IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 19, 2020, accepted August 2, 2020, date of publication August 10, 2020, date of current version August 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015510

Data-Driven Security for Smart City Systems: Carving a Trail

NADER MOHAMED[1], (Member, IEEE), JAMEELA AL-JAROODI[2], (Member, IEEE), IMAD JAWHAR[3], (Member, IEEE), AND NADER KESSERWAN[2]

[1]Department of Computer Science, Information Systems, and Engineering, California University of Pennsylvania, California, PA 15419, USA
[2]Department of Engineering, Robert Morris University, Moon, PA 15108, USA
[3]Faculty of Engineering, Al Maaref University, Beirut 1600, Lebanon

Corresponding author: Nader Mohamed (mohamed@calu.edu)

ABSTRACT Smart cities rely heavily on collecting and using data. Smart systems are implemented and deployed to provide intelligent features that help improve efficiency and quality of life. This creates a huge repository of data representing many aspects of smart city operations. Many data-driven applications can take advantage of this data to further improve the ''smartness'' of a smart city. At the same time, smart city systems, being very large-scale distributed systems and highly integrated with the physical infrastructure and residents of the city, pose immense security challenges as well. So why don't we take advantage of this data to improve security measures? In this paper we propose the use of data-driven security approaches to secure smart city systems. To illustrate the significance of this approach we first identify the different challenges for securing smart city systems given the unique characteristics of these systems. Then we discuss the benefits of using data-driven security. Furthermore, we categorize the different types of security applications (features) needed to help capitalize on the data needs and benefits. We also discuss the how these categories of applications can alleviate some of the challenges. In addition, we highlight possible future research directions to incorporate effective data-driven security in smart city systems.

INDEX TERMS Smart city, data-driven security, cybersecurity, security protection techniques, security policies, security management.

I. INTRODUCTION

The number of people living in cities has grown from 746 million people in 1950 to around 3.9 billion in 2014 [1]. This number is expected to grow to over 6 billion by 2050. Consequently, a number of cities are quickly growing into mega cities. For instance, the number of big cities, with more than 10 million people, will increase from 10 in 1990 to 41 in 2030. Accordingly, many issues will arise with respect to the administration of these mega cities and offering decent quality of life for their inhabitants. One feasible solution, some cities are either working on or considering, is transforming into smart cities [2], [3].

Developments in the Internet of Things (IoT), cyber-physical systems (CPS), cloud computing, fog computing, communication technologies and software technologies can contribute to effectively design and implement smart cities. Many exceptional prospects can be presented by these technologies to develop smart city applications such as smart energy grids, smart water networks, smart transportation systems, smart healthcare systems and smart public safety and security. These applications utilize the advantages in information and communication technology (ICT) to form smart city systems that offer enhanced quality of life for residents, improved city resources utilizations, and better sustainability. While smart city systems can offer many benefits, security threats are a major obstacle in realizing these benefits. As most smart city systems are networked, they can be highly exposed to many potential security risks. However, the risks associated with attacks on a smart city can be more harmful as they will impact humans and systems. The security attacks on smart city systems can lead to damages in city infrastructures, reduced quality of life, reduced resource utilization efficiency, and reduced sustainability. In extreme cases, it could lead to human harm as in injuries, death, and wrongful accusations.

While there are many security techniques developed to protect smart city systems, there are also more complex and advanced security attacks being developed every day. There are many challenges in rapidly identifying these attacks,

The associate editor coordinating the review of this manuscript and approving it for publication was Kaitai Liang.

analyzing them effectively, preventing them completely, and/or reducing their associated risks. One emerging approach that can provide many advantages is data-driven security [4]–[6]. Data-driven security relies on security data, data analytics, machine learning, data visualizations, and dashboards to improve systems and networks protection. Our work here aims to discuss the challenges of securing smart city systems in light of the unique characteristics these have. In addition, we provide a thorough discussion of the benefits of using data-driven security for smart city systems and categorize the different types of security applications that can benefit from this approach and discuss their issues and how they will take advantage of the approach to help resolve them. Finally we highlight some ideas and research directions for the future to further improve and incorporate data-driven security in smart city systems.

The rest of the paper is organized as follows. Section II covers some related work. Section III discusses the primary concepts of smart cities, their enabling technologies, data-driven security, and IoT Security. We then discuss different security challenges of smart cities in Section IV. Section V outlines the benefits of using data-driven security for smart city systems. Section VI defines the different types of security applications and issues of applying data-driven security for smart city systems. Section VII offers a big picture view for utilizing data-driven security for smart city systems and its possible impact. In Section VIII we suggest different areas open for future research and development efforts and advancements, then conclude the paper in Section IX.

II. RELATED WORK

Several researchers investigated and highlighted the importance of securing smart cities and the associated issues. General security and privacy issues and possible solutions in smart cities are studied by Zhang *et al.* [7], Elmaghraby and Losavio [8], Bartoli *et al.* [9], Kitchin [10], Ijaz *et al.* [11], Khatoun and Zeadally [12], Laufs *et al.* [13], Verma *et al.* [14], Cui *et al.* [15], Braun *et al.* [16], and Kitchin and Dodge [17]. None of these papers focused on studying data-driven security in particular. Others studied different aspects of smart city security including Biswas and Muthukkumarasamy [18] focusing on utilizing blockchain for smart cities, Wu *et al.* [19] with focus on defending against sophisticated attacks on wireless Sensor networks, Khan *et al.* [20] discussing cloud based smart cities data security and privacy management, Sen *et al.* [21] with focus on the role of software in smart city security and privacy, Baig *et al.* [22] focusing on digital forensic issues, Wang *et al.* [23] discussing data security and threat modeling, Chakrabarty and Engels [24] proposing a secure IoT architecture, Farahat *et al.* [25] focusing on protecting citizen data, Aloqaily *et al.* [26] focusing on intrusion detection for connected vehicles, and Hasbini *et al.* [27] outlining the information security management role in smart city organizations. In addition, Edwards [28] studied privacy, security and data protection in smart cities from the law perspective. Hassan *et al.* [29] used deep learning model to extract meaningful features, similar to discovering facts and patterns in data-driven security, to efficiently detect network intrusions.

There have been many studies recently on utilizing data-driven approaches for improving operations of different aspects of smart cities. Examples include general city management [30], urban water management [31], public transportation management [32], vehicular network improvements [33], rail transit safety [34], crisis response and disaster resilience [35], communication performance management [36], load forecasting in buildings [37], energy management [38]–[40], and in general city decision-making processes [41], [42]. These studies show the advantages of using data-driven approaches for smart cities. Yet, we have not seen any work specifically discussing the use of data-driven security solutions for smart cities.

Unlike other research papers, the main contributions of this paper are in identifying potential data-driven security applications for smart city systems, the benefits of data-driven security for such environments, and the issues associated with its use. In addition, we outline future research directions to further facilitate the effective use of data-driven security for smart city systems.

III. PRELIMINARY CONCEPTS

The three main players in this paper are smart city systems comprised of various infrastructures, humans and ICT, data-driven security solutions that can be used to protect smart city systems, and IoT security as it comprises the most common part of the infrastructure for smart cities. Therefore, in this section we provide some preliminary background information on them.

A. SMART CITIES AND THEIR ENABLING TECHNOLOGIES

The smart city goals are to improve the quality of life of the inhabitants, enhance the use of city resources, improve sustainability, and decrease the harmful on the environment. This idea depends a recipe of competence and optimization techniques, technological inventions, and both historic and live data to realize these goals [43]. The smartness level of a city can be assessed by numerous aspects including smart people, smart living, smart economy, smart mobility, smart environment, and smart governance. Several cities in different countries have decided to become smart cities for numerous motives. These include speedily growing populations, the lack of enough open spaces for expansion and development, inadequate resources, improved awareness on energy efficiency, environmental sustainability, and economic development.

The architecture of Smart Cities can be different depending on the types of applications a city may want to implement [24], [44]. Smart city systems may only target increasing utility efficiency (e.g. power, water, gas, etc.), while others aim to increase the safety and effectiveness of law enforcement [3]. In addition, many cities strive to achieve all possible benefits of becoming smart cities. These smart

city systems are enabled by the use of ICT [45]. Examples of ICT used to implement smart city systems include: the Internet of Things (IoT) to enable connecting different smart city physical devices and sensors [46]; the Internet of Services (IoS) to enable providing services for different smart city systems and organizations via the Internet [47]; cloud and fog computing platforms to provide scalable computation and data storage capabilities and other advanced services for different smart city systems [48]; Cyber-Physical Systems (CPS) to facilitate useful interactions between the cyber world and the physical world [49]; Unmanned Aerial Vehicles (UAV) to provide fast mechanisms for delivery, infrastructure inspections, environmental and security monitoring for smart cities [50]; and software solutions spanning features and functionalities that support the different operations of the smart city. These rely on some conventional methodologies; however, most require the use of smart software approaches powered by artificial intelligence, data analytics, machine learning, predictive analytics, and decision support.

Given all these heterogeneous components and the constant interactions between them, it is very difficult to efficiently and timely keep track of all of them and their activities. Add to this the high levels of physical world and human interaction in these systems and you will be facing a very complex and difficult job. This creates huge security and privacy concerns as components have different levels of capabilities to support effective security mechanisms. Moreover, the control and management of the different tiers and resources in smart city systems is typically distributed among varying entities. This introduces an additional layer of security concerns and limitations. Furthermore, having everything connected through networks (public and/or private), maintaining communication channels adds on the security burden as it exposes all the resources and applications in smart city systems to additional security threats.

B. DATA-DRIVEN SECURITY

Data-driven security is considered an application of data-driven decision making. Data-driven decision making is a process that requires collecting data based on defined measurable goals then discovering facts, patterns, correlations, insights and knowledge from this data. This knowledge is then used to develop or revise processes, activities, systems, policies, and strategies to benefit the owner of the data/system. The owner of the data can be a business, a private or public organization, or a governmental organization. The goals can be to increase business' profits, to improve customers' satisfaction about the products or services offered by an organization, increase students' retention rate in schools and universities, or provide basis for planning, management and enhancements of the systems in general.

Data-driven security aims to utilize collected data (general and specifically security data) to improve the security measures of an application, a system, or a complete organization or environment. The proper use of security data can help reduce possible security incidents, reduce security risks, discover and mitigate security incidents, and improve services' availability and quality. Data-driven security is a combination of concepts that work together to provide security decision making informed by data-collected and analyzed, rather than by intuition or general sense of what the correct way of protecting different systems. The base of data-driven security for a specific system is the data sets collected from that system and its surrounding and operating environment. These data sets may pass through different steps such as data clearing, filtering, anonymization, aggregation, organization, storage, exploration, analysis, discovery, and knowledge building. These will lead to generating security actions to improve the level of the security in the system.

In addition to collecting and appropriately processing and preparing relevant data to achieve Data-Driven Security, different advanced analysis techniques are needed. These include machine learning [51], [52], data mining [53]–[55], data visualization [56], [57], data analytics [58], in addition to many different security optimization techniques [59]–[61]. Moreover, diverse modeling and simulation techniques can be used to enable evaluating new security improvements before they are applied [62], [63]. Visualization techniques can also help improve risk discovery and response decisions, while using statistical models can create capabilities for forecasting and planning for possible future security incidents.

C. IoT SECURITY

The security mechanisms aim to secure communication protocols at various layers; physical, network and application, in a way that data in transit are confidential, reliable and available. The implementation of these security mechanisms is carried out by extending network protocols or developing new ones. At physical layer, the wireless protocol IEEE 802.15.4 defines different security suites to enable encryption that guarantees confidentiality and ensures the authentication of the data frame and the data integrity. The new version of Bluetooth Low-Energy (BLE, ver. 4.2) uses a short-range radio with a minimal amount of power to operate for a longer time. In addition, BLE provides replay protection, achieves message confidentiality by encrypting the payload portion of a frame. The family of Wi-Fi networks use WEP, WPA, or WPA2 protocols to implement authentication and encryption processes. These protocols use a 64- or 128-bit encryption to prevent attacks. For the LTE network, encryption and integrity algorithms were developed and standardized.

