Artificial Intelligence (AI) for Development Series

Report on AI and IoT in Security Aspects

July 2018

Work in progress, for discussion purposes
Comments are welcome!
Please send your comments on this paper at: gsr@itu.int by 30 July 2018

The views expressed in this paper are those of the author and do not necessarily reflect the opinions of ITU or its Membership.

This Introductory module was prepared by Gyu Myoung Lee, under the direction of the ITU/BDT Regulatory and Market Environment Division and ICT Applications and Cybersecurity Division and under close coordination with the Chief of the ITU/BDT Infrastructure, Enabling Environment, and E-Applications Department. We would like to thank the ITU General Secretariat and the ITU Standardization Bureau for their contributions.

Contents

1. Introduction

ITU BDT has launched an Artificial Intelligence (AI) for Development Series to help Information Communications Technology (ICT) regulators (NRAs) prepare for AI, digital transformation and the digital world.

The Series includes an overall framework that will set the scene with modules on:

- 5G development examining the investment and infrastructure requirement (to support digital transformation, AI, Internet of Things (IoT), etc.);
- the social and economic impact of digital transformation;
- AI regulation for governance;
- AI for society and the security and data protection aspects linked to IoT and AI.

The series also includes a roadmap of actions.

The module on AI and IoT in Security Aspects examines the relationship between AI and IoT and analyse the security aspects linked to AI as the key component for the full realization of IoT. It also addresses the potential roles of competent national authorities, such as NRAs in ensuring that security and data protection aspects are taken into consideration or the roll-out of IoT services and applications, using AI features (e.g., machine learning).

In this regard, this report examines the relevance of AI in the current and future development of IoT and how security should be addressed, including data protection and privacy.

It covers the following objectives:

- Analyse the relation between AI and IoT and how AI is instrumental to unleash the full potential of IoT;
- Identify the current landscape and threats surrounding AI-enabled IoT. What are the main challenges and the most common attack vectors?;
- Provide use cases and good practices on securing IoT based applications, highlighting the ones that are AI enabled.
- Analyse how standardization processes can facilitate the development of securing IoT based applications and what are the standardization requirements to ensure that AI can support such deployment. That part will take into account the work of the ITU-T Study Group 20.
- Identify key security, privacy and trust challenges for IoT and AI as well as provide indications on what solutions (policy and/or technical) can be put in place to address those challenges.
- Analyse the role of national authorities on developing regulations that would promote safe use of IoT based applications, with specific focus on how data processes by AI (e.g., through machine learning) can be protected, and how privacy can be ensured.

2. Key technical trends for digital transformation in the future Introduction to AI

This section provides key technical trends for digital transformation from literature and highlight the important of AI and IoT technology. It also addresses potential risks and threats while the number of new technologies continues to grow.

Digital transformation is the change associated with the application of digital technology in all aspects of human society. The transformation stage means that digital usages inherently enable new types of innovation and creativity in a particular domain, rather than simply enhance and support traditional methods[1]. Key trends for digital transformation includes IoT, AI, 5G, Edge, Block-chain, and others.

Figure 1 shows data-driven IoT applications to leverage the massive amounts of data. Areas of IoT applications have been extended to various domains including consumers and industries. It generates a large volume of data continuously. Therefore, data-driven IoT applications is becoming significantly important in order to leverage the massive amounts of data from devices with emergence of big data and AI technology. Big data technology enables to capture, storage and analysis of data based on data collection through IoT. On top of Big data, AI can contribute to support intelligent applications without human intervention through data-based learning.

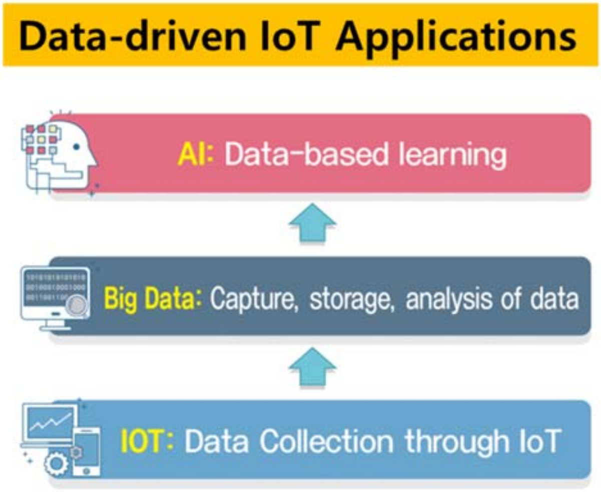
Data-driven IoT Applications
AI: Data-based learning
Big Data: Capture, storage, analysis of data
IOT: Data Collection through IoT

Figure 1. Data-driven IoT applications to leverage the massive amounts of data.

