, respectively.



(d)

FIGURE 1. (*continued*) The architectures of (d) the conventional LSTM model. FC: fully connected layer.

NASA uses constant current (CC) to charge the battery and then switches to the constant voltage (CV) when the voltage reaches 4.2 V. We assume that the voltage increases at a constant rate during the CC or CV condition. However, with the influences of the internal characteristics of the battery, voltage does not grow at a constant rate. We use the system noise **Q** to make up the deviation between the constant increasing voltage and real voltage.

The trajectory modeling equation is as follows:

$$\begin{cases} v_k = v_{k-1} + s_{k-1} \cdot \Delta t \\ s_k = s_{k-1} \end{cases}, \qquad (15)$$

where s_k and v_k indicate the increasing speed of the voltage and the prediction voltage at time k, respectively, and $\triangle t$ indicates the sampling interval. The parameters in the KF algorithm can be expressed by

$$\begin{cases} \mathrm{F} = \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix}, \\ \mathrm{H} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \\ \mathrm{X}_k = \begin{bmatrix} s_k \\ v_k \end{bmatrix}, \\ \mathrm{Q} = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}, \\ \mathrm{R} = 2. \end{cases} \qquad (16)$$

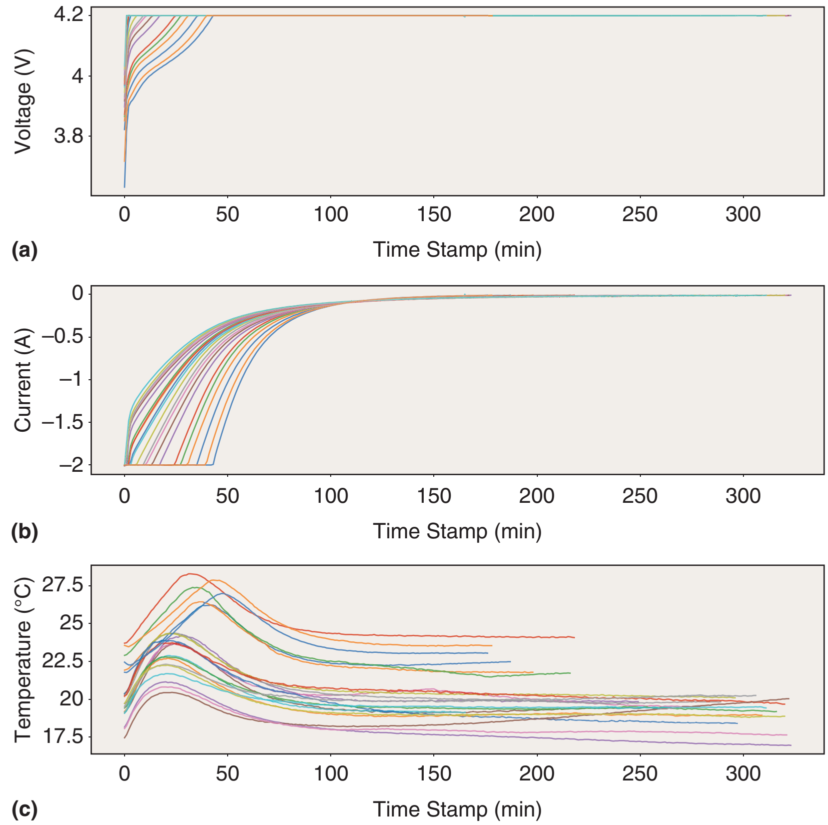


FIGURE 2. The NASA data set reference charging curve (a) charging voltage profile, (b) charging current profile, and (c) charging temperature profile.

Through the following algorithm, we can obtain the optimized prediction voltage:

$$V = V_{\text{LSTM-TD}} * w + V_{KF} * (1 - w), \qquad (17)$$

where $V_{\text{LSTM-TD}}$ indicates the first prediction voltage each round of the LSTM-TD model, and V_{KF} represents the prediction voltage of the KF algorithm. w, which represents the weight of two results, is between zero and one, and we set it to 0.5.

Data set

To simulate the real charging condition, we use the battery data from the data repository of NASA. In the test, the #14 battery goes through the charging and discharging cycles continuously to accelerate the aging process. The data of the charging process are presented in Figure 2, where different colors indicate different charging cycles of the aging process.

The LSTM-TD and conventional LSTM models are trained with the selected data from the $N - 1$ charging cycles, and we use the data of the one remaining cycle to test the prediction efficiency of the two models. The training data of the two models at time t are $\mathbf{X_t} = [V, I, \mathrm{T}]_k$, $k = t - 4, t - 3, \ldots, t$ where V, I, and T indicate voltage, current, and temperature, respectively. Due to the multistep-forward feature of the LSTM-TD model, there are multiple label data in a training process, $\mathbf{Y_t} = [V_{t+1}, \ldots, V_{t+5}]$. On the contrary, there is only one label data, the voltage at $t + 1$, for the conventional LSTM model in a training process, $Y_t = [V_{t+1}]$.

RESULTS AND DISCUSSIONS

The training processes were conducted on a surface book equipped with GTX1060 GPU and i7-8650u CPU.

Cumulative prediction time cost

Figure 3 presents that the LSTM-TD model takes much less time than the conventional LSTM model under the same prediction rounds. The cumulative prediction time increased from 0.16 to 0.9 s for the multistep-forward prediction approach when the prediction voltage numbers increased from 20 to 150. In contrast, the time cost of the conventional LSTM model rose from 0.57 to 3.79 s. Compared with the cumulative prediction time of the conventional LSTM model, the cumulative prediction time of our approach was reduced by 76.3%.

We used NVIDIA GTX 1060, whose computing capability is 6.1 teraflops, to conduct the experiments. We selected two of the most representative chips, EyeQ3 and EyeQ4 from Mobileye, to simulate the real time cost of the prediction. We used the following algorithm to estimate the real time costs of the predictions on EyeQ3 and EyeQ4:

$$t = \frac{C_{\text{gap}} * t_e}{\text{rounds}},$$

where t indicates the time required for the voltage prediction per round; "rounds" is the number of prediction rounds (that is, the ratio of the total number of predicted voltages to the number of predicted voltages in each round), t_e indicates the cumulative time cost of the prediction under the experimental environment, and C_{gap} represents the computing-power difference between the onboard and experimental environments.

The computing capabilities of the EyeQ3 and EyeQ4 are 0.256 and 2.5 tera operations/s, respectively. Table 1 displays the actual time costs of online voltage predictions for vehicles equipped with EyeQ3 and EyeQ4 chips.

From the prediction curves in Figure 4(h) it can be seen that the KF algorithm improves the prediction accuracy of the LSTM-TD model and hardly increases the overall running time. For optimizing 30 rounds, the KF algorithm only consumes 0.02 s. In Table 1, we use the ϵ symbol to represent the KF algorithm's time cost onboard, which is extremely small.

According to Table 1, it is obvious that the conventional LSTM model needs 3.03 s to finish the prediction. Meanwhile, the multistep-forward prediction approach only needs about $0.70 + \epsilon$ s.

BMS samples the characteristics of LIBs every second to observe the changes

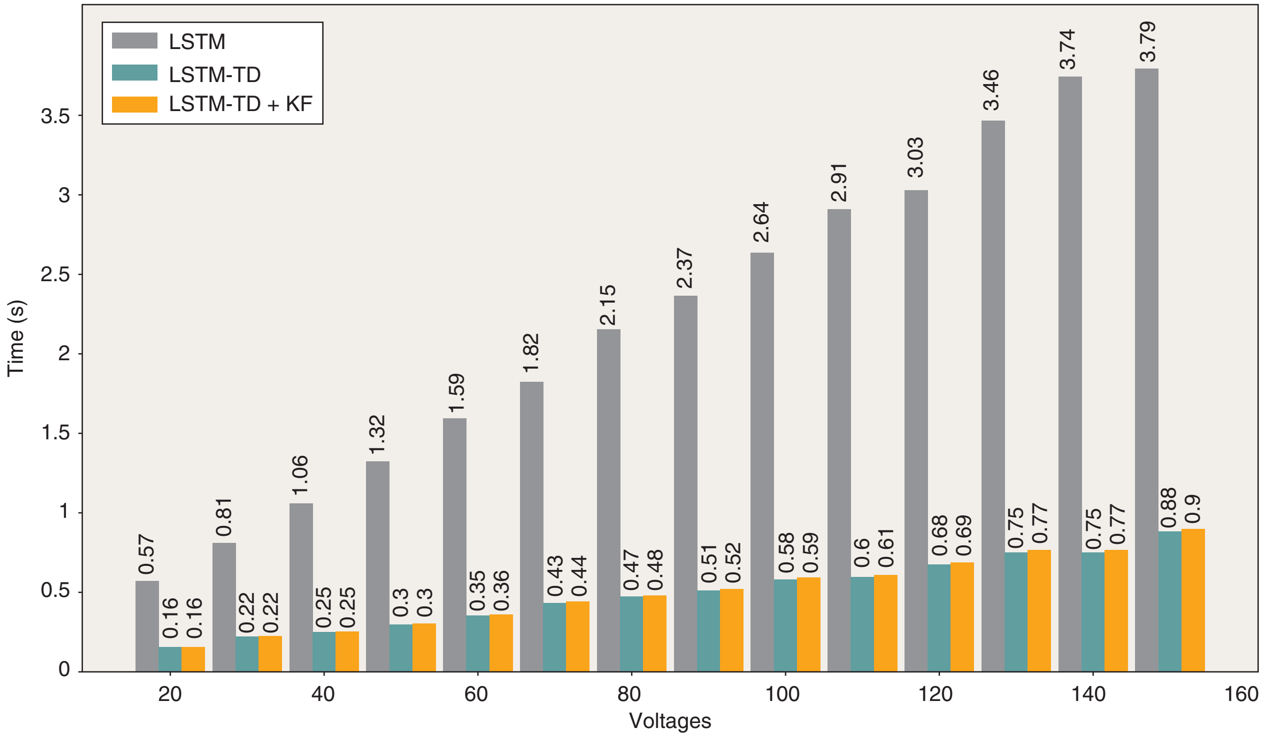


FIGURE 3. The time costs of the three approaches.

TABLE 1. The voltage-prediction time cost per round.

Model	Chip	C_{gap}	t_e	Rounds	t
LSTM	EyeQ3	24	3.79 s	30	3.03 s
LSTM	EyeQ4	2.5	3.79 s	30	0.32 s
LSTM-TD	EyeQ3	24	0.88 s	30	0.70 s
LSTM-TD	EyeQ4	2.5	0.88 s	30	0.07 s
LSTM-TD+KF[1]	EyeQ3	24 + a[2]	0.90 s	30	0.70 s + ϵ[3]
LSTM-TD+KF[1]	EyeQ4	2.5+a[2]	0.90s	30	0.07 s + ϵ[3]

[1] Combine the LSTM-TD model and KF algorithm.
[2] The equivalent computing-power ratio between CPUs.
[3] The onboard time cost of the KF algorithm.

for precise monitoring. In this case, the prediction had to be made within 1 s when the battery-sampling interval was 1 s; otherwise, the prediction would have been useless. Therefore, our approach can meet the demands for vehicles equipped with EyeQ3 chips, but the conventional LSTM model cannot.

Prediction accuracy and optimization

We used the mean absolute percentage error (MAPE) and the largest absolute percentage error (LAPE) to evaluate the accuracy of the models. The evaluation of the conventional LSTM and LSTM-TD models is illustrated in Table 2. The MAPEs of the conventional LSTM and LSTM-TD models were both less than 1%. However, the LAPE of the LSTM-TD model (1.94%) was slightly higher compared with that of the conventional LSTM model (0.23%). In the LSTM-TD model, only the last (fifth) predicted voltage in each prediction round used the complete input vectors while the first voltage only used the first of the five input vectors. Therefore, the first prediction voltage of each round had the largest deviation and caused the protruding points in Figure 4(b).

We further extracted the predicted voltages and true values of different series in each round to form a new comparison curve, and the results are displayed in Figure 4(c)–(g). In Figure 4(c), we pulled every first prediction voltage of each round and compared them with the corresponding voltages. In Figure 4(d)–(g), we extracted the second, third, fourth, and fifth prediction voltages and the true values in each round, respectively. In Figure 4, it can be seen that (c) has the largest deviation when compared with the five curves from (c) to (g). The MAPEs of different series of voltages in each round are presented in Table 3, which illustrates that the maximum MAPE is always the first predicted voltage of each round.

To improve the prediction accuracy, we used the KF algorithm to smooth the prediction curve of the LSTM-TD model by optimizing the first predicted voltage of each round. From Figure 4(a), it can be seen that the prediction curve

TABLE 2. The predictions' absolute percentage errors.

Rounds	LSTM (%)		LSTM-TD (%)	
	MAPE	LAPE	MAPE	LAPE
4	0.12	0.18	0.73	1.94
6	0.10	0.18	0.54	1.94
8	0.11	0.23	0.43	1.94
10	0.09	0.23	0.35	1.94
12	0.08	0.23	0.29	1.94
14	0.07	0.23	0.25	1.94
16	0.07	0.23	0.22	1.94
18	0.05	0.23	0.19	1.94
20	0.05	0.23	0.17	1.94
22	0.04	0.23	0.16	1.94
24	0.04	0.23	0.15	1.94
26	0.04	0.23	0.13	1.94
28	0.04	0.23	0.12	1.94
30	0.03	0.23	0.12	1.94

TABLE 3. The MAPE of the LSTM-TD model.

Rounds	MAPE (%)				
	1st	2nd	3rd	4th	5th
4	1.31	0.64	0.64	0.56	0.49
6	0.96	0.47	0.46	0.42	0.37
8	0.75	0.38	0.38	0.35	0.30
10	0.61	0.30	0.30	0.28	0.24
12	0.50	0.25	0.25	0.24	0.20
14	0.43	0.22	0.22	0.21	0.18
16	0.38	0.19	0.19	0.18	0.15
18	0.34	0.17	0.17	0.16	0.14
20	0.30	0.15	0.15	0.14	0.12
22	0.28	0.14	0.14	0.13	0.11
24	0.25	0.13	0.13	0.12	0.10
26	0.23	0.12	0.12	0.12	0.11
28	0.22	0.11	0.11	0.10	0.09
30	0.20	0.10	0.10	0.09	0.08

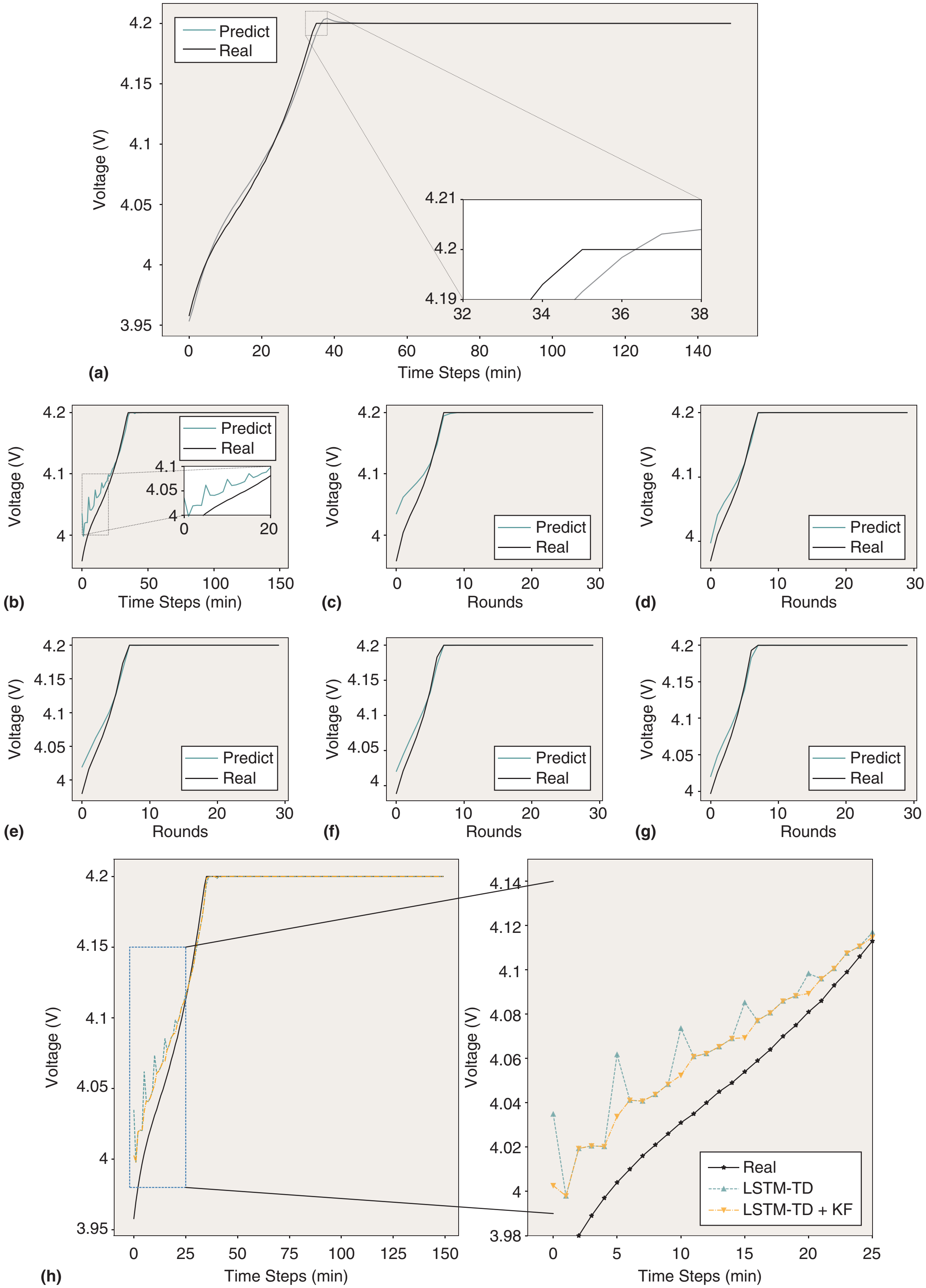


FIGURE 4. (a) The conventional LSTM model and (b) the LSTM-TD model prediction curves. The (c) first, (d) second, (e) third, (f) fourth, and (g) fifth prediction voltages of each round. (h) The KF algorithm-optimization curve.

of the conventional model was very smooth. Thus, we did not combine the conventional LSTM model with the KF algorithm.

The results optimized by the KF algorithm are displayed in Figure 4(h). It can be found that the protruding points in the prediction curve of the LSTM-TD model were eliminated. The LAPE of the LSTM-TD model was reduced from the original 1.94% to around 1.12%.

Limitations

Although our approach has a good performance in terms of efficiency improvement, there exist some potential threats during deployment. For example, a reliable optimization algorithm needs to be selected due to the characteristics of the LSTM-TD model. In our approach, we chose the KF algorithm to optimize the voltages. However, when the battery-sampling interval of BMSs is large and the voltage curve shows a nonlinear trend, the KF algorithm may not perform well.

In this article, we proposed an online multistep-forward voltage-prediction approach that consists of an LSTM-TD model and KF algorithm. The LSTM-TD model is responsible for voltage-trend prediction and efficiency improvement. For the first time, multi-output prediction technology has been used in the field of voltage prediction and has improved the performance of online voltage prediction. Moreover, our approach incorporates the KF algorithm to smooth the prediction curve of the LSTM-TD model, maintaining prediction accuracy. Compared with the conventional LSTM model, the experiments show that the prediction time of voltage prediction is shorter by 76.3%; it was reduced from 3.79 to 0.9 s. Overall, our approach greatly improves performance and expands the deployment scope of online voltage prediction.

ABOUT THE AUTHORS

YE NI is currently pursuing a master's degree in software engineering at Nanjing University, Nanjing, Jiangsu, 210093, P.R. China. His research interests include time-series prediction and the improvement of system efficiency. Ni received a bachelor's degree in electronic and information engineering from Nanjing Institute of Technology. Contact him at niyelinux@163.com.

ZHILONG XIA is currently pursuing a master's degree in software engineering at Nanjing University, Nanjing, Jiangsu, 210093, P.R. China. His research interests include autonomous driving system simulation. Xia received a bachelor's degree in software engineering from Jilin University. Contact him at epylice@smail.nju.edu.cn.

CHUNRONG FANG is a research assistant at the Software Institute at Nanjing University, Nanjing, Jiangsu, 210093, P.R. China. His research interests include BigCode and artificial intelligence testing. Contact him at fangchunrong@nju.edu.cn.

ZHENYU CHEN is a full professor at the Software Institute at Nanjing University, Nanjing, Jiangsu, 210093, P.R. China. His research interests include collective intelligence, deep learning testing and optimization, big data quality, and mobile application. Contact him at zychen@nju.edu.cn.

FANGTONG ZHAO is currently pursuing an M.S. at the University of Akron, Akron, Ohio, 44325, USA. Her interests include the solid-state composite electrolyte for lithium-ion batteries and 3D-printing fabrication. Contact her at fz12@zips.uakron.edu.

ACKNOWLEDGMENTS

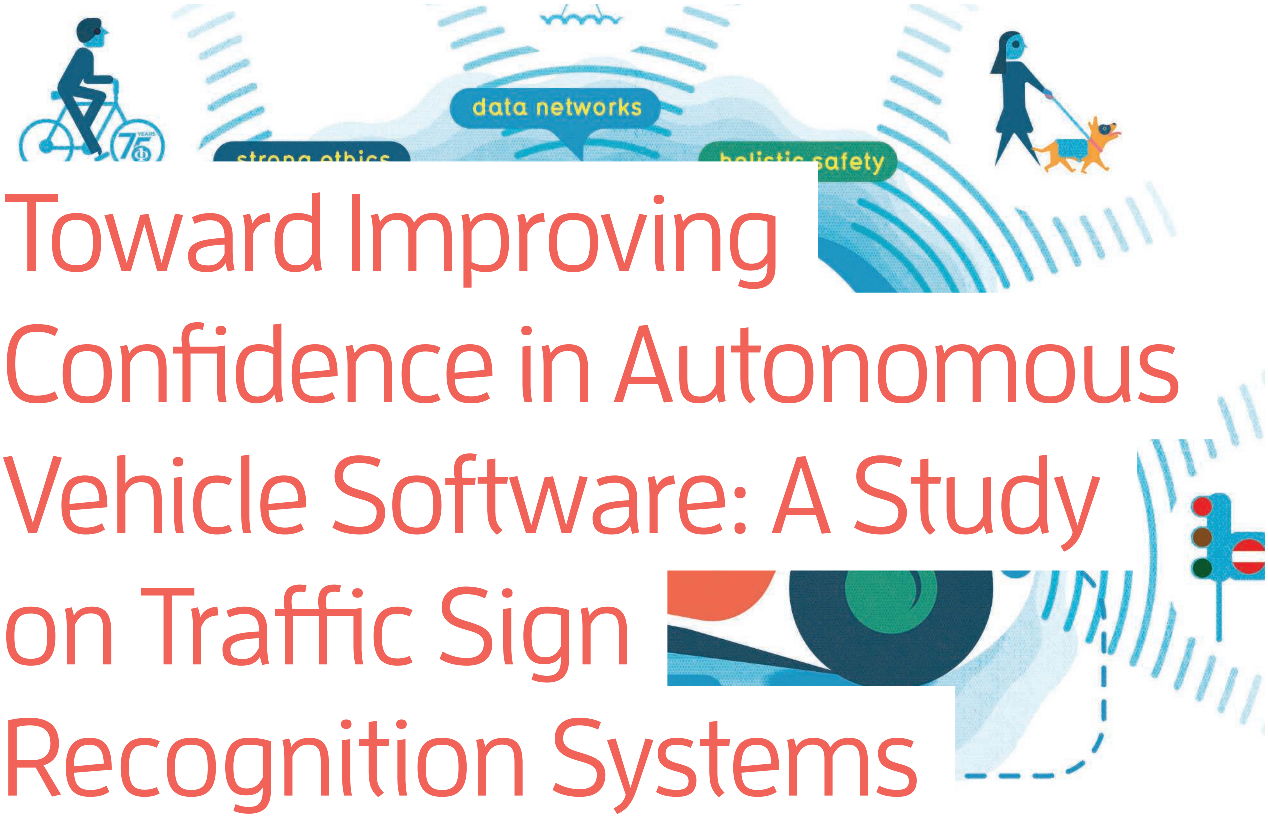
We would like to thank anonymous reviewers for insightful comments. We also thank the BAIC BJEV company for useful discussion and support during our experiments. This work is partially supported by the National Natural Science Foundation of China (Grant 61932012).

REFERENCES

1. K. W. E. Cheng, B. Divakar, H. Wu, K. Ding, and H. F. Ho, "Battery-management system (BMS) and soc development for electrical vehicles," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 76–88, 2010. doi: 10.1109/TVT.2010.2089647.
2. X. Hao, "Health status comparisons of lithium-ion batteries when fusing various features," *Int. J. Performability Eng.*, vol. 15, no. 1, p. 138, 2019. doi: 10.23940/ijpe.19.01.p14.138145.
3. J. Yan, G. Xu, H. Qian, and Y. Xu, "Robust state of charge estimation for hybrid electric vehicles: Framework and algorithms," *Energies*, vol. 3, no. 10, pp. 1654–1672, 2010. doi: 10.3390/en3101654.

4. V. Pop, H. J. Bergveld, D. Danilov, P. Notten, and P. P. Regtien, “State-of-the-art of battery state-of-charge determination,” *Meas. Sci. Technol.*, vol. 16, no. 12, p. R93, 2005.
5. J. Yang, B. Xia, W. Huang, Y. Fu, and C. Mi, “Online state-of-health estimation for lithium-ion batteries using constant-voltage charging current analysis,” *Appl. Energy*, vol. 212, pp. 1589–1600, Feb. 2018. doi: 10.1016/j.apenergy.2018.01.010.
6. H. Wang, “Estimation of battery health based on improved unscented Kalman filtering algorithm,” *Int. J. Performability Eng.*, vol. 15, no. 5, pp. 1482–1490, 2019. doi: 10.23940/ijpe.19.05.p25.14821490.
7. T. Hansen and C.-J. Wang, “Support vector based battery state of charge estimator,” *J. Power Sources*, vol. 141, no. 2, pp. 351–358, 2005. doi: 10.1016/j.jpowsour.2004.09.020.
8. E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, “Long short-term memory networks for accurate state-of-charge estimation of li-ion batteries,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6730–6739, 2017. doi: 10.1109/TIE.2017.2787586.
9. Y. Tian, R. Lai, X. Li, L. Xiang, and J. Tian, “A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter,” *Appl. Energy*, vol. 265, p. 114,789, May 2020. doi: 10.1016/j.apenergy.2020.114789.
10. T. Hao et al., “Edge AIBench: Towards comprehensive end-to-end edge computing benchmarking,” in *Proc. Int. Symp. Benchmarking, Measuring Optimization*, Springer-Verlag, 2018, pp. 23–30. doi: 10.1007/978-3-030-32813-9_3.
11. F. Yang, X. Song, F. Xu, and K. Tsui, “State-of-charge estimation of lithium-ion batteries via long short-term memory network,” *IEEE Access*, vol. 7, pp. 53,792–53,799, Apr. 2019. doi: 10.1109/ACCESS.2019.2912803.

Digital Object Identifier 10.1109/MC.2021.3094705



Toward Improving Confidence in Autonomous Vehicle Software: A Study on Traffic Sign Recognition Systems

Koorosh Aslansefat, University of Hull
Sohag Kabir, Amr Abdullatif, and Vinod Vasudevan, University of Bradford
Yiannis Papadopoulos, University of Hull

This article proposes an approach named SafeML II, which applies empirical cumulative distribution function-based statistical distance measures in a designed human-in-the-loop procedure to ensure the safety of machine learning-based classifiers in autonomous vehicle software.

The rise of artificial intelligence (AI) and the advancement of technologies have paved the way for autonomous systems such as autonomous vehicles to enter our everyday life. Such systems have the potential to make an enormous societal and economic impact. For instance, as mentioned in the Waymo safety report,[1] when human drivers are involved in driving, around 1.35 million lives have been lost due to traffic crashes worldwide in 2016 and US$836 billion have been lost annually due to loss of lives and injuries caused by crashes. For each person, there is a 67% chance of getting involved in drunk-driving crashes. In the United States, 94% of crashes involve human choice or error. Therefore, dependable and reliable autonomous vehicles can help to save lives and decrease economic losses by reducing the number of traffic crashes by eliminating human involvement in driving.

Autonomous vehicles are increasingly given autonomous decision-making power such that, while performing safety-critical tasks within human vicinity, they can

Digital Object Identifier 10.1109/MC.2021.3075054
Date of current version: 30 July 2021

autonomously make their own decisions and take actions with minimal human intervention. To be able to do so, an autonomous vehicle has to cooperate with other vehicles, roadside infrastructures (for example, traffic signs), smart traffic light systems, and so on. Consequently, using AI and machine learning (ML), such systems continuously learn from their operation and dynamically reconfigure in response to changes such as unexpected failures of components/subsystems, the continuous change in the context of operation, variable workloads, and physical infrastructures. A key challenge for software-intensive, AI-enabled self-adaptive autonomous systems is to provide assurance about their safety and reliability.

For traditional nonautonomous systems, assurance is provided through design and development activities including verification, validation, testing, conformance to standards, and certification. Safety assurances are often provided through safety arguments where safety goals are defined, and rationales for believing in these goals are designed to be dependent on a variety of assumptions. These assumptions may include aspects like failure semantics and failure rates of both hardware and software components, operating context, the efficiency of the human operator to respond to events, and so on.[2] In operation, the physical system and its operating environments are monitored to see if any of these safety assumptions are violated and thereby notify the users about the potential changes in the assurance and take necessary actions to achieve fail-safe behavior. For example, in a car, when transient errors in hardware-like sensors affect the functionality of the software like for cruise control, an error detection unit (monitoring function) can detect the error and degrade the system by appropriate warnings and allowing the driver to take over. Therefore, the integrity of the monitoring knowledge plays a crucial role in providing accurate runtime assurances.

The issue of continuous assurance provision is further complicated for autonomous systems where important pieces of evidence are collected through ML/AI components. Due to the black-box nature of these components, the confidence in the evidence provided by these components will directly affect the confidence in the overall assurance. For instance, consider the ML-based traffic sign recognition (TSR) system in an autonomous car, which is responsible for identifying different traffic signs and thus assisting in assuring safe driving. TSR for autonomous vehicles has several shortcomings; a survey of such shortcomings is available in Magnussen et al.[3] Therefore, it is likely that, in some cases, evidence/inputs received from a TSR could be misleading. If this misleading information is considered while providing safety assurance, then it is highly likely that, a false assurance could be provided, resulting in an autonomous vehicle driving with false assurance. In a worst-case scenario, this could lead to a catastrophic accident. Therefore, it is important to improve the confidence in the output generated by such software components in autonomous vehicles.

To address this issue with autonomous vehicle software, TSR in particular, in this article we have proposed a novel approach called "SafeML II," which has the following features:

- It ensures the safety of an ML-based TSR system using modified state-of-the-art empirical statistical distance measures and can work with a variety of distribution functions, especially exponential families.
- The implemented bootstrap *p*-value calculation in the SafeML II functions improves the accuracy and validity of its results.
- It utilizes a human-in-the-loop procedure that can use human intelligence and avoid catastrophic accidents.
- It is a model-agnostic approach that works with a variety of ML and deep learning classifiers.

The effectiveness of the approach is illustrated via an application to the real-world German Traffic Sign Recognition Benchmark (GTSRB) data set.

SAFETY ASSURANCE CHALLENGES OF AI/ML IN AUTOMOTIVE DOMAIN

In 2011, the International Organization for Standardization (ISO) proposed the ISO 26262 standard to regulate functional safety for road vehicles. It includes requirements and recommendations for the entire lifecycle of car manufacturing from the concept phase to operation and service. The main aim of ISO 26262 was to help the automotive industry address functional safety issues more systematically. However, it was defined without considering ML since the first version of ISO 26262 was published before the boom of AI. This eventually leads to a challenging issue today for car manufacturers and suppliers who are determined to incorporate ML for self-driving cars. Therefore, conventional safety assurance methods suggested by the ISO 26262 standard are insufficient or inapplicable for the assurance of ML.[4] Salay et al.[5] presented an analysis of ISO-26262 Part 6 methods with respect to the safety of ML models. Their assessment of the applicability of the software safety methods on ML algorithms (as software unit design) shows about 40% of software safety methods do not apply to ML models.

The AI community has recently produced several papers on the problems of "AI safety."[6] One of the more influential papers[7] identifies "concrete problems in AI" and, according to this paper, AI safety issues for autonomous vehicles can be categorized in five domains, including 1) safe exploration, 2) scalable oversight, 3) avoiding "reward hacking" and "wire heading," 4) avoiding negative side effects, and 5) robustness to distributional shift. Efforts have been made to assure safety and improve the safety performance of ML components in autonomous vehicles. For instance, the safety assurance process for ML models in safety-critical applications has been described, focusing on an explicit definition of safety requirements for ML components with respect to the safety requirements of the overall system.[8] The approach has been illustrated via an application to a pedestrian detection system in autonomous cars.

In 2019, Kläs et al.[9] emphasized the distributional shift in the data set and proposed an uncertainty wrapper based on the Wilson method for calculating confidence intervals. The conceptual idea was explained for the GTSRB example without reporting any experimental or numerical results. In other research, they improved their previous approach considering the impact of additional inputs like rain amount, wind direction, wind speed, and vehicle orientation on the confidence results specifically for TSR.[10,11] A year later, they proposed a framework for generating an uncertainty wrapper for data-processing models and their dataflow.[12] In all three research works, the drawback was the lack of a designed safety mechanism after measuring confidence. In the SafeML approach,[13] three different scenarios, including 1) repeating the measurement or requesting additional data, 2) providing a human-in-the-loop procedure, and 3) trusting the ML decisions and providing a confidence report, are considered based on the empirical cumulative distribution function (ECDF)-based statistical distance measures. The SafeML approach was not able to work with images particularly for the convolutional neural network (CNN)-based classifiers and more importantly the lack of consideration of *p* values of statistical distance measures in the procedure could lead to a wrong decision. In other words, there are some cases where a statistical distance exists, but based on an invalid associated *p* value it should not be considered for the confidence evaluation. In SafeML II, the ECDF-based statistical distance measure functions have been improved with a bootstrap-based *p*-value evaluation. It means that in the confidence evaluation of SafeML II, only the measured statistical distance value with a valid *p* value will be considered and the others will be dropped from the list. Moreover, by converting the images to flatten vectors, SafeML II is able to do a pixel-wise ECDF-based statistical distance measure and generate the confidence that will be explained in the next sections.

IT IS ASSUMED THAT THE DATA SET COVERS THE MAJORITY OF SITUATIONS, THE DATA SET LABELING HAS BEEN DONE PERFECTLY, AND THE DATA SET IS RELATIVELY BALANCED.