The routing protocols Internet Protocol (IPv6) extends IPv4 from 32 to 128 bits per IP address offering scalability for the IoT world. IPv6 also supports more-secure name resolution achieving network layer confidentiality, integrity and authentication through IPsec. IPv6 secures its transmission by employing cryptographically generated addresses (CGAs) to encrypt messages. CGA helps nullify neighbor/solicitation/advertisement spoofing, neighbor unreachability detection failure, DOS attacks, router solicitation, and advertisement and replay attacks. In the application

layer, a lightweight HTTP version Constrained Application Protocol (CoAP) is developed for constrained IoT devices. However, integrating these protocols in the current Internet infrastructure may add complexity in terms of management and protocols interoperability [64]. Additional work regarding IoT security is discussed in [65] where security issues and future challenges were outlined.

Under the Internet Engineering Task Force (IETF) guidance, several working groups (WGs) have been established aiming to standardize new IoT protocols or adapt the TCP/IP protocol stack to overcome the security challenges in IoT. The goal is to develop security mechanisms that prevent cyber-attacks like: Data encryption to protect data; access control and authentication system; secure routing; firewalls; intrusion detection system; anti-malware Solutions; and trust management system to provide trustworthy IoT system.

IV. CHALLENGES OF SECURITY IN SMART CITY SYSTEMS

A smart city is an integrated complex collection of resources from ICT, IoT, software and humans. This exposes smart city systems to many security risks and creates various challenges in addressing them. A smart city system's complexity arises from various directions: technological components used, large-scale distribution, heterogeneity, direct effects on the physical environment, and large user base, to name a few. As a result various challenges arise when considering security measures for smart city systems.

Since a lot of the smart city infrastructure relies on IoT devices, it is important to create an understanding of IoT security issues and how they impact the overall data-driven security in smart cities. Although, some of the IoT security breaches and vulnerabilities are common with the current Internet, IoT presents new security concerns that make it the ''Internet of Vulnerabilities'' [66]. For example, some of the primary IoT attacks are DY (Dolev-Yao) intruder, DoS/DDoS, physical attacks, privacy attacks, eavesdropping, data mining, and traffic analysis. Additionally, new types of attacks related to the constrained things characteristics (low power, low processing, etc.) are IoT specific. Therefore, the security and privacy issues, if not addressed, can hinder the use of IoT and the realization of smart city systems.

Several technology leaders, governments, and researchers are putting serious efforts to developed effective security solutions enabling wide IoT deployment. A secured IoT where data in transit is confidential, reliable, and available, creates a ''trust and interoperable ecosystem'' [65] and enables large-scale IoT deployment. Introducing trust mechanism into IoT can lead to higher and trustful collaboration between nodes, flexibility in dealing with changeable security conditions, reduction in management cost, and consistency across heterogeneous IoT domains. In addition, IoT devices provide the additional capability of collecting enormous amounts of data where they are deployed. This provides a rich dataset that can be used to facilitate data-driven security for smart cities.

Challenges of security for smart city systems stem from the same challenges identified for distributed systems, IoT-based systems and systems of systems in general. However, there are certain characteristics of smart city systems that make them more vulnerable to security attacks and a lot harder to protect.

One of the main characteristics of a smart city system is that it includes hundreds or even thousands of heterogeneous components connected and required to work together and interact with physical infrastructures and humans. Thus it is an important security requirement to protect all components to be able to provide the intended functions with good degrees of reliability, availability, and efficiency. In addition, many of these components come with limited resources and capabilities, thus including security mechanisms to them may be beyond their actual capabilities. Other components may be exposed to threats due to their location like open spaces or dangerous environments. Others may have good capabilities and resources, but many come with different interfaces and control tools. Some of these components are more difficult to protect as they are physically distributed all over the city and may be in difficult or exposed locations, in addition to their limited capabilities and resources. For example sensing devices in an infrastructure may be using different standards while collecting the same types of data, thus requiring additional methods for integration and for securing not only the exchange of, but also the transformation of the data. Another example is a wireless sensor network (WSN), which require specialized security mechanisms to securely integrate with and support smart city systems [19].

Another important characteristic is the various levels and types of integration techniques needed to build a complete smart city system from these heterogeneous components while maintaining adequate security measures. Even when technical components are protected individually, the integration of multiple components in a single system raises many additional security issues [67]. These issues arise due to using networking and communication tools and protocols to connect these components. A well protected component individually will be exposed to additional threats as soon as it is connected to other components. Threats can be because of network exposures or even from compromised or misbehaving components. For example, smart buildings can use a cloud-enabled building management system [68] to provide advanced buildings monitoring and control services to improve operational efficiency. Using a network to integrate these components and connect them to remote cloud services will add to the security concerns for building operations, occupants, and owners.

There is also the need for two-way control flow for most smart city systems. In general a smart city system collects data, processes it, makes decisions, and issues control commands to adjust specific operations in the system. This leads to additional threats as any interruptions, changes, delays or modifications to the control commands could lead to catastrophic results. For example, a delay in issuing a

fire alarm to initiate protection procedures like evacuations, notifying the fire department and initiating firefighting measures could cause catastrophic human and infrastructure losses. The control and feedback systems are usually attractive targets to intruders. Attacks like denial of service, spoofing, malicious data injection, and many others would disturb the smart city systems. Most of these malicious attacks and misbehaviors are discovered based on third party inspection and auditing [7].

Additionally, we have to consider the privacy issues arising from the use of smart city systems. While smart city systems can provide many advanced and efficient services, they can be associated with the risk of compromising residents' personal information. Smart city systems collect, store and use data from every part of the city including its residents. Even with the assumption that the smart city systems are operated correctly and there are no issues of misuse of this data, we still have the risks of security attacks that can expose the data in various ways. Data centers may be hacked, and data is stolen, data can be intercepted in transit, and it is also possible to hack individual components and expose collected data from these components. We have seen such incidents in various systems and the threat is a lot bigger when we are dealing with a smart city system that has data about each and every part of the city and its residents. Unauthorized access to the data is a big risk [69] and we also have to consider the data ownership issues and access control to reduce privacy invasion risks [70].

These unique characteristics of smart city systems pose a lot of challenges when considering their security. There are several challenges facing computer security in general [71]. Some are very similar for smart city systems and others become even more difficult to address due to the characteristics of these applications. The following is an adapted and extended list of security challenges pertaining specifically to smart city systems (may also apply to other large-scale distributed systems and systems of systems with direct connections with the physical world).

1. **Complexity**. Although security requirements can be stated in simple terms such as authentication, confidentiality, nonrepudiation and integrity, actually creating systems that can achieve all of these requirements is very complex. Many mechanisms and approaches are possible to use, and options and capabilities become so heterogeneous when considering a large-scale system like smart city systems. One method may work in some part of the systems, while it conflicts with other operations in other parts of the system. Unifying approaches across the whole systems is practically impossible.
2. **Human Factor**. Unlike many other distributed systems, smart city systems will be accessed, used and affected by people of all types. Human-system interactions are extensive and have two-way effects on the system and its users. Unfortunately, many security risks on any distributed system arise due to human errors [72]. The result? More exposure, higher security risks, and significant impact of security incidents. This effect is amplified by the enormous number of users in smart city systems and the huge discrepancies in their skill levels and security awareness. Therefore, introducing security measures must keep in mind this factor and create security mechanisms that will either minimize the user interactions, which counters the main goal of smart cities; or strengthen the security policies and procedures humans must follow to gain access, which could cause difficulties and inconveniences to many users.
3. **Physical Infrastructure and Resources**. Unlike many distributed systems, a smart city system is directly responsible for some infrastructure, environmental and human related services and activities. Security breaches are not limited to loss of operations or data, but also extend to affect all aspects of city life. A compromised smart traffic light system can wreak havoc on the streets. A faulty or compromised building entrance access control could allow unauthorized persons to enter. Compromised smart meters in an energy grid, could lead to overcharges on customers. Hacked ventilators, insulin pumps, heart monitors, or any other medical devices directly connected to patients could result in the death of patients. The severity of such security risks increases the need for strong security measures in smart city systems and, at the same time, further complicates these measures.
4. **Potential Attacks**. To create effective security measures, it is important to identify all possible types of attacks on these measures. In a smart city system, given their characteristics, there is a huge number of sources of threats and attacks that may happen in so many different ways. In addition, we have the complications described in points 2 and 3 above further increasing possible threats. It is very difficult (practically impossible), costly and time consuming to try and identify all possible attacks. For example, including encryption to protect data in transit. This requires making the encryption/decryption functions available on all devises in the system (think hundreds or thousands). How many possible attacks can happen to this set up?
5. **Placement**. The placement of the security measures plays an important role in how effective they can be. Physical and logical locations are important to evaluate for best results. Considering a smart city system, we have a huge number of possible locations, physical and logical, to consider. Will the security measures be more effective at the device level, on the network components or on the edge or cloud nodes, maybe all of them? It is a great challenge to test and estimate the effectiveness of such placements in a large system like a smart city system.
6. **Secret Information**. Security mechanisms and methods usually entail the use of some secret information. Authentication credentials, encryption keys, security certificates are some examples. Generating, distribut-

ing, monitoring and managing this information has been a challenge in small-scale distributed systems. As we grow the system size and user base to city wide coverage, handling secret information becomes an even bigger challenge. Smart city systems are highly distributed, managed by various entities, used by a huge user base and operated by a large number of personnel. Many questions arise like: how will the secret information be generated; who is responsible for generating and managing them; how different entities can agree on or exchange secret information; who is responsible for enforcing the secrecy and validation of this information?

7. **Monitoring**. Securing systems also requires constant monitoring of its operations, security measures effectiveness, unauthorized access, malicious attempts to access its components, unintentional errors leading to security risks, and proper application of all measures and policies by all system operators and users. The larger the system the bigger and more complex this task becomes. In addition, the growth in complexity is not always linear. Various aspects come into account such as ownership, responsibilities, compliance among other things and these can further complicate the monitoring processes by introducing more dimensions in the measurement and monitoring metrics.
8. **Disposal**. Data and system components that are not needed in a system pose an additional security threat. Keeping them means adding extra resources to store and protect them, while disposing of them requires very strict measures to ensure that they will not be exploited and become a security threat. The issue does not pertain only to the software components, but also to physical components. An example illustrating this issue is an article in INSIDEEVs [73] that discusses finding several old Tesla computers with personal data still on them and easily retrievable even with physical damage to the units. This issue grows in size and complexity when you consider the volume and variety of smart devises, sensors, and control components that have some private data on them. There are bound to be more similar problems coming up and these must be addressed.
9. **Patching Holes**. Taking point 4 in consideration, how much effort is needed to find every possible weak spot in the system and patch it to reduce attacks? How many of these weak spots will an attacker need to compromise the whole system? Imagine an army under siege barricading in a castle surrounded by high and strong walls for protection. The defense from inside relies heavily on finding and patching each and every possible weakness in the castle walls. It also requires securing all entrances and ensuring that only authorized people can go through. Furthermore, everything and everyone leaving the castle must be well protected as it leaves the castle and authenticated and verified when they come back. The enemy on the other hand, only needs to find one hole or weakness to penetrate the castle. This could be a tiny crack in a wall, a faulty lock, sloppy guards, or even internal collaborators. If you assume the smart city systems and their users are the army and you have a base security coverage (the wall) in place. How vast and complex fortifying and securing these walls can be? The walls are a lot bigger, there are significantly more exit and entry points to secure, there is huge movements of data, people and components, and there are thousands or millions of people involved. In addition, every part or function of this city may be owned and governed by different authorities and in many different ways. On the other hand, how many possible holes in these walls an attacker can exploit?
10. **Discrete Approaches**. Most approaches proposed for securing smart city systems offer individual solutions addressing very specific aspects or functions of smart cities. For example using data-driven methods to manage dynamic public transport systems, security of control systems, and data storage security. Except for a few articles discussing smart city systems security approaches and challenges, no one addresses the use of generic data-driven methods and tools to provide effective and efficient security measures for smart city systems as a holistic approach. As a result, many proposed approaches may interfere with each other or cause other security issues. Providing a holistic approach to address security issues in smart cities will help improve the effectiveness of the approaches and the management aspects as well.
11. **Tunnel Vision**. Security mechanisms are complex, and it is difficult to understand the full impact and need for each security measure until we gain a good understanding of the various aspects of all threats. In a smart city system, this issue becomes even more difficult due to the large number of components, which in turn increases threats and security risks. In addition, the introduction of discrete security measures amplifies this problem as each one brings the focus to a very specific view of security risks. Any security measure to be deployed must be considered in light of all possible interactions with and around it to ensure effective results. It is always possible to introduce a very strong security measure in one part of the system that could easily compromise the effectiveness or security of other components.
12. **Management**. Management of security policies, technologies and data has been and will continue to be a tough task for any system with any level of exposure to the outside world. Security management include tasks like risk assessment and mitigation, creating and enforcing effective policies, training users and ensuring their compliance, ensuring compliance with regulatory and standards requirements, and planning for the future. Generally many organizations were able to successfully incorporate security management tasks and controls