Big data can be thought of as an asset that is difficult to exploit. AI can be seen as a key to unlocking the value of big data; and machine learning is one of the technical mechanisms that underpins and facilitates AI. The combination of all three concepts can be called ‘big data analytics’[2].

Recently IoT technology has been shifted to creating value through analytics and action from connection, sensing and communications for connecting devices. In this regard, data analytics and learning techniques are very essential to support IoT applications with optimization and autonomy from relatively simple sending and remote control (see Figure 2).

[1] http://www.emptrust.com/blog/impact-of-digital-in-new-hire-onboarding

[2] ICO, “Big data, artificial intelligence, machine learning and data protection,” September 2017.

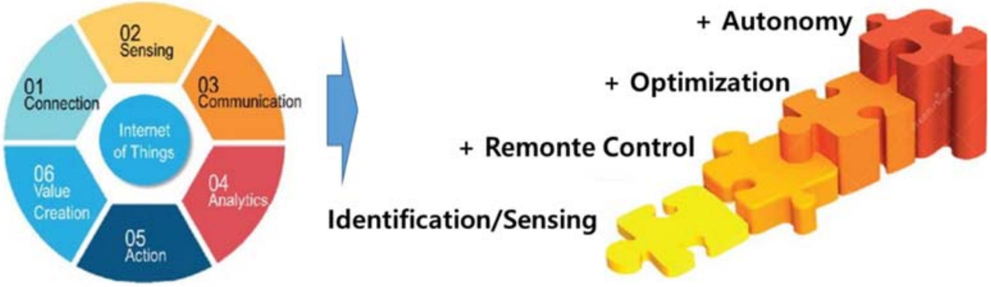

Figure 2. From connecting devices to creating value.

As one of key paradigm shifts, manufacturing processes are becoming increasingly digital, a new technical trend emerged recently; the digital twin as a sensor-enabled digital model of a physical object that simulates the object in a live setting. A digital twin can be defined, fundamentally, as an evolving digital profile of the historical and current behaviour of a physical object or process that helps optimize business performance. A digital twin is based on massive, cumulative, real-time, real-world data measurements across an array of dimensions[3]. These measurements can create an evolving profile of the object or process in the digital world that may provide important insights on system performance, leading to actions in the physical world such as changes in product design or manufacturing process.

The digital twin conceptual architecture in Figure 3 can rightly be thought of as an expansive or "under the hood" look at the enabling components that comprise the manufacturing process digital twin model, although the same basic principles may likely apply in any digital twin configuration. The conceptual architecture may be best understood as a sequence of six steps: create, communicate, aggregate, analyse, insight and act.

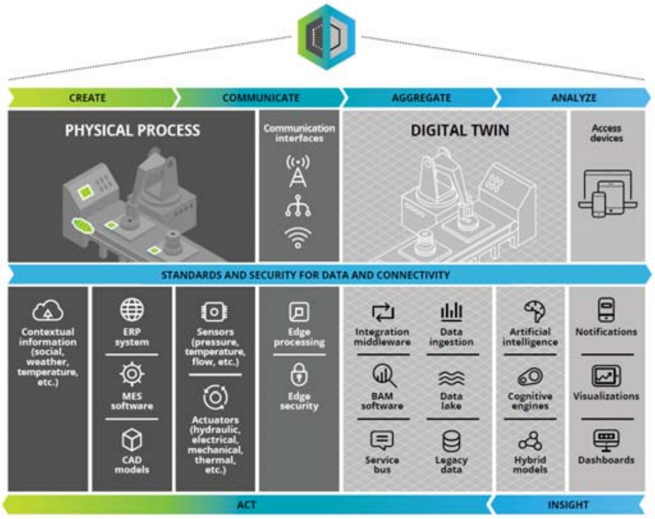

Figure 3. Digital twin conceptual architecture.

[3] "Industry 4.0 and the digital twin," Deloitte University Press, 2017.