ML SAFETY APPROACH

In this article, we extend the initial idea of SafeML[13] to propose SafeML II for 1) image-based classification problems and 2) dealing with outliers in data. Figure 1(a) illustrates the flowchart of SafeML II. It has two main phases: the training phase is an offline procedure and the application phase is an online procedure. In the training phase, the procedure starts with loading the trusted data set. It is assumed that the data set covers the majority of situations, the data set labeling has been done perfectly, and the data set is relatively balanced. Having loaded the trusted data set, a classifier will be trained with those data, and its performance will be evaluated

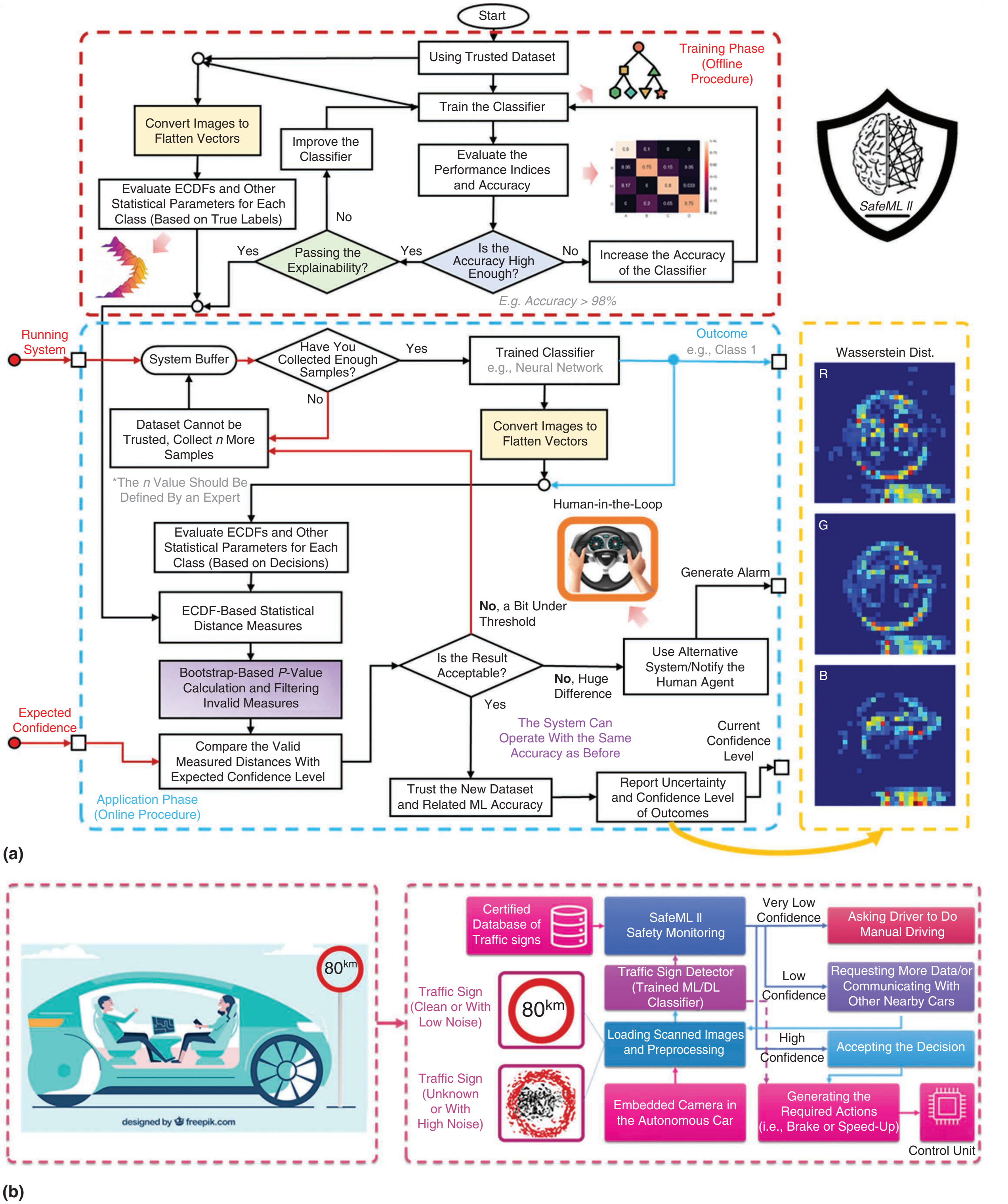


FIGURE 1. SafeML II flowchart and application block diagram. (a) A flowchart of the proposed SafeML II. (b) An example of using SafeML II for autonomous (self-driving) cars and their traffic sign recognition unit.

accordingly. In this part of the procedure, standard methods for cross-validation and explainability should also be considered. If the accuracy of the classifier and its explainability were high enough (for example, more than 95% accuracy), the classifier will be selected and the procedure goes to the next step. Otherwise, other classifiers or even data refinement will be needed to achieve a certain level of accuracy. Having selected the appropriate classifier, the statistical parameters of each feature in each class including cumulative distribution function, mean, and variance will be stored to be used for comparison in the next phase.

THE CONFIDENCE LEVEL WILL BE CALCULATED BASED ON THE AFOREMENTIONED COMPARISON AND WILL BE AGAIN COMPARED WITH THE EXPECTED CONFIDENCE THRESHOLD.

In the application phase, there would be a buffer to collect enough samples. The buffer size should be defined at design time by an expert in a way that the collected data contain the statistical characteristics of their class. Note that these upcoming data are not labeled. Having collected enough samples, the trained and tested classifier in the previous phase will be used, and based on its decisions, the data will be labeled. Based on classifier decision, the statistical parameters of buffered data will be collected and compared with training data set through ECDF-based statistical distance measures such as Kolmogorov–Smirnov (KS), Kuiper (K), Anderson–Darling (AD), Cramer–Von Mises (CVM), and Wasserstein (W).[14] Moreover, in the design time, an expected confidence threshold should be defined for each statistical distance measure. The confidence level will be calculated based on the aforementioned comparison and will be again compared with the expected confidence threshold. Three different scenarios have been considered: 1) when the confidence is a bit lower than the threshold, the system should collect more data; 2) when the confidence has a huge difference in comparison to the predefined threshold, then it is assumed that the upcoming data have not been seen by the classifier before and a human-in-the-loop procedure should be taken into consideration; and 3) when the confidence is higher than the predefined threshold, the results of the classifier will be accepted and a report of the statistical comparison will be stored in the system.

To have a better understanding of the idea, the illustrated example in Figure 1(b) is used. In this example, it is assumed that there is an autonomous vehicle and a specific module in the vehicle software for TSR based on the ML algorithm. The main task for the ML algorithm is to classify the upcoming images from the vehicle's embedded camera(s), and based on a lookup table a required action will be generated to be used in the control unit. It can simply be a brake or acceleration command. The main question is "How one can make sure that the decision is always correct?" The idea of SafeML II can be a solution to this question. As an example, consider there is an 80-km sign in the road, and the vehicle's embedded camera reads it. Most of the time it is expected to be a clear image, but in rare conditions, such as having a faulty camera, heavy rain, fog, or a cyberattack, the image may not be clear. In such rare cases, SafeML II can compare the images with the trusted data set and create confidence. For the very low confidence situation, it means that the input is not something that the trained ML algorithm has seen before and it is better to be handled by the driver (human-in-the-loop procedure). In autonomous vehicles that do not have a wheel to control, such as the Amazon Zoox, it is suggested that a human agent from the control center control the car remotely. It should be noted that the needed reaction time and possibilities to involve the human in the loop can be another research subject to be investigated in the future. When the confidence is low, SafeML II may ask for more data or communicate with surrounding cars and increase the level of confidence. If the confidence is high and the upcoming images are statistically similar to the trusted data set, the decisions can be accepted. Having a high confidence decision, the needed control command will be generated to be sent to the main control unit. All confidence reports should be stored in the system to be used for system improvement.

NUMERICAL RESULTS

In this section, numerical results comparing the proposed approach and existing approaches in the literature are presented for a GTSRB data set.[15]

The data set was released in 2011 and includes 43 different traffic signs. The data set is unbalanced, and the number of samples for some classes can be more than the others. Regarding the cross-validation, the hold-out method is used to split 80% of the data for training and 20% for validation. It should be noted that the data set has a separate folder for test data.

As mentioned earlier, SafeML II is a model-agnostic approach that can be used on top of any ML classifier regardless of its structure. In this article, a deep CNN classifier is used because of its reputation on image classification. The following structure is used as the configuration of CNN. The input has a 2D convolution layer (Conv2D) with a filter size of 32, a kernel size of 5 × 5, and the relu activation function. The second layer has another Conv2D with a filter size of 64, a kernel size of 3 × 3, and the relu activation function. Then, a max pooling layer with a size of 2 × 2 and a dropout layer with a rate of 0.25 is used. After that, another Conv2D layer with a filter size of 64, a kernel size of 3 × 3, and relu activation function is added. A max pooling with a size of 2 × 2 and a dropout with the rate of 0.25 is applied on top of it. A flattened and dense layer with a size of 256 and relu activation function with 0.5% dropout is used. Finally, for the output, a dense layer with the size of 43 and Softmax activation function is considered. Moreover, the Adaptive Moment Estimation (ADAM) optimizer and the cross-entropy loss function are used in the training procedure.

Using this configuration, the performance of the CNN classifier was 0.9797 on the test data set. The next level is to check whether the achieved accuracy is high enough or not. This part was not considered in the first version of SafeML, and it could reduce the precision of the proposed approach when a poor classifier is chosen in the offline phase. In the case of having a poor classifier, the loop should be repeated until reaching a certain level of satisfaction for accuracy. It is also possible to consider explainability approaches to make sure the trained classifier behaves reasonably and focuses on the right part of the image. Assuming that the level of achieved accuracy is acceptable for safety experts, the images of each class will be separated to the red, green, blue (RGB) matrix and converted to the flatten vectors accordingly. As the size of each image is 30 × 30, the equivalent vector will be 1 × 900. The ECDFs of each class will be generated and stored for use in the next phase. In the online phase, the buffer size is considered as 15. In a practical scenario, the buffer size should be defined by safety experts and designers. As there was no real time data, the test data are considered as the upcoming data, and we are going to see how the proposed approach will react to the wrong decisions. To have better visualization, class number three is chosen. This class is related to the 60-km speed limit sign, and it has 1,410 images in the training data set and 450 images in the test data set. Various risks can be considered for misclassification of this sign, such as having a lower speed and blocking the road or having a higher speed and increasing the probability of hitting pedestrians passing the street. The associated risk for the misclassification of each class can be investigated in a separate research study. The accuracy of the classifier for this class specifically was 0.9655. In other words, 435 images are detected correctly but 15 images are detected as the other classes. Based on the SafeML II procedure, the RGB matrix of test images are converted to flatten vectors and their ECDFs have been generated. Furthermore, using ECDF-based statistical distance measures such as KS, K, AD, CVM, and W, the statistical distances will be obtained. The first version of SafeML will jump to a comparison between statistical distance measures and the predefined expected confidence threshold. However, in SafeML II, a bootstrap algorithm with 1,000 iterations is used to obtain the *p* value and validate the measures.[16] Thus, the measures with a *p* value lower than 0.05 are stored and others will be omitted. The validated statistical distance measure can be compared with the expected confidence level. It should be noted that, for each ECDF-based statistical distance measure, there should be a particular expected confidence threshold predefined by a safety expert. The decision of the ML classifier is accepted and trusted if the distance measure is higher than the predefined threshold. Additionally, a report of the statistical distance measure will be stored in a database to be used for the further development of the system. In the situation that the statistical distance measure is 5% lower than the predefined threshold, the system may ask for further data. It should also be mentioned that in that situation, the autonomous vehicle can use other existing sources of information to validate the decision. For example, the autonomous vehicle can communicate with nearby vehicles or use GPS and preloaded map data. The mentioned percentage can also be changed based on the safety experts' and system designers' opinion. At the moment, there is no published standard to define these levels, but, in the future, these parameters can be defined using the published standards. The worst scenario is that the statistical distance measure is hugely different from the expected threshold, meaning the upcoming data has not been seen by the classifier before and there is a risk of missed classification. The SafeML II idea is to put the human in the loop and ask the driver to make the decision. It is assumed that the driver has enough time for making the decision. However, there might be some cases where the time is restricted and SafeML II cannot be used. As mentioned before, the autonomous vehicles that do not have the wheel-based driving capability, it is suggested that a human agent from the control center control the car remotely. The first row of Figure 2 illustrates the Wasserstein distance (WD) measure of the 60-km traffic sign (Class 3) for the RGB part of the images. As can be seen, the middle of the image has more statistical differences in all three color layers. Besides, the blue part of the image has less statistical distance in comparison to the red and green parts of the image. As can be seen, in the first layer, a previous version of SafeML is used, which has a lack of *p*-value-based distance validation while in the second row SafeML II is used that has the embedded *p*-value distance validation. Comparing the first and second row of Figure 2, it is clear that SafeML II has a better statistical distance representation and does not catch the background areas of the signs. The third row of this figure illustrates a sample image where the classifier has correctly detected the sign, while the fourth row shows a sample image where the classifier was not able to detect the sign correctly. However, it seems that it can be detected by a human with careful observation. Therefore, in these cases, the human in the loop can help the system to make the right decision and also learn it to make better decisions in the future. The AI system can be considered as a talented and clever child that needs to work in parallel with human and become mature over time. This figure also demonstrates how the ECDF-based W is calculated for a pixel in the image.

In Figure 2, it is shown how SafeML can be used for image-based classification problems and how it can provide a statistical representation and explanation between wrong predictions and the ground truth. It is also shown that *p*-value consideration can improve the statistical explanations (illustrated in the second row of Figure 2). The difference in results given by application of SafeML's four ECDF-based distance measures to two different data sets is illustrated in Figure 3(a)–(d). As can be seen, the KS distance measures the maximum value between two ECDFs. The KS distance cannot detect which ECDF has a higher value, while the Kuiper distance can measure two maximum up and maximum down. In a situation where two sets have the same mean value and different variances like spiral and circle benchmarks, the Kuiper distance has a better measure over the KS distance. As illustrated in Figure 3(c), the WD can somehow calculate the area between two ECDFs. Thus, the WD will be more sensitive to a change in the geometry of the distributions. The CVM distance has similar functionality to the WD, and it can perform faster. If we reduce the step size in the CVM algorithm, the results will be close to the WD ones. More detail on ECDF distance measures can be found in Aslansefat.[17] Figure 3(e) provides a comparison between true accuracy, estimated accuracy by SafeML II, and the Wilson interval confidence (WIC) bound from Kläs et al.[9] For the WIC, the z-score is chosen to be 3.29053 to gain a 99.99% confidence level. The WIC usually provides both upper bound and lower bound. To ensure the maximum safety level, only the lower bound is considered. From the existing 43 classes in the GTSRB, five safety-critical related classes have been chosen for the comparison. The results show that in most cases the W-based accuracy estimation has less error. For two cases, the W algorithm was not successful: for Class 11 (Cross Road Ahead), the AD estimation has less error, and for Class 13 (Yield), the low band WIC has better accuracy. It should be noted that the WD, CVMD, and AD are not always bounded between 0 and 1. However, based on our experiments, they are always correlated with accuracy. To clarify the effect of *p*-value consideration, the WD algorithm is selected as the best performing measure for GTSRB and its results with and without *p*-value consideration are compared with true accuracy as shown in Figure 3(f). The original SafeML[13] was successful for a feature-based data set. However, our experiments show that it is not always successful. For example, in GTSRB, the WD measure without *p*-value consideration has failed to detect true accuracy changes, while the WD with *p*-value consideration was successful. Generally, using ECDF-based distance measures with *p*-value consideration are more reliable and less noisy, especially when the results are going to be used for statistical explainability purposes.

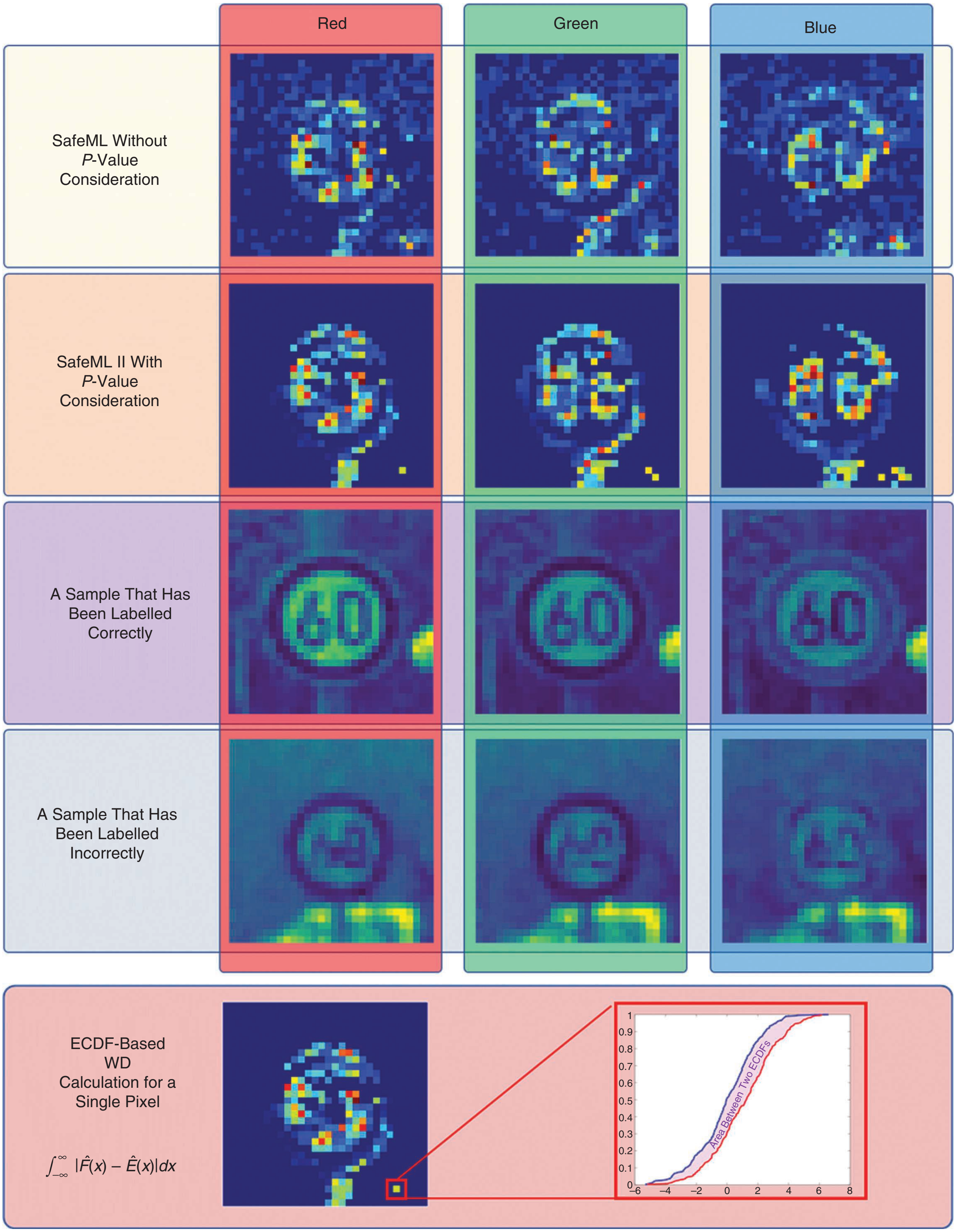


FIGURE 2. Sample results of SafeML II with Wasserstein distance (WD) and considering *p* values (Class 3).

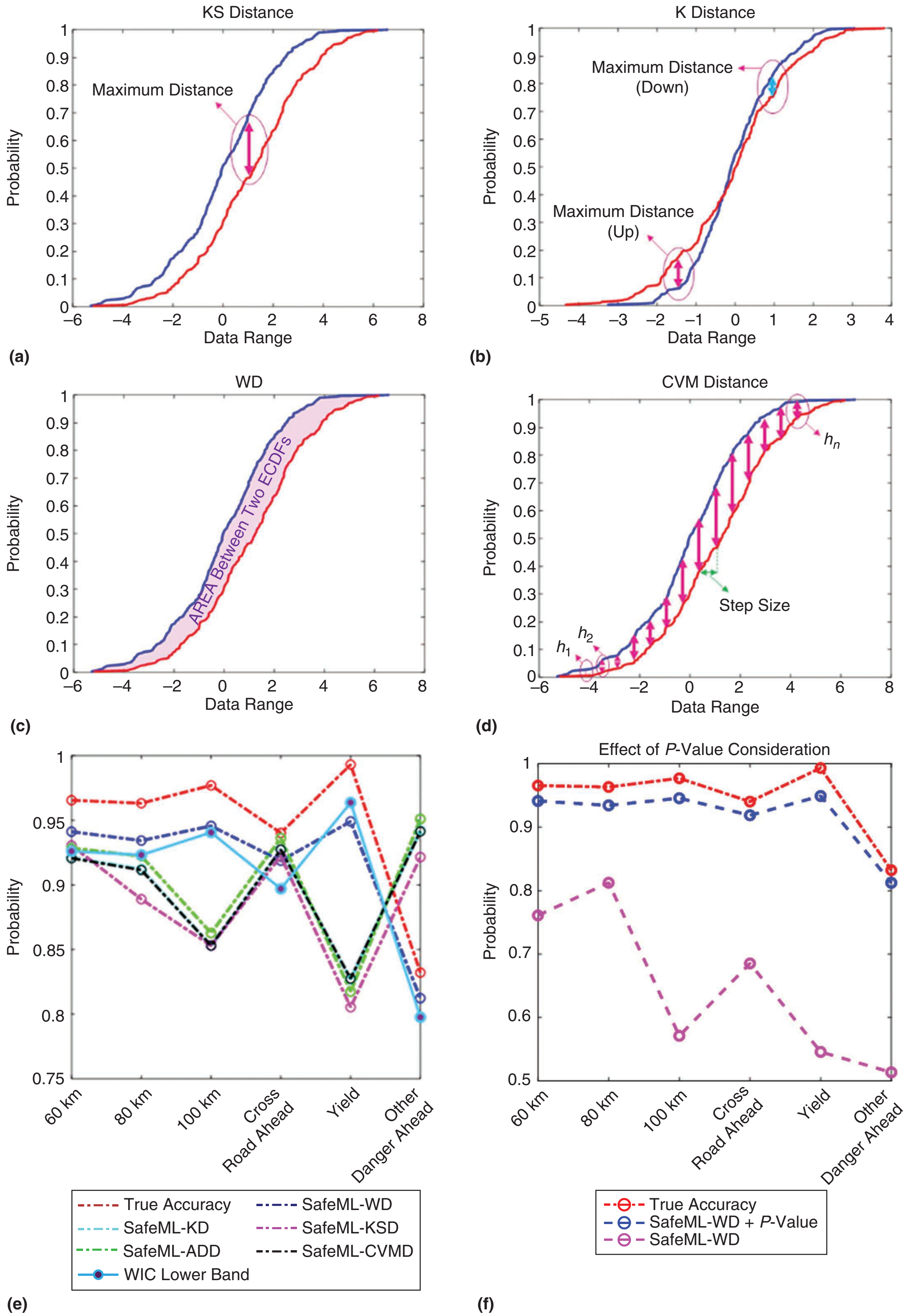


FIGURE 3. (a) The KS distance, (b) the K distance, (c) the WD, (d) the CVM distance, (e) a comparison between true accuracy, estimated accuracy by SafeML II and Klas et al.,[9] and (f) a comparison between true accuracy, WD with and without p-value consideration.

In this article, we have only focused on TSR; the idea can be integrated with other safety-related parts of autonomous vehicle software to cover wider safety perspectives. For example, it was explained how to build an integrated safety model and consider different components of a cooperative operation scenario of autonomous vehicles.[18] The results of SafeML II can be used as an input in the proposed safety model in that work to improve confidence in the provided assurance. It should be noted that the SafeML II concept has some limitations. For example, it can only work with ML classifiers, while having the SafeML II concept to work for prediction and regression algorithms is still an open research question. Moreover, we currently investigate what specific characteristics of a data set can lead to a better ECDF-based statistical accuracy estimation in run time. Due to the use of the buffering technique, in some time-critical applications, the proposed approach may not be able to handle a sudden shift of data efficiently within a very short period of time. Generally, for safety-critical systems, it is crucial to limit the possibility of making unsafe decisions and actions that may be caused by a sudden shift in the data. A potential solution to track sudden changes in the incoming data are to use the soft clustering models,[19] which offer a way to evaluate the changes through a natural measure by computing it directly from models. Moreover, in this article, we introduce the model-agnostic version of SafeML where we are unable to go inside any ML/DL algorithm. In our future research work, we will address the model-specific version of SafeML, where we will be able to utilize CNN's middle layer to avoid pixel-level alignment requirements.

ABOUT THE AUTHORS

KOOROSH ASLANSEFAT is a Ph.D. student in the Department of Computer Science and Technology at the University of Hull, Hull, HU6 7RX, U.K. His main research interests include artificial intelligence, Markov modeling, performance assessment, optimization, and stochastic modelling. Aslansefat received an M.Sc. in control engineering from the Shahid Beheshti University. He is a Member of IEEE. Contact him at k.aslansefat-2018@hull.ac.uk.

SOHAG KABIR is an assistant professor in the Faculty of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, U.K. His research interests include model-based safety assessment, probabilistic risk and safety analysis, fault tolerant computing, and stochastic modeling and analysis. Kabir received a Ph.D. in computer science from the University of Hull. Contact him at s.kabir2@bradford.ac.uk.

AMR ABDULLATIF is an assistant professor in the Faculty of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, U.K. His research interests include machine learning, safety assurance of machine learning based systems, predictive diagnostics, and online learning from data streams. Abdullatif received a Ph.D. in computer science and system engineering from the University of Genoa. Contact him at a.r.a.a.abdullatif@bradford.ac.uk.

VINOD VASUDEVAN is a Ph.D. student in the Faculty of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, U.K., and currently working as the lead engineer at Jaguar Land Rover. His research interests include machine learning safety/resilience and certification of autonomous vehicles. Nair received an M.B.A. from the University of Wales. Contact him at v.vasudevan@bradford.ac.uk.

YIANNIS PAPADOPOULOS is a professor in the Department of Computer Science and Technology at the University of Hull, Hull, HU6 7RX, U.K. His research interests include digital art and various aspects of philosophy and their interactions with science. Papadopoulos received a Ph.D. in computer science from the University of York. Contact him at y.i.papadopoulos@hull.ac.uk.

The rapid growth of artificial intelligence applications in various domains and particularly in autonomous vehicle software raises concerns in different perspectives, such as AI safety, AI responsibility, AI explainability and interpretability, human-in-the-loop AI, and AI trustworthiness. This article addressed the issue of distributional shift and its implications for the safety of ML or deep learning classification tasks in autonomous vehicle software. The article proposed SafeML II by extending

SafeML to improve its capabilities for the human-in-the-loop procedure and ECDF-based statistical distance measures, and applied them to image-based classification algorithms in a model-agnostic way. SafeML II improves the ECDF-based statistical distance measure functions using bootstrap-based *p*-value calculations. The proposed SafeML II approach is generic in nature; therefore, we believe it can be integrated with traditional safety assurance methods to enable them to provide assurance for ML/AI models and also to increase confidence in the provided assurance.

CODE AVAILABILITY

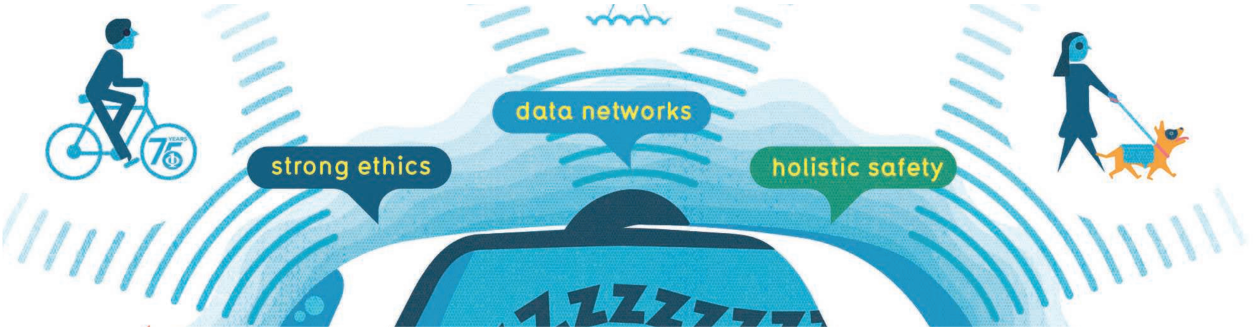
Regarding the research reproducibility, codes and functions supporting this article are published online at GitHub: https://github.com/ISorokos/SafeML. **C**

ACKNOWLEDGMENTS

This work was supported by the Secure and Safe Multi-Robot Systems (SESAME) H2020 Project under Grant Agreement 101017258 (www.sesame-project.org).

REFERENCES

1. N. Webb et al., "Waymo's safety methodologies and safety readiness determinations," 2020, arXiv:2011.00054.
2. S. Kabir and Y. Papadopoulos, "Computational intelligence for safety assurance of cooperative systems of systems," *Computer*, vol. 53, no. 12, pp. 24–34, 2020. doi: 10.1109/MC.2020.3014604.
3. A. F. Magnussen, N. Le, L. Hu, and W. E. Wong, "A survey of the inadequacies in traffic sign recognition systems for autonomous vehicles," *Int. J. Performability Eng.*, vol. 16, no. 10, pp. 1588–1597, 2020. doi: 10.23940/ijpe.20.10.p10.15881597.
4. Q. Rao and J. Frtunikj, "Deep learning for self-driving cars: Chances and challenges," in *Proc. 1st Int. Workshop on Softw. Eng. AI Autonomous Syst.*, 2018, pp. 35–38. doi: 10.1145/3194085.3194087.
5. R. Salay, R. Queiroz, and K. Czarnecki, "An analysis of iso 26262: Using machine learning safely in automotive software," 2017, arXiv:1709.02435.
6. P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012. doi: 10.1145/2347736.2347755.
7. D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," 2016, arXiv:1606.06565.
8. L. Gauerhof, R. D. Hawkins, C. Picardi, C. Paterson, Y. Hagiwara, and I. Habli, "Assuring the safety of machine learning for pedestrian detection at crossings," in *Proc. 39th Int. Conf. Comput. Safety, Reliabil. Security (SAFECOMP)*, Springer Nature, 2020, pp. 197–212. doi: 10.1007/978-3-030-54549-9_13.
9. M. Kläs and L. Sembach, "Uncertainty wrappers for data-driven models," in *Proc. Int. Conf. Comput. Safety, Reliabil., Security*, Springer, 2019, pp. 358–364. doi: 10.1007/978-3-030-26250-1_29.
10. L. Jöckel, M. Kläs, and S. Martínez-Fernández, "Safe traffic sign recognition through data augmentation for autonomous vehicles software," in *Proc. 2019 IEEE 19th Int. Conf. Softw. Quality, Reliabil. Security Companion (QRS-C)*, pp. 540–541. doi: 10.1109/QRS-C.2019.00114.
11. L. Jöckel and M. Kläs, "Increasing trust in data-driven model validation," in *Proc. Int. Conf. Comput. Safety, Reliabil., Security*, Springer, 2019, pp. 155–164. doi: 10.1007/978-3-030-26601-1_11.
12. M. Kläs and L. Jöckel, "A framework for building uncertainty wrappers for AI/ML-based data-driven components," in *Proc. Int. Conf. Comput. Safety, Reliabil., Security*, Springer, 2020, pp. 315–327. doi: 10.1007/978-3-030-55583-2_23.
13. K. Aslansefat, I. Sorokos, D. Whiting, R. T. Kolagari, and Y. Papadopoulos, "SafeML: Safety monitoring of machine learning classifiers through statistical difference measures," in *Proc. 7th Int. Symp. Model-Based Safety and Assessment*, Springer Nature, 2020, vol. 12297, pp. 197–211. doi: 10.1007/978-3-030-58920-2_13.
14. M. M. Deza and E. Deza, "Distances in probability theory," in *Encyclopedia of Distances*. Berlin: Springer-Verlag, 2009, pp. 1–583.
15. German traffic sign recognition benchmarks. https://benchmark.ini.rub.de/?section=gtsrb (accessed Jan. 20, 2021).
16. E. Gilleland, "Bootstrap methods for statistical inference. part ii: Extreme-value analysis," *J. Atmospheric Oceanic Technol.*, vol. 37, no. 11, pp. 2135–2144, 2020. doi: 10.1175/JTECH-D-20-0070.1.
17. K. Aslansefat. "How to make your classifier safe." Towards Data Science, 2020. https://towardsdatascience.com/how-to-make-your-classifier-safe-46d55f39f1ad (accessed Jan. 20, 2021).
18. S. Kabir et al., "A runtime safety analysis concept for open adaptive systems," in *Proc. Int. Symp. Model-Based Safety Assessment*, Springer, 2019, pp. 332–346. doi: 10.1007/978-3-030-32872-6_22.
19. A. Abdullatif, F. Masulli, and S. Rovetta, "Clustering of nonstationary data streams: A survey of fuzzy partitional methods," *Wiley Interdisciplinary Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1258, 2018. doi: 10.1002/widm.1258.



From Neuron Coverage to Steering Angle: Testing Autonomous Vehicles Effectively

Jack Toohey, Loyola University Maryland
M S Raunak, National Institute of Standards and Technology
Dave Binkley, Loyola University Maryland

A deep neural network (DNN)-based system is a black box of complex interactions, resulting in a classification or prediction. We investigate the use of realistic transformations to create new images for testing a trained autonomous vehicle DNN as well as their impact on neuron coverage.

On 8 October 2020, Waymo officially opened its driverless riding service in three Arizona cities. Many automobile and technology companies, including Tesla, Uber, Volkswagen, and Baidu, are not far behind. An autonomous vehicle typically employs a deep neural network (DNN), which learns to recognize objects in the vehicle's environment and makes split-second decisions based on current conditions.

There continues to be trust issues involving the safety and reliability of these systems.[1] Incidents such as the pedestrian fatality caused by an Uber sport utility vehicle in Tempe, Arizona,[2] exacerbate the situation. The primary approach for ensuring the reliability and correctness of these autonomous systems involves different software-verification activities, especially testing. The effective testing of any complex software system is challenging. Furthermore, autonomous vehicles are primarily data driven, statistical, and nondeterministic in nature, making them even more difficult to properly test and their reliability harder to ensure.

Digital Object Identifier 10.1109/MC.2021.3079921
Date of current version: 30 July 2021

Designing an effective test suite for verifying any system involves addressing two broad questions: 1) how to select the test cases and 2) how many test cases to select. When the source code is accessible, code coverage-oriented criteria that use the structure of the source code are commonly employed to address these questions. One can aim to ensure that all of the statements or all of the branches in the source code are covered, that is, are executed by at least one of the test cases in the test suite. These two criteria are known as *statement* and *branch coverage*, respectively. The motivation here is that a test suite is unlikely to reveal a bug located in a statement or a branch that has never been executed. Such coverage metrics are also helpful in providing concrete goals for selecting a sufficient number of test cases. One can set goals such as achieving 90% branch coverage to consider a test suite to be sufficient.