into their regular operations. Unfortunately, when we discuss smart city systems, we also have to start from an important fact, the system components are numerous, highly heterogeneous and interconnected. Smart components and different software applications in a smart city system will need to communicate, exchange information, and collaborate to achieve their goals. Managing a system like this becomes very complex and could easily have various incompatible policies, controls and management. All the challenges described before this one further adds to the complexity of the management process.

13. **Ownership and Control**. Another area affecting the management and effectiveness of security measures is ownership. When data and system components all belong to a single (or a group of federated) organizations, identifying ownership of the data and incorporating the proper privileges for access and control can be included to enhance the security of the data. It is not a simple task, but it is generally doable. However, a smart city system spans multiple entities who may own parts of the resources and data. These entities, most likely will have different rules defining ownership and control of data. They would also have varying levels of acceptance to share data and resources with other systems. Consolidating these rules and preferences to allow smooth access across the whole system is a great challenge.
14. **Costs versus Benefit**. Incorporating effective security measures in a software system adds to the cost of developing and operating it. Consider all the challenges we already discussed. How much will it cost to try to counter each and every one of them? This has led some organizations to not incorporate them until they face a real (and costly) security problem. This cost adds up very quickly when we consider smart city systems. Now we have to secure devices, software, networks, and end user access points, among many other aspects. Covering all possible risks and creating mitigation plans quickly adds up and projects become more costly. In addition, development time increases and usually delays delivery and operations of the systems. This increase in cost and time to market without showing any immediate benefits leads to leaving out some security measures to avoid the costs and speed up production.
15. **Performance**. Introducing security measures to a large system, means there will be needs for more energy, processing power, storage, and communication bandwidth. These requirements can negatively impact the overall performance of the system. Encryption and decryption for example add significant processing requirements of the data, thus leading to either having to spend more on additional resources or suffer a noticeable impact on the system's performance. Adding extra checks and firewalls on networks will lead to higher communication delays. With a system as large as a smart city system and attempts to address all security challenges and requirements, these performance issues become more noticeable and have bigger negative impact on overall system performance. As a result, some organizations may opt to minimize security measures for the benefit of better performance. In addition, a smart city system incorporates a large number of entities each with a different view of the balance between security and performance. Thus leading to possible issues when the system components belonging to different entities are integrated and are collaborating to achieve common tasks.

In light of these challenges, it is clear that the security of smart city systems is not only extremely difficult, but also time consuming and may affect the overall performance and usability of the systems. Yet we also have an advantage inherit to smart city systems that can support security efforts and help make them more effective and efficient, DATA. A lot of data is collected, generated and restructured through the different applications in a smart city system. A lot of this data is a by-product of other operations; however, there is a wealth of information among this data that can be extremely helpful for the security requirements. For example, a video streaming service may be monitoring and collecting IP addresses of the originating requests to improve the availability of the service close to the customers' locations. This same data can be used to analyze security aspects such as the attempts to thwart DDoS (Distributed Denial of Service) attacks or control users' access based on their location. Another example is the data collected by a smart building on the occupancy levels and locations of occupants for energy regulation purposes. This data can also be used to optimize access controls for applications within the building. Thus we are proposing the use of data-driven security approaches for smart city systems. We can use available data and introduce more data collection components to be the bases for these approaches.

Several of the challenges described can be addressed using data-driven security techniques. The following sections will further relate these challenges to the different types of security applications for smart city systems and how these can be alleviated using data-driven techniques.

V. DATA-DRIVEN SECURITY FOR SMART CITY SYSTEMS

Security of distributed systems can be approached using many different methods and technologies. In a smart city, a large number of distributed systems are integrated to form the ultimate mega-scale distributed system. This system and its sub-systems continuously interact and collectively record and generate huge amounts of data. This data can be the basis for adding more sophisticated security measures. Data-driven security can provide many advantages for securing smart city systems. The distributed components of smart city systems reside in various locations and communicate through various types of networks (private and/or public, dedicated and/or shared). The networks can also be wired, wireless, or a combination of both with various options in the types of wireless connectivity as well.

In all cases, all components, the networks and the applications, need to be protected from security attacks to ensure the reliability and availability of the smart city systems and protect the privacy and the wellness of its residents. To secure smart city systems, different security components must be added to the software and hardware infrastructure. These security components can be firewalls, intrusion detection systems (IDS), intrusion prevention systems (IPS), vulnerability management systems, hardware-based encryption/decryption components, and sophisticated authentication and authorization mechanisms. All of these security components along with the different smart city applications, devices, and networks will capture, process and generate large amounts of security-related data.

The security data can be utilized to apply data-driven security for protecting smart city systems. The collected data can be stored, analyzed to assist in understanding security risks, identifying better techniques for security and making smarter and more informed decisions to achieve more effective security solutions for the system. As more data is collected, there are more opportunities to reach more accurate and reliable deductions and improve overall reliability and availability. In addition, the data is also collected by some systems that are similar in their operational conditions or may have the same types of security risks. This collective data can facilitate better analytics and learning techniques using the larger set of data. This will lead to many advantages for smart city systems. Some of these advantages and benefits are the following.

A. SECURITY RISK ANALYSIS ENHANCEMENTS

The availability of security data from components of different smart city systems is important to reach fast conclusions. Having data from one smart application regarding the security status, possible breaches, and occurring incidents is good. However, if we can collect similar sets of data simultaneously from many smart applications, we will have a much larger data set to work with. This will lead to faster recognition of patterns that affect security risks and identification of conditions leading to a security incident. In addition, this will also lead to corrective actions not only for the breached components, but also for all other similar components to protect them from similar incidents. For example, if we collect detailed information about many DDoS and low-rate DDoS (LDDoS) [74] attacks that occurred in different smart city systems, it will provide a larger sample for analytics and could lead to identifying common preconditions that made these attacks possible and also find more accurate information about the sources and mechanisms used in the attacks. As a result, the response occurs faster and delivers more accurate and adaptive countermeasures.

In addition, this collection of data from multiple sources will also allow for comparisons with other applications and sub-systems in the smart city. For example if a DDoS attacks occur on several similar applications around the city, but not on some others, the data collected from both types can be analyzed closely to compare set up, configurations, security measures in use, and any other factors that may have created those differences. The data can be informative in terms of the factors that allowed, stopped or limited the attacks' effects. This can lead to improvements in all similar applications by applying the most effective security configurations and policies used.

B. SECURITY PROTECTION IMPROVEMENTS

As a larger security dataset is collected and analyzed, it is possible to utilize this dataset to improve the defense mechanisms to better protect smart city systems. This data can be also utilized to build better situation-based security systems for smart cities. Careful analysis of multiple smart applications security information can lead to better understanding of how these attacks work and what are the best ways to protect the applications from them. In addition, this provides a learning pattern for the security algorithms to identify and adapt to changes in possible security attacks.

Using data from multiple sources can also be used to fuel aggressive forecasting algorithms (sort of like weather forecasting). This type of analytics does not have to be in real-time or highly responsive. Therefore, with more data available over extended periods of time, the forecasts will become more accurate and could help prepare for possible risks or improve current measures to focus on the most eminent of these risks.

Developing a knowledge management methodology to collect and organize security data is important for forensics purposes [75]. However, it is also a very important component for predicting and protecting from future security attacks. Moreover, predictions of possible attacks can be further enhanced using advanced analysis, data mining and machine learning techniques. Various approaches can be used based on data [76] to articulate possible attacks and thus mitigate them or at the very least be better prepared for them ahead of time.

C. SECURITY MODELING, SIMULATION, AND VISUALIZATION CAPABILITIES

The collected security data can be utilized to create accurate virtual models representing different smart city applications/environments. These models can then be used to simulate different situations and security issues to identify problem areas [77]. For example, simulating a DDoS attack on the smart building management systems in a smart city and evaluating the impact on the buildings controls and residents. In addition, these models and simulations can help identify weak spots, provide insight on possible mitigation and protection activities, and create a safe environment to test these scenarios. The results can be used to study possible improvements in protecting smart city systems from security attacks.

Simulation models also help create different levels of abstractions that allow developers and managers focus on specific areas of the system in details or get a high-level view of a larger portion of the system with underlying details

hidden. This navigation capabilities into the different levels of details help to create better understanding of the whole system and at the same time appreciate the intricate details of each part. Moreover, it can help expose risks that may have not been visible through normal testing procedures. Another use for simulation models is to test additions or changes to the system or application. For example, an accurate simulation of the system can be compared to one with the modifications to see how much of the system is being affected by this new addition or modification. Even the introduction of hardware components or other resources can be simulated before implementing the changes.

Simulation models can also be used to experiment with new types of attacks, security protocols, and protection approaches. In addition, the models can be used to educate and train users. For example, creating a situation where a user can cause a security issue and showing the user what happens if that is actually done in production system. More drastic scenarios can be simulated to show accurate account of damages and losses if certain security attacks or breaches occur.

Visualizing security data is another important advantage of using data-driven security for smart cities. Visualization can be used to gain deeper understanding of risks and threats of different systems in smart cities [57], [78]. While it is difficult to monitor and understand security events using statistical analysis, it is easier to achieve these objectives within short times in some cases using visualization techniques for smart cities' security data. In addition, it is possible to uncover hidden patterns, correlations, and insights on security data and events. Visualization can be used to understand historical security events, current security events, or to understand and evaluate the outcomes and impact of future security improvements on smart cities systems.

D. SECURITY MANAGEMENT IMPROVEMENTS

Another benefit from data-driven security is the enhancements possible in smart city systems security management. Different data-driven techniques can be used to improve security management processes in smart cities. Collected data, analytics, and simulations can provide accurate insights into how well (or bad) management processes and use/access policies are working. They provide a more detailed view of all policies and their impact and allow for creating better policies and improving current ones. This can apply to various management activities such as auditing, upgrading and optimizing processes, handling emergency and recovery procedures, and creating more accurate authentication and authorization policies and access controls.

Security management include managing access controls, assets, incidents and business continuity. Using the available data, each policy can be accurately monitored and assessed in terms of its effectiveness and efficiency. Collected data can help study current security policies and procedures and identify weaknesses to improve them. It can also help monitor the resources being used to apply security measures and help find ways to optimize this usage. In addition, analyzing the security risks of an update in software and hardware components can be performed using available data and intelligent algorithms in addition to the simulation models.