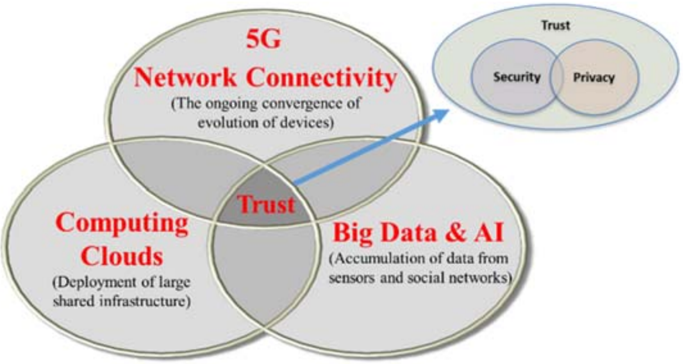
5G
Network Connectivity
(The ongoing convergence of evolution of devices)
Trust
Computing Clouds
(Deployment of large shared infrastructure)
Big Data & AI
(Accumulation of data from sensors and social networks)
Trust
Security
Privacy

Figure 4. Security, privacy and trust issues in technology convergence.

For the future, there are many emerging technologies such as 5G, cloud computing, big data and AI. These technologies will be integrated to support more advanced features and provide significant benefits in terms of technological and societal viewpoints. In the technology convergence environment, security, privacy and trust will be common issues to be controlled and managed as shown in Figure 4.

Security concerns the confidentiality, availability and integrity of data or information. Security may also include authentication and non-repudiation. Systems need a variety of methods to prevent behaviours with malicious intents. Security mainly concerns technological aspects such as the confidentiality, availability and integrity. It also includes attack detection and recovery/resilience.

Privacy concerns the expression of or adherence to various legal and non-legal norms. In certain contexts this is often understood as compliance with data protection laws regarding the right to private life. Although it would be highly complex to map into personal data protection, the globally accepted privacy principles give a useful frame: consent, purpose restriction, legitimacy, transparency, data security and data subject participation[4]. Users need the protection of their personal information related to their behaviours and interactions with other people, services and devices. Privacy mainly concerns user aspects to support anonymity and restrictive handling of personal user data.

Trust is broader concept that can cover security and privacy. Generally, trust presents the confidence and the assurance that entities, users, systems, data and process behave as they are expected to. Therefore, trust can be considered as a way of achieving extra security and privacy objectives. Trust is an important feature in the decision-making process not only used by humans in daily life but also by applications and services in the ICT environment.

It's essential to prepare 4th industrial revolution with ICT. The combination of AI, data and networks is beginning to emulate human intelligence. As explained in the introduction, key technologies to support intelligence for IoT, big data and AI will be integrated with networks for exchanging information through connectivity between all actors. These solutions will be applied to various fields such as drones and robots, etc. Basically this concept has key characteristics such as real-time response, data in everything, supporting human decision-making and self-evolution. To achieve these

[4] Robinson, N., Valeri, L., Cave, J., Starkey, T., Graux, H., Creese, S. and Hopkins, P.P. (2010), *The Cloud: Understanding the security, privacy and trust challenges*, EU CORDIS, November.

goals, the following technology areas as challenges will be of significantly important as shown in Figure 5.

- Intelligence with hyper-connected distributed intelligence.
- Interoperability with open collaboration.
- Security and privacy with block-chain trust.