It is generally accepted that a test suite that provides higher statement or branch coverage better tests a piece of software.[3] This type of coverage metric, however, is not suitable when testing DNNs.[4] As a result, researchers and practitioners have looked for other metrics and strategies to test DNN-based systems.[5] Consequently, newer metrics such as neuron coverage (NC), used by DeepXplore[6] and DeepTest,[7] have emerged. Although the inadequacy of traditional test-coverage criteria when applied to DNN-based systems is well established, the usefulness of new criterion such as NC is yet to be fully studied. Although DeepTest[7] has provided results in favor of using NC as an effective test-selection criterion, some newer studies have questioned its usefulness.[8]

In the domain of autonomous vehicle operations, one of the primary test inputs is the contiguous set of front-view images, and one of the computed outputs is the steering angle (SA) of the vehicle. Numerous research studies have been performed, looking for effective ways of training a DNN-based machine learning algorithm to convert images into specific actions.[9,10] Our focus in this article, however, is on the challenges of testing these systems. Testing a trained DNN requires the selection of test images. Based on the differing findings regarding NC's usefulness, further exploration is necessary to see how NC is impacted by different test image selection strategies and whether they lead to more effective testing of the underlying DNN model. This is what we investigate in this article.

Another challenge comes from the fact that one needs test images that have not been seen by the DNN before and for which the correct SA is known, which provides a test oracle. Synthetically creating new test images from the existing ones with known, expected driving angles addresses this problem. We utilize this approach in our study by deriving reasonable, synthetic test images from existing images. Specifically, we consider seven synthetic image transformations to gain insight into their effectiveness when testing autonomous vehicle software. The only image transformations considered are the ones that enable us to predict the expected SA, which solves the test oracle problem. Our work builds on the framework found in DeepTest.[7] We identify an important issue in that approach, update it, and consider new and refined image transformations, all to better understand the interplay between NC and the effective testing of autonomous vehicles.

Our research contributions presented in this article include showing that

- the use of new test images created by applying transformations to existing ones, both individually and in groups, increases NC
- some transformations are more effective than others at achieving higher NC
- there is a positive correlation between higher NC and the test suite's ability to extract output deviations, which suggests NC's potential as a measure of test-suite quality.

BACKGROUND AND RELEVANT CONCEPTS

Neural networks and NC

A neural network is a computing structure that attempts to mimic the structure and behavior of the human brain. At its core is a computing unit called a *neuron* (or a *perceptron*). The neurons are placed in layers with edges connecting one layer to the next. There are weights associated with the edges that connect neurons. Based on the inputs and the weights, a nonlinear activation function is used to decide when a neuron is activated and thus impacts the neurons in subsequent layers. Typically, a neural network will have a layer of neurons for accepting inputs as well as a layer for producing the computed output. A DNN includes additional neurons in a series of hidden layers, which enable it to be used for complex computations such as image classification. Users interact with the first and last layer of a DNN, which handle inputs and output, respectively. The neurons in the intervening hidden layers, through their connection weights and activation functions, learn complex decision mechanisms that contribute to the overall output.

There is natural appeal to the notion of transferring code coverage-based test-adequacy criteria for traditional software testing to NC-based test-adequacy criteria for DNN model testing.[6]

The idea here is to select a set of test cases (for example, a set of images) that maximize the activated neurons. The underlying argument is that if certain neurons never get activated during testing, then it is possible that there may be undiscovered erroneous behavior associated with their activation that has never been witnessed. Despite this appeal, the test effectiveness of complex structures such as DNNs may be uniquely different. Thus there is an ongoing need for the exploration of NC-based test-adequacy criteria.

Autonomous vehicles

An autonomous vehicle is a self-driving automobile that uses sensory devices such as cameras, infrared sensors, lidar, and GPS to navigate the world around it by making very fast decisions related to SA adjustment, acceleration, and braking. Central to this complex process is a DNN that learns different vehicle operation responses from a massive amount of training data. The images from the front-view camera are the primary data sources influencing the SA decision. As shown in Figure 1, a front-view image is fed into a trained DNN, which, in turn, predicts the SA that the self-driving car should maintain or adopt.

The output produced by a DNN is the result of a series of neurons being activated inside. The DNN's edge weights and neuron thresholds are determined during the training of the network. For a given input, the neurons of the neural network can be labeled either "active," when the weighted input value exceeds the threshold, or "inactive."

Metamorphic testing

Traditional software testing relies on the presence of a test oracle, which is capable of providing the correct output or expected behavior for a given test input. Automated software testing uses the oracle to identify failures of the software system where the output behavior deviates from that of the oracle. In the case of systems that are often termed *non-testable*[11] due to the absence of an oracle, such as cryptographic functions or scientific computations, metamorphic testing provides a way forward by producing two sets of related inputs whose outputs should be the same or should differ in a deterministically predictable way. Thus the two test cases can be used as pseudo-oracles for each other.[12] With DNN-based systems, such as autonomous vehicles, producing oracle output is possible but expensive. As a result, pseudo-oracles, such as those provided by metamorphic testing, should be utilized whenever possible. This approach has shown to be effective in testing DNN-based systems.[13]

In our case, we apply synthetic image transformations to create a set of new test images. The transformed images are metamorphic in nature, that is, while these transformations modify the image they do not change the expected behavior of the autonomous vehicle. For example, we do not expect the SA to change if we transform an image by increasing its brightness or reducing its contrast. The expected behavior can be determined from the original oracle behavior and the transformation applied. One of the metamorphic transformations that we introduce, "flip," requires that the oracle output be computed by turning the steering wheel in the opposite direction.

DeepXplore

DeepXplore[6] demonstrated that even a randomly selected set of test cases could achieve 100% statement coverage of a DNN while the share of internal neurons being activated was no more than 34%. This exemplified the inadequacy of traditional code coverage for testing a DNN. With the aim of exercising more of the DNN's internals, DeepXplore proposed NC as a metric to select effective test inputs. The testing framework used two DNNs trained for the same purpose (for example, classifying an image) as pseudo-oracles of each other and then generated test inputs that maximized both

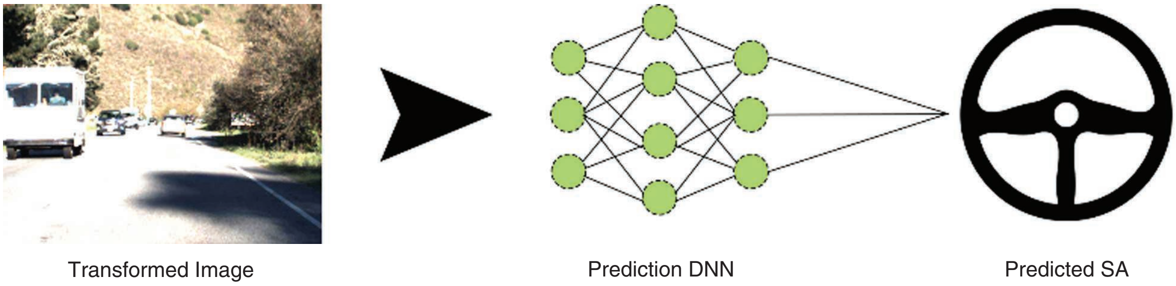


FIGURE 1. From image to SA output using a DNN.

NC as well as images for which the two DNNs produced different classifications. Their study showed this approach to be highly effective in discovering corner cases of erroneous DNN behavior.

DeepTest

The DeepTest work[7] builds on the NC idea, creating synthesized images to test DNNs trained for autonomous vehicle operations. DeepTest applies transformations to images to mimic changes in the natural environment, such as changes in sunlight, rain, fog, and so on. Most of the transformations they use are metamorphic, in that they do not change the autonomous vehicle's expected SA when compared to the original image. Thus, DeepTest could automatically detect whether the DNN was producing an erroneous behavior using the transformed images.

The DeepTest researchers worked with trained DNNs from the Udacity self-driving Challenge data set.[14] The tool detects previously inactive neurons being activated due to a transformed image. It uses a greedy search algorithm that repeatedly applies transformations to a single test image in an attempt to maximize activated neurons before moving on to the next test image.

A possibly unintended consequence of the greedy search used in DeepTest is that it achieves high NC at the cost of potentially over-transforming images. Repeatedly modifying a single image causes ever-increasing distortion, leading to the original scene being almost unrecognizable. An example of this is presented in Figure 2, where (b) is the result of repeatedly transforming (a). The fixation on increasing NC seems to void the metamorphic nature of the transformations. The use of such transformations in the experiments and the corresponding usefulness of the results become dubious. In our experiments, we build on the DeepTest framework but remove the repeated transformations of an image to avoid this scenario.

OUR APPROACH AND SETUP

We investigate the impact of the test images synthesized through transformations on NC and the predicted SA. We start with the basic DeepTest tool and modify it to avoid the over transformation scenario depicted in Figure 2. Instead, we reset the image back to its original, untransformed version before applying subsequent transformations.

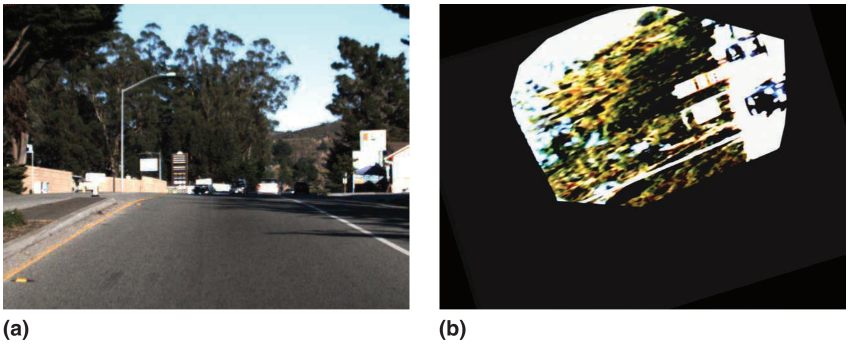


FIGURE 2. (a) An unaltered test image and (b) the same image transformed by the DeepTest algorithm.

Image transformations

Our study employs seven metamorphic image transformations. The chosen transformations allow us to predict the expected SA for the new images. Some of the transformations are more realistic in terms of being likely to occur while driving. For example, a change in brightness is caused all the time by clouds moving past the sun. We experimented with all the transformations used in the DeepTest study and found that some of the transformations (for example, shear) were not metamorphic and were prone to change the images toward unrealistic scenarios. Our initial exploration led us to choose four transformations: scale, contrast, brightness, and blur. We saw a similar problem with translating an image simultaneously in both the X and Y directions. We therefore separate it into two transformations, translate X and translate Y, and limit the range. This ensured a still-reasonable image after applying the transformation. Finally, we introduce a completely new metamorphic transformation, flip, which horizontally flips an image. Unlike the other six, which do not affect the predicted SA, flip requires changing its sign, that is, moving the SA in the opposite direction. Through flip, we aim to create new synthetic test images that are clear and with a known, expected SA.

Moreover, we are interested in observing the impact of NC when such pseudo new images are fed to the trained DNN. All of the transformations except flip take a parameter that dictates how much alteration to make to the image (for example, the number of pixels to translate the image by). Although it is not reported here, we also performed a systematic study of different parameter values used for transforming the images. Those experiments did not show any significant impact on NC or predicted

SA with particular value combinations or values outside our selected range. That work did find, however, that a small number of random parameter values provides sufficient diversity. Thus, we randomly select parameter values from slightly narrower ranges than used in the DeepTest experiments. The narrower range ensures that the correct SA is unaffected by the transformation.

The seven transformations and their corresponding parameter ranges used are as follows.

1. *Translate X*: This transformation shifts the image left or right by the given number of pixels. The range considered from left shift to right shift is [−*X* : *X*] − [−50 : 49].
2. *Translate Y*: This transformation shifts the image up or down by the given number of pixels. The range considered from up shift to down shift is [−*Y* : *Y*] − [−50 : 49].
3. *Scale*: This transformation shrinks or enlarges the image along both the *x*- and *y*-axes by a given percentage. The range of percentages considered is [0.5–1.9%].
4. *Contrast*: This transformation increases or decreases the contrast of the image by a given alpha value. The range considered is [0.5–1.9%].
5. *Brightness*: This transformation changes an image's brightness by a given bias parameter. The range considered is [−21–20].
6. *Blur*: This transformation blurs the image in one of three ways (chosen randomly) based on a parameter in the range [1–10].
7. *Flip*: This transformation flips the image across the vertical axis. There is no parameter here.

One of the contributions of our work is that, in addition to applying individual transformations, we also apply multiple transformations to create a new test image. We use the term *transformation group* to refer to one or more transformations. For example, when the transformation group {flip, contrast, translate Y} is applied to an image, the image is flipped, its contrast is adjusted, and it is translated along the *y*-axis. Although we do not present the details here due to space constraints, our experiments show that the application order of the transformations that make up a transformation group do not affect our NC results.

The metrics used in our study

To investigate the impact of different transformation groups, we must measure the increased NC associated with the application of a particular transformation group. We accomplish this using the metric isolation NC (INC) defined as follows.

- INC is computed relative to a set of *N* unaltered images that are first run through the model to establish a baseline NC. The specific transformation group being studied is then applied to each image before it is run through the model to identify the number of additional neurons activated above the baseline. Between the images, the coverage is reset to the baseline, thus isolating the contribution of each transformed image.

Our study also considers the SA predicted by the model. Each image in the data set includes the expected SA that should be produced by the DNN. In other words, we have a test oracle for the images, making it possible to consider the accuracy of the DNN's SA prediction. We do so by computing the SA deviation (SAD):

- *SAD* is defined as the absolute value of the difference between the predicted SA and the oracle SA for a given image.

Of particular interest here is the potential to examine the effectiveness of NC as a predictor of test-suite performance. Specifically, we are interested in studying whether test suites that produce higher NC also lead to the discovery of more anomalous behavior, as witnessed by greater SAD. This situation parallels a traditional test suite that provides greater code coverage being more likely to uncover more program faults. Should such a connection exist, then the test suites providing higher NC could be deemed better test suites.

Research questions

We are interested in better understanding the impact of transformation on INC and SAD. Our initial working hypothesis is that higher NC is indicative of a stronger test suite, and a stronger test suite is going to be more effective in discovering anomalies (weakness in the model). To explore this relationship, we consider two key research questions (RQs):

- RQ1: Do certain image transformations achieve higher NC than others?
- RQ2: What impact, if any, do transformed images have on the predicted SA?

RQ1 actually goes beyond simply asking if transformation increases NC and considers the relative impact of different transformation groups. Then, RQ2 factors in the consideration of SAD to

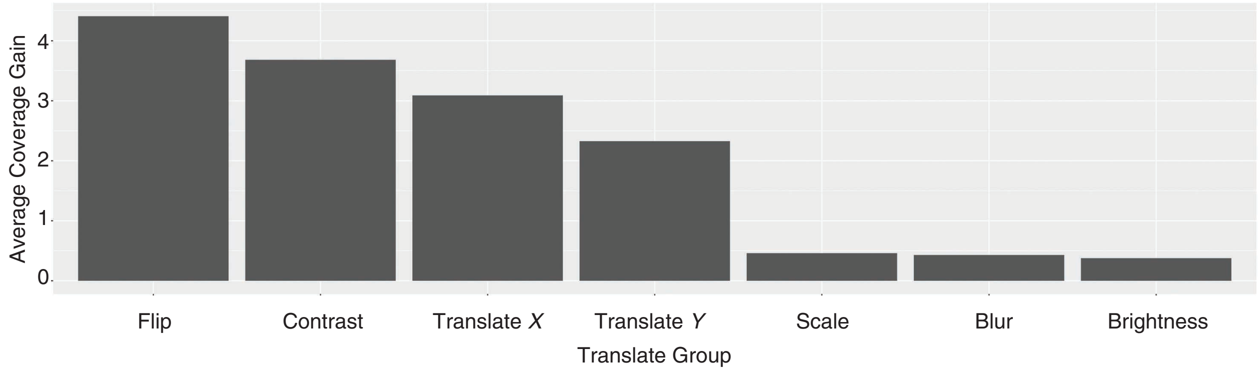


FIGURE 3. Transformation's impact on INC. The baseline includes activation of 14,871 neurons out of 18,899 total neurons.

evaluate the effectiveness of NC as a predictor of test-suite strength.

Experimental setup

Our experimental setup builds on the DeepTest[7] framework, written using Python version 2.7. We modified the framework such that image transformations are not repeatedly applied to the same image. Like DeepTest, we work with the Rambo model[15] from the Udacity self-driving challenge.[14] Rambo uses three separate convolutional neural networks for determining the SA. With each application of the model, we measured the number of its total 18,899 neurons that were activated during each SA computation.

TABLE 1. The transformation cover comparison.

Transformation	Mean neuron count increase	Group
Flip	4.41	*a*
Contrast	3.68	*ab*
Translate X	3.09	*bc*
Translate Y	2.33	*c*
Scale	0.46	*d*
Blur	0.43	*d*
Brightness	0.38	*d*

The Udacity self-driving challenge data include 5,000 still-frame images chopped from a 30-min video, which was taken by a car as it drove down the road. The images are from a front-view video feed and were thus taken from the point of view of the driver. In some of our experiments, we used a sample of 100 images selected by picking every 50th image.

RESULTS AND DISCUSSION

RQ1: Are all transformations created equal?

Going beyond the question "Does transformation increase NC?" we investigate whether certain transformation groups distinguished themselves. Such transformation groups are valuable if higher NC proves to be a useful metric for selecting test images.

Visually, Figure 3 shows the average INC gain for each transformation relative to the baseline. A statistical test using the analysis of variance[16] finds a strong difference (*p* value <0.0001); thus in Table 1, we show the results of Tukey's posthoc honest significance difference test[16] applied to the INC_7C_1 data (the additional NC of the individual transformations). Here ${}_7C_1$ denotes the combinations of seven things (our transformations) taken one at a time. Thus ${}_7C_1$ refers to each of the seven transformations considered individually. In the resulting groups, listed in Table 1, the transformations sharing a letter are not statistically separable. Although the existence of overlaps means that there is no simple order, it is clear from the data that flip and contrast are the top performers where flip outperforms all the other transformations except contrast. Next, translate X and translate Y are in the middle, where it is interesting that only translate Y can be separated from contrast. Finally, scale, brightness, and blur all produce notably inferior increases.

We also considered the transformation pairs of INC_7C_2, that is, the combinations of two transformations from the group of seven. As displayed in Table 2, the combination of flip and contrast with a mean of 8.45 distinguished itself from all of the other pairs in terms of increased NC. Although there is considerable overlap, the two other main groups become evident. First, in the middle are pairs that include translate X, and finally, at the bottom are groups with none of flip, contrast, or translate X. Looking at the combinations of the three transformations applied to the images, that is, from the INC$_7C_3$–produced data, there is a greater overlap between transformation groups, but all the triples with both flip and contrast come before those with one of the two, which come before those without either of the two transformations. Looking at INC$_7C_i$ for $i > 3$, this basic pattern continues, although as the number of transformations in the transformation groups increases, there is ever-greater overlap.

In summary for RQ1, not only does transformation bring increased NC, but flip and contrast stand out. Thus, where higher NC is the goal, these two transformations should be preferred. Intuitively, flip changes the images drastically compared to the others and therefore likely requires more neurons to be activated to process those images. Similarly, it is possible that contrast plays a relatively more important role in identifying the features that lead to the SA computation by the DNN.

RQ2: Is there a connection between INC and SAD?

Given transformation's impact on NC, a natural follow-up question is what impact, if any, does transformation have on the model's SA prediction? We intentionally limited our chosen transformations to preserve the known, expected SA (flip requires inverting the steering direction), which provides us with an oracle.

This leads us to investigate the relationship between INC and SAD. Recall that INC is the NC achieved by each image in isolation, and SAD is the absolute value of the difference between the model-predicted SA and the oracle SA. Here greater SAD is indicative of potentially anomalous behavior. Accordingly, if transformation leads to greater deviation, then it has proven effective in uncovering potential bugs or weaknesses in the model.

As a preliminary investigation, we first consider transformation's impact on SAD. Similar to the pattern seen in Figure 3 when using transformation groups of size one, flip leads to the largest mean SAD, followed by contrast, and then translate X. We also consider larger transformation groups where the same patterns perpetuate. For example, size-three transformation groups that include flip and contrast dominate the larger-mean SAD values. Of the two, flip more consistently leads to a greater deviation in SA. This pattern is reminiscent of the NC where flip is a consistent top performer. As a result, the data suggest a connection between NC and SAD.

Finally, we compare INC and SAD directly using linear regression.[16] We applied R's `lm` function using SAD as the response variable and INC as the explanatory variable. Aggregated over all seven transformations, `lm` yields the following relationship with a p value <0.0001:

$$\text{SAD} = 0.043 + 0.0071 \times \text{INC}.$$

The most relevant aspect of our research is that the slope of the regression line is positive, indicating that, although small, increased INC is associated with greater SAD. This result persists in the larger transformation groups. For example, none of $_7C_3'$*s* 35 transformation groups yields a line with a negative slope.

Digging deeper, we include the interaction term between the transformation group and NC. The model uncovers two interesting results. First, NC continues to have a positive coefficient (p value = 0.007). Second, flip differentiates itself (p value = 0.0164) where the slope for flip is three times steeper than the slope using the aggregated data.

In general, a stronger test suite for a system is one that reveals more errant behaviors or failures. Reflecting on our two RQs, in the domain of autonomous vehicle operations and especially in terms of SA prediction, a stronger test suite would reveal more and larger SADs as indicative of model weakness. If that stronger test suite also achieves higher NC, then NC could be argued to be a useful proxy for the strength of a test suite.

Based on our data, we see a small positive correlation between transformations that generate higher NC and those that reveal more and larger mean SADs. In other words, the data suggest that higher NC may be an indicator of test-suite strength.

However, care must be taken with this correlation as we do not know that the increase in NC is causing the increase in SADs. Regardless, it seems that the transformed images, likened to test cases in traditional software, have the potential to discover erroneous behaviors as manifested by their propensity to yield higher deviations in the predicted output.

Threats to validity

Our experiments and results have some limitations. We have worked with only one DNN model, Rambo.[15] Our initial set of images for testing the DNN was taken from the set used with DeepTest.[7] Increasing the external validity of our results by repeating our experiments with other DNNs, perhaps with other failure metrics, is an important part of our future work.

Another limitation of our analysis is that it does not isolate when the SAD rises to the level of grossly erroneous or failure inducing. Instead of arbitrarily selecting a threshold where we can say the steering angle is erroneous, we looked simply at the relationship between NC and SAD.

RELATED WORK

A number of studies have taken the approach of testing DNNs with synthetically created road images.[6,7,17,18] Among them, DeepRoad[17] argues that the images used in DeepXplore[6] and DeepTest[7] do not realistically represent real-world driving scenarios. DeepRoad alleviates the problem by using a generative adversarial network (GAN)-based technique[19] to synthesize realistic driving scenes. However, a GAN-based technique does not provide any guarantee of image creation preserving the desired SA. Contrary to DeepTest's findings, Harel-Canada et al.[8] did not find NC to be an effective metric to derive new test images. They argue that individual neurons do not represent the specific features of an image and thus the value in maximizing neuron activations is questionable.

TABLE 2. The transformation pair INC comparison.

Transformation	Mean INC increase	Group
Contrast + flip	8.45	*a*
Contrast + brightness	5.93	*b*
…		
Translate X + flip	4.27	*bcde*
Translate X + brightness	3.4	*cde*
…		
Translate Y + scale	2.27	*ef*
Brightness + blur	0.58	*f*
Scale + blur	0.55	*f*
Scale + brightness	0.54	*f*

Sun[20] proposed a new set of coverage criteria instead, inspired by the modified condition/decision coverage criterion of traditional software systems. The search for a more effective test image selection approach led the Deepgauge study[21] to propose a set of coverage criteria based on addressing different sections of the network as well as the range of values output by a neuron (boundary coverage).

We investigated the use of image transformation to create new test images and how these images impact the NC achieved by a DNN. We first found that although transformation increases NC, certain transformations achieve higher NC than others. In fact, our newly introduced image transformation, flip, which does not require a new oracle for testing, achieves higher NC than do other existing transformations. Furthermore, combinations of transformations also prove useful and often achieve even higher NC.

Our data show a small but positive correlation between NC and SADs. Parallel to the NC results, the positive correlation is strongest with flip. More importantly, we never found a negative correlation between NC and SAD among any applied transformations. This suggests that NC might act as a proxy for test-suite strength; however, there is a need for further investigation to discover whether NC is a consistent indicator of test-suite strength.

As DNNs become more integral to modern technology, a more thorough understanding of these sophisticated black-box systems is warranted. This includes not only empirical testing work, such as what was presented in this article, but also theoretical work aimed at providing a deeper understanding. For DNNs that take images as their input, having an effective metric for selecting good test images is an important aspect of their testing. The use of transformed images that do not require a new test oracle, combined with NC, may prove a good candidate for this role.

ACKNOWLEDGMENTS

Certain commercial equipment, instruments, or software are identified in this article to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the materials or equipment identified are necessarily the best available for the purpose. C

REFERENCES

1. Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: A survey," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 2, pp. 315–329, 2020. doi: 10.1109/JAS.2020.1003021.
2. E. Besic, N. Zych, and J. Iverson, "Preliminary report highway: HWY18MH010," National Transportation Safety Board, Washington, D.C., Mar. 2018. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Pages/HWY18MH010-prelim.aspx
3. H. Zhu, P. A. V. Hall, and J. H. R. May, "Software unit test coverage and adequacy," *ACM Comput. Surv.*, vol. 29, no. 4, pp. 366–427, Dec. 1997. doi: 10.1145/267580.267590.
4. J. Sekhon and C. Fleming, "Towards improved testing for deep learning," in *Proc. 2019 IEEE/ACM 41st Int. Conf. Softw. Eng.: New Ideas Emerging Results (ICSE-NIER)*, pp. 85–88. doi: 10.1109/ICSE-NIER.2019.00030.
5. J. M. Zhang, M. Harman, L. Ma, and Y. Liu, "Machine learning testing: Survey, landscapes and

horizons," *IEEE Trans. Softw. Eng.*, early access, Feb 2020. doi: 10.1109/TSE.2019.2962027.

6. K. Pei, Y. Cao, J. Yang, and S. Jana, "DeepXplore: Automated whitebox testing of deep learning systems," in *Proc. 26th Symp. Operating Syst. Principles (SOSP '17)*, 2017, pp. 1–18.
7. Y. Tian, K. Pei, S. Jana, and B. Ray, "DeepTest: Automated testing of deep-neural-network-driven autonomous cars," in *Proc. 40th Int. Conf. Softw. Eng., ICSE '18*, 2018, pp. 303–314. doi: 10.1145/3180155.3180220.
8. F. Harel-Canada, L. Wang, M. Gulzar, Q. Gu, and M. Kim, "Is neuron coverage a meaningful measure of testing deep neural network?" in *Proc. 28th ACM Joint Meeting European Softw. Eng. Conf. Symp. Foundations Softw. Eng. (ESEC/FSE)*, Nov. 2020, pp. 851–862. doi: 10.1145/3368089. 3409754.
9. M. Bojarski et al., "End to end learning for self-driving cars," 2016. [Online]. Available: http://arxiv.org/abs/1604.07316
10. P. M. Kebria, A. Khosravi, S. M. Salaken, and S. Nahavandi, "Deep imitation learning for autonomous vehicles based on convolutional neural networks," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 1, pp. 82–95, 2020. doi: 10.1109/JAS.2019.1911825.
11. M. D. Davis and E. J. Weyuker, "Pseudo-oracles for non-testable programs," in *Proc. ACM '81 Conf.*, 1981, pp. 254–257. doi: 10.1145/800175.809889.
12. T. Y. Chen, "Metamorphic testing: A simple method for alleviating the test oracle problem," in *Proc. 2015 IEEE/ACM 10th Int. Workshop on Automation SW Test*, pp. 53–54. doi: 10.1109/AST.2015.18.
13. Z. Q. Zhou and L. Sun, "Metamorphic testing of driverless cars," *Commun. ACM*, vol. 62, no. 3, pp. 61–67, Mar. 2019. doi: 10.1145/3241979.
14. "Udacity self-driving car challenge 2 datasets," GitHub, 2016. https://github.com/udacity/self-driving-car/tree/master/datasets/CH2 (accessed May 20, 2021).
15. "The Rambo DNN model for self-driving cars," GitHub, 2016. https://github.com/udacity/self-driving-car/tree/master/steering-models/community-models/rambo (accessed May 20, 2021).
16. R. L. Ott and M. Longnecker, *An Introduction to Statistical Methods and Data Analysis*, 6th ed. Pacific Grove, CA: Brooks/Cole, 2001.
17. M. Zhang, Y. Zhang, Z. Lingming, C. Liu, and S. Khurshid, "DeepRoad: GAN-based metamorphic testing and input validation framework for autonomous driving systems," in *Proc. 33rd ACM/IEEE Int. Conf. Automated Software Eng. (ASE 2018)*, Sept. 2018, pp. 132–142. doi: 10.1145/3238147.3238187.
18. X. Xie, L. Ma, F. Juefei-Xu, M. Xue, H. Chen, Y. Liu, and Z. Zhao, "DeepHunter: A coverage-guided fuzz testing framework for deep neural networks," in *Proc. 28th ACM SIGSOFT Int. Symp. Softw. Testing (ISSTA)*, July 2019, pp. 146–157. doi: 10.1145/3293882.3330579.
19. I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020. doi: 10.1145/3422622.
20. Y. Sun, X. Huang, and D. Kroening, "Testing deep neural networks," 2019. [Online]. Available: http://arxiv.org/abs/1803.04792
21. L. Ma et al., "DeepGauge: Multi-granularity testing criteria for deep learning systems," in *Proc. 33rd ACM/IEEE Int. Conf. Automated Softw. Eng.*, 2018, pp. 120–131. doi: 10.1145/3238147.3238202.

ABOUT THE AUTHORS

JACK TOOHEY is working toward his bachelor's degree in computer science at Loyola University Maryland, Baltimore, Maryland, 21210, USA. His research interests include artificial intelligence and embedded systems. Contact him at jrtoohey@loyola.edu.

M S RAUNAK is a computer scientist at the National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, 20899, USA. Raunak received a Ph.D. in computer science from the University of Massachusetts Amherst. His primary research interest includes developing and measuring verification and validation approaches for difficult-to-test systems such as cryptographic implementations, simulation models, and machine learning algorithms. He is a member of IEEE. Contact him at raunak@nist.gov.

DAVE BINKLEY is a professor of computer science at Loyola University Maryland, Baltimore, Maryland, 21210, USA. His research interests include the tools and techniques used to help software engineers understand and improve their code. Binkley received his doctorate from the University of Wisconsin–Madison. Contact him at binkley@cs.loyola.edu.

For the Common Defense

David Alan Grier, Djaghe, LLC

When a specialized article becomes part of our body of knowledge, it speaks not only for itself but also for an entire community of researchers.

No one is an island, we are told. This is true of individuals as well as of periodicals that support the body of technical knowledge. When a hacker invades our neighbor's cell phone, we have lost a bit of our security in the process. When a peer publication reports on new research, we are all enriched, including the flagship periodical of *Computer*. Perhaps this is illustrated best in the subjects of smartphones, mobile computing, and the challenge of mobile security.

The IEEE Computer Society is far from a monolithic organization. It consists of one dozen or so communities that are loosely connected by the von Neuman model of computing. You can get a sense of the scale and scope of these communities by scanning the titles of the Society's magazines: *IEEE Software*, *IT Professional*, *IEEE Pervasive Computing*, *IEEE Internet Computing*, and the rest. However, this process can be a little misleading. Small communities form in the Society long before they create the trappings of respectability: a conference, a transactions, a magazine. They can also be a little surprised when their work becomes the center of attention.

IEEE Computer Society members started forming a community to study mobile computing in the 1990s. They held their first conference in 1994 and quickly identified topics that would be central to the field, including topics that would later be described as the "location-aware system,"[5] "mobile-cloud services,"[10] and "mobile security."[2] The field grew quickly. In under than a decade, they were able to create both a transactions and a magazine. However, some members seemed a little surprised when

ARTICLE FACTS

- Article: "Mobile Security: Finally a Serious Problem"
- Author: Neal Leavitt
- Citation: *Computer*, vol. 44, no. 6, December 2011
- *Computer* influence rank: #40 with 8,370 views and downloads and 16 citations

Digital Object Identifier 10.1109/MC.2021.3055878
Date of current version: 30 July 2021

0018-9162/21©2021IEEE

EDITOR **DAVID ALAN GRIER**
Djaghe, LLC; grier@gwu.edu

smartphones began to emerge as the dominant example of mobile computing. "Mobile phones' prevalence gives them great potential to be the default physical interface for ubiquitous computing applications," explained a 2006 article in *IEEE Pervasive Computing Magazine*.[3] Within two more years, the iPhone and similar platforms demonstrated what these devices could do and how they could interact with Internet computing services that were starting to be called *cloud computing*.[9]

As a group, they were concerned about the integrity of these devices but also didn't want to swaddle the growth of this technology.

With the new devices came new opportunities, even before we really appreciated that cell phones were really small computers. As these phones started to become common, IEEE Computer Society authors started to express concern about viruses and malware. As a group, they were concerned about the integrity of these devices but also didn't want to swaddle the growth of this technology. "Although malicious attacks on mobile phones are somewhat inevitable," wrote one set of authors in 2005, "service providers and mobile phone manufacturers shouldn't restrain innovation and exploration."[4]

By 2010, the computing landscape had quite radically changed. Mobile phones were quickly becoming the dominant form of telecommunication. "Right now about 12% of the world's population pays a monthly cell phone bill,"[1] reported *IEEE Pervasive Computing*. This number indicated that "the number of wireless users has surpassed the number of people using standard wireline phones."[1] Furthermore, these devices promised a richer computing device than a mere phone disconnected from the wall. "Viewed collectively," explained another author, "the rapid evolution of mobile computing, location sensing, and wireless networking has engendered a new paradigm for computing."[8]

With this new computing environment came new threats. When hacked, the device would not only yield data but could also provide location and activity information. Hackers could also activate the device's microphone and eavesdrop on your activities—a threat that clearly bothered many of the early commentators.[6–8]

Like many of the articles in our list of influential publications, "Mobile Security: Finally a Serious Problem" (see "Article Facts") is actually the second major piece written by the author on that topic. Like many authors, he needed to revisit the subject and refine his ideas.[6,7] Leavitt's first article appeared in 2005. There are "a growing number of viruses, worms, and Trojan horses that target cellular phones," he wrote in his first discussion of the problem. "Although none of the new attacks has done extensive damage in the wild," he added, but that "it's only a matter of time before this occurs."[6]

Five years later, he was not so sanguine. After "years of warnings about mobile security, there finally appears to be a reason to worry," Leavitt explained. The "number and types of mobile threats—including viruses, spyware, malicious downloadable applications, phishing, and spam— have spiked in recent months."[7]

No cell phone is an island, of course. It connects us to data, to services, and to each other. No professional periodical is an island either. *Computer* was ultimately connected to *IEEE Pervasive Magazine*, which hosted an extensive discussion of mobile security—a discussion that was beyond the scale that could be supported by a general-purpose periodical.

Neal Leavitt, the author of "Mobile Security: Finally a Serious Problem" served as a link between the two communities. To *Computer*, he brought a careful analysis and taxonomy of the problem. It represented five years of careful observation, experience in the industry, and inventive thinking. This analysis could never capture the whole discussion of mobile security, but it was able to explain the nature of the problem to the general Society member, show what needed to be done, and demonstrate what others were doing. If an attack against one mobile device is an attack against all, then one clear article on mobile security could also serve as the first line of protection for all.

ACKNOWLEDGMENT

For these 2021 columns, "Body of Knowledge" takes its information from a report prepared by the IEEE Publications office on 20 November 2020, and the statistics were current as of that date. Other citation services can and do provide different numbers.