These improvements can be applied to various policies and procedures. User authentication policies, for example, can be assessed in terms of the number of breaches happening and the reasons for these breaches. Multiple sources will increase the accuracy of this data and help improve the policies such as adding multi-factor authentications for all users or limiting access to specific IP addresses or through a VPN (Virtual Private Network). Incident management procedures include activates like identifying possible incidents, planning mitigation or protection plans, and creating more secure policies. Using data-driven techniques can make minimizing incidents risks more efficient, highly adaptive, and possibly faster. Furthermore, achieving more efficient policies and procedures, using simulation models and forecasting techniques, and continuous monitoring of the system will lead to better capabilities to make the application or system more reliable and available to ensure business continuity.

E. SECURITY MANAGEMENT AUTOMATION

Several security policies and procedures require human intervention, which can slow down the reactions to security incidents and hinder the efforts to contain the incident or its damages. Many of these policies and procedures can be automated, if the right data is made available and efficient algorithms to make decisions based on this data are utilized. Forecasting possible risks in the near future can add another dimension to the decision-making process and create automated responses capable of handling many of the security risks without (or with minimal) human intervention. The result is faster detection, mitigation and adjustment in the applications and more effective and faster actions. The key to achieving more automation is continuous monitoring and analysis of activities relevant to security in the system, logins, abnormal access patterns, highly increased volume of requests are some examples that can be detected and acted upon automatically, while notifying system managers. Updates and reconfigurations can also be done via automatic configuration tools and scripts instead of manually applying them. The testing can also be automated to first take advantage of simulation models then running a large set of test cases in production and analyzing the results.

Many techniques for incident preparedness and response can be configured in different ways for effective results. These configurations can be optimized to efficiently and accurately protect smart city systems based on different criteria including the networks and systems capabilities, the protection techniques used, smart city applications types, and the heterogeneity, dynamicity, distribution of smart city components. Such optimizations usually require huge effort and time from skilled security professionals to identify different configuration issues, possible configuration solutions, system adjustments, operations and security procedures evaluation,

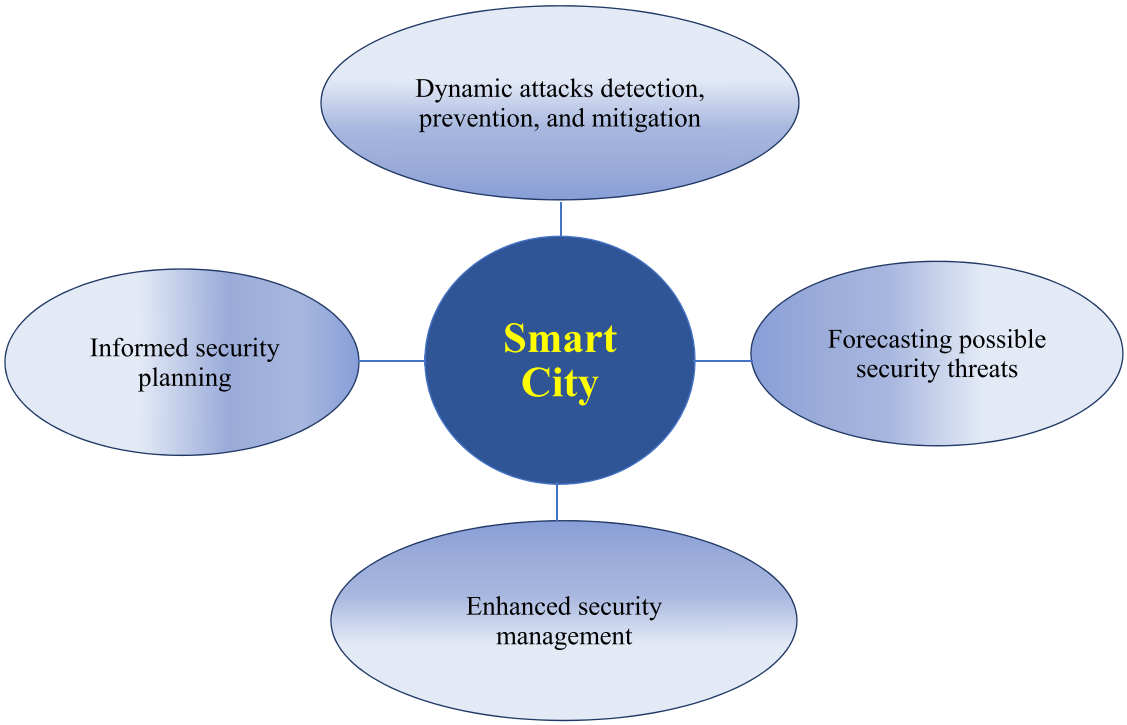


FIGURE 1. Applications of data-driven security for smart city systems.

and application of new configurations. With data-driven security, it is possible to automate several activities in this process. It is possible to build a closed loop control for automating the security improvements. It is possible to use collected security data to identify issues in current security configurations, create accurate virtual models representing different smart city security components and use them to generate and evaluate optimized configurations and procedures, then applying them in the system. One main advantage of this approach is that the cycle will be automatically repeated to reach a stable and optimized security conditions.

Collected data from the continuous update cycle can also be used to create incident response plans and procedures. Every time an incident occurs, the preconditions, activities and impact of the incident is recorded. This information can help feed the control loop with additional information for the next set of adjustments. The security configurations can also be augmented with tools and techniques to detect and respond to security incidents. Many of the response activities can be automated and measured in the same closed loop. Therefore, the security configurations can be automatically adjusted whenever a new incident occurs in the smart city system. With this advantage, it is possible to have more adaptive security mechanisms for the dynamic smart city systems.

VI. APPLICATIONS AND ISSUES OF DATA-DRIVEN SECURITY FOR SMART CITY SYSTEMS

There are many types of applications for data-driven security approaches for smart city systems. However, there are also several issues that need to be addressed for optimal utilization. We will discuss the applications and issues then introduce ideas for possible solutions for these issues. Utilizing data-driven security approaches offers many benefits for smart city systems security. However, the issues associated with these approaches are creating different obstacles impeding the advances in this direction. These issues are not the same for all applications of data-driven security approaches. However, they change based on the different application categories of data-driven security as shown in Figure 1. In addition, Table 1 provides a summary of these issues, identifies the different types of applications relevant to these issues and attempts to provide possible solutions approaches for the different applications categories.

A. APPLICATIONS FOR DYNAMIC ATTACKS DETECTION, PREVENTION, AND MITIGATION (DYNAMIC DETECTION)

Interactive and real-time smart city applications require immediate and adaptive responses to security incidents. These applications need to handle different security situations in real-time. One example is situation-based intrusion detection and prevention systems (IDPSs). These systems must be continuously monitoring the conditions and respond immediately when a threat is detected. The issues of dynamic data-driven security for detection, prevention, and mitigation are mainly due to the distributed nature of smart city systems and the requirements of real-time actions and feedback. Security methods in this type of applications also need interactive and real-time data collection, effective handling of

TABLE 1. Data-driven security applications categories for smart city systems with their issues and possible solution approaches.

Data-Driven Security Applications Category	Issues	Possible Solutions
1. Dynamic Detection	• Interactive data collection • Distributed data • Real-time operations and controls • Need for fast threats recognitions • Need for fast mitigation actions	• Policy-based management • Multi-agent decision making • Stream-based data analysis and classification • Parallel processing
2. Forecasting	• Real-time continuous data collection • Access to correct parts of current and past data • Data filtering and organization • High computational and networking demand on resources	• Utilizing efficient streaming services • Using predesigned and efficient data collection, organization and aggregation services • Integrating fog and cloud computing services to support the required functionalities effectively
3. Management	• Creating a comprehensive knowledgebase • Managing data and policies • Decisions-making • Solution optimizations • Solution evaluations	• data mining and machine leaning • Modeling attackers, attacks, users, network risks, and defensive strategies • Applying operations research • Applying simulations • Utilizing high performance computing resources
4. Planning	• Big data storage and processing issues • Data trustworthiness • Data sharing • Knowledge building and processing • Planning optimizations • Decisions evaluations	• Big data management • Big data mining and machine learning • Applying blockchain technology for forming credit-based systems for data sharing • Applying operations research • Applying modeling and simulations methods

distributed data, fast attacks detection and mitigation actions. In addition, it is required to have an interactive and effective dashboard that can help view the security status of the applications. Analyzing huge amounts of data in real-time and producing usable results requires a lot of resources and may not always be accurate or effective.

One possible solution is to use a policy-based or rule-based management approach for applying different security policies as needed. This approach can provide fast and efficient security mechanisms for protecting smart city systems and reducing the bourdon on the detection components. Policy-based management also creates a more manageable set of data to be monitored and analyzed to detect and react to threats. A different approach is to use real-time stream analysis services to quickly classify data and identify risks or threats. Other possible approaches involve using multi-agent decision making and parallel processing to make security decisions faster, more efficient, and more accurate.

B. APPLICATIONS FOR FORECASTING POSSIBLE SECURITY THREATS (FORECASTING)

Forecasting applications rely heavily on the availability of a rich repository of past and current security data. In addition, they use artificial intelligence and machine learning to find variabilities in the smart city system and predict attacks or threats. The forecast results can be used to enhance current monitoring and protection procedures and policies; create and deploy suitable counter measures; identify and patch vulnerable areas; and plan to minimize possible damages if such attack occurs. Here too we find several issues in analyzing past and current aspects and providing effective and accurate forecasts. One of the main difficulties is identifying the data needed for analysis and producing the required results. Data is collected and generated at very high rates and in huge volumes. Yet, not all of it is relevant, or useful for the specific forecasting needs. Therefore, it is important to recognize the important types of data, apply data filtering and organization techniques first. The resulting data sets will still be large and may also be dynamically changing over time. Therefore, the analysis and forecasting techniques require compute intensive algorithms that need high performance computing infrastructures. As a result, it is not always feasible to apply these techniques unless the return is equally beneficial. Furthermore, forecasting techniques need to be detailed and specific to certain applications, thus requiring redesign and customization for different applications and data sets.

Some possible activities that may help improve the situation are creating generic data organization and aggregation solutions that will help put the needed data in a more effective format for the analysis. Another possible approach is to design and deploy a knowledgebase system with facilities to apply ETL (Extract, Transform, Load) techniques to make the data useable by different forecasting techniques in different applications. Along with that, integrated cloud and fog components and services can support the different parts of the process in more efficient ways. For example, aggregation and organization can be done on incoming data streams at edge locations using fog nodes. Resulting data can then be streamed to the cloud for proper storage, organization and analysis.

C. APPLICATIONS FOR ENHANCED SECURITY MANAGEMENT (MANAGEMENT)

Effective security measures in smart cities cannot be achieved without proper management of the different security related systems. Here there are several opportunities to utilize security data to enhance management and system configuration processes. In addition, this data can help improve risk analysis and management and access control for smart city systems. The common factor in this type of applications is the need for intensive data analytics to arrive at the most useful solutions. Therefore, accurate responses are needed, yet there is usually more time available to get these responses. Many management activities are done in the background and can be implemented as needed and over different periods of time. For example, applications analyzing collected data to identify week spots and offer solutions can take longer as the data sets grow, and the accuracy requirements increase. The issues are mainly related to the complexity of building effective knowledge, decision-making, optimization, and evaluation techniques for the applications. However, security approaches here are generally not real-time and can operate independently from the operational components of the system.

Possible solutions for these issues include using data mining, machine leaning, operation research, and simulation techniques coupled with the integration with powerful and high-performance computing infrastructures like the cloud. This will provide the necessary methods and resources for intensive analysis of security data and find new observations and trend. Using these we can optimize and evaluate new security management and configuration solutions. Furthermore, it is important to be able to model attackers, attacks, users, network risks, and defensive strategies for effective security applications in this category [23].