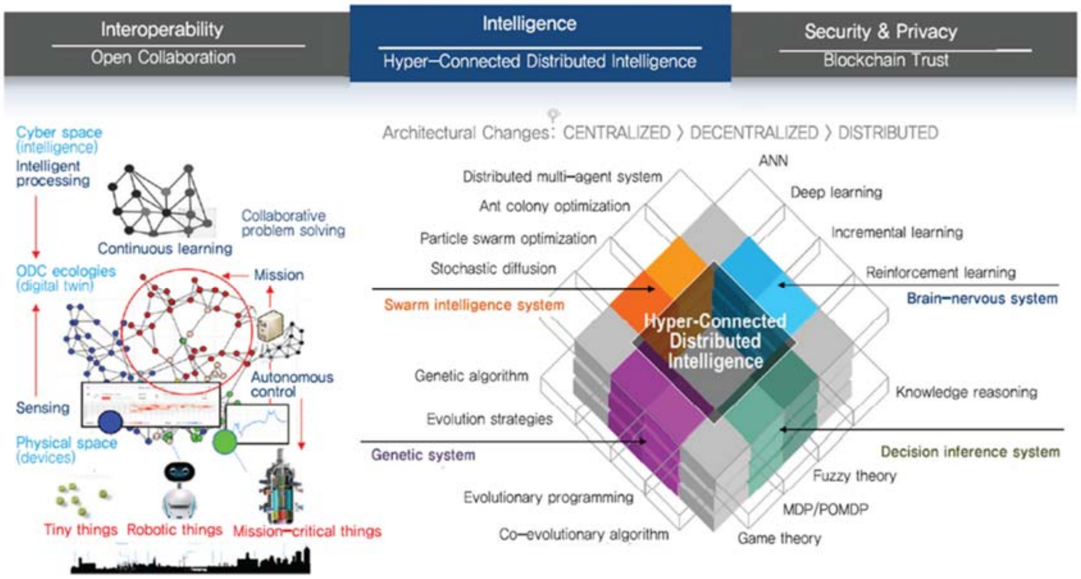

Figure 5. Challenges for 4th industrial revolution.

3. AI and Machine Learning

3.1 Introduction to AI and Machine Learning

In the summer of 1956, the dream of AI pioneers was to construct complex machines that possess the same characteristics of human intelligence. This is the concept that we think of as “General AI” - fabulous machines that have all the human senses and the ability to reason and think just like we do. General AI machines have remained in movies and science fiction novels. AI, as we currently know it, falls into the concept of “Narrow AI” which is defined as technologies that are able to perform specific tasks assigned by humans. Examples of narrow AI are: image classification, and face recognition on Facebook. These technologies exhibit some facets of human intelligence. Where does this intelligence come from? The answer is machine learning.

Machine learning is an approach to empower AI. At its most basic form, it is the practice of using algorithms to parse data, learn from it, and then make a decisions or predictions based on the collected data. It leverages algorithms to automatically model and find patterns in data, and these algorithms are heavily based on statistical and mathematical optimization. The optimization process often involves finding the smallest or largest value (minima or maxima) of a function. In a nutshell, machine learning is all about AI automatically learning a highly accurate predictive or classifier model, or finding unknown patterns in data, by leveraging optimization techniques. Machine learning algorithms are used primarily for the following types of output: Clustering (Unsupervised), Two-class and multi-class classification (Supervised), Regression: Univariate, Multivariate (Supervised), Anomaly detection (Unsupervised and Supervised), and Recommendation systems (aka recommendation engine).

Deep learning is a technique for implementing machine learning by using artificial neural networks. These artificial neural networks have discrete layers, connections, and directions of data propagation. Based on these general understanding of AI and machine learning, this report induces various leaning techniques.

Figure 6 illustrates basic concepts of AI and machine learning. Machine learning is the capability of a machine to learn without explicitly being programmed. On the other hand, AI is the capability of a machine to imitate intelligent human behaviour. For typical big data domain, it is concerned to ‘analyse’ on what happened and what will happen. With AI and networking, it is mainly concerning ‘act’ on which action should I take.

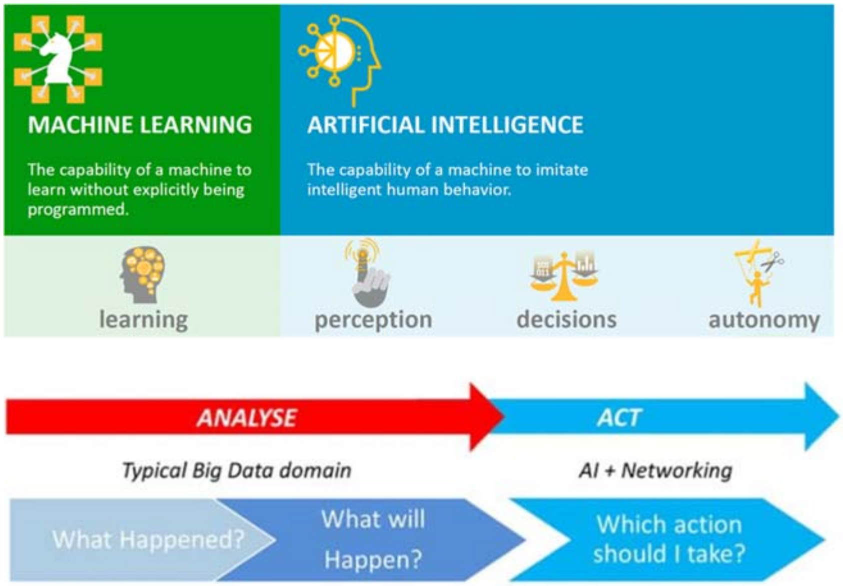
MACHINE LEARNING
The capability of a machine to learn without explicitly being programmed.
ARTIFICIAL INTELLIGENCE
The capability of a machine to imitate intelligent human behavior.
learning
perception
decisions
autonomy
ANALYSE
ACT
Typical Big Data domain
AI + Networking
What Happened?
What will Happen?
Which action should I take?