REFERENCES

1. A. Applewhite, "What knows where you are?" *IEEE Pervasive Comput.*, vol. 1, no. 4, p. 4, 2002. doi: 10.1109/MPRV.2002.1158272.
2. N. Asokan, "Anonymity in a mobile computing environment," in *Proc. 1994 1st Workshop on Mobile Comput. Syst. Appl.*, pp. 200–204. doi: 10.1109/WMCSA.1994.9.

3. R. Ballagas, J. Borchers, M. Rohs, and J. Sheridan, “The smart phone: Ubiquitous input device,” *IEEE Pervasive Comput.*, vol. 5, no. 1, p. 1, 2005. doi: 10.1109/MPRV.2006.18.
4. D. Dagon, T. Martin, and T. Starner, “Mobile phones as computing devices: The viruses are coming!” *IEEE Pervasive Comput.*, vol. 3, no. 4, pp. 11–15, 2004. doi: 10.1109/MPRV.2004.21.
5. A. Herzberg, H. Krawczyk, and G. Tsudik, “On travelling incognito,” in *Proc. 1994 1st Workshop on Mobile Comput. Syst. Appl.*, pp. 205–211. doi: 10.1109/WMCSA.1994.29.
6. N. Leavitt, “Mobile phones: The next frontier for hackers?” *Computer*, vol. 38, no. 4, p. 4, 2005. doi: 10.1109/MC.2005.134.
7. N. Leavitt, “Mobile computing: Finally a serious problem,” *Computer*, vol. 44, no. 6, pp. 11–14, Dec. 2011. doi: 10.1109/MC.2011.184.
8. C. Patterson, R. Muntz, and C. Pancake, “Challenges in location-aware computing,” *IEEE Pervasive Comput.*, vol. 2, no. 2, p. 2, 2003. doi: 10.1109/MPRV.2003.1203757.
9. R. Want, “iPhone: Smarter than the average phone,” *IEEE Pervasive Comput.*, vol. 9, no. 3, p. 3, 2010. doi: 10.1109/MPRV.2010.62.
10. S. Zdonik, M. Franklin, R. Alonso, and S. Acharya, “Are ‘disks in the air’ just pie in the sky?” in *Proc. 1994 1st Workshop on Mobile Comput. Syst. Appl.*, pp. 12–19. doi: 10.1109/WMCSA.1994.45.

DAVID ALAN GRIER is a principal with Djaghe, LLC, Washington, D.C., 20003, USA. He is a Fellow of IEEE. Contact him at grier@gwu.edu.

Digital Object Identifier 10.1109/MC.2021.3094706

EDITOR MARK CAMPBELL
EVOTEK; mark@evotek.com

Conversational Artificial Intelligence: Changing Tomorrow's Health Care Today

Mark Campbell, EVOTEK
Mlađan Jovanović, Singidunum University

Conversational artificial intelligence (AI) is making inroads into health-care administrative automation and continual care today, and its greatest potential lies in the preventive, therapeutic, and diagnostic care of tomorrow. Although conversational AI won't replace human caregivers, will those who use it replace those who don't?

The health-care industry is predicated on invention. Recent global events have placed an unprecedented demand on health-care ingenuity, way beyond the stress, sweat, and courage endured. Today's evolving health-care innovation methods create many avenues for emerging technologies to improve lives—conversational artificial intelligence (AI) is perhaps the most exciting health-care innovation outside of biotechnology labs.

"Many health-care companies are passionate about innovation but less experienced with fast, iterative software development processes. This is starting to change in a meaningful way, but we are now seeing a more agile innovation cycle, like that used to develop software, built on an adaptable and nimble experimentation platform," observes Evan Macmillan, cofounder of Gridspace, a leader in conversational AI for the health-care industry.[1] Two key examples are Pfizer/BioNTech's messenger RNA (mRNA)-based COVID-19 vaccine platform, which can quickly adapt to the new variants,[2] and CureVac's MRNA printer, which can print myriad genetic fingerprints to combat not just COVID-19 but also Ebola, Zika, and Lassa.[3]

Conversational AI has grown rapidly in sophistication and adoption over the past several years. Recently, Microsoft acquired conversational AI pioneer Nuance,[4] which, among other offerings, provides deep learning voice transcription services for health care. Other conversational AI platforms, such as Amazon Web Services' Alexa,[5] Apple

Digital Object Identifier 10.1109/MC.2021.3083155
Date of current version: 30 July 2021

Siri, Google Assistant, and Microsoft, offer easy-to-build, off-the-shelf services that accelerate the development and deployment of generalized conversational applications.[6] Although these general-purpose platforms allow for easy "assembly" of rudimentary virtual assistants, they are difficult to customize. This is because their internal models, training data, and data pipelines are inaccessible, and their sequential workflow and data-sharing structures increase response latency, which creates awkward user-agent dialogue delays.

There are other purpose-built conversational AI tools emerging that allow engineers under the hood to develop highly sophisticated applications for case-specific dialogues. One such example is Gridspace Grace, which presents a virtual assistant dialogue almost indistinguishable from a real human. Grace was recently used to help schedule COVID-19 vaccines for people without online access by calling them directly, "talking" them through the scheduling process, and answering any questions they might have. Today, conversational AI is having its greatest impact in two key areas: administrative automation and continuous care.

ADMINISTRATIVE AUTOMATION

Conversational AI has been proven an effective method in replacing human communication steps in administration workflows, such as insurance coverage verification, identify verification, scheduling, information gathering, and cross-organization communications. Once the communication steps are automatic, the entire administrative workflow can be automated using common robotic process automation tools. This frees resources and budget away from repetitive noncognitive tasks to more critical caregiving.

Patient data privacy and sovereignty and regulatory compliance will soon be competitive advantages as well as imperatives for health-care providers. Conversational AI offers a confidential and trustworthy alternative to traditional human-in-the-loop information gathering. Experiments with computational linguistics have shown since the 1960s that people will often trust the discretion of a nonjudging and discrete synthetic assistant over a human.[7] The accuracy, privacy, and efficiency that conversational AI will bring to clinical admission, discharge, and feedback will continue to change the public face of health care.

> "Many health-care companies are passionate about innovation but less experienced with fast, iterative software development processes."

CONTINUAL CARE

Hospital stays are expensive. In-hospital patient care is often extended for observation and continual care, despite studies showing that patients often recover faster at home than in a hospital.[8] Conversational AI, coupled with smart health monitoring, allows patients to transition from hospital to home care earlier, which saves costs and increases recovery rates. In many situations, the combination of smartphones, mobile applications, Internet-attached medical devices, the Internet of Things (IoT), and conversational AI all provide a platform for continued outpatient care that rivals inpatient care.

From posttreatment recovery to wellness checks, a "virtual caregiver" platform equipped with conversational AI and connected devices will revolutionize elderly care, digital therapy, and continual care. "Imagine taking home a few IoT devices that can talk to both you and your doctor from home in natural language," posits MacMillan.[1]

Dr. Milica Pejović-Milovančević, a child psychiatrist at Belgrade's Institute of Mental Health, notes that "although conversational technologies lack the human touch, they can successfully automate many activities in mental health-care provision[s]. Examples include diagnostics, reaching out and contacting new patients, progress monitoring, reminders, and initiating small talk to elicit important information from patients during depression treatment."[9]

CONCEPTUAL ARCHITECTURE

The breakneck maturation speed of conversational AI across various industries and applications in recent years means that one can develop the conceptual architecture for a virtual caregiver built on interdependent layers, as presented in Figure 1:

- The data layer contains the sources of structured (for example, personal health records) and unstructured data (such as health training data).
- At the knowledge layer, health information is retrieved or created from the existing sources or the virtual caregiver's data layer. This layer contains medical knowledge bases (that is, ontologies) and a live repository of patient-related information that is built and updated using automated data analysis and interpretation tools.
- The information generated at the knowledge layer is input for the service layer, where automated health-care decisions are made for prevention, diagnosis, and therapy.
- Once the decisions for users are ready, they are communicated to a dialog layer, which accomplishes the following tasks:
  - It processes the input information from the service layer or

from end users using natural language understanding (NLU) techniques.

  - It creates conversation threads, either by following predefined dialog rules or with probabilistic machine learning tools selected driven by the dialog management module.
  - It outputs from the dialogue management module produce responses using natural language generation tools.
- The dialog layer responses are then delivered to the presentation layer and on to the patient via an array of presentation methods (for example, text, image, voice, and multimode).

The conversational threads take a round-trip journey through the layers where the patient and/or the connected medical device requests are ingested, interpreted, and enriched with health-specific information from the data, knowledge, and service layers. This information is fed to the dialog layer, which delivers the response to the patient via the presentation layer.

THREE KEY CLINICAL USE CASES

Prevention

Conversational AI is being applied in three settings: prevention, therapy, and diagnostics. Various caregivers help maintain the specific aspects of human health. They do so by facilitating the desirable well-being behaviors connected with that particular health aspect(s) to avoid degradation and decline. The examples include physical activity, nutrition, and regular sleep. The immediate goal behind the different applications is sustained user engagement in these behavior, while the critical long-term goal is that the behaviors become a person's self-care habits or everyday routines.

Caregivers act as a companion, advisor, or coach who tries to establish deeper social bonds with their users. This way, it is assumed that users will perceive them as peers with authority, follow their inputs, and engage in target behaviors. Over time, the caregivers learn about their collocutors and personalize their suggestions to the user's needs and preferences. Forksy is an example of automated nutrition advising.[10] A female persona conducts conversations on food-related topics, keeps a diary of the user's food intake, and provides a diet schedule with nutrition recommendations and a fitness program. Ally is an embodied, virtual fitness coach that facilitates an active lifestyle.[11] It promotes a healthy way of life by offering and following personalized workout plans for weight management.

Therapy

The advancements in natural language processing technologies have enabled conversational AI to conduct more human-like conversations. Also, the ability to recognize and interpret the user's mental and emotional states during and from spoken or written discourse have made these technologies more humane and trusted. They have become a natural fit for certain medical practices, such as cognitive behavioral therapy and medication monitoring.

Conversational AI offers a confidential and trustworthy alternative to traditional human-in-the-loop information gathering.

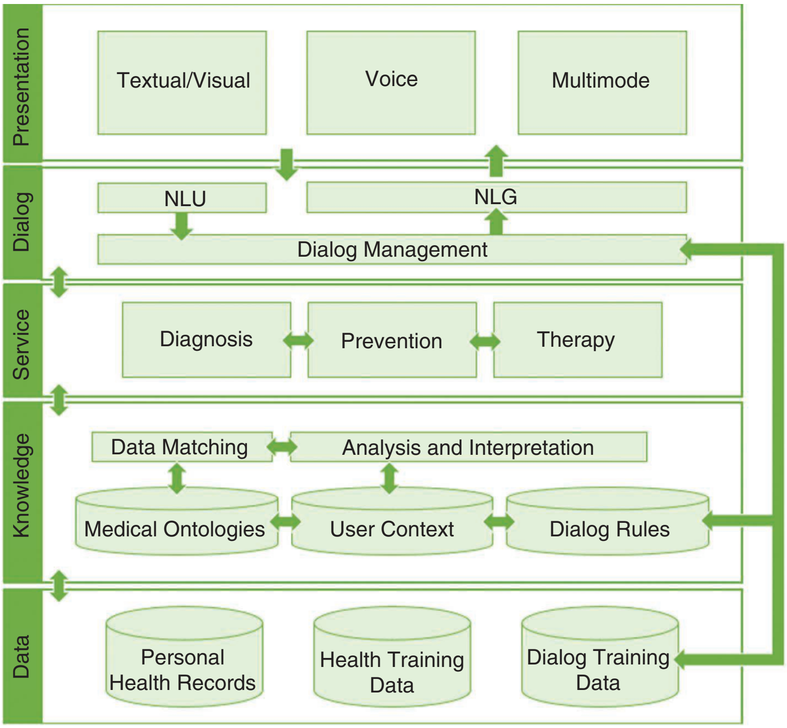


FIGURE 1. A conceptual conversational AI architecture for health care. NLU: natural language understanding; NLG: natural language generation.

Woebot[12,13] is a digital therapist that guides its users through open conversations to estimate their current mood (for example, small talk about current events and happenings in the patient's life). It then initiates structured dialog sessions, such as mirroring/paraphrasing what users have previously said or labeling/acknowledging how they might feel at the moment. The former shows the person that it is listening and understands them, while the latter tries to counterbalance negative emotions and reinforce positive ones ("You look like …" or "It seems like …"). The approach attests to positive health outcomes—neutralizing anxiety/depression[12] and reducing substance use[13] in younger people. Florence is a virtual nurse that supports users in specific treatment/rehabilitation activities.[14] It speaks with its users to remind them to take their prescribed medications(s), keep track of their health (body weight and women's health), and provide real-time information on nearby doctors/pharmacies.

In mental therapy, virtual caregivers can reduce the overburden on doctors of routine patient check-ins that complement face-to-face treatments. Dr. Pejović-Milovančević observes that, "At the moment, we have even 25 such controls per day, while the norm is up to 12." Moreover, she states, "Digital communicators are technical means to make the conversation ongoing. This is crucial for helping some technology-literate populations, such as children with speech-linguistic and nonverbal developmental issues (that is, autism)."[9]

Diagnosis

Some conversational applications can recognize symptoms of illnesses to inform users about the changes/declines in their health and recommend a course of action. They conduct a structured Q&A dialog to elicit the necessary user information. The collected data are a basis for making a diagnosis. The information is mainly subjective (self-reported by users) and may influence an accurate diagnosis. The course of action is often a medication program or doctor visit. For example, Symptomate is a diagnostic caregiver that screens its patients and provides the diagnosis for a range of conditions (roughly 700).[15] The screening is done through a conversational Q&A session from which the virtual caregiver establishes a diagnosis as a ranked list of possible conditions with a confidence level (for example, strong, moderate, or weak).

The screening is done through a conversational Q&A session from which the virtual caregiver establishes a diagnosis.

The different virtual caregivers help people quickly find information regarding their health. Even the demographic groups with a lower technological literacy, such as older adults, primarily use commercial voice assistants (such as Amazon Alexa) to search for health-related information online.[16] These caregivers extract keywords from conversations with their users and query the available knowledge bases to find answers. They can deliver responses in a textual or document form. Some caregivers, such as Gridspace Grace, automate health-care customer services, including medical advising on COVID-19.[17]

As Dr. Pejović-Milovančević highlights, "Conversational AI provides opportunities for teletherapy by engaging patients' families (that is, parents) to work with their children under the continuous supervision of doctors so they do not feel alone."[9]

A LOOK INTO THE FUTURE

There are several key areas where conversational AI in health care will grow and mature.

- *Admission avoidance*: Conversational AI can provide the autonomous screening and treatment of outpatient ailments, thus reducing the number of hospital admissions.
- *Explainability*: Conversational AI systems typically use black-box models to take a set of inputs to produce dialogue. Exposing the details of this process in a patient-friendly manner will become crucial as virtual caregivers increase their role in diagnosing a user's current health and prognosis and making treatment recommendations.
- *Transparency*: Conversational AI's capabilities in health care still fall short of human medical professionals. As this gap shrinks in the coming years, there will be increased demand for transparency on where the conversational AI limits lie and where consulting a professional is needed. Added transparency on how patient information is collected and used will also help foster trust in virtual caregivers.
- *Continuous health care*: As mentioned previously, conversational AI holds great potential for continual care, which shortens hospitalizations by providing autonomous care once the patient transitions back home. However, in the future, this will not just continual be but constant—assisting the patient "in sickness and in health" with nutrition, prevention, diagnostics, exercise, real-time checkups, and alerts.
- *Automated application creation*: Advancements in automated application development will democratize health-care solutions

creation process by enabling medical experts to develop, deploy, and test applications without involving designers and programmers.[8]

- *Multiparty communications*: The systems involving patients, health-care professionals, and families through mediated or separate peer-to-peer multiparty dialogs, while challenging, are feasible and will take on a central role in health-care communications in the future.[18]
- *Integration with legacy systems*: Aside from the practical concerns regarding security and privacy, a better connectivity of digital medical records, medical devices, hospital procedures, and health-care staff will reduce the burden on health-care services and provide faster and more effective care.

Although conversational AI is making rapid inroads in the health-care industry, particularly in administrative automation and continual care, its greatest potential could be in preventive, therapeutic, and diagnostic care. This space is maturing rapidly, and even though the future looks bright for virtual caregivers, there are significant hurdles yet to be cleared. At present, conversational AI cannot replace health-care professionals, but doctors who use it will replace those who do not.

REFERENCES

1. M. Campbell, interview with E. Macmillian, Apr. 28, 2021.
2. P. Wnuk, "Pfizer invests in mRNA vaccine powerhouse BioNTech," PharmaPhorum, Surrey, U.K. Aug. 16, 2018. [Online]. Available: https://pharmaphorum.com/news/pfizer-invests-mrna-vaccine-biontech/
3. "Epidemic group invests $34 million in potential vaccine printer tech." Reuters, Feb. 27, 2019. https://www.reuters.com/article/us-health-vaccines-curevac-idUSKCN1QG1MD (accessed May 6, 2021).
4. B. Dickson. "Why Microsoft's new AI acquisition is a big deal." TechTalks, Apr. 15, 2021. https://bdtechtalks.com/2021/04/15/microsoft-nuance-acquisition/ (accessed May 25, 2021).
5. M. Turea. "10 ways Alexa is revolutionizing healthcare." Healthcare Weekly, Apr. 8, 2021. https://healthcareweekly.com/alexa-in-healthcare/
6. P. Cox and J. M. Geyer, "Voice assistants in healthcare," Smart Business - Great Medicine, May 16, 2020. https://www.smartbusinessgreatmedicine.com/voice-assistant-technology-healthcare/
7. J. Weizenbaum and J. Weizenbaum, "Computational linguistics," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966. doi: 10.1145/365153.365168.
8. D. M. Levine et al., "Hospital-level care at home for acutely ill adults: A randomized controlled trial," *Ann. Internal Med.*, vol. 172, no. 2, pp. 77–85, Jan. 21, 2020. doi: 10.7326/M19-0600.
9. M. Campbell, interview with M. Pejović-Milovančević, May 10, 2021.
10. C. Wiedemann. "Die Therapeuten in der Tasche (The therapists in your pocket)." Frankfurter Allgemeine, Apr. 26, 2018. https://www.faz.net/aktuell/stil/quarterly/koennen-chatbots-einen-echten-therapeuten-oder-coach-ersetzen-15549081.html (accessed Apr. 21, 2021).
11. J. Zhang, Y. J. Oh, P. Lange, and Y. Fukuoka, "Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: Viewpoint," *J. Med. Internet Res.*, vol. 22, no. 9, p. e22845, Sept. 30, 2020. doi: 10.2196/22845.
12. K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, p. e19, 2017. doi: 10.2196/mental.7785.
13. J. J. Prochaska et al., "A therapeutic relational agent for reducing problematic substance use (Woebot): Development and usability study," *J. Med. Intern Res.*, vol. 23, no. 3, p. e24850, Mar. 23, 2021.
14. "Researcher develops a chatbot that already is a reference in healthcare," Phys.org, Aug. 31, 2017. https://phys.org/news/2017-08-chatbot-healthcare.html (accessed Apr. 29, 2021).
15. "Feeling sick? There's an app for that! – The big symptom checker review." The Medical Futurist, Apr. 11, 2019. https://medicalfuturist.com/the-big-symptom-checker-review/ (accessed May 13, 2021).
16. A. Pradhan, A. Lazar, and L. Findlater, "Use of intelligent voice assistants by older adults with low technology use," *ACM Trans. Comput.-Hum. Interaction*, vol. 27, no. 4, pp. 1–27, Sept. 2020. doi: 10.1145/3373759.
17. "Grace, the best conversational AI virtual agent in the market," EVOTEK, Apr. 2, 2021. https://www.youtube.com/watch?v=17X1kJbNLMA (accessed May 7, 2021).
18. R. J. Moore and R. Arar, *Conversational UX Design*. New York: ACM Books, 2019.

MARK CAMPBELL is the chief innovation officer for EVOTEK, San Diego, California, 92121, USA. Contact him at mark@evotek.com.

MLAĐAN JOVANOVIĆ is with Singidunum University, Belgrade, 11000, Serbia. Contact him at mjovanovic@singidunum.ac.rs.

Educate Back Better: A Perspective from Industry

Dejan Milojicic, Hewlett Packard Labs

COVID-19 has accelerated and scaled out remote education beyond our wildest imagination. How can we retain benefits, eliminate downsides, and build education better for the long-term future?

COVID-19 has had a detrimental impact on human existence. It dramatically affected the way we live, work, and get educated. COVID-19 has also necessitated acceleration in the development of vaccines, hospitalization, remote work, supply chains, and numerous other areas. The reverberations of COVID-19 have caused tectonic shifts in industry and consequently in the ways that individuals from industry are educated. Similarly, whole education systems, from kindergarten through university, have had to quickly switch to remote teaching.

Education had already started changing prior to the COVID-19 pandemic, from partnerships among industry and academia around industry instructors for select courses (mostly at flagship universities), to industry funding curricula (e.g., what IBM has been doing in the New York area or Hewlett Packard Enterprise in Silicon Valley and Houston), many online programs, to recent hiring policies that do not require any academic credentials. COVID-19 accelerated some of this and made it obvious that it can happen at a completely different scale than what was thought before. The challenge will be to prevent technology-facilitated education from creating an even greater gap in education for those with less access to technology.

Many industry verticals in the developed world were fortunate in that the technologies and infrastructure required for their employees to work remotely already existed. This enabled the shifts to happen very quickly and business disruptions to be kept to a minimum. However, adopting tools for in-person study to remote education are only the tip of the iceberg. The problems caused by lack of necessary Internet infrastructure run deep from sub-Saharan Africa to the poor population in Latin America as well as parts of sparsely populated regions of Australia, Canada, and Russia.[1] Going up the stack, school educators have not been retrained, curricula have not been adjusted for remote education, the critical topics for new times are

Digital Object Identifier 10.1109/MC.2021.3083915
Date of current version: 30 July 2021

completely nonexistent, and the whole system is inadequate for the needs of the future workforce.

"Build Back Better" is a strategy and movement to better prepare us for future disasters, pandemics, and dramatic changes to the way humanity lives today. COVID-19 has exposed all these problems and emphasized them. Solutions require thinking outside of the box to find new solutions to an old but now magnified problem. We need not only patch up our education system but also prepare it for many years to come and for other possible catastrophic scenarios.

HOW HAS THE LANDSCAPE CHANGED?

The world has become increasingly more interconnected, as labor can now be hired from anywhere in the world and the types of jobs have also changed to enable this.[2] Similarly, education can be obtained remotely. How does this affect education and students? In the past, schools were primarily educating students for the local labor market and "traditional jobs" for which the demand is dwindling and may soon exceed supply. Now suddenly there are no more limits on where graduating students can be hired. Similarly, competition is no longer just local schools but any remote school in the world.

New science and technology advances are dramatic, and they have radically outpaced existing curricula in most countries. Revising curricula is not trivial, and the only way out is through close work with local governments to subsidize the education system and leverage industry to support schools. Education is no longer only a matter of schools but now has become deeply intertwined with the education system as a whole. Governments and industry are ever more influencing or supporting education, and this is further influenced by societal, economic, and environment factors (see Figure 1).

Tools nowadays enable anybody to work remotely. There are still jobs that require physical presence, but many that traditionally required physical presence are evolving. Zoom, Teams, Skype, WebEx, Google Meet, and many other tools for collaboration enable secure remote meetings. A plethora of other tools enable virtual boards and document exchanges, and the workforce can still rely on the traditional Microsoft and Google suites of tools. Finally, many tools exist for teaching remotely and conducting tests, lab experiments, exams, grading, and much more, just as if you were in an in-person class. What was earlier an exception at best today is a common approach.[3]

INDUSTRY NEEDS

Most of the people who exit schools go on to work in industry (e.g., see examples and classification for the United States[4]); therefore, it is most important to understand the nature of demand (industry) to best prepare supply (schools). Industry has not always been served sufficiently well by schools, triggering corporations and even whole nations to deal with it in different ways. At the same time, industry hasn't engaged academia as openly as it should have. A more positive symbiotic relationship is needed.

In Germany, for example, vocational schools were very successful in preparing students for their ultimate occupation. However, the way it was traditionally done wouldn't be economically feasible in areas of low population density. Leveraging remote learning technologies might make it generally feasible. Large corporations establish their own training programs to prepare new employees for the specific job they will be pursuing. Professionals are used to learn by doing, but

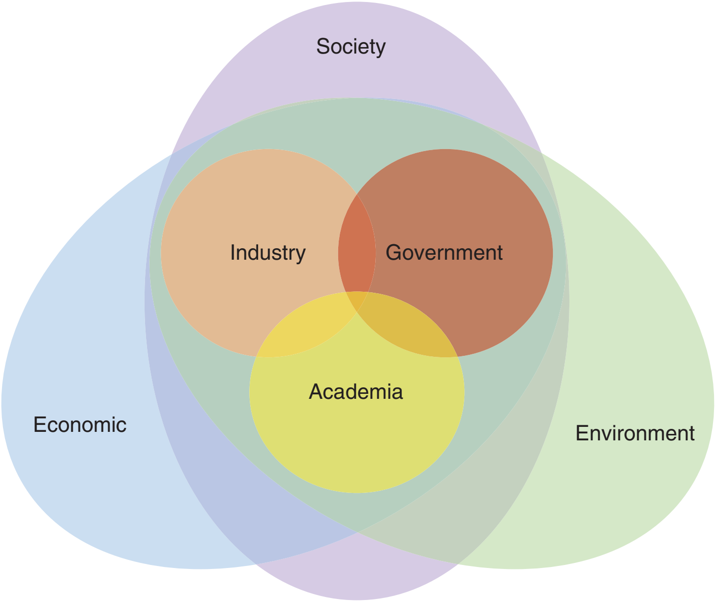


FIGURE 1. Education is ever more so deeply intertwined with industry and legislature (governments), and all three are motivated by economic, societal, and environmental factors. Economic factors are often the ones used to base decisions on, particularly if there are hard costs involved. Societal factors include diversity, equity, and inclusion; reverse migration from cities to rural areas; globalization of the workforce; and so on. Environmental factors are likely to play a bigger role in the future, given global concerns about climate change, hunger, space travel and space pollution, and so on.

as technology becomes more and more complex, months and sometimes years are wasted until a well-performing employee can start to contribute fully. Focused education, complementing experiential experience, can go a long way.

WHAT IS NEEDED?

What does industry need from new employees, from schools, legislatures, and from professional organizations like the IEEE Computer Society?

Industry would prefer a new employee who is already equipped to start working quickly, one who also has sufficiently broad knowledge to adjust as market and industry needs shift quicker than ever before. Students need to be well versed in the tools that are used in the workforce and able to contribute both remotely and in in-person work-place scenarios. New employees should be proactive in learning new skills and be aware of how to achieve them. Industry would also like quicker ways of establishing that recent graduates have the necessary skills. Recruiting/interviewing tends to be a heavyweight process for jobs that require special skills. The hiring process also needs improvements in terms of diversity.

Industry could benefit substantially from schools that can teach new and current employees updated skills, such as how to predict new trends in artificial intelligence, machine learning, and ethics. Education should be with purpose, for specific broad domains of knowledge, supplemented by special vertical courses, potentially delivered with the help of industry. A healthy exchange of visiting professors to industry and teaching professionals can help to cross these boundaries.

Industry prefers elastic scaling of the workforce, growing it as the business works well and shrinking it at times of lower business demand or changes in focus. It is not trivial to grow teams in new areas, hence the need for reeducation for repurposing the existing workforce. Governments could facilitate more effective programs and curricula for a future elastic workforce.

Industry expects that professional societies such as the IEEE Computer Society can continue to support and educate its members in the lifelong learning process. This could be achieved through publications and events that have more practical and immediate value to employee training rather than academic papers.

EDUCATION OF THE FUTURE

Given the changing educational landscape and ever-evolving industry needs, what should future education look like?

While remote education was evolving even before COVID-19, the pandemic has proven change possible on a global scale. Even postpandemic, education will retain a substantial amount of online teaching, deferring to in person only for courses that require deep interaction, such as innovation, design, and troubleshooting.

Education will organically move away from entirely independent systems and become integrated into a global ecosystem that respects the needs of individual local regions. Education will continue to maintain a holistic, broad approach, just enough to provide a sufficient base, but the rest will be with purpose to fulfill the needs of emerging industry verticals, markets, and new technologies.

Part of this broad knowledge will come from home schooling and self-education, but the focus will be on the use and application of knowledge. "Educating by doing" will increase in importance, especially complemented with practical assignment projects and internships, which all have been proven possible remotely even if not desired (interns like to travel to new sites).

Synergy between highly successful schools and industry in some unique areas such as Stanford in Silicon Valley, Santa Clara University, and San José State University will continue, but this will become more global. For example, schools will start satisfying the needs beyond immediate geographical proximity, serving the entrepreneurial community globally. Similarly, some professors and students may continue to work remotely.

Diversity, equity, and inclusion (DEI) will be at the core of all educational activities. Behavior will not be confined only to the walls of one room and go unnoticed. Recordings of unacceptable statements during lectures have raised DEI awareness and reduced tolerance for any unacceptable behavior. Statistics of diversity participation in schools has enabled industry to hire underrepresented minorities much more easily, but there is still future progress to be made in that direction.

LITMUS TEST FOR AN IDEAL HIRE

I am both a technologist and a manager, and I frequently hire new employees and, even more so, interview candidates. The best litmus test of my whole theory behind "Educate Back Better" is proved via the "eat your own dog food" model. My own ideal new hire

- has both broad and deep skills in at least one new area (see Figure 2), as evidenced by prior results
- has the ability to learn quickly, as evidenced by learning in a variety of areas and via multiple internships
- can adapt to new needs of the company and the market, as evidenced by evolving his/her own technical agenda during education or prior work

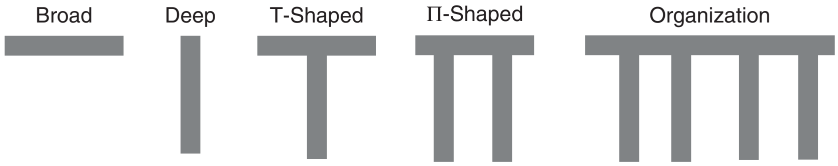


FIGURE 2. The shapes represent people with broad knowledge versus those with deep knowledge. People who have both deep and broad knowledge are rare (T-shaped) and those with two areas of depth (Π-shaped) are even more rare. However, an organization has the ability to build teams with a portfolio of technology areas that meet its needs.

- fits well with the culture and social styles of the organization, as evidenced by extracurricular activities
- fulfills self-purpose while working in the company, not just to do the job, and matches the higher goals and purpose that organization has, as evidenced by the opening statement of his/her résumé
- approaches work with a healthy dose of excitement and fun not just as a job, usually evidenced in the interview, remote or in person
- understands international and cultural differences and fully respects DEI, as evidenced by shared values.

In summary, education, just like work, and the workforce are undergoing substantial changes due to COVID-19. A large number of these changes will continue even after the pandemic is over. Education will be assisted more by industry and governments, weaving into a global ecosystem. New graduates should be much more prepared for awaiting jobs, but the traditional hiring criteria of doing the job well will be enhanced by DEI, broader purpose, and the common good. All of this will be increasingly more practiced in a connected world.

ACKNOWLEDGMENTS

I am indebted to Martin Arlitt and Ada Gavrilovska for their deeply insightful comments and to the "Future of Workforce" team (Nita Patel, Tom Coughlin, Paolo Faraboschi, Eitan Frachtenberg, Phil Laplante, Roberto Saracco, and Martin Arlitt) for motivating discussions. Their valuable inputs substantially influenced the content and presentation of this article.

REFERENCES

1. M. Arlitt et al., "Future of workforce," IEEE Comput. Soc. Rep., to be published.
2. "Will the tech workplace ever be the same again?" IEEE Spectrum Job Site. https://jobs.ieee.org/jobs/content/Will-the-Tech-Workplace-Ever-Be-the-Same-Again-2020-07-07 (accessed June 23, 2021).
3. D. Milojicic, "Autograding in the Cloud: Interview with David O'Hallaron," *IEEE Internet Comput.*, vol. 15, no. 1, pp. 9–12, Jan. 2011. doi: 10.1109/MIC.2011.2.
4. "Overview of BLS statistics by industry." U. S. Bureau of Labor Statistics. https://www.bls.gov/bls/industry.htm (accessed June 23, 2021).

DEJAN MILOJICIC is a distinguished technologist at Hewlett Packard Labs, Palo Alto, California, 94304, USA. He is a Fellow of IEEE. Contact him at dejan.milojicic@hpe.com.

Digital Object Identifier 10.1109/MC.2021.3094707

COMPUTING'S ECONOMICS

COPYRIGHT ISTOCKPHOTO, CREDIT: SITTIPONG PHOKAWATTANA

The Economics of Blockchain-Based Supply Chain Traceability in Developing Countries

Nir Kshetri, University of North Carolina at Greensboro

Shirking, cheating, and other misbehaviors are pervasive in many developing countries, increasing the costs of economic exchanges. Blockchain-based solutions may address these problems and help developing world-based economic actors to engage in exchange relationships.

According to The Netherlands-based market intelligence platform Blockdata, traceability and provenance constituted the most popular blockchain use case among the world's biggest brands in 2020.[1] This use case of blockchain is especially important in developing countries, where most serious environmental, social, and governance issues can be found.[2] Vulnerable smallholder farmers who grow subsistence and cash crops and workers in artisanal and small-scale mines (ASMs) in these countries are often exploited by powerful supply chain actors.

Due to social issues such as human rights violations and environmental injustice, some developing countries have lost the trust of developed world-based multinational enterprises.[3] To take an example, until 2019, the German manufacturer of luxury automobiles BMW sourced its cobalt needs from many countries, including the Democratic Republic of the Congo (DRC). In 2019, BMW announced its plans to avoid the DRC and source cobalt from Morocco and Australia instead, starting in 2020, for the production of electric vehicles. BMW's decision to avoid cobalt from the DRC was based on several factors, including sustainability.[4] Due to

Digital Object Identifier 10.1109/MC.2021.3082835
Date of current version: 30 July 2021

the weak rules of law, ineffective and corrupt law enforcement, and inefficient judiciary systems, many actors shirk, cheat, and engage in opportunistic behavior with impunity in countries such as the DRC.

Blockchain-based solutions have the potential to address these types of problems and help developing world-based economic actors engage in exchange relationships. Sustainability-related indicators can be measured with emerging technologies such as artificial intelligence (AI) and machine vision, using data from Internet of Things devices, remote sensing satellites, and other sources. Blockchain can help establish the authenticity of measurement data. Blockchain-based smart contracts, which execute automatically when certain conditions are met, can address some of the deficiencies of existing problems associated with contract laws and their enforcement in these countries. All of these features can increase efficiency and help developing world-based economic actors to build and gain the trust of the counterparties in economic transactions.

In light of these considerations, in this article, we look at how blockchain can help developing world-based economic actors engage in exchange relationships. We illustrate this by considering several blockchain-based traceability solutions implemented in developing countries.