D. APPLICATIONS FOR SECURITY PLANNING (PLANNING)

Security planning include utilizing security data for future security planning in smart cities including proposing new security protection methods and strategies, new security systems, new security architectures, and new regulations and policies that can provide long-term security improvements for smart city systems. These activities usually take a long time and require extensive analysis of the security data collected over long periods of time. In addition, modeling and simulation techniques become very important in this type of security planning as new tools, scenarios and cases can be modeled and simulated to understand how they work and what their effects will be. Furthermore, many optimization techniques can be introduced here to further improve future plans for securing smart city systems. The main issues in this category are mainly related to dealing with big data, the complexity of data trustworthiness, data sharing among different organizations within a smart city, knowledge building and processing, planning optimizations, and new solutions/decisions evaluations.

One way to address these issues is using efficient big data management, mining, and learning tools and services for the applications. Another solution is using blockchain to apply credit-based systems to enable data sharing and trust management [79], [80]. In addition, operations research and simulation techniques can be very helpful to optimize and evaluate any solutions for security planning in smart cities. For this type of analysis high performance computing infrastructure is necessary. Integrating different analysis models and scenario-based evaluations becomes helpful. In addition, user interfaces are very important to allow humans to interact with and use the generated results for planning and management.

VII. THE BIG PICTURE AND POSSIBLE IMPACT

The discussion of the benefits and application categories for data-driven security for smart city systems leads to the need for an understanding of how they can help alleviate some of the challenges identified earlier. Different data-driven security application categories can address a subset of the challenges or in some cases set the stage to create other security applications. For example, creating a security application to monitor incoming requests to a system will help detect possible DDoS attack, but will also provide a dataset that can be used to create forecasting or automation techniques. Table 2 provides an overview of how these challenges may be addressed. Each row identifies a challenge, the data-driven security application category or categories that can help alleviate it, and a brief explanation of the techniques or approaches possible.

Data-driven security applications for smart city systems developed within each of the four categories summarized earlier in Table 1 can take advantage of available resources and create more innovative methods relying on the wealth of data available in smart city systems. Applications within the same category can also collaborate effectively across multiple systems in a smart city to achieve better and faster results. Furthermore, data-driven security applications across categories can be integrated to collaborate and provide more effective security solutions. Information, knowledge, and proposed actions can be aggregated among these application categories as shown in Figure 2 to improve and optimize security mechanisms, while reusing already available data and resources.

One example is facilitating the collaborations between applications of data-driven security for interactive attacks detection, prevention, and mitigation and applications of data-driven security for better security configuration and management. The first (level 2) offers continuous and fine-grain data collection creating a detailed view of past and current activities and events (level 3). This data can be the basis for the knowledgebase needed for configuration and management applications (level 4). Aggregated and filtered data can be shared with the configuration and management applications, which use their techniques to effectively use this data and enhance security processes. New configurations,

TABLE 2. An overview of how the challenges in security for smart city applications can be alleviated by applying different categories of data-driven security applications.

	Challenge	Application Category	Possible Use/Benefit
1.	Complexity	• Dynamic Detection • Forecasting • Management • Planning	The organization of security features and capabilities across the different categories will help create a more systematic approach to solutions at different level. Thus separating the concerns and reducing the overall complexity.
2.	Human Factor	• Dynamic Detection	Human computer interactions are the forefront of any software system. Focusing on dynamic monitoring and controls of these interactions in real time will help mitigate many possible errors and risks associated with the users. Collected data could be used to inform the other three categories.
3.	Physical Infrastructure and Resources	• Dynamic Detection	Similar to the human interactions, most interactions with the physical parts of the smart city system is at the peripherals and in real-time. Applying security features to monitor and control these interactions must also be in real time. Collected data could be used to inform the other three categories.
4.	Potential Attacks	• Dynamic Detection • Forecasting	Security attacks must be detected and handled in real time; thus, we need security application Shandling these dynamically. However, being able to forecast possible attacks can help improve the detection capabilities.
5.	Placement	• Dynamic Detection • Management • Planning	Effectively all types of applications can benefit in terms identifying the placement of different security applications at the most effective locations. Forecasting, however, is the least affected here.
6.	Secret Information	• Management	Creating, securely distributing, monitoring, and authenticating secret information is mainly handled by management applications. Users, components and other systems are merely using this information. This separation allows for focused efforts, better integration and enhanced controls among the different entities involved.
7.	Monitoring	• Dynamic Detection	Security threats may occur anywhere and anytime and in many forms; therefore, real-time security monitoring applications are always needed. Collected data could be used to inform the other three categories.
8.	Disposal	• Management	Like secret information, unused sensitive data need to be disposed of securely. The security management applications should be responsible to handle this requirement, thus relieving other applications from the burden.
9.	Patching Holes	• Management • Planning	Finding and patching security weaknesses and problems is general a management task. It can also be part of the planning applications where new resources and capabilities may be added for better functionality. Dynamic detection will feed this step with continuous real time data to help find such holes and furcating results can improve the planning for improved protections.
10.	Discrete Approaches	• Dynamic Detection • Forecasting • Management • Planning	Applying independent security applications in different parts of a smart city system creates conflicts, integration problems, and possibly undermine other applications of features. Organizing these applications and creating clear interaction models and interfaces across all applications will help create a more holistic approach to security. Data collected from the real-time security applications, will inform the other applications seamlessly to achieve overall security improvements.
11.	Tunnel Vision	• Dynamic Detection • Forecasting • Management • Planning	This can be reduced by having a complete picture of the smart city system and its security requirements. A holistic approach will help identify and manage the effects of certain security features on others and on the system's overall performance. Data collected and analyzed in all categories will inform the developers and create a better understanding of these links.
12.	Management	• Dynamic Detection • Forecasting • Management • Planning	Managing the security of smart city systems is as complex as creating it. Management efforts can be simplified and streamlined through automation. Automation is based on collected data (mainly in Dynamic Detection category), analysis and decision-making capabilities that are informed by the other categories of security applications. Visualization is also a helpful tool to manage security applications. Similarly, this requires a lot of current and past data.
13.	Ownership and Control	• Management • Planning	Challenges posed by the smart city application collaboration across multiple systems owned and operated by different entities require suitable management tools and planning for better collaboration. These two categories provide the necessary functionality using data collected in Dynamic Detection applications.
14.	Costs versus Benefit	• None	This is an issue beyond the technical realm of security applications. However, security managers can use the data collected and analyzed by available security applications to create cost-benefit analysis and risk exposure analysis to justify the need for more security applications.
15.	Performance	• Dynamic Detection	Intuitively, we think security measures negatively impact the performance of the system. However, closer examination of what each security applications category does, there are many opportunities to enhance the overall performance of security applications for real-time features by using the analysis from the other application categories. Utilizing more resources in the background for resource heavy functions like optimization and forecasting processes will help improve the performance of the real time applications in different ways.

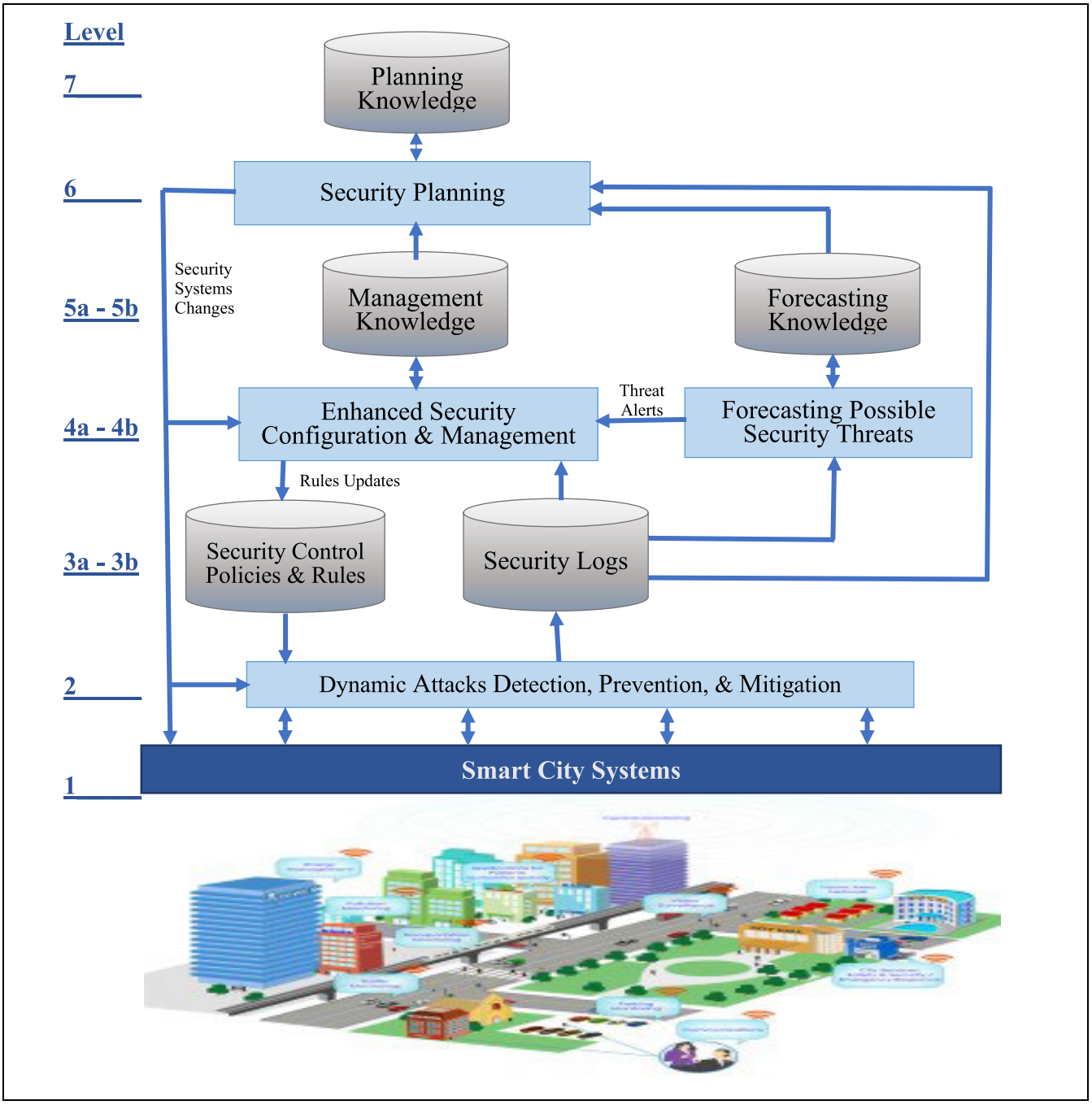


FIGURE 2. Information and action flow among different data-driven security applications categories.

policies, procedures will be generated based on real data and delivered to level 2 through 3a. In addition, the results of these updates will be used to adjust and enhance real-time security measures for all applications, which in turn generate more data for the upper level.

Many of these activities can also be achieved automatically through this collaboration, while some may still require human intervention. For example software changes and policy updates can be fully automated, but replacing devices, adding new physical connections or components to the system cannot. Automating security management is accomplished by including effective closed loop controls in a cycle to apply, evaluate and enhance operations and configurations. A major issue is the required collaboration and data exchange across many applications and systems across the smart city. Logistics, privacy, ownership and regulations can affect the efficiency and effectiveness of data sharing, which is the main requirement for these applications.

TABLE 3. A summary of different data-driven applications and their relationships.