Figure 6. Artificial intelligence and machine learning.

3.2 Supervised learning

Supervised learning deals with datasets that are labelled[5]. That is, the dataset includes the features by which the event/object/thing is defined as well its desired output. In supervised learning, the data contains the response variable (label) being modelled, and with the goal being that you would like to predict the value or class of the unseen data.

Supervised learning is a machine learning technique for learning a function from training data. The output of the function can be a continuous value (called regression) or a class label of the input object (called classification). The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples (i.e., pairs of input and target output). The followings are the most common supervised algorithms: Decision Trees, Bagging, Boosting, Random Forest, K-NN, Linear Regression, Naive Bayes, Neural Networks, Logistic Regression, Relevance Vector Machine (RVM), and Support Vector Machine (SVM).

- Random Forest – An ensemble learning method for classification, regression which utilizes multitudes of decision trees and outputting of the mode of the classes (classification) and mean prediction (regression).
- Support Vector Machines – Another powerful set of Machine Learning algorithms used for classification of anomalies.
- K-nearest neighbour (K-NN) – Used heavily in pattern recognition and classification, they use the majority votes to align a data point to a specific class.

A good example of supervised leaning is classification for malicious email detection. The classification algorithm would be supplied a large dataset of emails as well as the labels ‘malicious’ or ‘safe’.

[5] Matt Lewis, “Rise of the machines: machine learning & its cyber security applications,” NCC Group Whitepaper, 2017.

3.3 Unsupervised learning

Unsupervised learning deals with unlabelled datasets. One of the most common application is to find groups/subsets within datasets, with applications being many and varied.

Unsupervised learning involves learning from data, but without the goal of prediction. This is because the data is not given with a target response variable (label), or someone chooses not to designate a response. The primary goal is to discover patterns, deep insights, understand variation, find unknown subgroups (amongst the variables or observations), from the data.

The two most commonly used techniques in unsupervised learning are principal component analysis (PCA) and clustering. PCA is the approach to learn what is called a latent variable model. Other notable latent variable modelling approaches include expectation-maximization algorithm (EM) and method of moments.

The followings are the most common unsupervised algorithms: BIRCH, Hierarchical, K-Means, DBSCAN, OPTICS, Mean-Shift, and Gaussian Mixed Model; they belong to Anomaly Detection Algorithms: Local outlier factor, and Isolation Forest; and they belong to Deep Learning Algorithms: Restricted Boltzmann machine, SOM, Autoencoder, and Generative Adversarial Networks.

3.4 Reinforcement Learning

Reinforcement learning is a reward-based learning system. Unlike most forms of machine learning, the learner of reinforcement learning is not told which actions to take. Instead, the learner must discover which actions yield the most reward by trying them. Actions may affect not only the immediate reward but also all subsequent rewards. Trial and error search, and delayed reward are the two most important distinguishing features of reinforcement learning. Reinforcement learning learns from immediate interaction with the environment, so it is different from supervised learning (learning from examples provided by a knowledgeable external supervisor).

3.5 Classification, clustering and regression

Machine learning problems can be categorised by their output or aim:

- Classification: inputs are mapped to user-specified outputs, such as emails to ‘malicious’ or ‘safe’.
- Clustering: inputs are grouped into clusters. The definitions of clusters are not known beforehand, unlike classification.
- Regression: a technique from the field of statistics used to estimate or predict outputs from a continuous – rather than discrete – set.
- Dimensionality reduction: the conversion of datasets with vast numbers of dimensions into datasets with fewer dimensions, resulting in more concisely-conveyed data for classification or regression tasks.