BLOCKCHAIN-BASED SOLUTIONS' EFFECTS ON THE COSTS OF EXCHANGE

Among the key factors that influence the costs of exchange are 1) the costs of measuring and 2) the costs of enforcement.[5] Regarding 1), measuring the dimensions and attributes of the goods and services being exchanged or the performance of agents is not an easy task.[5] Blockchain-based traceability systems can provide new methods for accurate measurements to describe precisely what the parties engaged in a transaction are exchanging and what performance characteristics can be expected.

Regarding 2), in a society characterized by perfect contract enforcement, a neutral third party impartially evaluates disputes and awards compensation to the party affected by a violation. In such a situation, opportunism, shirking, and cheating are not attractive options. However, the real world is far from ideal. The high costs of measurement often make it difficult to determine if a contract has been violated and, if so, who violated it. Many developing economies have a weak rule of law and lack well-developed court systems and state's coercive power to enforce judgments. It is difficult to employ complex contracting as a formal governance tool.

Several blockchain solutions have been launched that have the potential to address measurement and enforcement issues. In Table 1, we look at the effects of blockchain-based traceability systems on the costs of measuring and enforcement.

MEASURING

In the previous BMW example, the company's decision not to source cobalt from the DRC was due to the difficulty in determining whether the mining companies' practices were sustainable. Nongovernmental organizations and activists are promoting corporate social responsibility by naming and shaming companies that are responsible, knowingly or unknowingly, for human rights violations and child abuse.

Some blockchain-based traceability solutions have been launched to

TABLE 1. The effects of blockchain-based traceability systems on measuring and enforcement.

Effect	Explanation	Examples
Facilitating measurements	Measure the attributes of goods and services/performance of agents that otherwise cannot be measured	Circulor's system captures data related to cobalt's origin, attributes, and supply chain participants' actions
Lowering costs of measurement	By automating, costs of measuring can be reduced	In the solutions of Circulor, Bext360, and BlocRice, costs to small commodity producers are lower than the alternatives
Increasing accuracy of measurement	In many cases, compared to humans, machines can provide more accurate and objective measurements of the attributes of goods and services or the performance of agents	Bextmachines use machine vision and AI to analyze coffee cherries and coffee parchment
Strengthening contract enforcement with transparency and documentary evidence	A higher degree of authenticity can be achieved in documentary evidence, such as contract documents, which can make contract enforcement more effective and less costly	eMin tool to benefit migrant workers in the seafood industry Parties in BlocRice's contract include organic farmers and rice exporters in Cambodia and buyers in The Netherlands.

address measurement challenges in the context of sustainability practices in developing countries. With such solutions, it is possible to measure the attributes of goods and services or the performance of agents that otherwise cannot be measured and establish the authenticity of measurement data.

The U.K.-based traceability-as-a-service provider Circulor operates a blockchain platform to monitor cobalt from the DRC used in electric vehicle batteries.[6] Circulor combines blockchain with AI to perform due diligence, detect data anomalies, and identify actions that may need additional investigation. The data captured include the cobalt's origin, attributes (for example, weight and size), the chain of custody, and information to establish supply chain participants' compliance with globally recognized guidelines.[7]

In July 2020, Volvo Cars' venture capital investment arm Volvo Cars Tech Fund teamed up with other investors to further develop Circulor's traceability system.[8] The new funding was intended to train and improve Circulor's machine learning models to distinguish between children and adults working in the mines with a high level of accuracy, using the data obtained from aerial imagery of mining.[9]

Also, some blockchain-based solutions are more affordable than alternative technologies used in establishing and demonstrating supply chain traceability. Table 2 presents three such solutions. For instance, the International Tin Industry Association Tin Supply Chain Initiative's (ITSCI's) bagging and tagging system is an established traceability program, which was started in response to the Dodd–Frank Wall Street Reform and Consumer Protection Act, which requires U.S. companies to vet their supply chains.[10] Countries that are covered under this legislation include South Sudan, Uganda, Rwanda, Burundi, Tanzania, Malawi, Zambia, Angola, Congo, the Central African Republic, and the DRC. The ITSCI does not use blockchain. Complaining about the high costs of the ITSCI at a 2019 mining forum in Kigali, Rwanda, the chief executive officer of Rwanda Mines, Petroleum and Gas Board, demanded that "the cost of traceability and due diligence must be reduced to make it affordable and fair."[11] An ASM producing 0.5 ton per month is required to pay US$780–1,080/year to use ITSCI traceability. Circulor has said that its system will change the business models of ASMs by shifting traceability costs from miners to end users.[12] Its mobile app is free for small companies, whereas companies further up the supply chain pay and use more complicated interfaces.[13]

Some blockchain-based traceability systems automate measurements. Such systems replace labor-intensive activities, such as physical inspection and paper work, which can reduce the costs of measuring. In the coffee industry, such audit tasks are performed by intermediaries such as certification agencies, which are estimated to cost as high as US$0.91 per pound of coffee.[14] For instance, in the agriculture sector, a key role of middlemen is to reduce the measurement cost problem.[15] Middlemen are more likely to visit the farm during cultivation and harvest compared to urban wholesalers and bigger traders.[16]

An example of a company providing a blockchain-based traceability system that has eliminated labor-intensive audit tasks is Denver, Colorado, based startup Bext360. Its kiosks in Uganda and Ethiopia evaluate coffee beans using a Coinstar-like device known as a *bextmachine*, which employs smart image recognition technology machine vision, AI, and blockchain to grade and track coffee beans. Bext360's systems store data related to the time, date, and location of transactions and the amount paid. The systems also record indicators related to sustainable sourcing and satellite images to determine if producers are polluting water.[17]

As another example, in 2018, the charitable organization Oxfam launched the Blockchain for Livelihoods From Organic Cambodian Rice (BlocRice) project in Cambodia. It uses blockchain to improve Cambodian small-scale rice farmers' bargaining and negotiating power by storing relevant data on a blockchain system. BlocRice is arguably a lower cost social certification mechanism compared to alternatives such as FairTrade.[18]

Blockchain systems would allow for accurate and objective measurement of the attributes of the goods and services being exchanged or the performance of agents. This aspect is important because smallholder farmers often get paid low wages due to the subjective quality assessment by powerful value chain actors, such as middlemen and industrial buyers.[19] In the bextmachine example, machine vision and AI, rather than human beings, measure the quality of coffee. Bextmachines take a 3D scan of each bean's outer fruit to analyze coffee cherries and coffee parchment.[20] Farmers who supply bigger and riper cherries are paid more.

TABLE 2. Some blockchain solutions to trace/track commodities.

Blockchain solution	Implemented in	Commodities traced/tracked	Cost performance in relation to available alternatives
Circulor's platform	The DRC	Cobalt	ASMs do not pay for traceability
Bext360	Ethiopia and Uganda	Coffee	Because of automated processes, costs are lower than those of certification agencies
BlocRice	Cambodia	Rice	Lower costs compared to FairTrade

The bextmachines link the output to cryptotokens, which represent the coffee's value. New tokens are automatically created when the product passes through the supply chain. The values of tokens increase at each successive stage of the supply chain.[14]

ENFORCEMENT

Blockchain-based solutions can also help address human rights issues, such as slave labor and the exploitation of migrant workers. For instance, the marine fishing industry exhibits a high propensity to use "slave" or underpaid labor due to its huge size and the lack of enforcement mechanisms. Most of the workers in the Thai fishing industry are migrants from Cambodia and Myanmar. These workers are paid about 25% lower than the Thai minimum wage. They often sign a contract in their home country, but that changes when they arrive in Thailand.[21]

The blockchain solutions provider Diginex has been working with the International Organization for Migration and the antislavery organization the Mekong Club to ensure ethical recruitment of migrant workers by increasing the transparency of workers' contracts.[22] Its blockchain-based mobile app eMin stores copies of employment contracts for workers in the aquaculture sector.[21] Workers can access their contracts on the Ethereum blockchain, which can be used as a basis for claiming the rights and benefits offered when they were recruited.[21]

Likewise, in BlocRice, key parties involved in the contracts include agricultural cooperatives, organic farmers, and rice exporters in Cambodia and buyers based in The Netherlands. The term of the contract is that the exporter will pay farmers the market price plus a premium.[23] Such a condition would guarantee a market for the rice and reduce some uncertainties for farmers.[23] Each farmer receives a digital identity, which can be used to log onto a website to see details such as shipment weights and prices. The information is available in both Cambodian and English.[24]

ENABLERS AND INHIBITORS OF BLOCKCHAIN DEPLOYMENT IN DEVELOPING COUNTRIES

Many factors have facilitated the implementation of blockchain-based traceability systems in the developing world. Because of the increasing competition in the area of enterprise blockchain, there are several quick and easy options to develop blockchain projects in supply chains. For instance, enterprise blockchain solutions based on Hyperledger Fabric, which are used in Circulor's traceability system, are offered by a number of technology companies, such as IBM, Amazon Web Services, SAP, Oracle, and Microsoft. For instance, SAP provides Hyperledger Fabric on its cloud platform, and Microsoft offers this solution on Azure.[25] Companies do not need to worry about infrastructure, storage, and networking costs. A ready-made blockchain platform also has sharing, encryption, a consensus algorithm, and a peer-to-peer network.[25]

Such systems are becoming affordable, and their ease of use is improving. Using Circulor's system, for instance, small mining companies do not see an increase in their workload. The mobile app is easy to use.[26] Once the miners open the app, there are three buttons on the front page. A step-by-step process is presented by clicking "Start." The process begins with facial recognition. The next process involves the scanning of a tag to enter the details of minerals such as cobalt.[26]

Blockchain's future potential is even greater. For instance, blockchain-based smart contracts in combination with automated payments would be the game changer in the agriculture sector. Such systems can make it possible for small-scale farmers to be automatically paid when the produce is delivered.

Many challenges remain, however. First, a large proportion of populations living in rural areas in least developed countries (LDCs), which are low-income countries that perform poorly in human assets and face high economic vulnerability, lacks connectivity (Figure 1). The penetration rates of mobile devices such as smartphones are low

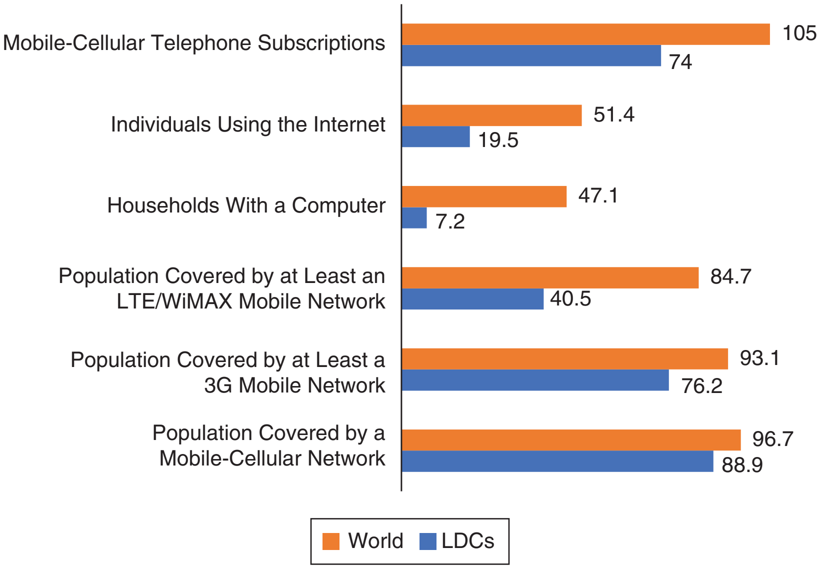


FIGURE 1. A comparison of connectivity indicators in LDCs and the world (in percentage of the population). Data for computer use and Internet use are for 2019. Others are for 2020. (Data source: International Telecommunications Union.)

in these countries. For instance, in 2019, only 10–20% of the farmers had smart phones in Cambodia.[13] These factors increase the costs of establishing and operating traceability systems. Some farmers are illiterate and thus cannot take advantage of blockchain-based traceability systems.

The implementation of smart contracts also requires high-quality data, such as those related to the weather. In the absence of such data, blockchain-based solutions such as BlocRice provide only a small improvement over the current alternatives. The availability of risk-sharing and risk-transfer mechanisms such as farm insurance is critical to improve the livelihood and development of farmers. In addition, the regulatory frameworks to support blockchain-based innovations such as smart contracts are lacking.

Blockchain-based solutions have been designed to measure sustainable practices, which would help reward responsible and ethical behaviors and penalize unethical and irresponsible ones. Such solutions can help economic actors in developing countries to engage in exchanges as well as reduce the costs of and maximize the benefits from an exchange relationship. Since a key role of middlemen in these countries is to reduce the measurement cost problem, automated measurements of blockchain-based traceability systems can eliminate intermediaries. This would allow economic actors with limited resources, such as smallholder farms and ASMs, to increase their incomes. C

REFERENCES

1. A. Fenton. "Blockchain traceability overtakes payments among major corporations." Cointelegraph, 2020. https://cointelegraph.com/news/blockchain-traceability-overtakes-payments-among-major-corporations (accessed May 15, 2021).
2. "Briefing transparency." Sedex, 2013. https://www.sedex.com/briefing-transparency/
3. W. Clowes. "BMW to source cobalt directly from Australia, Morocco mines." Bloomberg, Apr. 24, 2019. https://www.bloomberg.com/news/articles/2019-04-24/bmw-to-source-cobalt-directly-from-mines-in-morocco-australia (accessed Mar. 11, 2021).
4. "Making mining safe and fair: Artisanal cobalt extraction in the democratic republic of the congo," World Economic Forum, Geneva, Switzerland, White Paper, Sept. 2020. [Online]. Available: http://www3.weforum.org/docs/WEF_Making_Mining_Safe_2020.pdf
5. D. C. North, "Dealing with a Nonergodic World: Institutional Economics, Property Rights, and the Global Environment," *Duke Environ., Law, Pol. Forum*, vol. 10, no. 1, pp. 1–12, 1999.
6. N. Rolander. "Volvo cars goes for blockchain tech to avoid unethical cobalt." Bloomberg, 2019. https://www.bloomberg.com/professional/blog/volvo-cars-goes-for-blockchain-tech-to-avoid-unethical-cobalt/ (accessed May 15, 2021).
7. R. Wolfson. "Volvo adopts Oracle's blockchain for its supply chain — Here's why." Cointelegraph, 2019. https://cointelegraph.com/news/volvo-adopts-oracles-blockchain-for-its-supply-chain-heres-why (accessed May 15, 2021).
8. "Volvo Cars Tech Fund invests in blockchain technology firm Circulor," Volvo Cars, Gothenburg, Sweden, 2020. [Online]. Available: https://www.media.volvocars.com/global/en-gb/media/pressreleases/269598/volvo-cars-tech-fund-invests-in-blockchain-technology-firm-circulor
9. M. Kapilkov. "Startup helps reduce child labor in Africa & aspires to work with Tesla." Cointelegraph, 2020. https://cointelegraph.com/news/startup-helps-reduce-child-labor-in-africa-aspires-to-work-with-tesla (accessed May 15, 2021).
10. "iTSCi joint industry traceability and due diligence programme." ITSCI, 2016. https://www.itsci.org/wp-content/uploads/2017/01/iTSCi-Booklet-2016-.pdf (accessed May 15, 2021).
11. J. Bizimungu, Rwanda protests mineral traceability scheme The New Times; 2019, https://www.newtimes.co.rw/news/rwanda-protests-mineral-traceability-scheme
12. C. Mwai. "Will blockchain fix the mineral traceability woes?" The New Times, 2018, https://www.newtimes.co.rw/news/will-blockchain-mineral-traceability (accessed May 15, 2021).
13. N. Kshetri, *Blockchain and Supply Chain Management*. New York: Elsevier 2021.
14. "World's first blockchain coffee project." Moyee Coffee Ireland, 2018. https://moyeecoffee.ie/blogs/moyee/world-s-first-blockchain-coffee-project (accessed Mar. 11, 2021).
15. Y. Barzel, "Measurement cost and the organization of markets," *J. Law Econ.*, vol. 25, no. 1, pp. 27–48, 1982. doi: 10.1086/467005.
16. G. K. Abebe, J. Bijman, and A. Royer, "Are middlemen facilitators or barriers to improve smallholders welfare in rural economies? Empirical evidence from Ethiopia," *J. Rural Stud.*, vol. 43, pp. 203–213, Feb. 2016. doi: 10.1016/j.jrurstud.2015.12.004.
17. C. Zhong. "Innovator BanQu builds blockchain and bridges for traceability, small farmers' livelihoods." Greenbiz, 2019. https://www.greenbiz.com/article/innovator-banqu-builds-blockchain-and-bridges-traceability-small-farmers-livelihoods (accessed Mar. 11, 2021).
18. J. Hallwright and E. Carnaby. "Complexities of Implementation: Oxfam Australia's Experience in Piloting Blockchain." Frontiers, 2019. https://www.frontiersin.org/articles/10.3389/fbloc.2019.00010/full (accessed Mar. 11, 2021).

19. L. T. Phelan, “Adding value to smallholder forage-based dual-purpose cattle value chains in Nicaragua, in the context of carbon insetting,” International Center for Tropical Agriculture, San Diego, CA, 2015. [Online]. Available: http://ciat-library.ciat.cgiar.org/articulos_ciat/biblioteca/Adding_Value_to_Smallholder_Forage-Based_Dual-Purpose_Cattle_Value_Chains_in_Nicaragua_in_the_context_of_Carbon_Insetting-Thesis-Lisette%20Phelan.pdf

20. Z. Cadwalader. “Trace your coffee using blockchain.” Sprudge, 2018. https://sprudge.com/132380-132380.html (accessed Mar. 11, 2021).

21. “Detecting modern slavery in the supply chain.” Word on the Streets, 2020. https://wordonthestreets.net/Articles/560874/Detecting_modern_slavery.aspx (accessed Mar. 11, 2021).

22. T. Dao. “Companies ink deal to use blockchain for protecting Thai aquaculture sector workers.” Seafood Source. https://www.seafoodsource.com/news/aquaculture/companies-ink-deal-to-use-blockchain-for-protecting-thai-aquaculture-sector-workers (accessed Mar. 11, 2021).

23. “Can blockchain help rice farmers fight poverty?” Oxfam. https://cambodia.oxfam.org/can-blockchain-help-rice-farmers-fight-poverty (accessed Mar. 11, 2021).

24. K. Cottrill. “Blocrice makes a case for blockchain in smallholder farming.” Chain business Insights, 2019. https://www.chainbusinessinsights.com/insights-blog/blocrice-makes-a-case-for-blockchain-in-smallholder-farming (accessed Mar. 11, 2021).

25. “How much does it cost to build a blockchain project?” DevTeam. Space, 2020. https://www.devteam.space/blog/how-much-does-it-cost-to-build-a-blockchain-project/ (accessed Mar. 11, 2021).

26. M. Bennett. “Blockchain app for miners in Rwanda ensures the minerals in your iPhone are conflict-free.” Diginomica, 2019. https://diginomica.com/blockchain-app-for-miners-in-rwanda-ensures-the-minerals-in-your-iphone-are-conflict-free (accessed Mar. 11, 2021).

NIR KSHETRI is a professor at the Bryan School of Business and Economics, the University of North Carolina at Greensboro, Greensboro, NC, 27412, USA, and the “Computing’s Economics” column editor for *Computer*. Contact him at nbkshetr@uncg.edu.

Digital Object Identifier 10.1109/MC.2021.3094708

Making Open Source Project Health Transparent

Sean P. Goggins, University of Missouri

Matt Germonprez and Kevin Lumbard, University of Nebraska Omaha

We explore the Community Health Analytics for Open Source Software project and how it plays an integral role in the automation of key measures to make the state of open source readily observable.

We are long past discussions about the benefits of proprietary software versus open source software (OSS). The world we live in is built on OSS. The software exists in complex infrastructures and supply chains, often alongside proprietary programs. It is more ubiquitous and complex than ever, and it is continuing to grow. Corporations have embraced OSS in a way few could have imagined 40 years ago. For example, automotive-grade Linux is deployed in dozens of vehicle models, and Kubernetes enables massive on-demand scaling for a wide range of online firms. All this begins with a project or Git repository, where bugs are fixed, features are added, and discussions about the inclusion of contributions take place. When OSS projects are used by others, they become part of a larger supply chain composed of many released versions that delivers value to software consumers. The projects and products depend on yet another layer of OSS: the dozens to thousands of libraries imported by each software component. OSS's ubiquity is making it more visible, and its complexity is making it more difficult to manage. In this article, we illustrate how a five-year project, Community Health Analytics for Open Source Software (CHAOSS), helps improve the effectiveness of the people who build, maintain, contribute to, and consume OSS in an interconnected world.

As Gonzalez-Barahona[1] describes in an earlier column, the history of free and open source software (FOSS) is largely the story of computer scientists laboring to

Digital Object Identifier 10.1109/MC.2021.3084015
Date of current version: 30 July 2021

0018-9162/21©2021IEEE

EDITOR DIRK RIEHLE
Friedrich Alexander-University of Erlangen Nürnberg;
dirk.riehle@fau.de

make core computing functions available across an expanding number of architectures. The growth of contemporary software infrastructure is the cousin of early FOSS work, and it persists with increasing corporate sponsorship through paid contributors and organizations, such as the Linux Foundation. The motivation for OSS infrastructure work is moving beyond the core value of open production, increasingly centering on the open sourcing of corporate intellectual property that a firm, or a collection of firms, determines is essential and nonmarket differentiating. If we all need it, why not share the cost of maintenance and evolution?

Individuals and organizations from a significantly more diverse set of domains than the early days are also working on OSS projects and delivering OSS products. The United Nations Children's Emergency Fund maintains a portfolio of more than 800 OSS projects aimed at attaining policy goals, social good, and greater diversity in the OSS contributor pool. Scientists researching diseases, biological and plant genomes, and pharmacological treatments depend, to a growing extent, on a substantial OSS ecosystem. Journalists are using OSS to close information and skill gaps within individual, increasingly resource-starved newsrooms to fulfill their roles, and even video game designers are starting to build their work using OSS engines, such as Godot (https://godotengine.org/). If you use a computer, drive a car, or purchase groceries, it is a virtual certainty that a number of OSS products interact to ensure your success. The realization that our world is constructed on OSS illuminates our need to understand how our built environment shapes our lives[4] and how we can continue to structure and maintain a world we want to inhabit.

FROM THE EDITOR

Hello everyone, and welcome back to the "Open Source Expanded" column! We are well into the open source community theme now. In this instance, Sean Goggins and colleagues from the Linux Foundation's Community Health Analytics Open Source Software project discuss how to identify healthy open source project communities and determine if something is going wrong. Practical metrics are important for any open source project leader, and our experts have a story to tell. Enjoy, and as always, stay safe and healthy, and keep on hacking. — *Dirk Riehle*

Diversity, ubiquity, and complexity within each OSS application adds a responsibility for computer scientists and others to be aware of the health and sustainability of their projects and of those they depend on. One result of our collective recognition of OSS complexity awareness is the formation and development of the Linux Foundation's CHAOSS project (https://chaoss.community), which is the product of active engagement from hundreds of OSS maintainers, contributors, corporations, and domains of construction and use. In the remainder of this article, we describe the CHAOSS approach for addressing OS health and sustainability, the project's core focus areas, and how the project's software (Augur) plays an integral, ethically grounded role in the automation of key measures that make project growth, risk, value, and potential transparent at today's OSS scale.

OSS HEALTH

CHAOSS develops tools to support consistency for OSS maintainers and other stakeholders in their individual assessments of projects and ecosystem health and sustainability. While mundane in appearance, critical historical gaps have been closed through the project. For example, OSS health and sustainability metrics originate from earlier tools focused on measuring commit activity, which the authors outline in a recent review of the literature.[2] Activity metrics are helpful but incomplete for understanding contemporary health and sustainability questions that organizations, individuals, and foundations ask about OSS project portfolios that often number in the thousands, usually include dependencies on many other OS projects, and influence corporate valuations.[3] As OSS began experiencing something of a Cambrian explosion, the need for CHAOSS reached critical mass in 2017.

CHAOSS recognizes that the growing complexity associated with OS work, when mitigated through research that boosts visibility beyond activity metrics, is likely to accelerate innovation by increasing the adoption of shared, essential resources in a larger number of cases. For example, within organizational boundaries, key contributors are known. However, in OS projects, the dynamics of contributor turnover can create uncertainty about incorporating work into commercial products. The effects of that turnover and the limits to its visibility heighten perceptions of risk. The lack of visibility, like other issues, can have direct and lasting impacts on the work found in OS projects.

Most organizations (for example, corporations, nonprofits, collaboratives, and universities) that engage in OS projects rely on a small number of experts who use heuristics to assess opportunities and approximate the value and risk of participation. Prior to the CHAOSS project, metric definitions were idiosyncratic; the tools were a bricolage of homegrown and small-scale OSS projects. Organizations had difficulty consistently understanding the return on their OS investments, especially more strategic ones that crossed ecosystems and included competitors. Metrics and tooling that reach beyond activity measures make OS projects work and the subsequent evolution of OSS more visible. The growing uptake of standardized metrics and software from CHAOSS is helping organizations assess risk and value in ways that overcome the useful but haphazard assessments that are commonplace.

In a drive to make OS projects more sustainable, the CHAOSS project has published more than 55 metrics and tools aimed at lifting the veil of complex interdependencies that encumber sustainable growth. In the sections that follow, we define the scope of five core working groups and their focus areas that increase the visibility of multiple dimensions related to project health. While metrics are organized within the working groups that develop them, you may recognize alternate, potentially more useful structural presentations, and we welcome those suggestions on our mailing list (https://lists.linuxfoundation.org/mailman/listinfo/chaoss). Our working group structure and the developed metrics are illustrated in Figure 1.

Five core working groups

CHAOSS metrics for code development activity and quality as well as issue resolution, efficiency, and community growth are devised by the Evolution Working Group (https://github.com/chaoss/wg-evolution). Long-standing metrics focused on commit activity are developed and maintained primarily within this group. Most of the activity metrics that are not produced there are part of the Common Working Group (https://github.com/chaoss/wg-common), which also creates measurements that are of interest to multiple other groups. The Risk Working Group (https://github.com/chaoss/wg-risk) maintains metrics focused on license coverage, Construction Industry Institute best practices, and hazards pertaining to maintainer diversity. More recently, it has defined metrics, measures, and resource lists for understanding the increasing complexity of dependencies between projects. Dependency concerns are especially prominent in the work of OS program offices, community managers, and project maintainers.

The Value and the Diversity, Equity, and Inclusion (DEI) Working Groups develop and maintain metrics that are more difficult to derive solely from Git platforms, issue trackers, and electronic project communication. For example, the Value Working Group (https://github.com/chaoss/wg-value) develops metrics for assessing project popularity, labor investment, and OSS as a social good. The DEI Working Group (https://github.com/chaoss/wg-diversity-inclusion/) crafts metrics focused on event inclusivity and that aim to raise awareness of project practices, such as mentorship and managing burnout, and other factors that have been shown to foster or erode inclusiveness. These two groups help CHAOSS move beyond measuring health and to advance the sustainability of OSS as a whole.

GROWING OSS CONTRIBUTION

CHAOSS has generated concrete metrics for more than 10,000 OS projects and implemented reporting systems tailored to the needs of several dozen corporate organizations. The most prominent types of analysis desired by project maintainers relate to the retention of contributors and the responsiveness of maintainers to contributions. In some cases, projects have focused on competitive analysis along the lines of maintainer responsiveness, recognizing that faster replies are more likely to keep contributors engaged. Figure 2 illustrates a competitive analysis of maintainer responsiveness, and Figure 3 describes the integration of several CHAOSS metrics focused on contributor retention.

To make OS projects more sustainable, CHAOSS focuses on making work visible. Our members apply theories of organizational development and analytical tools that advance visibility to support the exponential growth and increasing interdependency that OSS is undergoing. As CHAOSS cofounders, maintainers, and board members, we apply our deep, embedded fieldwork in concert with machine learning and network science to facilitate an energetic response to these changes. CHAOSS recognizes that software engineering is part of adapting to this phase shift. In that context, one risk to future software engineering practices is that social, organizational, and technical responses to change will unintentionally replicate existing approaches that are incommensurate to a problem. For example, more of the work is performed by a less diverse collection of people, who build OSS for pay, than we find in other professions. Advancing inclusivity in OS work and the now ubiquitous impact OSS has on society are growing focuses of the CHAOSS project.

DEI event badging

The reshaped world of OSS demands that CHAOSS not only standardize metric definitions and tool kits but innovate with programs that strive to increase the workforce through greater diversity and inclusion. Our DEI event badging program (https://chaoss.community/diversity-and-inclusion-badging/) has already recognized 21 major initiatives in the Kubernetes, cloud-native, and other OSS communities (https://github.com/badging/event-diversity-and-inclusion), following an open peer-review process modeled after *Journal of Open*

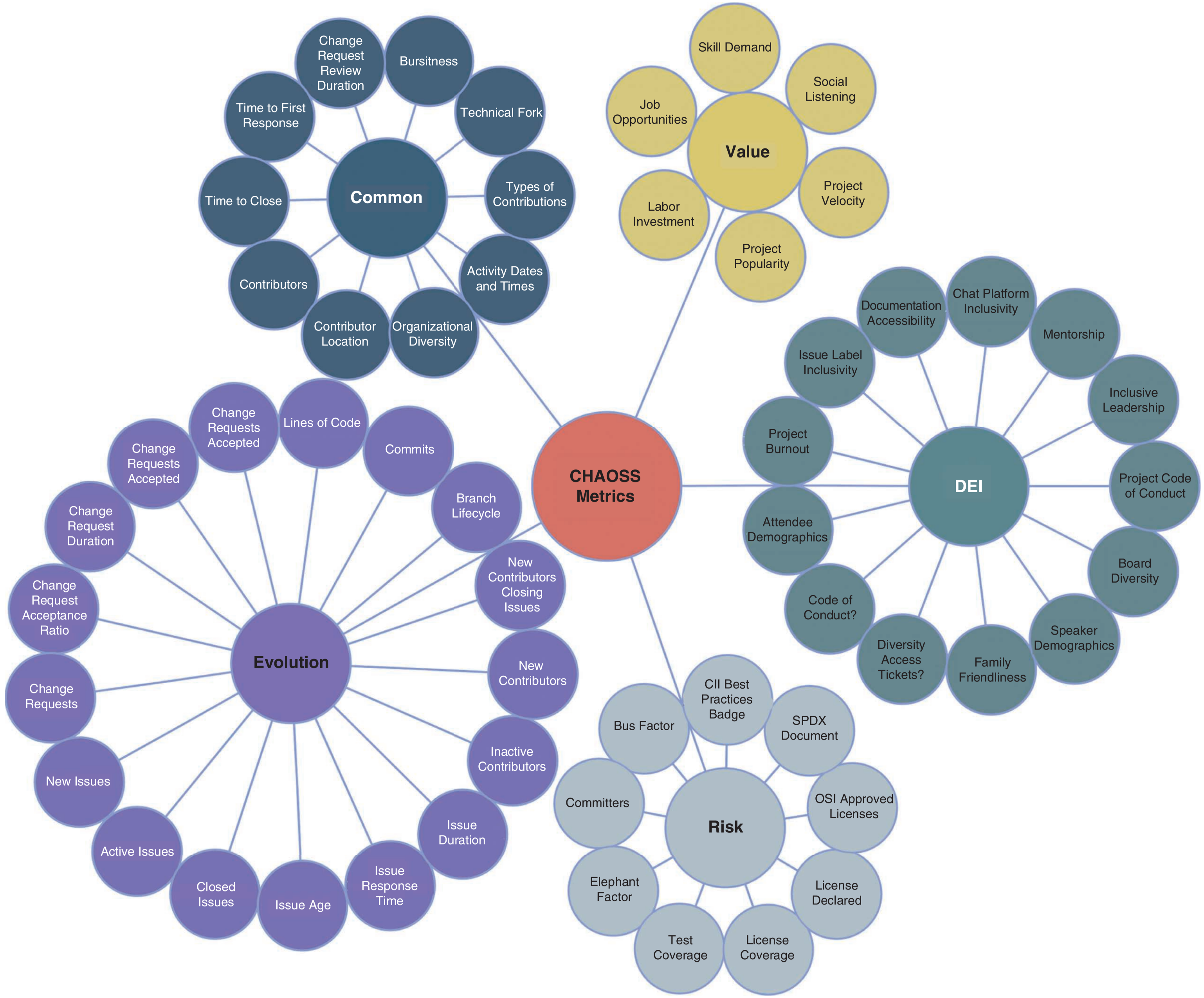


FIGURE 1. The 57 CHAOSS health and sustainability metrics organized by five working groups: Risk; Evolution; Value; Common; and Diversity, Equity, and Inclusion (DEI). CII: Construction Industry Institute; SPDX: Software Package Data Exchange; OSI: Open Source Initiative.

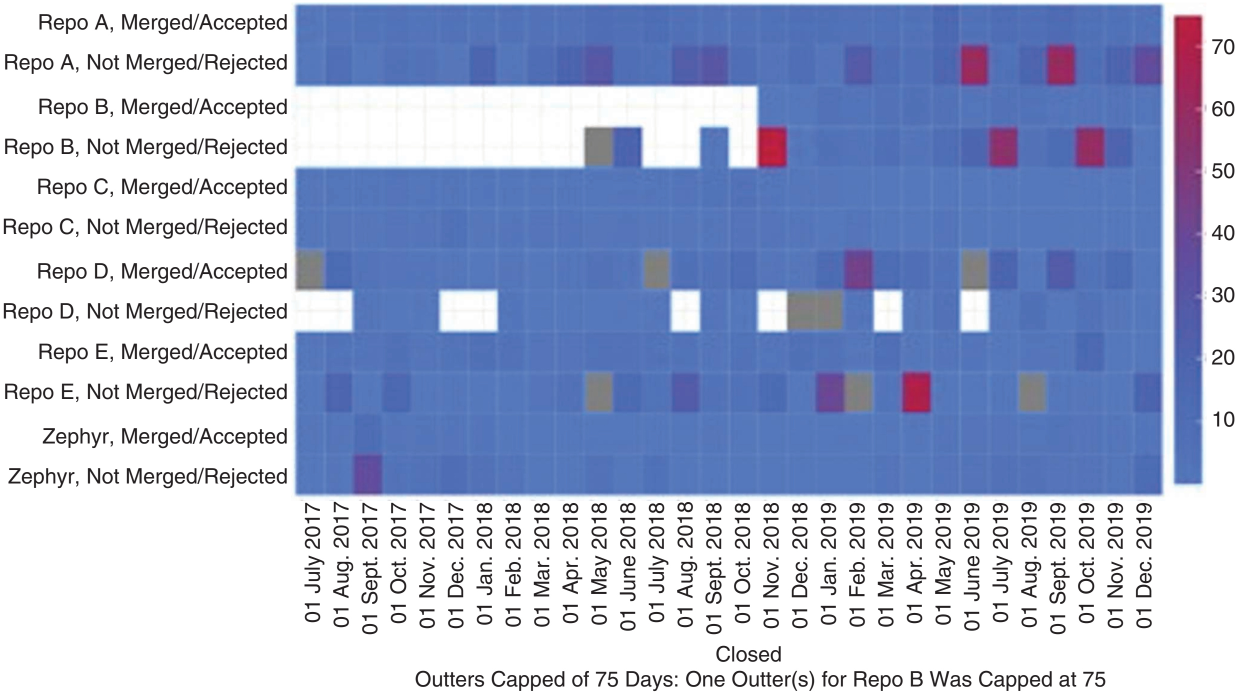


FIGURE 2. Zephyr is a real-time operating system for critical infrastructure. This analysis compare's Zephyr's pull request responsiveness to that of competing products through time, using a heat map-style visualization. The competitors have been anonymized to maintain a focus on Zephyr.