Application Category Factors	Dynamic Detection	Management	Forecasting	Planning
Application Main Purpose	Continuous monitoring running systems and collecting security data for immediate response to security incidents.	Utilizing security data to effectively configure and manage security policies, procedures and mechanisms better system protection.	Finding variabilities, trend and changes in security data to predict future security threats	Utilizing security data for planning future security measures like proposing new protection methods, strategies, policies, architectures, and regulations.
Application Inputs	Live security data and security control policies and rules.	Big data collected over time like logs from this and other applications' and data related to possible threats in addition to previous management knowledge.	Big data collected over time, security and usage logs, security incidents history and previous forecasting knowledge.	Big data collected over time from applications' security logs, collected management knowledge, and data regarding past and forecasted threats.
Application Outputs	Security control messages for different smart city components and security application logs.	Security control rules and possibly aggregated data for security planning applications.	Reports on possible threats and their characteristics for management and build forecasting knowledge.	Plans for: new security resources; their capacities/locations; changes in current resources, and changes or additions of policies, procedures, and regulations.
Processing Requirements	Low	High	Very High	Very High
Storage Requirements	Small	Large	Very Large	Huge
Response Time	Real-time	Non-real-time online or possibly offline	Offline or possibly non-real-time online.	Offline
Evaluation Capability	Not required	Required	Required	Required
Forecasting Capability	Limited	Low- to mid-level	Mid-level to intensive	Long processing times and intensive
Knowledge-base	Not required	Required	Required	Highly Required
Learning Capability	Not available	High Decision making	Mid-level Statistical and predictive analysis	Very High Knowledge extraction
Users	Operational staff	Middle management	Middle management	Middle management Executives
Automation	Highly possible except for some physical aspects	Hybrid depending on methods and algorithms used	Mostly automated	Initially automated, but final outcomes depend on human analysis and input

Another collaboration opportunity can be created between the applications of data-driven security for forecasting security events (level 4b). Forecasting requires access to a detailed data covering long time periods (from 3b) that is being populated from level 2. The forecasting process itself is complex and time consuming; however, the results it generates are very important and informative. The knowledge generated (in 5b) will inform the planning applications (level 6) to assess plans and adjust for foreseeable needs. In addition, forecasting applications will pass some direct information to the management applications (level 4a) to help make short term decisions based on forecasted information. Table 3 provides a summary and comparison of the four application categories of data-driven security for smart city systems. The table shows the characteristics of and the relationships among these applications' categories.

From another perspective, we can identify the main components in a smart city as its residents, resources, services, and economy. These components must be secured for cities to effectively and safely operate as smart cities. It is important to secure these components directly or indirectly through several security functions that can be enabled using the data-driven approach. These security functions are usually handled by security engineers who work to implement and apply security functions to maintain the smart city security goals. Security engineers work to manage security risks, ensure safety, maintain the availability of smart city systems, improve security operations, reduce operational [and residents] inconveniences, optimize security resource utilization, and improve security future planning. These security functions are enabled with the availability of advanced security tools such as security tools for forecasting, decision support,

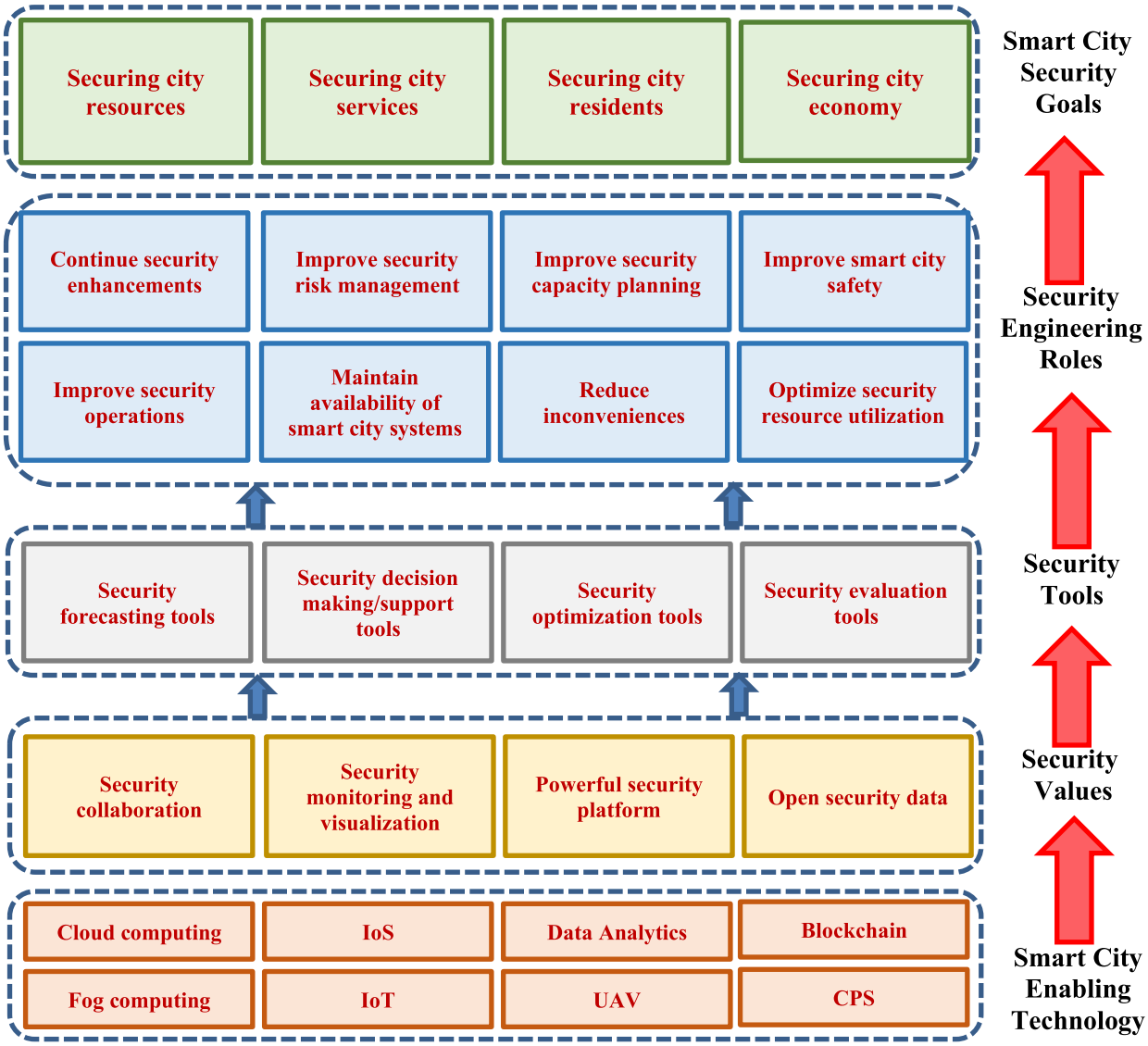


FIGURE 3. An impact architecture of data-driven security on smart cities.

optimization, and evaluation. Creating these tools based on data-driven approaches will highly benefit the security applications and smart city systems. Data is already being collected and used for several other purposes in smart city systems. Adding new features for security using this data and additional specifically provisioned data for security is a natural transition. Data-driven security tools can be deployed and operated using smart city infrastructure and available advanced technologies such as cloud and fog computing, IoT, IoS, CPS, blockchain, and data analytics. The infrastructure is mostly in place and already capable of supporting different levels of resources and processing requirements. Therefore, it will be possible to apply intelligent algorithms for optimizations, decision support and automation in general. Figure 3 provides an impact architecture of data-driven security in smart city systems.

VIII. FUTURE DIRECTIONS

As smart city systems continue to grow and become more important, security techniques for smart city systems will need to be developed to match the growth. Current security solutions and approaches provide different level of protection. However, the complexity and sheer size of smart city systems in addition to their direct impact on humans and physical infrastructures dictate the need for a lot more to be done. Data-driven security approaches offer some good possibilities and open up more research directions.

1. **Data Science**. An important research area that needs to be studied and linked to data-driven security. This field has made great strides and research has promising results in terms of optimized methods and algorithms to better use and manage data. Many of these methods can be applied as is or adapted for security purposes.

Some examples of data science techniques are context representation, reasoning, using graph grammars, using stream-based classification [4], and rules evaluation. An example is adapting techniques in data science for intrusion detection and prevention systems (IDPSs) and in dealing with challenging situations such as advanced persistent threats (APTs) [81] and slow DoS attacks [82].

2. **Big Data Analytics**. Another area providing great insights on the best practices and methods to handle big data effectively. Data-driven security planning relies heavily on big data and we can benefit a lot from this field. Various infrastructures have been designed along with the needed tools and algorithms to handle big data. For example, the Hadoop ecosystem [83], [84], non-relational (NoSQL) databases [85], in-memory databases [86], blockchain [18] and data lakes [87]. All of these offer some optimized infrastructures and operational methods that can be utilized to enhance the performance of data-driven security techniques. In addition, many of the approaches used can be further refined to specifically fit the data-driven security needs.
3. **Modeling, simulation and visualization**. Many techniques are currently available in general applications. These can be researched and adapted for security data and security analysis purposes. These techniques can be designed for creating and validating accurate system security models, simulating different aspects of the security functionalities, and visualizing the data in ways that facilitate better understanding and utilization of it. An example is using risk-based approaches to security metrics [4] by using specific simulations of the working parts of the security system.
4. **Data trustworthiness**. Data sharing's most important requirement is the trustworthiness. When various entities need to share data across systems and applications, it is important that the sharing and usage are done safely and securely. To achieve that, entities must be able to trust the others and the data they share first. Research in this direction has led to several approaches, many of which are working well in other context. Here, we need to find ways to establish and build trust in shared data in more innovative and efficient ways to facilitate efficient large-scale data-driven security. Techniques like provenances-based, formal policy analysis, and security experiments reproducibility need to be investigated and adapted. In addition, new more tailored techniques need to be created. Furthermore, new technologies like blockchain can be enablers for trust establishment in data-driven security [79].
5. **Blockchain.** Security of smart cities require trustworthy mechanisms across multiple entities, organizations, users and infrastructures. Blockchain [18] is an emerging technology facilitating encryption capabilities, non-repudiation, and transparency of transactions. These features are needed to facilitate security mechanisms as well. For example, when blockchain is used to automatically log every access and action in a smart city system, it will be easy to recognize and isolate illegitimate activities and security attacks. In addition, different entities can safely exchange information through the encrypted capabilities of block chain and at the same time can trust these transactions. The research in blockchain has moved forward in various ways. However, there is more to be accomplished in terms of the efficiency and scalability of blockchain on a large scale, interoperability and sustainability of its methods [88].
6. **Human Behavior**. Another area in need for in depth research is the relationships between human behavior and the vulnerability to different types of security attacks [89]. There are needs in two directions: one is developing effective models to collect, organize and analyze relevant behavior data; and the other is using the results of the analysis to enhance the data-driven security designs and policies to reduce vulnerability. Available examples are human behavior-based security solutions that incorporate human behaviors [90] and the use of context-aware techniques to enhance security measures [91]. This direction of research will benefit different sectors in smart city systems such as systems administrators, residents, development companies, and governments. In addition, this will create a better understanding of how well (or not) a security system will be effective and usable.
7. **Smartness**. Artificial intelligence, predictive analytics, and machine learning tools are supporting progress in various fields including security. However, what we currently have is just the tip of the iceberg. A lot more research is needed to create intelligent data-driven security applications and enhance current ones. In addition, the rising use and reliance on smart devices is establishing another foundation to explore and utilize to enhance security and enrich the data sets available to achieve more advances in data-driven security techniques.
8. **New Technologies**. Researchers also need to start thinking outside the box and coming up with new ideas to further improve the security of smart city systems. Whether it is new hardware components, networking models, security protocols, or new methods for data analysis and utilization. We currently have tremendous technological growth and tons of advanced smart technologies that promise a lot more benefits.

The ultimate goal of a smart city is to optimize operations and quality of life without jeopardizing the security and privacy of this city and its people. Advancing research in these and possible many other areas in technology, will help get us closer to this goal.

IX. CONCLUSION

As development of smart city systems intensify, it is imperative that they are matched by development in security

approaches to protect them. Smart city systems are generally data-driven and decision based. Therefore, data-driven security is a logical and effective approach that leverages the data collected in smart cities for security purposes. In this paper we propose using data-driven approaches to create security applications for smart city systems.