These are classification and regression algorithms: Decision Trees, Bagging, Boosting, Random Forest, K-NN, Linear Regression, Naive Bayes, Neural Networks, Logistic Regression, Relevance Vector Machine (RVM), and Support Vector Machine (SVM). These are cluster algorithms: BIRCH, Hierarchical, K-means, DBSCAN, OPTICS, Mean-shift, and Gaussian Mixed Model.

3.6 General application areas of AI

Classification learning systems are the area of the application which gives a new dimension for machine learning. AI systems have been widely applied in different domains: computer programming, game playing (AlphaGo), image recognition, speech recognition, medical diagnosis, agriculture, physics, email management, robotics, music, mathematics, natural language processing, etc.

The most recent AI trend is autonomous cars. An autonomous car (driverless car, self-driving car, robotic car) is a vehicle that is capable of sensing its environment and navigating without human input. Autonomous cars can detect surroundings using a variety of techniques such as radar, LIDAR, global positioning system (GPS), odometry, and computer vision. Google's self-driving car is an example of autonomous car project. For creating autonomous car, the system must be equipped with a strong AI[6].

IoT applications are generating data collected from various domains and industrial sectors. The data generated provides insights from the environments and applications that generated it. AI techniques provide the framework and tools to go beyond analytics of real time monitoring and automation use cases for IoT and move to IoT platforms that use concepts from AI and apply them to specific IoT use cases to provide smarter decision-making. AI-enabled IoT applications add a new layer of functionality and access, creating the next generation of smart homes/buildings, smart vehicles and smart manufacturing by providing intelligent automation, predictive analytics and proactive intervention[7].

In the IoT context, AI will support companies in finding the smart data and analyse the trends and patterns for better decision-making based on defined set of rules.

The AI techniques will enable cognitive systems to be integrated with IoT applications creating optimized solutions for each individual application. Cognitive IoT technologies will allow embedding intelligence into systems and processes, allowing businesses to increase efficiency, find new business opportunities, and to anticipate risks and threats so they can better address them. The IoT applications will gather and integrate data from many types of sensors and other sources, reason over data, and learn from the interactions, while creating communities of devices that share information. The information collected can be interpreted and managed by people, IoT applications or IoT platforms using cognitive systems in order to generate new and better services and use cases.

The data generated by edge devices combined with the unstructured data available from sources ranging from news Web sites and social networks can be combined using cognitive IoT capabilities at the edge at the cloud level.

The use on AI, swarm intelligence and cognitive technologies together with deep learning techniques for optimising the IoT services provides by IoT applications in smart environments and collaboration spaces will create solutions capable of transforming industries and professions.

[6] https://www.normshield.com/machine-learning-in-cyber-security-domain-1-fundamentals/

[7] Ovidiu Vermesan, et al., "IoT digital value chain connecting research, innovation and deployment," 2016.

4. IoT and Security – Security framework in Cyber-Physical Systems

This section discusses IoT and related security issues. For this, it presents a security framework in cyber-physical system (CPS) in order to understand the overall features and technical issues for securing IoT environment.

4.1 Cyber-Physical Systems for the IoT

The IoT can be characterised as a cyber physical system (CPS). A CPS conceptual model[8] is shown in Figure 7. This figure is presented here to highlight the potential interactions of devices and systems in a system of systems (SoS) (e.g., a CPS infrastructure). A CPS may be as simple as an individual device, or a CPS can consist of one or more cyber-physical devices that form a system or can be a SoS, consisting of multiple systems that consist of multiple devices. This pattern is recursive and depends on one's perspective (i.e., a device from one perspective may be a system from another perspective). Ultimately, a CPS must contain the decision flow together with at least one of the flows for information or action. The information flow represents digitally the measurement of the physical state of the physical world, while the action flow impacts the physical state of the physical world. This allows for collaborations from small and medium scale up to city/nation/world scale.

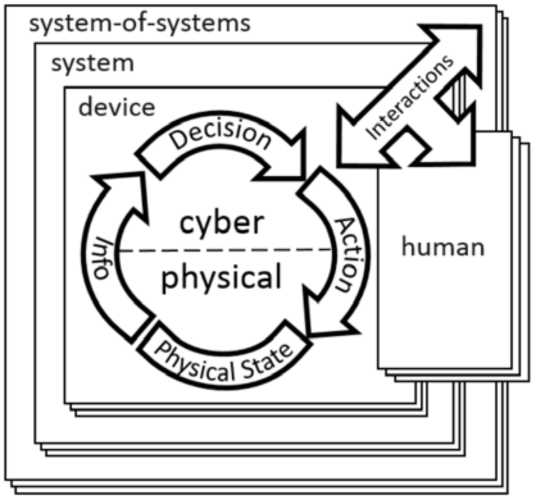

Figure 7. CPS conceptual model.