Source Software (https://joss.theoj.org/). Such rapid success illustrates a wider recognition of DEI as central to managing the changed nature of OSS. The badging program has several tiers based on the number of CHAOSS DEI metrics attained, in much the same way the Core Infrastructure Initiative (https://bestpractices.coreinfrastructure.org/en) issues badges, as shown in Figure 4.

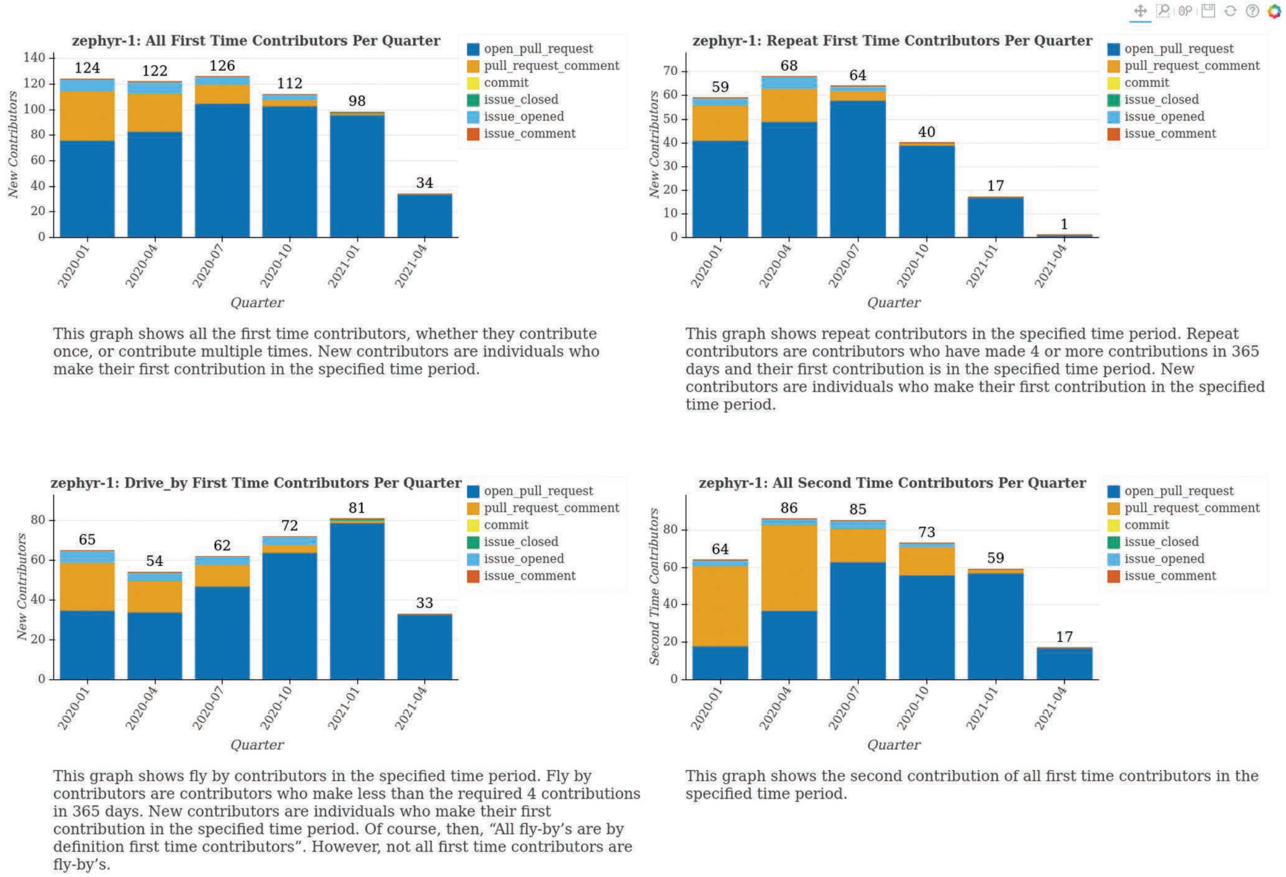
TOOLS FOR SCALING HEALTH AND SUSTAINABILITY AWARENESS

Developing and distributing common definitions for OSS health metrics includes collaboration among hundreds of OSS organizations that recognize the urgent need to move beyond activity metrics, yet a dictionary of sorts is of limited use without implementation in software. CHAOSS tools, such as GrimoireLab (https://github.com/chaoss/grimoirelab) and Augur (https://github.com/chaoss/augur; https://github.com/chaoss/augur-community-reports), implement CHAOSS metrics and present them in ways that enable maintainers, contributors, and other stakeholders to draw inferences about the relative health and sustainability of their projects by using indicators whose consistency, if not perfection, can be trusted.

Analyzing information within ecosystems

Prior to the CHAOSS project's introduction of Augur, OSS metrics collection and persistence focused on the analysis of individual efforts and predefined collections of initiatives, using definitions that were specific to each tool. Augur's design supports the phase shift in the number of projects and the dependencies between them by collecting lists of repositories where dependencies are managed and where contributors and maintainers work, beyond the predefined scope of analysis. This type of "snowball collection" of basic information makes it possible for OS program offices, community managers, scientists, and other stakeholders to take a peek at parts of their infrastructure that were not visible before CHAOSS consistently defined metrics and developed tools to address contemporary challenges.

CHAOSS recognizes that ecosystems are defined by the goals and questions of each OSS stakeholder. A science OS ecosystem is typically bounded by a particular field that develops and uses OSS. Corporations often participate in and manage numerous interconnected ecosystems. With a consistent taxonomy of metrics, each stakeholder is enabled, through the flexible and well-defined data structures Augur implements, to combine and present the most critical information for decision making at any point in time. CHAOSS and Augur can answer questions such as, "Where are the most vulnerable dependencies across 11,000 OSS projects in one program office?" and "What are the most vulnerable dependencies in each project?"



This graph shows all the first time contributors, whether they contribute once, or contribute multiple times. New contributors are individuals who make their first contribution in the specified time period.

This graph shows repeat contributors in the specified time period. Repeat contributors are contributors who have made 4 or more contributions in 365 days and their first contribution is in the specified time period. New contributors are individuals who make their first contribution in the specified time period.

This graph shows fly by contributors in the specified time period. Fly by contributors are contributors who make less than the required 4 contributions in 365 days. New contributors are individuals who make their first contribution in the specified time period. Of course, then, "All fly-by's are by definition first time contributors". However, not all first time contributors are fly-by's.

This graph shows the second contribution of all first time contributors in the specified time period.

FIGURE 3. New contributors can be understood from many perspectives. Zephyr maintains a constantly updating instance of Augur that can generate a report (http://zephyr.osshealth.io:5222/api/unstable/contributor_reports/new_contributors_stacked_bar/?repo_id=26222) at any time.

Linking ecosystems for corporatized open Source. Rapid OSS growth also means fast changes to the ways corporations bound the different ecosystems in their spheres. Applying CHAOSS metrics enables OSS leaders' real-time awareness of software that may become part of an ecosystem they are engaged in or soon will be. Risk awareness, in contrast, focuses on the OSS ecosystem as it is now, with the software bill of materials Augur provides, using the Software Package Data Exchange standard and FOSSology scanners; details about file-level license declaration completeness and diversity; and reasonably clear information about the organizations and contributors that are most essential for a project's health and sustainability. Through support from programs, such as the Google Summer of Code, and growing partnerships with the Open Source Security Foundation, CHAOSS and Augur are on the leading edge of defining essential indicators of OSS health: software security, software bills of material, development time, and runtime dependencies.

Augur: Making subtle project changes Transparent, with ethical artificial intelligence. Dependencies, licensing, emerging ecosystems, and software bills of material, while complex, are being incorporated into CHAOSS metrics and Augur by using discrete, discoverable data. There are, of course, a number of OSS projects that have fractured or declined despite their critical importance. Often in these cases, the reasons are subtle and, with the growth of OSS, more challenging to identify early enough for intervention. Typically, early warning signs can be found in project communications on mailing lists, issue trackers, and merge request discussions. The volume of those communications enables "trouble in paradise" issues that go unnoticed for too long.

Machine learning, artificial intelligence, computational linguistics, topic modeling, clustering, and statistical anomaly detection are useful for identifying project health concerns earlier. In our experience working with OSS teams using Augur's seven computational models, applying these technologies is most effective if two conditions are mutually understood. First, no one model is sufficient for identifying projects experiencing challenges or in the early signs of remarkable success. An integrated analysis using all seven models provides more useful information. Second, human judgment, clear communication, and the protection of individual privacy are paramount if CHAOSS metrics emerging from these tools are to be useful and accepted by the maintainer and contributor communities. In our view, metrics and tools leveraging these technologies are unethical if humans are removed from the interpretation of results.

USEFUL SIGNALS FOR A HIGH-VELOCITY OSS WORLD

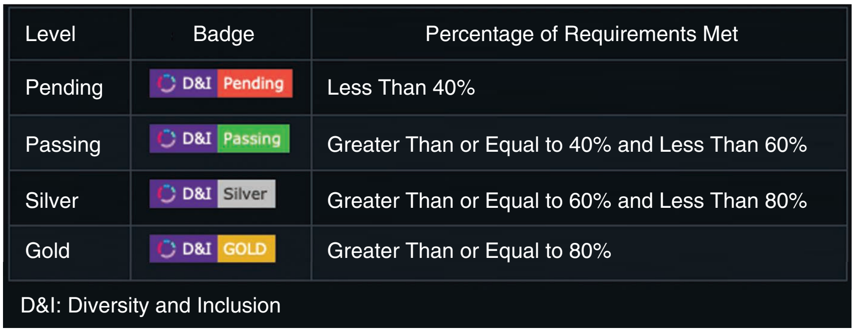
The issues of OSS health and sustainability are nearly as old as OSS itself. In our four years focused on CHAOSS, we have observed nearly daily a range of customs, communication patterns, and contribution cycles, and we have seen the effects that changing alliances have on goals and the questions OSS leaders need answered. In some cases of applying CHAOSS metrics, within weeks, questions that were never asked before emerge and require answers. At times, these sudden shifts have financial motivations; in other cases, legal rulings and the unexpected growth of new technologies drive them. Candidly, many discussions about the shortage of OSS engineers reference Ostrom's "overgrazed commons" metaphor to illustrate the constraint. Yet, activity metrics alone show how the OS contributor community remains insufficient for the work at hand, without suggesting remedies.

Based on emerging needs, the CHAOSS project will monthly identify and develop new signals. Often, Augur and other CHAOSS tools respond rapidly to implement those metrics to support decision making. To sustain our efforts, the partnerships we have formed help reseed our commons, but the volume of work continues to outpace the availability of OSS engineers. To continue developing useful metrics and tools, and for OS in general, it is becoming increasingly clear that one of our necessities is to identify and remove shared obstacles, which take many forms, such as barriers to careers in OSS. There are known impediments of context and distance that are basic features of OSS and that will likely remain in the future. However, others may be more manageable.

We think part of the upcoming contribution of CHAOSS can involve working with current and future projects to classify goals and related activities into categories of obstacles that provide shared utility. Not every project is at the same point in its evolution or focuses its questions on the same CHAOSS metrics. In fact, there is a good deal of organization-, ecosystem-, and maintainer-focused curiosity that drives the application of CHAOSS metrics through a "context" filter. In each case, prior work and lessons inform how we identify and classify obstacles. Briefly, during activities when CHAOSS contributors seek and explore obstacles in a general sense, they attempt to map risks to project health. Some activities in this category include health and sustainability metrics, including software licensing and regulation, and the increasing number of dependencies, such as when an OSS project imports a library from another.

Reflecting on our work with partners on DEI in OSS, we have observed and been told of project communication patterns that welcome newcomers and others that do not. We have gathered a number of oral narratives on the CHAOSS podcast that illustrate how efforts to help people recognize relationships between their motivations is an effective tool to make newcomers recognize, frankly, that OSS exists and that they can be paid for it. The increasing interest in understanding obstacles that impact DEI health is a reflection of what we should all know: contributors are the lifeblood of every OSS project, and OS health and sustainability require creating a diverse, equitable, and inclusive environment.

The future of OS metrics, as found in CHAOSS and through advanced tooling, such as Augur, seems virtually assured by the incredible growth of OSS. Through our work, we have assisted a number of OS projects, talked with hundreds of contributors and maintainers, and occasionally ventured out of that bubble. Health and sustainability for OSS is tied very clearly to many aspects of our lives as people: fishing, home improvement, travel, close relationships, and all the things that bring us joy. One CHAOSS member summed up this occasionally overlooked interconnection well, and we close with that thought: "Open source helps power virtually every piece of technology in our lives. The only way open source technology will equitably serve all of us is if we center DEI in the design and development of that technology."



Level	Badge	Percentage of Requirements Met
Pending	D&I Pending	Less Than 40%
Passing	D&I Passing	Greater Than or Equal to 40% and Less Than 60%
Silver	D&I Silver	Greater Than or Equal to 60% and Less Than 80%
Gold	D&I GOLD	Greater Than or Equal to 80%

D&I: Diversity and Inclusion

FIGURE 4. There are four levels in the CHAOSS DEI event program. Each represents a progressively higher percentage of the DEI metrics that have been met. To date, 21 major OS conferences have achieved one of the badge levels.

ACKNOWLEDGMENTS

This work was supported through funding from the Alfred P. Sloan Foundation as well as Mozilla. Jesus Gonzalez-Barahona and the GrimoireLab initiatives are founders of the CHAOSS project, and their software remains the most widely used for OS metrics.

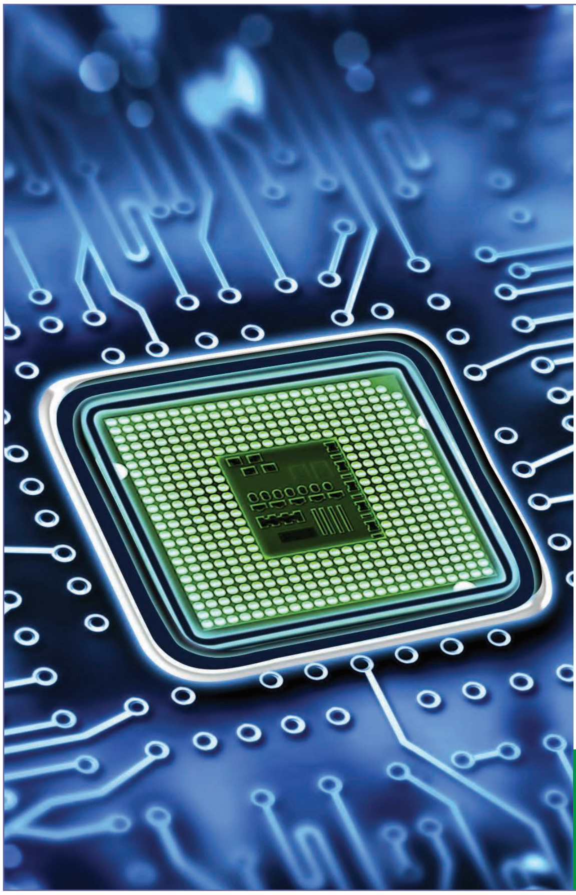
REFERENCES

1. J. M. Gonzalez-Barahona, “A brief history of free, open source software and its communities,” *Computer*, vol. 54, no. 2, pp. 75–79, 2021. doi: 10.1109/MC.2020.3041887.
2. S. Goggins, K. Lumbard, and M. Germonprez, “Augur: Architecture for capturing, reshaping, and socially contextualizing open source software communities,” in *Proc. ACM SoHeal Conf. Int. Conf. Softw. Eng.*, 2021. doi: 10.5281/zenodo.4627236.
3. M. Germonprez, G. J. P. Link, K. Lumbard, and S. Goggins, “Eight observations and 24 research questions about open source projects: Illuminating new realities,” *Proc. ACM Human-Comput. Interact.*, vol. 2, pp. 1–22, Nov. 2018. doi: 10.1145/3274326.
4. S. W. Goldhagen and A. Gallo, *Welcome to Your World: How the Built Environment Shapes our Lives*. New York: Harper, 2017.

SEAN P. GOGGINS is a professor of computer science at the University of Missouri, Columbia, 65211, Missouri, USA; cofounder of the Linux Foundation’s Community Health Analytics Open Source Software project; and creator of the University of Missouri’s master’s program in data science. Contact him at outdoors@acm.org.

MATT GERMONPREZ is the Mutual of Omaha Distinguished Chair of Information Science and Technology and a professor of information systems and quantitative analysis in the College of Information Science and Technology, University of Nebraska Omaha, Omaha, 68182, Nebraska, USA. Contact him at mgermonprez@unomaha.edu.

KEVIN LUMBARD is a Ph.D. candidate at the University of Nebraska Omaha, Omaha, 68182, Nebraska, USA. Contact him at klumbard@unomaha.edu.



Digital Object Identifier 10.1109/MC.2021.3094709

Cyber Resilience: by Design or by Intervention?

Alexander Kott, U.S. Army DEVCOM Army Research Laboratory

Maureen S. Golan, U.S. Engineer Research and Development Center and Credere Associates

Benjamin D. Trump, U.S. Engineer Research and Development Center and University of Michigan

Igor Linkov, U.S. Engineer Research and Development Center and Carnegie Mellon University

The term "cyber resilience by design" (RBD) is growing in popularity. But what is the other resilience, not by design? In this article, we explore the differences and mutual reliance of RBD and resilience by intervention.

The term "cyber resilience by design" (RBD) appears ever more frequently in the literature[1–3] and is the topic of this article. This article is the third of a series in this column on cyber resilience—an important attribute of system dependability. The previous two were by Linkov et. al.[12] and Kott and Linkov.[9]

Before we move any further, let's orient ourselves with respect to our lexicon. Here, by "cyber resilience," we refer to the ability of a system or network of computing and communicating devices to minimize and mitigate a degradation caused by a successful cyberattack.[8,10] This is not the only definition of "cyber resilience," but we see it as an important one and will settle on it for the sake of this article. But, what does "by design" mean?

Opinions differ. Some use the term "by design" when arguing that systems must be designed and implemented in a provable mission-assurance fashion, with systems' intrinsic properties ensuring that cyberadversaries are unable to cause meaningful degradations.[5] Others envision autonomic resilience,[4] where cyber resilience is provided by reflexive, no-thinking-required actions of appropriate elements within the systems. Yet others[7] recommend that systems should include built-in, autonomous intelligent

Digital Object Identifier 10.1109/MC.2021.3082836
Date of current version: 30 July 2021

agents responsible for thinking and acting toward the continuous observation, detection, minimization, and remediation of cyberdegradations. In all of these cases, the qualifier "by design" indicates that the source of resilience is somehow inherent in the structure and operation of a system.

But what, then, is the other resilience, not by design? Clearly, there has to be another type of resilience. Otherwise, what's the purpose of the qualifier "by design?" Indeed, while it is mentioned less frequently, there is an alternative form of resilience called *resilience by intervention* (RBI).[11] Unlike in the case of RBD, in RBI the effects of an external actor are the necessary source of resilience.

To use a psychological analogy, most humans have some innate abilities to emotionally cope with adverse effects and to return to a precrisis state. Many cases, however, require intervention by professionals, for example, therapists.[14] In the case of a human, the distinction seems clear enough: resilience through self-coping mechanisms residing in the mind and body reflects RBD while the engagement of an external actor (like a therapist) is analogous to RBI. However, does this distinction make sense in the cyberworld, where systems are typically distributed and networked and have fluid, permeable, and uncertain boundaries? How would we even define an "external" actor?

THREE EXAMPLES

To make our discussion more concrete, let's introduce three examples. For the first example, Ms. Johnson owns a self-driving car. She purchased a service in which a third-party provider remotely and continuously monitors and assesses the cyberhealth of her car and issues alerts to her if a compromise is detected. Then, Ms. Johnson can take her car to any shop of her choice with cyber-repair capabilities to have the compromise eliminated.

For the second example, Ms. Johnson owns a self-driving car. The car includes a resident software agent responsible for continuously monitoring all of the events on the car's multiple computing devices and related internal and external networks to detect and mitigate cybercompromises and to ensure that the car can drive even when it is cybercompromised.

For the third example, Ms. Johnson often gets around by calling a self-driving cab from Robo-Cab Inc. This company owns a fleet of self-driving cars, controlled and serviced from the company's depot. A depot-based system monitors each car, safely disables one when a compromise is suspected, and sends a response team to the vehicle. In addition, every few hours, each car returns to the depot for a thorough checkup and potential reimaging.

Clearly, in each case, we find some provisions for cyber resilience. But is it RBI or RBD? Before we proceed with answering this question, let's offer two perspectives on what differentiates RBI and RBD.

INTEGRATION AND AUTHORITY

We find that the differences between RBI and RBD of a given system are more clearly seen in light of two considerations. The first is the degree of the integration of the resilience-producing elements with the rest of the system. On one hand, RBD is likely to be found where strong links exist between the resilience elements and the components that are responsible for producing the primary functions of the system. On the other hand, if the links are established occasionally and selectively based on circumstances, we are probably looking at RBI.

We should mention that the term *system* in this article refers to a sociotechnical system consisting of technical artifacts as well as humans (if and when they are integral parts); that is, the term encompasses the human, software, and hardware partitions of an information system. Therefore, when we refer to various capabilities of a system, those capabilities are allocated among the humans, software, and hardware. Thus, when we refer to "intervention" and "external actor," the intervention may be performed by a strictly technical system (e.g., the external actor being an automated cybermitigation tool), by humans, or by a sociotechnical system that combines humans and technical means.

The observation that tight integration is associated with RBD seems to be true regardless of the specific mechanisms by which resilience is established. In the case of a system designed with innate cyber-resilience properties, as advocated by Jabbour,[5] the resilience mechanism is literally indistinguishable from the system's primary functional elements, implying RBD. In the case of a built-in, autonomous intelligent agent responsible for cyber resilience,[7] the agent is tightly integrated within the system in an RBD fashion.

This is also true regardless of the nature of the system, be it, for example, technical or biological. For instance, psychological coping mechanisms are inherent in a human and, as such, illustrate RBD. A therapist, on the other hand, is clearly not integral to a human patient, implying RBI.

In addition, the notion of integration does not necessarily imply spatial colocation. In a distributed system, various functions, including those that provide resilience, may reside, say, on different computers that are networked but separated by thousands of miles. This separation in itself tells us little about RBD or RBI. Instead, we should explore how closely and persistently these functions rely on each other.

The second perspective that helps differentiate between RBD and RBI is that of authority. When actions associated with resilience are taken (that is, actions of resisting, minimizing, and mitigating a degradation caused by a successful cyberattack), who has the authority to direct these actions? Who decides whether, when, where, and how these actions are performed?

When these decision rights are internal to a system, it does not require the permission (or direction) of an external actor to change system properties or reorganize itself. It is important for this authority to quickly seize opportunities to minimize disruptions or to recover from them. A system with an internal decision-making authority is able to interpret new information within itself or its environment and adjust accordingly. The system benefits from avoiding potential inefficiencies or delays by waiting for an external actor to yield permission or direction for change.

Conversely, a system that lacks internal decision authority has to rely on an external intervention to recover or adapt system properties or organizing characteristics. Such systems may also have advantages. It is possible that external intervention comes with greater decision-making abilities and may yield greater control over the intended future of a disrupted or adapted system. External and internal authorities may coexist or conflict. We will return to this point shortly.

On one hand, if the decision authority resides primarily outside of the system, we suspect RBI. If, on the other hand, the decisions are made primarily within the system (whether by humans or by a technical decision-making element), it is more likely to be RBD.

Referring again to the emotional crisis example, in self-coping, the decisions inevitably reside with the human experiencing the crisis. However, when a therapist intervenes, the patient may express his or her wishes, but the therapist reserves the ultimate authority on whether and how to provide a treatment. This is characteristic of RBI.

BACK TO THE THREE EXAMPLES

Now we are ready to return to Ms. Johnson and her means of transportation. Let us use the two perspectives we just discussed (integration and authority) to explore whether our three examples represent RBD or RBI.

In the first example, the integration links between Ms. Johnson and her car (together they are "the system") and the cyberservice providers are rather weak. If Ms. Johnson is unhappy with the service or finds a better price, she will readily switch to another provider. She probably does not adapt her routine or her car to the needs of the provider. She would prefer to keep her option to choose between multiple providers. Likewise, the providers are unlikely to customize their services to meet Ms. Johnson's unique requirements. They might try to entice her into a long-term contract, but, otherwise, Ms. Johnson is just one of many customers. Their mutual reliance is minimal.

And what about authority? Ms. Johnson may express her wishes, but ultimately the service providers will decide how they provide the service, when they will provide it (for example, they might say "sorry, this week we are too busy"), and which specific processes and procedures they will use. Ms. Johnson and her car do not have much authority over these decisions. Let's conclude: Integration is minimal and authority is external to the system. This is RBI.

In the second example, the resilience-producing component (a software agent responsible for the monitoring and mitigation of compromises) is fully integrated into the overall system. It is customized or custom-designed for the given car and for the pattern of operations characteristic of consumers like Ms. Johnson. It is entirely dedicated to the car and to Ms. Johnson, and it does not have much purpose outside of that system. The integration is tight.

The authority for making decisions about how to monitor the cyberstate of the car, how to detect and analyze the compromise, and what mitigating actions to take all reside with the agent. Granted, the agent will attempt to ask for Ms. Johnson's approval in appropriate situations, but, in most cases, the agent will have to act autonomously because Ms. Johnson will be unavailable or unable to make a relevant decision. Here, we have a tight integration of resilience-producing elements into the system, and the decision-making authority resides internally in the system. This is RBD.

However, although rare, some impacts to this cybersystem may be too large, outside of the scope of the software agent's mitigation or repair capabilities, or causing a myriad of cascading impacts that require Ms. Johnson to seek specific service capabilities elsewhere. In these circumstances, Ms. Johnson would not have any prior commitments or contracts to any service providers, and the integration would be loose. And, although the monitoring service would have authority over the vehicle under situations it would be capable of mitigating, should Ms. Johnson require to seek services elsewhere during exceptional events, the single-job provider would have the onus of recovery and adaptation placed on it. The provider would likely collaborate or seek system history from Ms. Johnson's cybermonitor, but a majority of the authority would be external to the system. This scenario therefore is RBI.

The third example differs significantly from the previous two. Importantly, the system under consideration is different. It consists of the fleet of cars and the depot. Ms. Johnson is not a part of the system; she is an external consumer of the services produced by it.

All of the activities related to cyber resilience are integrated within the depot-cars system. Note that the system is highly distributed and dynamic: the depot itself may consist of multiple physical facilities, and its cyber-related systems may reside on multiple computers, anywhere in the world. The cars travel over a wide geographic area, constantly changing their locations. Nevertheless, all of the cyber-resilience components and their activities are integrated within this distributed system. The depot's resilience-producing components (including human specialists) are designed to support the fleet of cars, and the cars cannot operate (at least for this business model) without the depot support.

The authority for making cyber-resilience-related decisions also resides substantially within the depot-cars system (including its human employees). The decisions on how to resist, minimize, and mitigate a degradation caused by a successful cyberattack are made by automated or human elements within the overall system, even if a decision may be produced in a distributed fashion among multiple elements of the system. Thus, we have tight integration and internal authority, implying RBD. See Table 1 and Table 2 for details of the examples.

DESIGN AND INTERVENTION SHOULD BE COMPLEMENTARY

Needless to say, RBD and RBI, although different, are not incompatible. Both approaches may coexist. A given system may have provisions for both RBD and RBI, although this coexistence should be carefully orchestrated. In particular, a clear protocol should be established for the handover of responsibilities between RBD mechanisms to RBI mechanisms and back. If both RBI and RBD mechanisms need to operate simultaneously, a coordination protocol should ensure that their respective actions do not produce undesirable interference. The interactions of RBD and RBI should be specific to each system and situation.

Furthermore, both RBD and RBI have their disadvantages. For example, relying strictly on RBI (as in Ms. Johnson's first example) may be risky. If RBI-based mitigation of a cybercompromise cannot be provided rapidly, Ms. Johnson could find herself speeding on a highway in a dangerously

TABLE 1. A comparison of illustrative examples and their implications for cyber-resilience implementation.

	Example one	Example two	Example three
Overview	Ms. Johnson owns a self-driving car and has a contract with a third-party continuous monitoring provider.	Ms. Johnson owns self-driving car with a resident software agent for continuous monitoring.	Ms. Johnson uses a self-driving car service with a depot-based continuous monitoring system.
System	Ms. Johnson and her car	Ms. Johnson and her car with monitoring	Robo-Cab, Inc. (fleet of cars, depot, monitoring)
Goal	Ms. Johnson and her car should maintain safe mobility and quick recoveries from cyberthreats.	Ms. Johnson and her car should maintain safe mobility and seamless recoveries from cyberthreats.	Robo-Cab should maintain safe mobility for customers and seamless recoveries from cyberthreats.
Environment	Third-party monitoring, alerts, and choice of cyber-repair shop	Any self-driving car shop/garage	Ms. Johnson and her mobility needs
Integration	Loose: There is a weak link between Ms. Johnson and the repair process.	Tight: There is a tight autonomous response by the software agent (representing Ms. Johnson) and repair process. Loose: Certain disruptions require one-time contracts with specialty shops.	Tight: The fleet, depot, and support are uniquely designed for the distributed system.
Authority	External: Service providers decide the path forward and are responsible for repair outcomes.	Internal: The resident software agent holds the responsibility for repairs. External: The specialty shops are responsible for repairs.	Internal: Robo-Cab's cyberphysical-human team holds the responsibility for repairs.
Resilience type	RBI	RBD and RBI	RBD
Advantage	The response is more threat- and impact-dependent; resources are used as needed.	The response is more immediate, and the consequences may be less severe	The response is more immediate, and the consequences may be less severe
Disadvantage	There may be too slow of a response.	The cyberthreat may adapt to the response and overcome it.	The cyberthreat may adapt to the response and overcome it.

TABLE 2. A comparison of the risk-management approaches (that is, cybersecurity) of RBD and RBI for cybersystems.

	Risk management	RBD	RBI
Objective	Harden individual components	Design components to be self-reorganizable	Rectify disruption to components and stimulate recovery by external actors
Capability	Predictable disruptions, acting primarily from outside the system components	Either known/predictable or unknown disruptions, acting at a component or system level	Failure in the context of societal needs; there may be a constellation of networks across systems
Consequence	Vulnerable nodes and/or links fail as a result of a threat	Degradation of critical functions in time and capacity to achieve system's function	Degradation of the critical societal function due to cascading failure in interconnected networks
Actor	Either internal or external to the system	Internal to the system	External to the system
Corrective action	Either loosely or tightly integrated with the system	Tightly integrated with the system	Loosely integrated with the system
Stages/ analytics	Prepare and absorb (the risk is a product of a threat, vulnerability, and consequences, and is time independent)	Recover and adapt (explicitly modeled as time to recover system function and the ability to change system configuration in response to threats)	Prepare, absorb, recover, and adapt (explicitly modeled as the ability to recover and secure the critical societal function and needs through the constellation of relevant systems)

uncontrollable car. For such emergency situations, it would be far safer to provide her car with an onboard RBD mechanism (like in the second example), even if it is less capable then a comprehensive cyber-resilience service by intervention. As we discussed elsewhere,[7] Ms. Johnson does live in an environment where fast, brutal cyberattacks exist and demand extremely fast responses. Criminals or irresponsible pranksters could take control of Ms. Johnson's car, and the consequences could be tragic.[13] That's why her car would benefit from having an onboard intelligent autonomous agent capable of taking the necessary resilience actions, that is, an RBD component.

Similarly, relying only on RBD comes with its own risks. For example, if a cyberattacker is able to overcome the capabilities (inevitably limited) of an RBD mechanisms, external intervention is a necessity. Such an intervention may come from multiple external actors, depending on the specific challenge of a cybercompromise. Even if most of the external actors found themselves unable to handle the complex compromise, chances are relatively high that in a multiplicity of external actors, at least one could be found who would be able to deal with the challenge.

Provisions for adding RBI to RBD when a situation dictates are necessary because, sooner or later, such a situation would arise. This conclusion is by no means unique to cyber resilience. For example, in the cases of supply-chain, engineering, and environmental resilience, the use of resilience analytics facilitates the implementation of corrective actions from within a system and/or external to it for a unified approach that maintains the necessary critical function despite inevitable disruption.[6,11,15–17]

To conclude, the terms "cyber resilience" and "resilience by design" are gaining popularity. Occasionally, they seem to be misused as synonyms for or as "cooler" substitutes for tired terms like *cybersecurity*, with little concrete meaning behind them. This should not be the case. RBD is a particular type of cyber resilience (not to be confused with cybersecurity), distinct from RBI and perhaps other types of resilience yet to be identified. Both types require clear definition and differentiation. To distinguish RBD and RBI, it is useful to examine how tightly resilience-producing elements are integrated into overall systems and whether the authority for making resilience-related decisions resides within or outside those systems. Neither RBD nor RBI are panaceas, and careful integration of the two is likely to produce superior resilience. C

ACKNOWLEDGMENT

This study was funded in part by the U.S. Army Engineer Research and Development Center FLEX project on Compounding Threats.

REFERENCES

1. G. Ahmadi-Assalemi, H. Al-Khateeb, G. Epiphaniou, and C. Maple, "Cyber resilience and incident response in smart cities: A systematic literature review," *Smart Cities*, vol. 3, no. 3, pp. 894–927, 2020. doi: 10.3390/smartcities3030046.