We first established the sources of complexity in smart city systems and their security requirements based on their characteristics. We also briefly outlined several challenges imposed due to these characteristics. Examples of these challenges are the management, monitoring, potential attacks, ownership and control, cost vs. benefits, among others. These challenges further complicate security requirements and methods. Therefore, conventional security measures are not enough to provide active and effective protection for smart city systems, hence our opposed approach.

We then discussed the benefits of utilizing the data-driven security approaches for protecting smart city systems. These benefits include fast security risk analysis, security protection improvements, security modeling and simulation, and security management improvement, and automation. In addition, there are many applications of data-driven security in smart city systems, which were arranged in four categories: data-driven dynamic attacks detection, prevention, and mitigation; data-driven forecasting of security threats; data-driven enhanced security management, and data-driven security planning. Applications in each of these categories and those integrating across multiple categories play important roles in protecting smart city systems. We also outlined several issues to be addressed for each category and proposed possible approaches for solutions.

This work shows that data-driven security is a strong candidate to help achieve high levels of security for smart city systems. However, there is still a lot to do to make this approach more effective, efficient and practical. Thus, we included a discussion of the impact of these approaches and possible future directions for research. These include investigating different areas of promise such as data science, big data, data trustworthiness, smart techniques, human behavior, and data modeling. In addition, there is always the need for more innovations and inventions for better data-driven security for smart city systems.

REFERENCES

[1] *World Urbanization Prospects—The 2014 Revision*, United Nations, New York, NY, USA, 2014.
[2] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl, "Understanding smart cities: An integrative framework," in *Proc. 45th Hawaii Int. Conf. Syst. Sci.*, Jan. 2012, pp. 2289–2297.
[3] R. Khatoun and S. Zeadally, "Smart cities: Concepts, architectures, research opportunities," *Commun. ACM*, vol. 59, no. 8, pp. 46–57, Jul. 2016.
[4] B. Thuraisingham, M. Kantarcioglu, K. Hamlen, L. Khan, T. Finin, A. Joshi, T. Oates, and E. Bertino, "A data driven approach for the science of cyber security: Challenges and directions," in *Proc. IEEE 17th Int. Conf. Inf. Reuse Integr. (IRI)*, Jul. 2016, pp. 1–10.
[5] J. Jacobs and B. Rudis, *Data-Driven Security: Analysis, Visualization and Dashboards*, Hoboken, NJ, USA: Wiley, 2014.
[6] S. Tan, D. De, W.-Z. Song, J. Yang, and S. K. Das, "Survey of security advances in smart grid: A data driven approach," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 397–422, 1st Quart., 2017.
[7] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and privacy in smart city applications: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 122–129, Jan. 2017.
[8] A. S. Elmaghraby and M. M. Losavio, "Cyber security challenges in smart cities: Safety, security and privacy," *J. Adv. Res.*, vol. 5, no. 4, pp. 491–497, Jul. 2014.
[9] A. Bartoli, J. Hernández-Serrano, M. Soriano, M. Dohler, A. Kountouris, and D. Barthel, "Security and privacy in your smart city," in *Proc. Barcelona Smart Cities Congr.*, vol. 292, 2011, pp. 1–6.
[10] R. Kitchin, "Getting smarter about smart cities: Improving data privacy and data security," Data Protection Unit, Dept. Taoiseach, Dublin, Ireland, Tech. Rep. 7242, 2016. [Online]. Available: http://mural.maynoothuniversity.ie/7242/1/Smart
[11] S. Ijaz, M. Ali, A. Khan, and M. Ahmed, "Smart cities: A survey on security concerns," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 2, pp. 612–625, 2016.
[12] R. Khatoun and S. Zeadally, "Cybersecurity and privacy solutions in smart cities," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 51–59, Mar. 2017.
[13] J. Laufs, H. Borrion, and B. Bradford, "Security and the smart city: A systematic review," *Sustain. Cities Soc.*, vol. 55, Apr. 2020, Art. no. 102023.
[14] A. Verma, A. Khanna, A. Agrawal, A. Darwish, and A. E. Hassanien, "Security and privacy in smart city applications and services: Opportunities and challenges," in *Cybersecurity and Secure Information Systems*. Cham, Switzerland: Springer, 2019, pp. 1–15.
[15] L. Cui, G. Xie, Y. Qu, L. Gao, and Y. Yang, "Security and privacy in smart cities: Challenges and opportunities," *IEEE Access*, vol. 6, pp. 46134–46145, 2018.
[16] T. Braun, B. C. M. Fung, F. Iqbal, and B. Shah, "Security and privacy challenges in smart cities," *Sustain. Cities Soc.*, vol. 39, pp. 499–507, May 2018.
[17] R. Kitchin and M. Dodge, "The (in) security of smart cities: Vulnerabilities, risks, mitigation, and prevention," *J. Urban Technol.*, vol. 26, no. 2, pp. 47–65, 2019.
[18] K. Biswas and V. Muthukkumarasamy, "Securing smart cities using blockchain technology," in *Proc. IEEE 18th Int. Conf. High Perform. Comput. Commun., IEEE 14th Int. Conf. Smart City, IEEE 2nd Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, Dec. 2016, pp. 1392–1393.
[19] J. Wu, K. Ota, M. Dong, and C. Li, "A hierarchical security framework for defending against sophisticated attacks on wireless sensor networks in smart cities," *IEEE Access*, vol. 4, pp. 416–424, 2016.
[20] Z. Khan, Z. Pervez, and A. Ghafoor, "Towards cloud based smart cities data security and privacy management," in *Proc. IEEE/ACM 7th Int. Conf. Utility Cloud Comput.*, Dec. 2014, pp. 806–811.
[21] M. Sen, A. Dutt, S. Agarwal, and A. Nath, "Issues of privacy and security in the role of software in smart cities," in *Proc. Int. Conf. Commun. Syst. Netw. Technol.*, Apr. 2013, pp. 518–523.
[22] Z. A. Baig, P. Szewczyk, C. Valli, P. Rabadia, P. Hannay, M. Chernyshev, M. Johnstone, P. Kerai, A. Ibrahim, K. Sansurooah, N. Syed, and M. Peacock, "Future challenges for smart cities: Cyber-security and digital forensics," *Digit. Invest.*, vol. 22, pp. 3–13, Sep. 2017.
[23] P. Wang, A. Ali, and W. Kelly, "Data security and threat modeling for smart city infrastructure," in *Proc. Int. Conf. Cyber Secur. Smart Cities, Ind. Control Syst. Commun. (SSIC)*, Aug. 2015, pp. 1–6.
[24] S. Chakrabarty and D. W. Engels, "A secure IoT architecture for smart cities," in *Proc. 13th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2016, pp. 812–813.
[25] I. S. Farahat, A. S. Tolba, M. Elhoseny, and W. Eladrosy, "Data security and challenges in smart cities," in *Security in Smart Cities: Models, Applications, and Challenges*. Cham, Switzerland: Springer, 2019, pp. 117–142.
[26] M. Aloqaily, S. Otoum, I. A. Ridhawi, and Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities," *Ad Hoc Netw.*, vol. 90, Jul. 2019, Art. no. 101842.
[27] M. A. Hasbini, T. Eldabi, and A. Aldallal, "Investigating the information security management role in smart city organisations," *World J. Entrepreneurship, Manage. Sustain. Develop.*, vol. 14, no. 1, pp. 86–98, 2018.
[28] L. Edwards, "Privacy, security and data protection in smart cities: A critical EU law perspective," *SSRN Electron. J.*, vol. 2, no. 2, p. 28, 2016.
[29] M. M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, and G. Fortino, "A hybrid deep learning model for efficient intrusion detection in big data environment," *Inf. Sci.*, vol. 513, pp. 386–396, Mar. 2020.

[30] M. Fitzgerald, "Data-driven city management: A close look at Amsterdam's smart city initiative," *MIT Sloan Manage. Rev.*, vol. 57, no. 4, May 2016.

[31] S. Eggimann, L. Mutzner, O. Wani, M. Y. Schneider, D. Spuhler, M. M. de Vitry, P. Beutler, and M. Maurer, "The potential of knowing more: A review of data-driven urban water management," *Environ. Sci. Technol.*, vol. 51, no. 5, pp. 2538–2553, Mar. 2017.

[32] P. Horazdovsky, V. Novotny, and M. Svitek, "Data-driven management of dynamic public transport," in *Proc. Smart City Symp. Prague (SCSP)*, May 2018, pp. 1–5.

[33] N. Cheng, F. Lyu, J. Chen, W. Xu, H. Zhou, S. Zhang, and X. Shen, "Big data driven vehicular networks," *IEEE Netw.*, vol. 32, no. 6, pp. 160–167, Nov. 2018.

[34] Y. Chen, X. Zeng, and T. Yuan, "Data-driven safety model on urban rail transit signal system," in *Proc. Int. Symp. Intell. Transp. Smart City*. Singapore: Springer, 2019, pp. 145–154.

[35] C. Yang, G. Su, and J. Chen, "Using big data to enhance crisis response and disaster resilience for a smart city," in *Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2017, pp. 504–507.

[36] W. Tu, "Data-driven QoS and QoE management in smart cities: A tutorial study," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 126–133, Dec. 2018.

[37] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Identifying services for short-term load forecasting using data driven models in a smart city platform," *Sustain. Cities Soc.*, vol. 28, pp. 108–117, Jan. 2017.

[38] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 215–225, Apr. 2016.

[39] N. Mohamed, J. Al-Jaroodi, and I. Jawhar, "Service-oriented big data analytics for improving buildings energy management in smart cities," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 1243–1248.

[40] N. Petrovic and D. Kocic, "Data-driven framework for energy-efficient smart cities," *Serbian J. Electr. Eng.*, vol. 17, no. 1, pp. 41–63, 2020.

[41] R. Matheus, M. Janssen, and D. Maheshwari, "Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities," *Government Inf. Quart.*, Feb. 2018, Art. no. 101284. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0740624X18300303

[42] J. Al-Jaroodi and N. Mohamed, "Service-oriented architecture for big data analytics in smart cities," in *Proc. 18th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGRID)*, May 2018, pp. 633–640.

[43] N. Mohamed, S. Lazarova-Molnar, and J. Al-Jaroodi, "Cloud of things: Optimizing smart city services," in *Proc. 7th Int. Conf. Model., Simul., Appl. Optim. (ICMSAO)*, Sharjah, UAE, Apr. 2017, pp. 1–5.

[44] A. Gaur, B. Scotney, G. Parr, and S. McClean, "Smart city architecture and its applications based on IoT," *Procedia Comput. Sci.*, vol. 52, pp. 1089–1094, Jan. 2015.

[45] I. Jawhar, N. Mohamed, and J. Al-Jaroodi, "Networking architectures and protocols for smart city systems," *J. Internet Services Appl.*, vol. 9, no. 1, p. 26, Dec. 2018.

[46] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[47] J. Cardoso, K. Voigt, and M. Winkler, "Service engineering for the Internet of services," in *Proc. Int. Conf. Enterprise Inf. Syst.* Berlin, Germany: Springer, 2008, pp. 15–27.

[48] N. Mohamed, J. Al-Jaroodi, I. Jawhar, S. Lazarova-Molnar, and S. Mahmoud, "SmartCityWare: A service-oriented middleware for cloud and fog enabled smart city services," *IEEE Access*, vol. 5, pp. 17576–17588, Dec. 2017.

[49] K. M. Alam and A. El Saddik, "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems," *IEEE Access*, vol. 5, pp. 2050–2062, 2017.

[50] N. Mohamed, J. Al-Jaroodi, I. Jawhar, A. Idries, and F. Mohammed, "Unmanned aerial vehicles applications in future smart cities," *Technol. Forecasting Social Change*, vol. 153, Apr. 2020, Art. no. 119293.