CPSs enable the physical world to merge with the virtual world by integrating computation and physical processes. A CPS facilitates tight integration between computation, communication, and control in its operation and interactions with the environment in which it's deployed. Now, in addition to embracing cyber and physical features, interest is growing in harnessing human and social factors in CPSs. For instance, one recently proposed cyber-physical-social system (CPSS)[9] embraces the cyberspace-enabled parallelism: a real system and its artificial counterparts run in parallel and interactively through cyberspace. Another proposal is a human-in the-loop CPS, which infers users' intent by measuring human cognitive activity through body and brain sensors. These studies take human and social features as important elements in CPSs mainly from the system automation and

[8] NIST, "Framework for cyber-physical systems" Release 1.0, May 2016.

[9] "A Data-Centric Framework for Cyber-Physical-Social Systems", IEEE IT Professional, Nov.-Dec. 2015.

control perspective. However, to grasp the fuller potential of a CPSS, data-centric realization is necessary.

Data-driven development will likely be a promising software paradigm in the coming decades. This will also lead to a revolution in the design and development of cyber-physical-social applications and services. With data-driven CPSS (D-CPSS), we can leverage the cross-space, multimodal data from heterogeneous data sources to better characterize the target (for instance, an event or object). The combined effects of tri-space data will also nurture numerous novel applications or services in urban environments. From a data-centric viewpoint, each CPSS follows a generic life cycle, consisting of data collection, processing, and usage. A four-layered architecture of a D-CPSS that follows this life cycle is shown in Figure 8. The resource management and cooperative sensing layers deal with data collection, whereas the data pre-processing and data analysis layers deal with data processing.

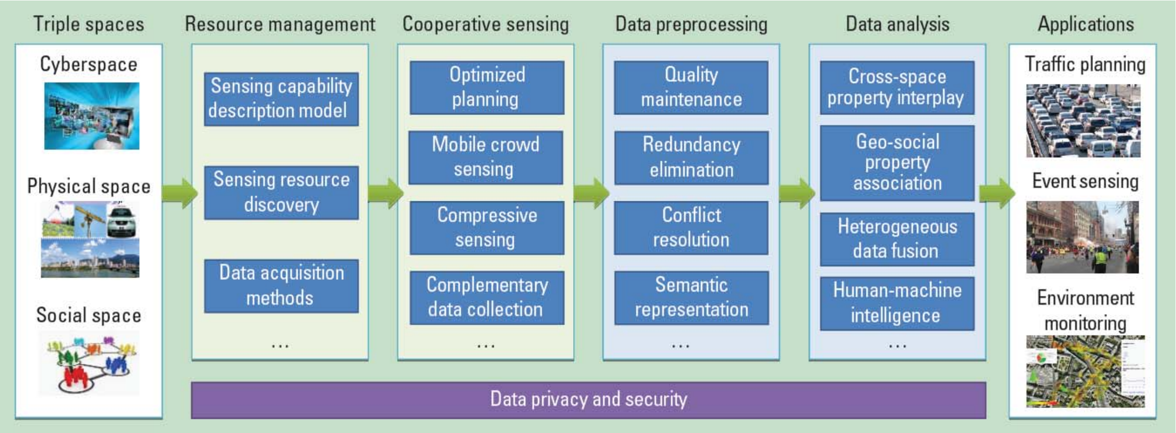

Figure 8. A layered architecture for data-driven cyber-physical-social systems (D-CPSS).

4.2 Potential risks in Cyber-Physical-Social Systems

There have been a larger number of emerging technologies been integrated in ICT infrastructures. These ICT infrastructure remain lacking in terms of having a proper platform and frameworks to ensure security and privacy, especially without the support of international standards. ICT systems, applications and services are significantly getting broader in scope, which require collaborations among various kinds of technologies and types of entities. ICT infrastructure including systems, applications and services is basically considered in CPS architecture in which the physical domain mainly consists of physical devices which interwork with each other through information and communication networks. The cyber domain is responsible for the delivery, storage and processing of data and information. While the social domain has become popular among people for sharing and showing their knowledge and become a new medium for connecting people in cyberspace.