2. S. Bagchi et al., "Vision paper: Grand challenges in resilience: Autonomous system resilience through design and runtime measures," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 155–172, July 2020. doi: 10.1109/OJCS.2020.3006807.
3. S. N. G. Gourisetti, S. Mix, M. Mylrea, C. Bonebrake, and M. Touhiduzzaman, "Secure design and development cybersecurity capability maturity model (SD2-C2M2): Next-generation cyber resilience by design," in *Proc. Northwest Cybersecurity Symp.*, 2019, pp. 1–9. doi: 10.1145/3332448.3332461.
4. S. Hariri, M. Eltoweissy, and Y. Al-Nashif, "BioRAC: Biologically inspired resilient autonomic cloud," in *Proc. 7th Annu. Workshop on Cyber Security Inf. Intell. Res.*, 2011, p. 1. doi: 10.1145/2179298.2179389.
5. K. Jabbour, "The Post-GIG era," *Cyber Defense Rev.*, vol. 4, no. 2, pp. 117–128, 2019. [Online]. Available: https://cyberdefensereview.army.mil/CDR-Content/Articles/Article-View/Article/2017729/the-post-gig-era-from-network-security-to-mission-assurance/
6. M. S. Golan, B. D. Trum, J. Cegan, and I. Linkov, "The vaccine supply chain: A call for resilience analytics to support COVID-19 vaccine production and distribution," in *COVID: Risk and Resilience*, New York: Springer International Publishing, 2021. [Online]. Available: https://arxiv.org/abs/2011.14231
7. A. Kott and P. Theron, "Doers, not watchers: Intelligent autonomous agents are a path to cyber resilience," *IEEE Security Privacy*, vol. 18, no. 3, pp. 62–66, 2020. doi: 10.1109/MSEC.2020.2983714.
8. A. Kott and I. Linkov, Eds., *Cyber Resilience of Systems and Networks*. New York: Springer-Verlag, 2019.
9. A. Kott and I. Linkov, "To improve cyber resilience, measure it," *Computer*, vol. 54, no. 2, pp. 80–85, 2021. doi: 10.1109/MC.2020.3038411.
10. National Research Council. *Disaster Resilience: A National Imperative*. The National Academies Press, Washington, D. C., 2012.
11. I. Linkov, B. D. Trump, M. Golan, and J. M. Keisler, "Enhancing resilience in post-COVID societies: By design or by intervention?" *Environ. Sci. Technol.*, vol. 55, no. 8, pp. 4202–4204, 2021. doi: 10.1021/acs.est.1c00444.
12. I. Linkov, S. Galaitsi, B. D. Trump, J. M. Keisler, and A. Kott, "Cybertrust: From explainable to actionable and interpretable artificial intelligence," *Computer*, vol. 53, no. 9, pp. 91–96, 2020. doi: 10.1109/MC.2020.2993623.
13. T. Ring, "Connected cars: The next target for hackers," *Netw. Security*, vol. 2015, no. 11, pp. 11–16, Nov. 2015. doi: 10.1016/S1353-4858(15)30100-8.
14. E. K. Rynearson, Ed., *Violent Death: Resilience and Intervention Beyond the Crisis*. New York: Routledge, 2006.
15. D. Simchi-Levi and E. Simchi-Levi, "We need a stress test for critical supply chains," Harvard Business Review, Apr. 28, 2020. https://hbr-org.cdn.ampproject.org/c/s/hbr.org/amp/2020/04/we-need-a-stress-test-for-critical-supply-chains
16. B. D. Trump, I. Linkov, and W. Hynes, "Combine resilience and efficiency in post-COVID societies," *Nature*, vol. 588, no. 7837, p. 220, 2020. doi: 10.1038/d41586-020-03482-z.
17. I. Linkov, J. Keenan, and B. D. Trump, *COVID-19: Systemic Risk and Resilience*. Amsterdam: Springer-Verlag, 2021.

DISCLAIMER

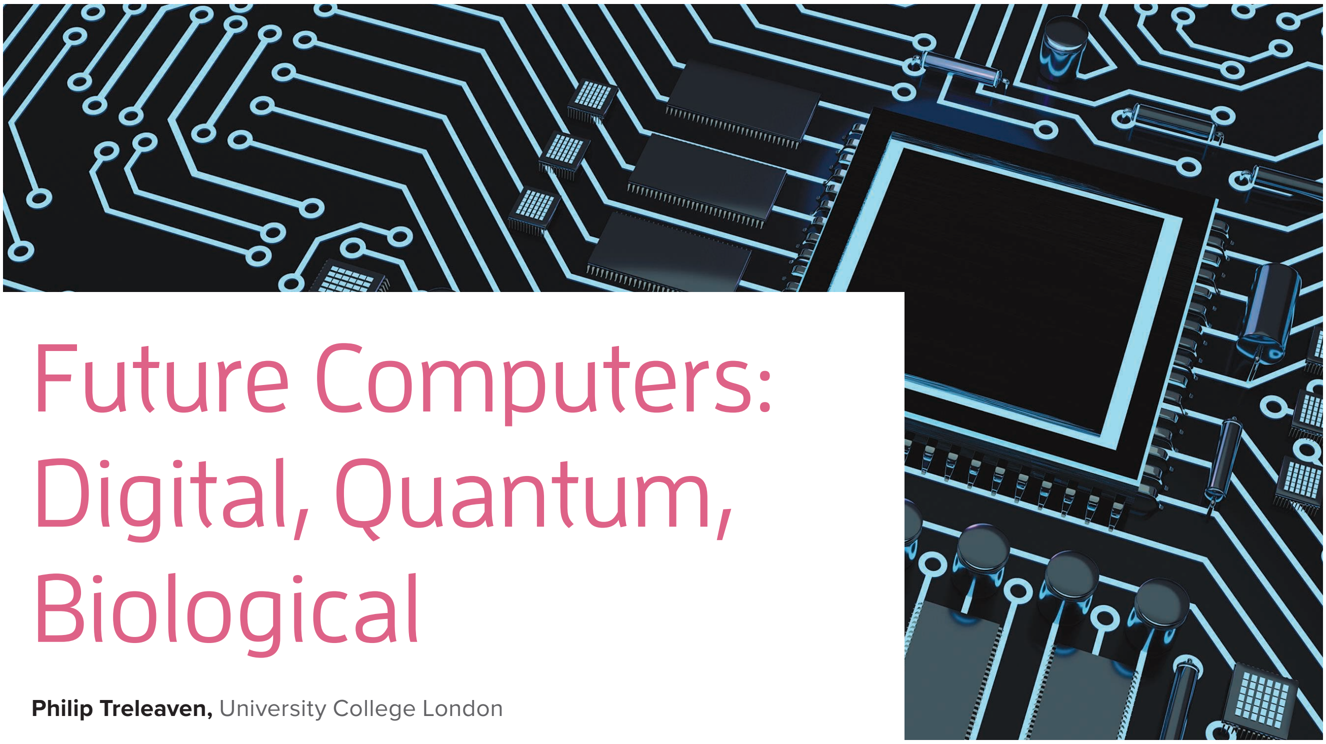
The views and conclusions contained herein are those of the authors and should not be interpreted as representing the policies or endorsements, either expressed or implied, of their employers. The U.S. government is authorized to reproduce and distribute reprints for government purposes, notwithstanding any copyright annotations thereon. Certain commercial entities, equipment, or materials may be identified in this document to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the authors or their employers, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

ALEXANDER KOTT is the chief scientist at the U.S. Army Combat Capabilities Development Command Army Research Laboratory, Adelphi, Maryland, 20783, USA. Contact him at alexander.kott1.civ@mail.mil.

MAUREEN S. GOLAN is a research engineer with Credere Associates and the U.S. Army Corps of Engineers' Engineer Research and Development Center, Concord, Massachusetts, 01742, USA. Contact her at maureengolan@gmail.com.

BENJAMIN D. TRUMP is a research social scientist with the University of Michigan and the U.S. Army Corps of Engineers' Engineer Research and Development Center, Concord, Massachusetts, 01742, USA. Contact him at benjamin.d.trump@usace.army.mil.

IGOR LINKOV is a senior science and technology manager with Carnegie Mellon University and the U.S. Army Corps of Engineers' Engineer Research and Development Center, Concord, Massachusetts, 01742, USA. Contact him at Igor.linkov@usace.army.mil.



Future Computers: Digital, Quantum, Biological

Philip Treleaven, University College London

Quantum computers offer huge potential performance, and biological computers can revolutionize pharmaceuticals. However, simple "engineering" descriptions and standardized approaches are needed. We provide "layperson" descriptions of quantum and biological computer architectures, comparisons with digital computers, and discussions of industry-standard models.

The digital (von Neumann control-flow, stored-program) computer model has underpinned technological progress for 60 years.[13] However, supporting a sequential program-execution architecture, devices are reaching their physical limits in terms of miniaturization, scalability, and performance.[2] This is driving interest in new computer architectures. Valuable insights can be gained by comparing digital and quantum computers, also with biological information processing.[1]

The von Neumann model is the industry standard because it is common to control-flow computers and procedural languages. What we require are general-purpose, scalable models embedded in computers and the associated programming languages:

- *digital computer*: a parallel variant of the traditional binary, control-flow, stored-program computer model, such as multi-instruction-multidata (MIMD)-stream computers
- *quantum computer*: information processing using quantum bits (qubits), quantum entanglement, and "configurable" quantum logic gates for processing

Digital Object Identifier 10.1109/MC.2021.3083430
Date of current version: 30 July 2021

- *biological computer*: information processing using cells (protein synthesis) as well as DNA, proteins, and RNA to create new cells (compare with hardware) or enact computational operators.

The key to progress is, first, simple "engineering" descriptions of quantum and biological computers and, second, common (industry-standard) models embedded in computer architectures and programming languages. A standard architecture for quantum computers may deliver major computational benefits. In turn, specifying a general-purpose architecture for biological computers (compare with cell and molecular biology) has the potential to unlock an understanding of information processing in nature and our ability to engineer cells programmed at the DNA/protein level for applications such as immunology.

COMPUTATION MECHANISMS

Computation is defined as the controlled transformation of information, sensitive to the representation's properties, determining the behavior of information processing.[5] The profound distinction is that natural (biological) computation uses physical structures that directly transform themselves, whereas artificial (human-made) computation uses abstract structures that interpret a model or simulation.

We propose four foundational engineering paradigms underpinning computation:

- *Information*: the way information is encoded
  - discrete/digital represents information by a series of values of a physical quantity, such as binary (01) or DNA [adenine, thymine, guanine, cytosine (ATGC)].
  - continuous/analog represents information by a variable physical quantity, such as qubits or voltage
- *Structure/processing*: how computation is represented and processed
  - an interpreted model, that is, an abstract model simulated by a digital program/data
  - a direct-transformed physical structure, such as DNA/proteins
- *Control*: the method by which information is selected for processing
  - explicit, such as instruction addresses
  - pattern matching, for example, with proteins
- *Communication*: represents how information and state changes propagate
  - multicast or broadcast, such as shared memory
  - unicast or point to point, such as messages.

DIGITAL COMPUTERS

Digital binary, stored-program, control-flow computers (see Figure 1) comprise an addressable memory containing data and instructions and a CPU that interprets instructions. Being able to write data and then execute them as instructions is a powerful basis for general-purpose computation. The CPU comprises an arithmetic and logic unit (ALU) and a program counter defining the memory address of the next instruction(s) to be executed.

In the late 1940s, a number of digital stored-program computer architectures were proposed, but the von Neumann architecture became the industry standard model, embedded in

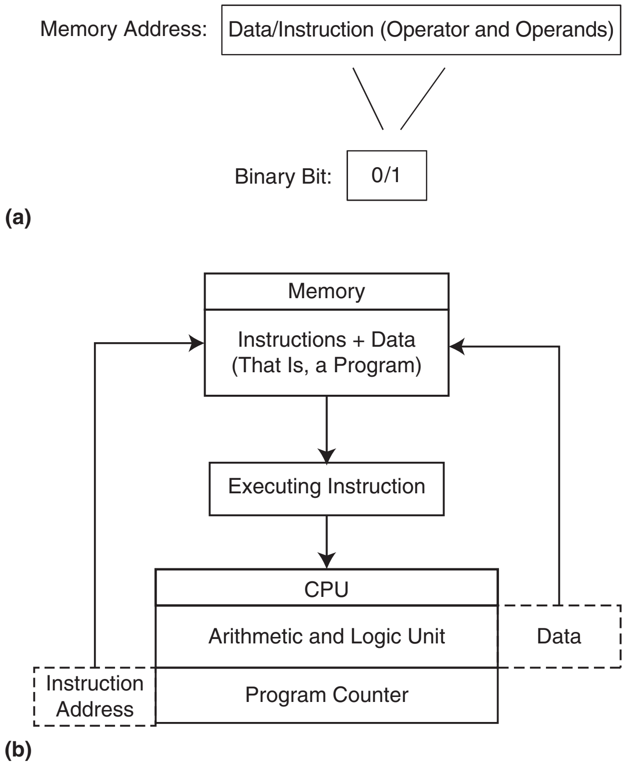


FIGURE 1. Digital computer: (a) information representation and (b) computer architecture.

computers and procedural languages. The model's instructions comprise an operator (ALU or control) and operands (data or memory addresses). With an ALU instruction, the program counter automatically increments. With a control instruction, the memory address overwrites the program counter.

As a benchmark for our discussion of quantum and biological computers, we present a scalable control-flow model. For a general-purpose, parallel, control-flow computer, we need an inherently MIMD architecture (compare with von Neumann) model embeddable in parallel computers and procedural languages (Figure 2). For instance, each computer's local memory might form part of a shared address space, and control instructions could specify multiple (next) instruction addresses.

Programming model

This parallel computer model can be retrofitted into procedural languages by including the concepts in Figure 2 of processor (p1, p2), shared address space (p1.x, p2.x), "fork" control (||), and possibly some form of synchronization statement. Next, we provide a simple but broadly accurate "engineering" description of a quantum computer.

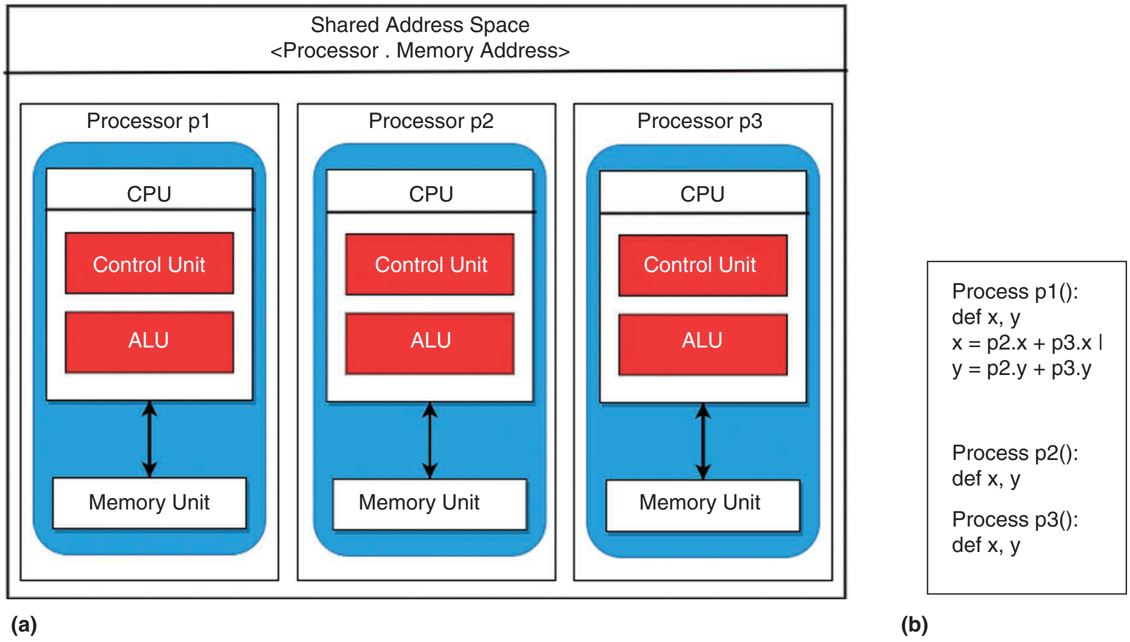
QUANTUM COMPUTERS

Quantum computers, typically gate based,[11] are analogous to programming at the circuit logic level or configuring a field programmable gate array. We start with some definitions:

- *qubit*: the basic unit of quantum information, representing simultaneously two values/states (0 and 1)
- *quantum register*: a system comprising multiple linked qubits and the quantum analog of the classical processor register (that is, entanglement)
- *quantum gate*: a basic quantum operator and a circuit processing a small number of qubits
- *quantum circuit*: a model in which a quantum computation is a configuration of quantum gates
- *quantum parallelism*: the ability of a quantum computer to process, in parallel, all values of a quantum register in a single computation (that is, superposition)
- *quantum instructions*: specifications for configuring the quantum gates to perform a quantum computation
- *quantum programming*: the process of assembling sequences of instructions, called quantum programs, capable of running on a quantum computer.

Physicists discuss quantum computers in terms of quantum states that are specified by probability amplitudes, and two continuous variables are needed to uniquely specify the state of a qubit.[11] However, it is popular to contrast a qubit (0 and 1) with binary (0 or 1).

A colloquial analogy is to consider information as represented by a "coin." Binary information is the coin's head or tail. A qubit is depicted as a coin balancing or spinning on its edge on a table, representing both 0 and 1. Storing a value in a qubit is "spinning" our coin between 0 and 1 (that is, in limbo). Reading a qubit is equivalent to "banging the table," causing the "coin" to fall as 0 or 1.



```
Process p1():
def x, y
x = p2.x + p3.x |
y = p2.y + p3.y


Process p2():
def x, y

Process p3():
def x, y
```

FIGURE 2. An example of the general-purpose parallel architecture model: (a) computer architecture and (b) programming language.

Reading "freezes" both the output and also the stored value (that is, the state) of the qubit (0 or 1). It also discards information on how the value was represented by the qubit (for example, polarity, spin, and so on). Due to a qubit being able to support two states at the same time, one qubit stores two states [(0), (1)], two qubits store four states [(00), (01), (10), and (11)], and eight qubits store 256 states [(00000000) . . . (11111111)]. Importantly, when a value is stored in a qubit, a probability can be added to influence the output. Continuing our coin analogy, this is like adding a bias (compare with a biased coin), changing the probability of an outcome.

Four key concepts in quantum computers are as follows:

- *Superposition*: A qubit represents information as 0 and 1.
- *Entanglement*: Qubits are linked together as a register, that is, linked behavior.
- *Interference*: Controlling quantum values' probabilities and amplifying the signals leads toward the right answer.
- *Coherence/decoherence*: Since qubits lose information over time, estimating the "shelf life" of information is important.

With quantum entanglement, a group of qubits, or a register, is formed so that actions performed on one qubit affect the others. (Entanglement is created, for example, by microwave pulses or using a laser to split photons into pairs.) With quantum interference, adding a probability influences the output toward the right result.

Quantum processing is performed by quantum logic gates configured by quantum instructions. Qubit data are input and passed through the quantum gates, producing outputs. Configured gates provide a mapping/transformation function. For a quantum computer, the inputs to the quantum gates are (0 or 1), and the outputs to the quantum memory (0 and 1) are modified by a probability.

The simplistic quantum computer model (Figure 3) follows digital computers with the qubit memory based on qubit registers (compare with memory words) containing data and gate instructions in the same memory. The addressing structure for memory access requires a separate processing unit to select data and gate instructions for processing using these addresses.

Designing a quantum computer can be guided by our computation mechanisms from the "Computation Mechanisms" section:

- *Information*: Qubits represent computation.
- *Structure/processing*: Quantum entanglement is used to build structure, and a quantum processor interprets the computation.
- *Control*: A quantum control mechanism might use explicit memory addresses.
- *Communication*: Qubit memory supports multicast.

Quantum computers can be subdivided as follows:

- *Quantum hybrid*: Current quantum computers are hybrids comprising a traditional digital computer for program control together with a quantum coprocessor comprising qubit memory and quantum logic gates for execution. The digital computer is used to "write" the qubit memory, configure the quantum gates, and select the desired result.
- *Quantum unitary*: A future general-purpose quantum computer arguably needs the following:
  - the qubit memory to store data and gate instructions (compare with shared memory)
  - a qubit addressing mechanism identifying a specific qubit
  - a set of universal quantum gate operators for all programs

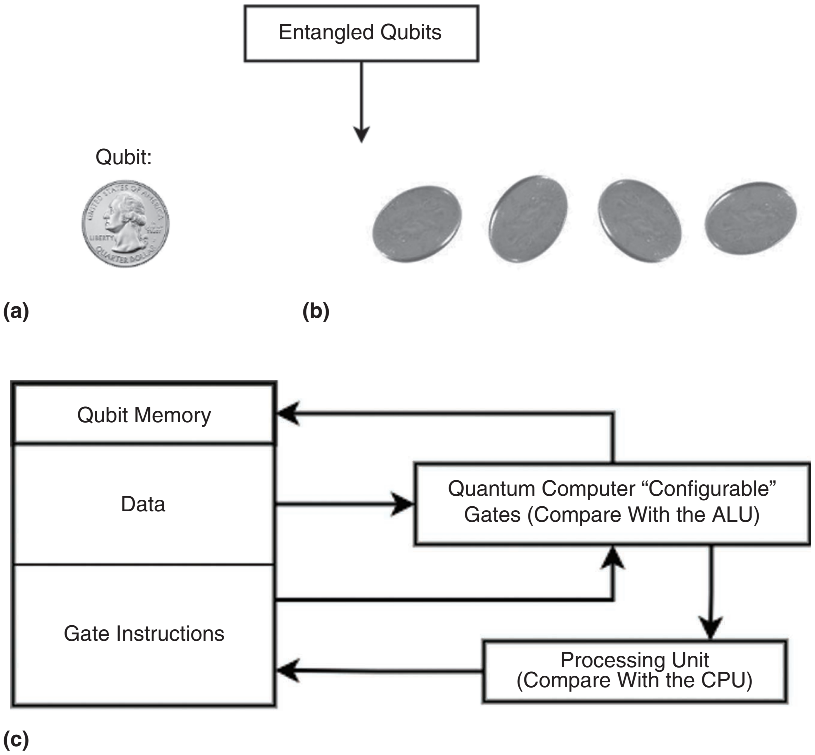

Entangled Qubits
Qubit:
(a)
(b)
Qubit Memory
Data
Gate Instructions
Quantum Computer "Configurable" Gates (Compare With the ALU)
Processing Unit (Compare With the CPU)
(c)

FIGURE 3. A quantum computer: (a) one qubit is both 0 and 1, and (b) four qubits represent all 2^4 possible permutations. (c) The quantum computer architecture.

- a processing unit for the gate instructions to configure the quantum gates (compare with a CPU)
- outputs that store a superposition value in the qubit memory
- addressing mechanisms for selecting proceeding gate instructions (compare with a program counter).

A quantum computer has three sources of power:

- *Qubit correlations*: Information is stored in correlations using entanglement.
- *Qubit parallelism*: A group of qubits, connected by entanglement and guided by inference, tests all combinations of values in parallel (compare with slot machine dials).
- *Gate parallelism*: Output flows through a network of quantum logic gates in parallel.

Programming model

Importantly, a quantum algorithm and program is independent of the need for quantum computer hardware. That said, what is required for quantum computers is a common "industry-standard" model.[10] This model then needs to be embedded in a quantum hardware architecture and corresponding quantum programming system. Already, quantum programming systems exist, including quantum instruction sets; quantum software development kits; and quantum programming languages, which can be divided into imperative, such as quantum instruction language (Quil), QCL, Q#, and Silq, and functional, such as quantum flow chart/quantum programming language and QML.[12]

When designing a quantum hybrid coprocessor, many of the engineering challenges (addressing qubits, configuring quantum gates, and writing a distribution into a qubit) can be handled by a digital computer. Exploiting qubit parallelism (that is, superposition) for performance using digital technology is less obvious compared to quantum gates.[7]

Designing a general-purpose quantum unitary computer raises a number of interesting questions: first, for quantum memory, how qubits and qubit registers represent data, gate instructions, and addresses; second, for quantum gate instructions, how qubit values can specify gate configurations; third, for quantum addressing, how to access a specific qubit register and qubit, which may require an address tuple; and fourth, for quantum processing, how all possible combinations of entangled qubits are tested. Finally, a "left-field" solution for quantum computers may be optical devices[11] due to noninterference or biological devices using a DNA/protein approach and pattern-matching control.

BIOLOGICAL COMPUTERS

The reassuring fact about biological information processing is that we know it works.[4] The key distinction of biological over digital computers, as discussed, is that natural information processing transforms the physical structure (compare with hardware) and/or its environment. Figure 4 illustrates biological information processing:

- *DNA*: DNA comprises genes (and noncoding DNA) and is composed of four nucleotide bases: A, T, G, and C. Genes contain the information needed to make proteins. DNA is analogous to a stored executable program (compare to instructions + data).
- *RNA*: RNA comprises a copy of DNA constituting a gene, composed of four nucleotide bases: A, G, C, and uracil (U). RNA often acts as the control mechanism, binding to "complementary" DNA fragments and modifying their activity (regulating gene expression).
- *Proteins*: These comprise one or more chains of amino acids, of which there are 20 different types. Proteins appear analogous to an executing program.

The A, T, G, and C bases of DNA; A, G, C, and U bases of RNA; and 20 amino acids of proteins are akin to jigsaw puzzle pieces. The sequence of bases in the DNA determines the sequence of bases in the RNA molecule, which, in turn, determines the sequence of amino acids making up the protein. Information encoded in their sequential structure flows unidirectionally from DNA through RNA to proteins (known as the *central dogma*). Computation in biological computers centers on molecular (or molecular assembly) 3D shape pattern matching and processing.

The journey from gene to protein consists of two major steps:

- *Transcription*: Information stored in a gene's DNA is transferred by messenger RNA (mRNA), which acts as a "blueprint" for creating proteins.
- *Translation*: A ribosome uses mRNA and transfer RNA (tRNA) to assemble a protein. Processing is driven by powerful pattern-matching (control) mechanisms that manipulate the DNA structure.

The flow of information from DNA to RNA to proteins (that is, protein synthesis) is one of the fundamental principles of biology. Enzymes (that is, polymerases) assemble DNA and RNA, and ribosomes produce proteins for which they receive mRNAs. An mRNA molecule is (approximately) a copy of a DNA sequence that describes the order of amino acids in a chain that (when properly folded) makes up the corresponding protein. A tRNA of a given shape transports one specific amino acid and attaches to a specific fragment of mRNA—this high degree of specificity of shapes ensures that, in general, only the correct amino acid can be attached to the growing chain.

The design of a general-purpose biological computer guided by our

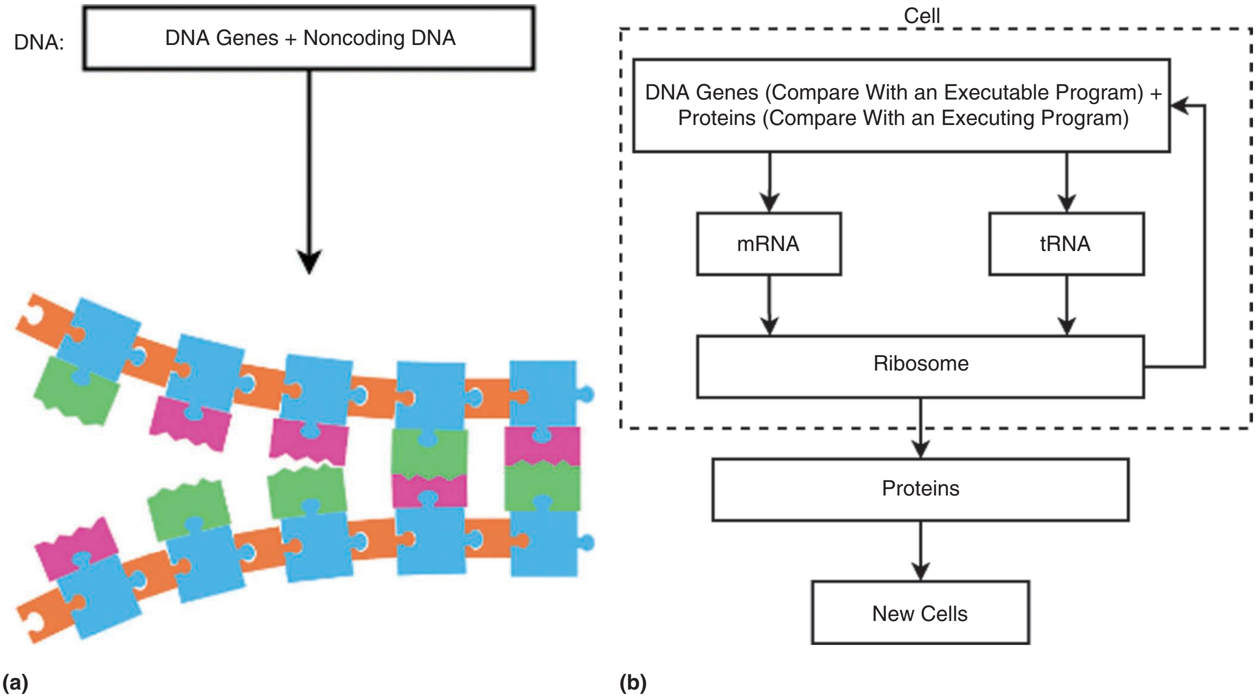

DNA:
DNA Genes + Noncoding DNA
Cell
DNA Genes (Compare With an Executable Program) +
Proteins (Compare With an Executing Program)
mRNA
tRNA
Ribosome
Proteins
New Cells
(a)
(b)

FIGURE 4. A biological computer: (a) information representation and (b) computer architecture. mRNA: messenger RNA; tRNA, transfer RNA.

computation mechanisms (see the "Computation Mechanisms" section) comprises the following:

- *information*: represented by physical quantities (the ATGC of DNA, AGCU of RNA, and 20 amino acids of proteins)
- *structure/processing*: a direct-transformed physical structure (by ribosomes)
- *control*: a pattern-matching mechanism based on the information content (by polymerases)
- *communication*: not clear if this is multicast, unicast, or both.

To give a specific example, the pioneering work on mRNA therapeutics for vaccines is likely to drive biological computer development. An RNA vaccine consists of an mRNA strand that codes for a disease (that is, a specific antigen). Once the mRNA strand in the vaccine is inside the body's cells, the cells use the genetic information to produce the antigen. To create an mRNA vaccine, scientists use a synthetic equivalent of the mRNA that a virus uses to build its infectious proteins. The cells read the mRNA as "instructions" to build that antigen protein, and the immune system detects these viral proteins and starts to produce a defensive response to them. We can fast forward to "general-purpose" mRNA therapeutics.

An excellent illustration of biological "machine code" is the 4,284-character COVID BNT162b2 mRNA code,[3] shown in Figure 5. This code is uploaded to a DNA printer that converts the digital characters to actual biological DNA and then RNA. Once inside a cell, the RNA is used to produce proteins typical for the virus, which prompts the immune system to develop defenses against it. The machine code in Figure 5 exploits the existing DNA and protein "program" in the cell, which, in turn, leads to the activation of other programs in nearby immune cells.

Programming model

As discussed, further progress toward a general-purpose biological computer and programming language requires a standard model where we can specify DNA, proteins, RNA strings, and possibly a template library of customizable DNA and proteins. In terms of the standard model, data/instructions comprise DNA, RNA, and proteins; operators support transcription and translation; and a control mechanism is based on pattern-matching control execution (for example, gene editing via Cas9/CRISPR).

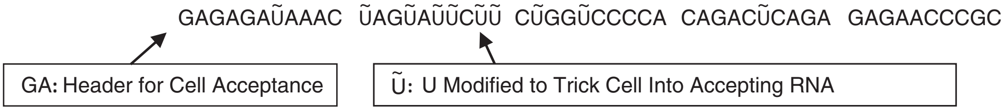

GAGAGAŨAAAC ŨAGŨAŨŨCŨŨ CŨGGŨCCCCA CAGACŨCAGA GAGAACCCGC
GA: Header for Cell Acceptance
Ũ: U Modified to Trick Cell Into Accepting RNA

FIGURE 5. The first 50 characters of the COVID BNT162b2 mRNA "machine code."

	Digital	Quantum	Biological
Information	Discrete/digital with information represented by a physical quantity (for example, binary)	Continuous/analog with information represented by a variable physical quantity (for example, a qubit)	Discrete/digital with information represented by a physical quantity (for example, the DNA bases A, T, C, and G)
Structure/ Processing	An abstract model simulated by program/data and interpreted	An abstract model simulated by program/data and interpreted	Direct structure transformed
Control	Explicit, such as instruction addresses	Explicit, probably using addresses	Pattern matching
Communication	Multicast or broadcast	Multicast or broadcast	Multicast and/or unicast

FIGURE 6. A comparison of digital, quantum, and biological computers.

Important questions arise: first, what chemicals (compare with hardware) exist to create a biological computer; second, how the pattern-matching control mechanism might work; and third, what a high-level programming language will look like.

As discussed, the industry-standard von Neumann model has underpinned technological progress, but it is reaching its physical limits. This has spurred massive interest in quantum computers. Likewise, understanding information processing in natural systems, such as cells and plants, can lead to a breakthrough in (DNA) programmable biological computers.[5,6] (We should also revisit MIMD-stream digital parallel computers.) Using our four computational mechanisms discussed in the "Computation Mechanisms" section, Figure 6 presents a summary comparison of digital, quantum, and biological computer models.

In conclusion, general-purpose, scalable architecture models for digital, quantum, and biological computers can provide powerful new systems. More importantly, this research should contribute to an understanding of biological information processing.[8] The starting point is simple "engineering" descriptions of quantum and biological computers. ■

ACKNOWLEDGMENTS

Dr. Martin Schoernig is acknowledged for extensive discussions and contributions to this article. Prof. Stephen Emmott, Dr. Alexei Kondratyev, Dr. Cat Mora, and Sarvesh Rajdev are thanked for reviewing numerous drafts of this article.

REFERENCES

1. G. Bassel, "Information processing and distributed computation in plant organs," *Trends Plant Sci.*, vol. 23, no. 11, pp. 30,184–30,185, 2018. doi: 10.1016/j.tplants.2018.08.006.
2. C. Bennett and R. Landauer. "The fundamental physical limits of computation." Scientific American, June 2011. www.scientificamerican.com/article/the-fundamental-physica l-limits-of-computation/ (accessed 2021).
3. "Reverse Engineering the source code of the BioNTech/Pfizer SARS-CoV-2 Vaccine." https://berthub.eu/articles/posts/reverse-engineering-source-code-of-the-biontech-pfizer-vaccine/ (accessed 2021).
4. D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 27, no. 6538, pp. 307–312, 1995. doi: 10.1038/376307a0.
5. P. J. Denning. "Ubiquity symposia." 2011. https://ubiquity.acm.org/symposia2011.cfm?volume=2011 (accessed 2021).
6. L. Kari and G. Rozenberg, "The many facets of natural computing," *CACM*, vol. 51, no. 10, pp. 72–83, Oct. 2008.
7. M. Lanzagortaa, and J. Uhlmannb, "Is quantum parallelism real?" in *Proc. SPIE*, 2008. [Online]. Available: https://www.researchgate.net/publication/252477910_Is_quantum_parallelism_real. doi: 10.1117/12.778019.
8. M. Mitchell, "Ubiquity symposium: Biological computation," *Ubiquity*, vol. 2011, p. 3, Feb. 2011. [Online]. Available: http://delivery.acm.org/10.1145/1950000/1944826/a1-mitchell.pdf?ip=128.16.12.209&id=1944826&acc=OPEN&key=BF07A2EE685417C5%2ED93309013A15C57B%2E4D4702B0C3E38B35%2E6D218144511F34378&—acm—=1546851025_35a94fcfcdef910b6f064e1ad8b0b5a7 doi: 10.1145/1940721.1944826.
9. Optical computing. Wikipedia, 2021. https://en.wikipedia.org/wiki/Optical_computing
10. R. Smith, M. Curtis, and W. Zeng, "A practical quantum instruction set architecture," 2017. [Online]. Available: https://arxiv.org/abs/1608.03355
11. "Quantum computing." Wikipedia, 2021. https://en.wikipedia.org/wiki/Quantum_computing (accessed 2021).
12. "Quantum programming." Wikipedia, 2021. https://en.wikipedia.org/wiki/Quantum_programming (accessed 2021).
13. "von Neumann architecture." Wikipedia, 2021. https://en.wikipedia.org/wiki/Von_Neumann_architecture (accessed 2021).

PHILIP TRELEAVEN is with University College London, London, WC1E 6BT, U.K. Contact him at p.treleaven@ucl.ac.uk.