[51] N. Shone, T. Nguyen Ngoc, V. Dinh Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 1, pp. 41–50, Feb. 2018.

[52] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proc. 9th EAI Int. Conf. Bio-Inspired Inf. Commun. Technol. (BIONETICS)*, 2016, pp. 21–26.

[53] R. Sahani, C. Rout, J. C. Badajena, A. K. Jena, and H. Das, "Classification of intrusion detection using data mining techniques," in *Progress in Computing, Analytics and Networking*. Singapore: Springer, 2018, pp. 753–764.

[54] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, and P.-N. Tan, "Data mining for network intrusion detection," in *Proc. NSF Workshop Next Gener. Data Mining*, 2002, pp. 21–30.

[55] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE Access*, vol. 2, pp. 1149–1176, 2014.

[56] M. Kolomeec, A. Chechulin, A. Pronoza, and I. V. Kotenko, "Technique of data visualization: Example of network topology display for security monitoring," *JoWUA*, vol. 7, no. 1, pp. 58–78, 2016.

[57] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 8, pp. 1313–1329, Aug. 2012.

[58] A. A. Cárdenas, P. K. Manadhata, and S. P. Rajan, "Big data analytics for security," *IEEE Security Privacy*, vol. 11, no. 6, pp. 74–76, Nov./Dec. 2013.

[59] D. Han, Y. Mo, and L. Xie, "Convex optimization based state estimation against sparse integrity attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 6, pp. 2383–2395, Jun. 2019.

[60] B. Mao, Y. Kawamoto, and N. Kato, "AI-based joint optimization of QoS and security for 6G energy harvesting Internet of Things," *IEEE Internet Things J.*, early access, Mar. 23, 2020, doi: 10.1109/JIOT.2020.2982417.

[61] T. Godquin, M. Barbier, C. Gaber, J.-L. Grimault, and J.-M.-L. Bars, "Placement optimization of IoT security solutions for edge computing based on graph theory," in *Proc. IEEE 38th Int. Perform. Comput. Commun. Conf. (IPCCC)*, Oct. 2019, pp. 1–7.

[62] M. E. Kuhl, M. Sudit, J. Kistner, and K. Costantini, "Cyber attack modeling and simulation for network security analysis," in *Proc. Winter Simul. Conf.*, Dec. 2007, pp. 1180–1188.

[63] M. Ficco, M. Choraś, and R. Kozik, "Simulation platform for cybersecurity and vulnerability analysis of critical infrastructures," *J. Comput. Sci.*, vol. 22, pp. 179–186, Sep. 2017.

[64] O. Salman, I. Elhajj, A. Chehab, and A. Kayssi, "IoT survey: An SDN and fog computing perspective," *Comput. Netw.*, vol. 143, pp. 221–246, Oct. 2018.

[65] M. Frustaci, P. Pace, G. Aloi, and G. Fortino, "Evaluating critical security issues of the IoT world: Present and future challenges," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2483–2495, Aug. 2018.

[66] M. A. Jan, P. Nanda, X. He, Z. Tan, and R. P. Liu, "A robust authentication scheme for observing resources in the Internet of Things environment," in *Proc. IEEE 13th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Beijing, China, Sep. 2014, pp. 205–211.

[67] N. Mohamed and J. Al-Jaroodi, "A middleware framework to address security issues in integrated multisystem applications," in *Proc. IEEE Int. Syst. Conf. (SysCon)*, Orlando, FL, USA, Apr. 2019, pp. 1–6.

[68] N. Mohamed, S. Lazarova-Molnar, and J. Al-Jaroodi, "CE-BEMS: A cloud-enabled building energy management system," in *Proc. 3rd MEC Int. Conf. Big Data Smart City (ICBDSC)*, Mar. 2016, pp. 1–6.

[69] Z. Khan, Z. Pervez, and A. G. Abbasi, "Towards a secure service provisioning framework in a smart city environment," *Future Gener. Comput. Syst.*, vol. 77, pp. 112–135, Dec. 2017.

[70] S. Clever, T. Crago, A. Polka, J. Al-Jaroodi, and N. Mohamed, "Ethical analyses of smart city applications," *Urban Sci.*, vol. 2, no. 4, p. 96, Sep. 2018.

[71] W. Stallings and L. Brown, *Computer Security: Principles and Practice*, 4th ed. Harlow, U.K.: Pearson Education Limited, 2018.

[72] S. Kraemer and P. Carayon, "Human errors and violations in computer and information security: The viewpoint of network administrators and security specialists," *Appl. Ergonom.*, vol. 38, no. 2, pp. 143–154, Mar. 2007.

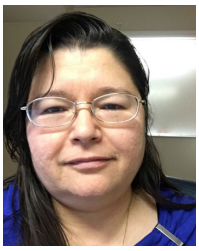
[73] G. H. Ruffo. (2020). Evidence emerges Tesla doesn't erase personal data from replaced components and they're winding up for sale online. INSIDEEVs. Accessed: May 26, 2020. [Online]. Available: https://insideevs.com/news/419525/tesla-data-leak-personal-info-ebay/

[74] K. Hayawi, Z. Trabelsi, S. Zeidan, and M. M. Masud, "Thwarting ICMP low-rate attacks against firewalls while minimizing legitimate traffic loss," *IEEE Access*, vol. 8, pp. 78029–78043, 2020.

[75] E. Bajramovic, K. Waedt, A. Ciriello, and D. Gupta, "Forensic readiness of smart buildings: Preconditions for subsequent cybersecurity tests," in *Proc. IEEE Int. Smart Cities Conf. (ISC)*, 2016, pp. 1–6.

[76] N. Sun, J. Zhang, P. Rimba, S. Gao, L. Y. Zhang, and Y. Xiang, "Data-driven cybersecurity incident prediction: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1744–1772, 2nd Quart., 2019.
[77] Y. Zhang, S. Eisele, A. Dubey, A. Laszka, and A. K. Srivastava, "Cyber-physical simulation platform for security assessment of transactive energy systems," 2019, *arXiv:1903.01520*. [Online]. Available: http://arxiv.org/abs/1903.01520
[78] C. P. Udeagwu, S. Sotiriadis, E. Asimakopoulou, N. Bessis, and M. Trovati, "Analysis of techniques for visualizing security risks and threats," in *Proc. 10th Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput. (PGCIC)*, Nov. 2015, pp. 584–590.
[79] J. Al-Jaroodi and N. Mohamed, "Blockchain in industries: A survey," *IEEE Access*, vol. 7, pp. 36500–36515, 2019.
[80] N. A. Dawit, S. S. Mathew, and K. Hayawi, "Suitability of blockchain for collaborative intrusion detection systems," in *Proc. 12th Annu. Undergraduate Res. Conf. Appl. Comput. (URC)*, Apr. 2020, pp. 1–6.
[81] C. Tankard, "Advanced persistent threats and how to monitor and deter them," *Netw. Secur.*, vol. 2011, no. 8, pp. 16–19, Aug. 2011.
[82] E. Cambiaso, G. Papaleo, G. Chiola, and M. Aiello, "Slow DoS attacks: Definition and categorisation," *Int. J. Trust Manage. Comput. Commun.*, vol. 1, nos. 3–4, pp. 300–319, 2013.
[83] J. Y. Monteith, J. D. McGregor, and J. E. Ingram, "Hadoop and its evolving ecosystem," in *Proc. 5th Int. Workshop Softw. Ecosyst. (IWSECO)*, vol. 50, 2013, pp. 50–61.
[84] S. Landset, T. M. Khoshgoftaar, A. N. Richter, and T. Hasanin, "A survey of open source tools for machine learning with big data in the Hadoop ecosystem," *J. Big Data*, vol. 2, no. 1, p. 24, Dec. 2015.
[85] A. B. M. Moniruzzaman and S. A. Hossain, "NoSQL database: New era of databases for big data analytics-classification, characteristics and comparison," *Int. J. Database Theory Appl.*, vol. 6, no. 4, pp. 1–14, 2013.
[86] Q. Cai, H. Zhang, W. Guo, G. Chen, B. C. Ooi, K.-L. Tan, and W.-F. Wong, "MemepiC: Towards a unified in-memory big data management system," *IEEE Trans. Big Data*, vol. 5, no. 1, pp. 4–17, Mar. 2019.
[87] C. Mathis, "Data lakes," *Datenbank-Spektrum*, vol. 17, no. 3, pp. 289–293, 2017.
[88] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics Informat.*, vol. 36, pp. 55–81, Mar. 2019.
[89] M. Ovelgönne, T. Dumitraş, B. A. Prakash, V. S. Subrahmanian, and B. Wang, "Understanding the relationship between human behavior and susceptibility to cyber attacks: A data-driven approach," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 4, 2017, Art. no. 51.
[90] J. Siminoff and M. J. Mitura, "Behavior-aware security systems and associated methods," U.S. Patent 16 001 627, Jun. 6, 2018.
[91] M. J. Covington, P. Fogla, Z. Zhan, and M. Ahamad, "A context-aware security architecture for emerging applications," in *Proc. 18th Annu. Comput. Secur. Appl. Conf.*, 2002, pp. 249–258.

NADER MOHAMED (Member, IEEE) received the Ph.D. degree in computer science from the University of Nebraska-Lincoln, Lincoln, NE, USA. He was a Faculty Member with the Stevens Institute of Technology, Hoboken, NJ, USA, and United Arab Emirates University, Al Ain, United Arab Emirates. He also has several years of industrial experience in information technology. He is currently a Professor of computer science and information systems with the California University of Pennsylvania, California, PA, USA. His current research interests include cybersecurity, middleware, industry 4.0, cloud and fog computing, networking, and cyber-physical systems.



JAMEELA AL-JAROODI (Member, IEEE) received the M.Ed. degree in higher education management from the University of Pittsburgh, Pittsburgh, PA, USA, and the Ph.D. degree in computer science from the University of Nebraska-Lincoln, Lincoln, NE, USA. She was a Research Assistant Professor with the Stevens Institute of Technology, Hoboken, NJ, USA, and an Assistant Professor with United Arab Emirates University, United Arab Emirates. She was also an Independent Researcher in computer and information technology. She is currently a Professor and a Coordinator of software engineering concentration with the Department of Engineering, Robert Morris University, Pittsburgh. She is involved in various research areas. Her research interests include middleware, software engineering, security engineering, cyber-physical systems, smart systems, distributed and cloud computing, and UAVs and wireless sensor networks.

IMAD JAWHAR (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from The University of North Carolina at Charlotte, USA, the M.S. degree in computer science, and the Ph.D. degree in computer engineering from Florida Atlantic University, USA. He was a Faculty Member with Florida Atlantic University for several years. He was with Motorola, as an Engineering Task Leader involved in design and development of IBM PC based on software used to program the world's leading portable radios, cutting-edge communication products and systems, and providing maximum flexibility and customization. He was also the President and the Owner of the Atlantic Computer Training and Consulting, USA, that trained thousands of people and conducted numerous classes in the latest computer system applications. Its customers included small and large corporations, such as GE, Federal Express, and International Paper. He is currently a Professor and the Chairman of the Computer Engineering with the Faculty of Engineering, Al Maaref University, Beirut, Lebanon. He has published numerous papers in international journals, conference proceedings, and book chapters. His current research interests include cyber-physical systems, wireless networks and mobile computing, sensor networks, routing protocols, and distributed and multimedia systems. He is a member of ACM and ACS Organizations. He served on numerous international conference committees and reviewed publications for many international journals, conferences, and other research organizations, such as the American National Science Foundation (NSF).

NADER KESSERWAN received the Ph.D. degree in information systems and engineering from Concordia University, Montreal, QC, Canada. He is currently an Assistant Professor of software engineering concentration with the Department of Engineering, Robert Morris University, Pittsburgh, PA, USA. His research interests include model-based testing, the IoT security, wireless sensor networks, and software engineering.