DATA

EDITORS
NORITA AHMAD American University of Sharjah, nahmad@aus.edu
PREETI CHAUHAN IEEE Senior Member, preeti.chauhan@ieee.org

NASA

2021 State of the Practice in Data Privacy and Security

Preeti S. Chauhan, IEEE Senior Member
Nir Kshetri, University of North Carolina at Greensboro

The data privacy and security landscape is changing drastically due to new laws and regulations, the COVID-19 pandemic, and other emerging trends. This article looks at the current state of these issues.

The last year has been a turning point for data privacy and security. While there were releases of new data privacy laws such as the California Consumer Privacy Act (CCPA) and New York data privacy bill, the COVID-19 pandemic has altered the premise of data privacy in ways one would not have previously imagined. The increased reliance on social media and video communication platforms to stay connected, almost complete work-from-home transition across all global organizations over the last year, and massive adoption of online commerce and home-delivery services exposed major vulnerabilities across the board that increased the risk of data security breaches.

Not only that, the monumental data collection efforts by governments, health agencies, and organizations to support contact tracing, health screening, and vaccination record tracking for public health purposes has, in fact, made people more vulnerable to theft and/or leakage of their private information. The need for public and private institutions as well as individuals at large to employ strong data security measures has never been more critical. This article goes over the state of data privacy and security in 2021, including the latest trends, best practices, and threats.

DATA PRIVACY AND SECURITY DEFINITIONS AND SCOPE

The terms *data privacy* and *security* are often used interchangeably, but they actually mean vastly different things. One's private data may or may not remain secure or unknown to unintended users. An example would be when we inadvertently leak our private information due

Digital Object Identifier 10.1109/MC.2021.3083916
Date of current version: 30 July 2021

to a lack of appropriate data security measures, such as a weak password. In general, a more convenient method to store private data also makes the user more vulnerable to data security breaches.[1]

While data privacy and security can apply to any type of data, personally identifiable information (PII) is the most often discussed. PII includes information that can help trace the identity of a person by itself or in combination with other information directly or indirectly related to the individual. The scope of PII has been evolving and expanding, from driver's licenses, Social Security numbers, addresses, and so on to online personal data, social media posts, and IP addresses, among others.

Data privacy governs how data are collected, used, archived, shared, and deleted in accordance with the law. Recently, data privacy laws have continued to be developed and implemented all around the world. For example, the European Union (EU) enacted the General Data Protection Regulation (GDPR) in 2018 to govern the collection of personal information, including phone numbers, biometric data, IP addresses, and so on. Ireland, Australia, Denmark, Norway, Canada, Portugal, France, Brazil, Switzerland, and Iceland, among other countries, have strong privacy laws with a simple focus—the right of an individual to be left alone.

Security, on the other hand, is related to how information is protected.[2] It includes technical safeguards used to ensure the confidentiality, integrity, and availability of data.[3] Let's take an example of patient data management at hospitals to understand the difference between data privacy and security. It is common for patients to share their personal information with health-care providers. If the hospital protects the data against leaks and thefts, it is maintaining both data privacy and security.

However, if the hospital sells a patient's private information to a third party without the individual's consent, that is a breach of data privacy. In such cases, security measures are not of much use since the authorities with data access are allowing the privacy invasion.[4]

DATA PRIVACY CONTROLS VERSUS DATA SECURITY

Data privacy controls limit the sharing of nonessential information and ensure data governance systems are in place. The first step of data governance is the identification of business initiatives, mapping of the personal data assets and data flow, stakeholder identification for the data governance teams, and assessment of the security and privacy readiness.

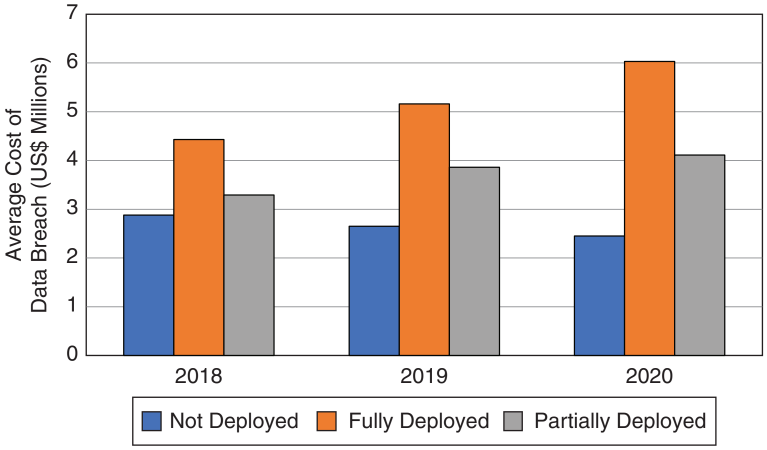


FIGURE 1. The average total cost of a data breach by security automation deployment level.[5]

The data governance team identifies risk tolerance, drives alignment on privacy policies, and develops plans for closing security gaps and data breach governance. The team also identifies any third-party vendors and associated data-sharing agreements.

In the execution phase, data assets are cataloged, and privacy policies and controls are implemented. This is also when the data subject access requests, integration of third-party vendors, and staff training on new processes takes place. The last and most important step is to measure and monitor the data governance process, course-correct as needed, and periodically test the data breach response processes.[5]

Data security tools and measures are utilized to prevent leaks and hacks of private user data. Organizations need to start by identifying the sensitive data and classifying them according to the data category and level of sensitivity. The usage policy for data needs to be set up to identify who has access to data according to data classification; the access time frame; and rules around the data usage, which should be allowed based on an individual's need to know—be it read-only or full access.

Data should be protected both physically and with endpoint security systems such as antivirus software, antispyware, pop-up blockers, and firewalls. Multifactor authentication is a very useful and powerful tool to prevent data hacks by providing a second layer of authentication for data access.[6]

An IBM study on security automation states that businesses without security automation experience an average cost of US$6.03 million in data breaches, which is more than double the average cost of US$2.45 million for companies with fully deployed security automation (Figure 1). Given the increased data security risks with remote and hybrid work environments, it is expected that the security adoption will continue to grow in 2021 and beyond.[5]

2021 STATE OF DATA PRIVACY AND SECURITY

Data privacy and security continues to evolve and has changed tremendously as a result of COVID-19 situation around the world. New macro trends of working from home, increased reliance on social media to stay connected due to lockdowns and sheltering in place, cryptocurrency growth and the resulting ransomware attacks, and so on have contributed to the emerging trends in data privacy and security risks.

Contact tracing

Contact tracing as a means to identify COVID-19 exposure risks for the community was a major initiative from governments and health agencies throughout the world. However, privacy concerns have resulted in lower adoption rates of these centralized contact-tracing tools. The applications upload the anonymized user data to a remote server, which matches the data with the user contacts, should a person start to develop COVID-19 symptoms. Besides privacy, a couple of other reasons—low GPS location precision and the discontinuation of the tracing efforts among states—severely limited the effectiveness of these tools.

The alternative technology—decentralized applications that broadcast rotating, randomized Bluetooth identifiers—has gained more traction and is also better at preserving the user's privacy. The contact-tracing collaboration by Apple and Google based on such Bluetooth identifiers has demonstrated that the data-sharing efforts can be implemented without tracking user locations or collecting PII data.

This emerging model might achieve wider adoption by other companies for non-COVID-related applications as well as to understand user preferences and activities to support their businesses. However, this will come with new privacy and data security concerns, which will need to be thought through ahead of time.

Remote/hybrid work

Remote/hybrid work has emerged as a major change owing to COVID-19 pandemic social distancing requirements. Major organizations have driven people, processes, and culture to adapt to the new reality. Some of the challenges have been to determine secured technologies to conduct confidential meetings in a remote workspace and manage confidential data outside of remote places. There are increased vulnerabilities in the form of phishing email attacks, unauthorized access through unsecured remote-access tools, hacking of video conference tools, and so on. There is a need to do periodic risk assessments, perform routine monitoring, and secure the tools enabling the remote work.

COVID-19-related medicalt and personal information

To keep the on-site business running, companies have developed new processes for COVID-19 testing as well as employee data collection to enable contact tracing. These measures enable the timely release of warnings for people who might have come in contact with a COVID-19-affected individual at the workplace. The user data collection is expected to continue throughout 2021 with the addition of the COVID-19 vaccination program. It will require organizations and companies to have systems in place to enable secure data collection, storage, and release of individuals' medical and COVID-19-related data.

Biometrics

Biometrics are physical or human characteristics that can be used to digitally identify a person, typically to give access to devices, data or systems, and so on. Examples include fingerprints, facial patterns, and voice, among others.

In 2019, the Illinois Supreme Court released Illinois's Biometric Information Privacy Act (BIPA), which states that collecting biometric information without a release or sharing the biometric information with a third party without consent would be a violation. Individuals can allege a violation of their rights to qualify as an aggrieved person and, in turn, be entitled to seek monetary compensation and injunctive relief under the act.

BIPA litigations related to biometric timekeeping-/access-related gaps have been impacted by COVID-19. Biometric data in the form of thermal scanners, facial recognition tools, and so on are being collected to have COVID-19 screening programs in place. In late 2020, there was a lawsuit filed on behalf of a company's employees, alleging that their consent was not obtained for the employer's COVID-19 screening program, which required the workers to undergo facial geometry and temperature scans to enter the company warehouses. It was alleged that the company violated BIPA by making its employees go through the above screening program, without their consent. The states of Washington, Texas, and California have similar privacy laws like BIPA, while Arizona, Idaho, Massachusetts, and New York are in the process of proposing similar legislation.[7]

Ransomware

Ransomware is a malware attack that uses asymmetric encryption to hold a user's or organization's critical data for ransom using a pair of keys to encrypt and decrypt a file. The attackers make the decryption key available to the victim upon payment, failing which the data are lost forever. While ransomware attacks used to be different from hacking, wherein the user data gets stolen, nowadays, nearly half of ransomware attacks do steal data before encrypting systems, which makes them a full cybersecurity incident.

Ransomwares have increased 62% globally and by a 158% spike in North America since 2019.[8] A large part of the dramatic rise in 2020 has been due to the work-from-home policies implemented by organizations worldwide when COVID began, which have opened up a lot of security vulnerabilities for many organizations. In May 2021, Colonial Pipeline was held for a ransom of US$5 million by the DarkSide hacking group.[9] The company ended up paying the ransom in cryptocurrency, which has become a preferred means of ransom payment. Cryptocurrency is a virtual currency that is secured by cryptography, decentralized, and based on blockchain-distributed ledger technology, all of which allows hackers to hide their identities through the use of mixes and tumbler services.

DATA SECURITY TECHNOLOGY, TOOLS, AND TRENDS

Many organizations and individuals are devoting more resources to improving their defenses against cyberthreats. According to Juniper Research's report *The Future of Cybercrime & Security: Enterprise Threats & Mitigation 2017–2022*, global cybersecurity spending will reach US$135 billion in 2022.[10] Cybersecurity Ventures estimates even higher global cybersecurity spending: US$1 trillion for the five-year period from 2017 to 2021, or an average of US$200 billion annually.[11]

Organizations' awareness of emerging threats and several other factors have led to increased cybersecurity spending. First, major countries have issued new regulations that emphasize strong cybersecurity measures. For instance, China's Cybersecurity Law, enacted in 2017, requires financial services firms to have IT infrastructures that meet various specifications. They also need to pass standard cybersecurity tests and have certifications. Failure to comply can result in heavy fines and even criminal charges.[12]

A second factor is the shifting customer mindset. Customers expect companies to give a high priority to cybersecurity. According to "RSA Data Privacy & Security Report," based on a survey in Europe and the United States, 62% of the respondents said that they would blame the company, instead of cybercriminals, if their data were breached.[13] Finally, the evolution to a digital business strategy has stimulated cybersecurity spending.[14] Companies are gathering more data on consumers. Effective measures to secure sensitive information and maintain privacy are important to build and retain trust in brands.

Firms are deploying advanced and sophisticated tools, such as artificial intelligence (AI), machine learning (ML), and blockchain to fight cyberattacks. AI's use in cybersecurity has gained prominence in recent years. For instance, AI can analyze large numbers of documents, server logs, and other information to identify, classify, and present possible cyberthreats. Doing so is difficult and time consuming for human cybersecurity analysts. AI programs can generate real-time reports of cyberthreats, which can help the cybersecurity team to identify and resolve them quickly.[15]

However, due to challenges, such as too many false positives of AI systems, some analysts have recommended using multiple AI algorithms to fight cybercrimes. Resistant AI, a cybersecurity company in Czech Republic, uses up to five different ML modules to make a decision.[15]

Likewise, blockchain has the potential to significantly strengthen organizations' cybersecurity practices. For instance, a party can cryptographically sign transactions, and, by verifying the cryptographic signatures, the recipient can ensure that the transaction originated from a trusted source. There is no need to store sensitive information with third parties. Many interlocked computers hold identical information, and, if one computer's blockchain updates are breached, the discrepancy is noticed by all computers, and the system rejects it.

THREATS

Organizations and individuals face multiple and diverse cyberthreats. According to "The Economic Value of Prevention," a report from the Ponemon Institute, phishing, Domain Name System-based attacks, viruses, bots, distributed denial-of-service, and ransomware are among the most common types of cyberattacks facing organizations.

These attacks are growing at alarming rates. For instance, by 2019, there were about 980 million malware programs, and 350,000 new malware types were detected every day.[16] In 2019, SonicWall recorded 9.9 billion malware attacks.[17] Kaspersky Lab detected more than 482 million phishing attempts in 2018, compared to 236 million such attempts in 2017.[18]

According to the 2020 Internet Crime Report released by the U.S. Federal Bureau of Investigation's Internet Crime Complaint Center, the agency received 791,790 cybercrime complaints in 2020 compared to about 300,000 in 2019. The reported losses from cybercrime in 2020 were US$4.2 billion. The top three categories of crimes in 2020 were phishing, nonpayment/nondelivery, and cyberextortion.[19]

Increased Digitization and Social Media Use

More than 60% of the world's population is online, and 2020 marked the year in which more than half of the world's population had used social media. As of April 2021, the biggest social media company, Facebook, had 2.8 billion monthly active users.[20] Due to the huge size and high-quality information, social media is an attractive target for cybercriminals.

Social media users have been victims of high-profile privacy violations and security breaches. A study of the cloud-based email security vendor Vade Secure found that Facebook was the second-most impersonated brand in phishing attacks. Among the 25 most impersonated brands in the fourth quarter of 2019, three were social media websites.[21]

As a recent example, the media widely reported in April 2021 that a user in a hacking forum published the personal information of more than 533 million Facebook users from 106 countries. The exposed data included the phone numbers, Facebook IDs, full names, locations, birthdates, biographies, and, in some cases, email addresses of the victimized social media users.[21]

Social media companies have also been found to engage in illegal sharing of personal information. In 2020, South Korea's information protection regulator, the Personal Information Protection Commission, fined Facebook US$6.1 million. From mid-2012 to mid-2018, Facebook allegedly shared 3.3 million South Korean users' personal information with as many as 10,000 companies without users' consent.[22]

COVID-19-Led Increase in Data Privacy and Security Risks

While broad public support existed for protective measures against COVID-19, concerns have been raised about the intrusiveness of such measures on data privacy.[23] Many COVID-19 tracking apps perform poorly in privacy and security. A study published in *Nature Medicine* in May 2020 analyzed 50 such apps, including 20 issued by government agencies in developing and developed countries. The analysis found that only 16 had indicated that they would make users' data anonymous, encrypt and secure them, and report only in an aggregated format.[24]

Likewise, an analysis by the independent watchdog agency International Digital Accountability Council (IDAC) of 108 COVID-19 apps across 41 countries found that many of the apps failed to follow best privacy and security practices.[25] IDAC's report, published in June 2020, found that many apps were using third-party software development kits, which raised the possibility that data could have been shared with outside organizations without users' consent. The apps lacked transparency about the information collected, and some failed to encrypt transmitted information.[26]

Moreover, systemic cyberrisks are posed due to remote working in the COVID-19 environment.[23] Lower safeguard standards are likely when people work from home.[12] For instance, employees working remotely may use their own devices, such as phones, laptops, and tablets. Unlike devices issued by organizations, personal devices are less likely to be patched for the latest vulnerabilities. Consumer-level systems focus more on ease of use and often lack options for customization. Enterprise-level systems, on the other hand, are designed to protect larger organizations and come with additional resources and features to strengthen security.[28]

Social control mechanisms that protect workers from dangerous cyberattacks do not operate in remote working. For instance, employees' in-person interactions with coworkers and supervisors may shield them from unsafe cyberpractices at work, which are not available in a remote working environment.[29]

DATA PRIVACY REGULATIONS

To improve the legal clarity and certainty around data privacy and strengthen cybersecurity, many jurisdictions are revamping their regulatory systems. In this section, we provide a brief overview of data privacy and security regulations worldwide.

The EU

The EU's GDPR, which is viewed as the world's most comprehensive data privacy legislation, went into effect on 25 May 2018. GDPR has made it mandatory to notify the relevant supervisory authority of any data breach within 72 h of becoming aware of the event. The number of such notifications as well as the fines imposed for privacy violations and data breaches have increased dramatically (Figure 2).

Many new privacy laws have been inspired, at least in part, by GDPR, which has dramatically changed the processes that organizations need to follow to track consumers' online behaviors and process that data. Under GDPR, companies are required to obtain specific legal bases to use customers' data or track their behaviors. Many companies choose consent as the legal basis.[31] GDPR requires companies to have an explicit opt-in consent from customers to keep personal data.

In a context of international comparison, GDPR's Article 45 is of special interest; it gives the European Commission (EC) the power to determine whether a country outside the EU provides an adequate level of data protection. If a non-EU country meets the adequacy standard, data flow from the EU to that country is treated in the same manner as intra-EU transmissions of data.[32]

The United States

Compared to the EU, the United States provides more autonomy to businesses regarding the way they disclose and store personal information. Organizations' data privacy and security frameworks are subject to a number of federal laws and regulations enacted to protect the privacy, security, and confidentiality of specific categories of data and information. In addition to federal laws, there are about 47 different state laws regarding how people should be notified in the case of a data breach involving personal information.[33]

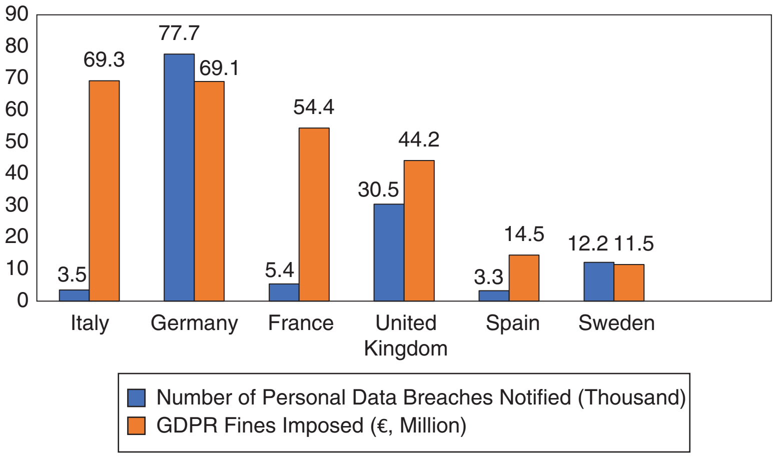


FIGURE 2. The top six economies imposing GDPR fines (25 May 2018 to 27 January 2021).[30]

Among most notable state-level legislation in the United States, the CCPA became effective on 1 January 2020. It is largely modeled after GDPR but is less stringent. Both GDPR and the CCPA strongly emphasize transparency. For example, to comply with the CCPA, a business is required to include a section in its privacy policy that describes the rights of consumers. It must provide clear instructions regarding how consumers can opt out of the sale of their personal information.

In November 2020, the California Privacy Rights Act (CPRA) was passed, which will replace and build on the CCPA. The CPRA will take full effect in 2023 and will give users new rights, such as the right to correct inaccurate information, right to have personal information collected subject to data minimization and purpose limitations, and right to receive notice from businesses planning on using sensitive personal information and ask them to stop.[34]

Among other U.S. states, Virginia's governor signed the Consumer Data Protection Act (CDPA) into law in March 2021, which will take effect January 2023. Like the CPRA, the CDPA requires companies to publish privacy policy notices that explain the ways they use, collect, and share personal data. Without consumers' affirmative consent, companies cannot collect and process their personal data. The CDPA also gives individuals the right to ask whether a company is storing and processing their personal information. In addition, they can request the deletion and correction of their personal data. Virginia consumers also have the right to opt out of the sale of their personal data as well as the use of such data to create targeted advertising.[35]

Asia

As of May 2021, Japan was the only Asian country that was granted an Adequacy Decision by the EC. South Korea enacted the Personal Information Protection Act (PIPA) in 2011. In addition, the country has sector-specific data privacy legislation.[36] The PIPA requires private- and public-sector entities collecting information that identifies a specific person to meet strict compliance requirements.[37]

In China, the 2012 Online Data Protection Regulation bans the sale and distribution of personal information without the owner's consent. It requires Internet service providers to ensure the security of personal data and prevent misuse as well as provides consumers the right to seek the deletion of personal data posted without consent and sue for violations. However, the Chinese government's state power allows it to get unlimited access to citizens' personal information for surveillance.[38]

As of April 2021, India lacked a comprehensive data privacy law to protect personal data. In December 2019, India's lower house of the bicameral parliament, Lok Sabha, introduced the Personal Data Protection Bill, 2019. It specifies how the data of Internet users are stored, processed, and transferred. The bill was tabled as of April 2021.[39]

Latin America

After GDPR's implementation, significant amendments and improvements in privacy laws were made by major Latin American countries. As of May 2021, the EC had recognized Argentina and Uruguay as jurisdictions that provide adequate protection. Argentina changed its data protection laws enacted to align with GDPR. In 2018, a bill was proposed that contains key provisions of GDPR, such as a requirement for governmental agencies processing sensitive and big data to appoint a data protection officer and standards for the lawfulness of data processing.[40] Likewise, the Brazilian General Data Protection Law (Lei Geral de Proteção de Dados) was ratified in the Congress in mid-2018 and became effective on 18 September 2020.

Africa

In most countries in Africa, underdeveloped regulatory regimes fail to provide adequate data privacy protection. As of early 2021, about half of Africa's 53 countries had adopted some form of data privacy regulations.[41] Likewise, in 2014, the African Union's convention on cybersecurity and personal data was adopted. As of May 2021, 14 nations had signed it, and eight had ratified and/or accessed the conventions.[27]

Data privacy and security are and will continue to cause an evolving conversation around the world. The increased threats of data leakages and thefts from high digitization and social media usages as well as new vulnerabilities resulting from the hybrid work culture need to be continuously monitored and addressed.

The complexities around privacy require action and active participation from individuals, businesses, and government agencies. It is important to improve legal clarity and certainty around data privacy and security through regulations at the state and federal levels in the United States as well as around the world.

At the individual and business levels, it is necessary to remain cognizant of cybersecurity threats and vulnerabilities and actively work to address these with advanced and sophisticated tools, such as AI, ML, and blockchain, to improve privacy controls and ward off cyberattacks. In these unprecedented times, the definition, scope, and impact of data privacy and security have gone through tremendous changes, and it will require individual, corporate, and government actions to ensure that the user data are handled in a manner that does not compromise privacy and security. C

ACKNOWLEDGMENT

We thank *Computer*'s editor in chief, Jeff Voas, for suggestions on a previous version of this article.

DISCLAIMER

The views, thoughts, and opinions expressed in the article belong solely to the authors, and not necessarily to their employer or organizations, committees, or groups that the authors may be a part of.

REFERENCES

1. S. Ritter. "Data privacy Vs. data secrecy: The danger of worrying about the wrong issue." Forbes, Sept. 3, 2019. https://www.forbes.com/sites/forbestechcouncil/2019/09/03/data-privacy-vs-data-secrecy-the-danger-of-worrying-about-the-wrong-issue/?sh=3fe3585b3ce9
2. S. Symanovich. "Privacy vs. security: What's the difference?" Norton, Jan. 18, 2020. https://us.norton.com/internetsecurity-privacy-privacy-vs-security-whats-the-difference.html (accessed Mar. 20, 2021).
3. K. Schwartz. "Data privacy and data security: What's the difference?" IT-Pro Today, May 2, 2019. https://www.itprotoday.com/security/data-privacy-and-data-security-what-s-difference (accessed Mar. 20, 2021).
4. L. O. Gostin, "Health information privacy," *Cornell Law Rev.*, vol. 80, no. 3, pp. 451–528, 1995.
5. F. Velez. "What's in store for data privacy in 2021?" CPO Magazine, Jan. 18, 2021. https://www.cpomagazine.com/data-privacy/whats-in-store-for-data-privacy-in-2021/ (accessed Mar. 2, 2021).
6. "9 data security best practices for 2021." Team LoginRadius, Dec. 09, 2020. https://www.loginradius.com/blog/start-with-identity/2020/12/data-security-best-practices/ (accessed Mar. 20, 2021).
7. M. T. Costigan. "Top 10 for 2021 – Happy data privacy day!" Workplace Privacy Report, Jan. 28, 2021. https://www.workplaceprivacyreport.com/2021/01/articles/written-information-security-program/top-10-for-2021-happy-data-privacy-day/ (accessed Mar. 20, 2021).
8. C. Matthews. "Bitcoin extortion: How cryptocurrency has enabled a massive surge in ransomware attacks," MarketWatch, May 15, 2021. https://www.marketwatch.com/story/bitcoin-extortion-how-cryptocurrency-has-enabled-a-massive-surge-in-ransomware-attacks-11621022496 (accessed Mar. 20, 2021).
9. "Ransomware soars with 62% increase since 2019," Security Magazine, Mar. 16, 2021. https://www.securitymagazine.com/articles/94831-ransomware-soars-with-62-increase-since-2019#:~:text=Ransomware%20reaches%20new%20heights%20with,to%20earn%20an%20easy%20payday (accessed Mar. 20, 2021).
10. "Cybercrime & the internet of threats 2018," Juniper Research, 2017. https://www.juniperresearch.com/whitepapers/cybercrime-the-internet-of-threats-2018 (accessed Mar. 20, 2021). (accessed Mar. 20, 2021).
11. B. Sterling. "Global cybercrime. Costs a trillion dollars Maybe 3." Wired, July 19, 2017. https://www.wired.com/beyond-the-beyond/2017/07/global-cybercrime-costs-trillion-dollars-maybe-3/ (accessed Mar. 20, 2021).
12. N. Kshetri, *Cybersecurity Management: An Organizational and Strategic Approach*. Toronto: The Univ. of Toronto Press, 2021
13. M. Nadeau. "General Data Protection Regulation (GDPR) requirements, deadlines and facts." CSO, 2018. https://www.csoonline.com/article/3202771/data-protection/general-data-protection-regulation-gdpr-requirements-deadlines-and-facts.html (accessed Mar. 20, 2021).
14. "Gartner forecasts worldwide security spending will reach $96 billion in 2018, up 8 percent from 2017." Gartner, 2017. https://www.gartner.com/en/newsroom/press-releases/2017-12-07-gartner-forecasts-worldwide-security-spending-will-reach-96-billion-in-2018 (accessed Mar. 20, 2021).
15. J. Kahn. "Cybercriminals adapt to coronavirus faster than the A.I. cops hunting them." Fortune, Apr. 30, 2020. https://fortune.com/2020/04/30/cybercriminals-adapt-to-coronavirus-faster-than-the-a-i-cops-hunting-them (accessed Mar. 20, 2021).
16. B. Jovanović. "Malware statistics – You'd better get your computer vaccinated." DataProt. 2019. https://dataprot.net/statistics/malware-statistics/ (accessed Mar. 20, 2021).
17. "2020 Sonicwall cyber threat report: Threat actors pivot toward more targeted attacks, evasive exploits." Sonicwall, San Jose, CA, Feb. 4, 2020. [Online]. Available: https://www.sonicwall.com/news/2020-sonicwall-cyber-threat-report/
18. "Phishing attacks more than doubled in 2018 to reach almost 500 million." Kaspersky, 2019. https://www.kaspersky.com/about/press-releases/2019_phishing-attacks-more-than-doubled-in-2018 (accessed Mar. 20, 2021).
19. "FBI Releases the internet crime complaint center 2020 internet crime report, including COVID-19 scam statistics," FBI National Press Office, Washington, D.C., Mar. 17, 2021. [Online]. Available: https://www.fbi.gov/news/pressrel/press-releases/fbi-releases-the-internet-crime-complaint-center-2020-internet-crime-report-including-covid-19-scam-statistics
20. Q. Wong. "Facebook says iPhone users will start seeing new privacy prompt today." CNET, Apr. 26, 2021, https://www.cnet.com/news/facebook-says-iphone-users-will-start-seeing-new-privacy-prompt-today/ (accessed May 20, 2021).
21. A. Holmes. "533 million Facebook users' phone numbers and personal data have been leaked online." Business Insider. Apr. 3, 2021, https://www.businessinsider.com/stolen-data-of-533-million-facebook-users-leaked-online-2021-4 (accessed May 20, 2021).
22. S. Ikeda. "South Korean regulator fines Facebook for privacy violations; social media giant shared personal data without user consent," CPO Magazine, Dec. 3, 2020. https://www.cpomagazine.com/data-privacy/south-korean-regulator-fines-facebook-for-privacy-violations-social-media-giant-shared-personal-data-without-user-consent/ (accessed May 20, 2021).

23. D. Mikkelsen, H. Soller, and M. Strandell-Jansson, "Privacy, security, and public health in a pandemic year," McKinsey & Company, New York, June 15, 2020. [Online]. Available: https://www.mckinsey.com/business-functions/risk/our-insights/privacy-security-and-public-health-in-a-pandemic-year

24. T. Sharma and M. Bashir, "Use of apps in the COVID-19 response and the loss of privacy protection," *Nature Med.*, vol. 26, no. 8, pp. 1165–1167, May 26, 2020. [Online]. Available: https://www.nature.com/articles/s41591-020-0928-y. doi: 10.1038/s41591-020-0928-y.

25. E. Reuter. "Report: COVID-19 apps fall short in privacy, security." Med City News, June 8, 2020. https://medcitynews.com/2020/06/report-many-covid-19-apps-fall-short-in-privacy-security/ (accessed May 20, 2021).

26. "Privacy in the age of COVID: An IDAC investigation of COVID-19 apps," Digital Watchdog, June 5, 2020. https://digitalwatchdog.org/wp-content/uploads/2020/06/IDAC-COVID19-Mobile-Apps-Investigation.pdf (accessed May 20, 2021).

27. "List of countries which have signed, ratified/acceded to the African Union Convention on Cybersecurity and Personal Data Protection." https://au.int/sites/default/files/treaties/29560-sl-AFRICAN%20UNION%20CONVENTION%20ON%20CYBER%20SECURITY%20AND%20PERSONAL%20DATA%20PROTECTION.pdf (accessed May 20, 2021).

28. J. Kelly. "Consumer vs. enterprise security: There is a difference." Law Technology Today, Sept. 5, 2019. https://www.lawtechnologytoday.org/2018/09/consumer-vs-enterprise-security/ (accessed May 20, 2021).

29. J. Boehm, J. Kaplan, and N. Sportsman "Cybersecurity's dual mission during the coronavirus crisis," McKinsey & Company, New York, Mar. 25, 2020. [Online]. Available: https://www.mckinsey.com/business-functions/risk/our-insights/cybersecuritys-dual-mission-during-the-coronavirus-crisis?cid=other-eml-alt-mip-mck&hlkid=34fdd9f7b5154c8d88ee3d25ac6cfd3f&hctky=2762145&hdpid=ffacc134-7120-4a94-ba44-bd4b2fe75bcc

30. "DLA Piper GDPR fines and data breach survey: January 2021." DLA Piper. https://www.dlapiper.com/en/uk/insights/publications/2021/01/dla-piper-gdpr-fines-and-data-breach-survey-2021/

31. D. Meyer. "GDPR attacks: First Google, Facebook, now activists go after Apple, Amazon, LinkedIn." ZDNet. May 29, 2018. https://www.zdnet.com/article/gdpr-attacks-first-google-facebook-now-activists-go-after-apple-amazon-linkedin/ (accessed May 20, 2021).

32. "Adequacy decisions: How the EU determines if a non-EU country has an adequate level of data protection," European Commission, Brussels, Belgium. https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en (accessed May 20, 2021).

33. R. King. "New EU cyber security directive to impact U.S. companies." The Wall Street Journal, Feb. 7, 2013. http://tinyurl.com/cjujzgl (accessed May 20, 2021).

34. "California privacy rights act: An overview." Privacy House Clearing House, Dec. 10 2020. https://privacyrights.org/resources/california-privacy-rights-act-overview (accessed May 20, 2021).

35. A. Nicodemus. "More than a CCPA clone? Virginia passes nation's second comprehensive privacy law." Compliance Week, Mar. 3, 2021. https://www.complianceweek.com/data-privacy/more-than-a-ccpa-clone-virginia-passes-nations-second-comprehensive-privacy-law/30104.article (accessed May 20, 2021).

36. "Data protection laws of the world: South Korea." DLA Piper. https://www.dlapiperdataprotection.com/index.html?t=law&c=KR (accessed May 20, 2021).

37. H. Chan. "Pervasive personal data collection at the heart of South Korea's COVID-19 success may not translate," Thomson Reuters. Mar. 25, 2020 https://blogs.thomsonreuters.com/answerson/south-korea-covid-19-data-privacy/

38. E. Feng, "In China, a new call to protect data privacy," NPR, Washington, D. C., Jan. 5, 2020. [Online]. Available: https://www.npr.org/2020/01/05/793014617/in-china-a-new-call-to-protect-data-privacy

39. A. Chakravarty and A. Sivasubramanian. "The Privacy question in India's drone regulation." Jurist, Apr. 14, 2021. [Online]. Available: https://www.jurist.org/commentary/2021/04/chakravarty-sivasubramanian-privacy-drone/

40. M. Egan, "Data privacy reform gains momentum in Latin America," IDB, New York, Feb. 12, 2019. [Online]. Available: https://blogs.iadb.org/conocimiento-abierto/en/data-privacy-reform-gains-momentum-in-latin-america/

41. J. Daniel. "Data protection laws in Africa: What you need to know." CIO, Feb. 14, 2021. https://www.cio.com/article/3607734/data-protection-laws-in-africa-what-you-need-to-know.html (accessed May 20, 2021).

PREETI S. CHAUHAN is a technical program manager at Google, Sunnyvale, California, 94089, USA. She is a Senior Member of IEEE. Contact her at preeti.chauhan@ieee.org.

NIR KSHETRI is a professor of management in the Bryan School of Business and Economics at the University of North Carolina at Greensboro, Greensboro, North Carolina, 27412, USA. Contact him at nbkshetr@uncg.edu.

75 YEARS 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 12 magazines and 17 journals. 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 three software developer credentials. For more information, visit www.computer.org/certification.

BOARD OF GOVERNORS MEETING

14, 15, 16, and 18 June 2021, virtual

IEEE

revised 2 April 2021

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:** help@computer.org
Los Alamitos: 10662 Los Vaqueros Cir., Los Alamitos, CA 90720; **Phone:** +1 714 821 8380; **Email:** help@computer.org

MEMBERSHIP & PUBLICATION ORDERS

Phone: +1 800 678 4333; Fax: +1 714 821 4641; Email: help@computer.org

Digital Object Identifier 10.1109/MC.2021.3055689