While ICT infrastructure has grown in size and complexity, mechanisms to ensure security and privacy have not been able to keep in pace. Consequently, ICT infrastructures are at risk and vulnerabilities to a wide–range of threats at all component, device and system levels. Attackers will seek to exploit the device's physical, cyber, and social vulnerabilities by conducting various types of attacks targeted at the physical, cyber, social domains of CPS. There are many potential risks in ICT infrastructures as follows.

Risk at Physical Domain:

Smart devices and sensors have been more and more integrated into ICT infrastructures which are usually unattended by humans communicate with others via different media such as wired, wireless, broadband. However, these physical components are usually resource constrained with limited computational capability and security mechanisms. Thus, they are extremely vulnerable to both external and internal attacks.

- Natural threats[10]: Earthquakes, hurricanes, floods, and fire could cause severe damage to the physical components and computer systems of ICT infrastructures. While few safeguard measures can be implemented against natural disasters, disaster recovery plans like backup and contingency plans are the best approaches to secure systems against natural threats.
- Physical Attacks: this kind of attack tampers with hardware components and device protocols such as insertion of valid authentication tokens into a manipulated device, inserting and booting with fraudulent or modified software, and environmental/side-channel attacks, both before and after of the device's deployment. The risk is also from access attacks that unauthorized persons gain access to networks or devices to which they have no right to access by means of eavesdropping[11], spoofing, packet sniffer and network ports scanning (reconnaissance attacks)[12].

<u>Risk at Cyber Domain:</u>

The cyber risks cover a large number of areas including cyber security, information security, data provenance, and privacy in which vulnerabilities, threats and cyber-attacks are analysed and managed. Cyber security and privacy mechanisms should protect the services, hardware resources, information and data, both in transition and storage, ensure both cyber networks and services are protected against unauthorized access from within the devices and externally.

- Cyber/Information Security Attacks[13]:
    - Attacks on the Core Network: Threats at mobile network operators such as impersonation of devices, traffic tunneling between impersonated devices, mis-configuration of the firewall in the modem, router, and gateways could be the target of several kinds of attacks such as denial of service (DoS). They may also include changing the device's authorized physical location in an unauthorized fashion or attacks on the radio access network, using a rogue device.
    - Configuration Attacks include fraudulent software update/configuration changes, mis-configuration by the owner, subscribers or users, mis-configuration or compromise of the access control lists.
    - Compromise of Credentials comprise of brute force attacks on tokens and (weak) authentication algorithms, physical intrusion, or side-channel attacks, and malicious cloning of authentication tokens.
    - User Data and Identity Privacy Attacks include eavesdropping for other users or devices data sent over the systems; masquerading as other user/subscribers device; users network ID or other confidential data revealed to unauthorized third parties.
    - Access Attacks: unauthorized persons gain access to networks or devices to which they have no right to access. There are two different types of access attack: the first is physical access, whereby the intruder can gain access to a physical device. The second is remote access, whereby the intruder gain access via IP-connected devices.

[10] H. G. Brauch, "Concepts of Security Threats, Challenges, Vulnerabilities and Risks," *Copying with Global Environmental Change, Disasters and Security*, vol. 5, pp. 61-106, 2011.

[11] G. H. I. Naumann, "Privacy features of european eid card specifications," *Network Security*, vol. 4, pp. 9-13, 2008.

[12] S. R. H. C. S. Ansari, "Packet sniffing: a brief introduction," IEEE Potentials, vol. 21, pp. 17-19, 2002.

[13] C.Wilson, "Botnets, cybercrime, and cyberterrorism:Vulnerabilities and policy issues for congress," CRS Report for Congress DTIC Document, Washington DC, 2008.

- Privacy Attacks[14]: Privacy protection in IoT has become increasingly challenging due to large volumes of information easily available through remote access mechanisms.
  - Data mining: enables attackers to discover information that is not anticipated in certain databases.
  - Cyber espionage: using cracking techniques and malicious software to spy or obtain secret information of individuals, organizations or the government.
  - Eavesdropping: listening to a conversation between two parties
  - Tracking: User's movements can be tracked by the devices unique identification number (UID). Tracking user's location would allow for attackers to pinpoint the user's location in situations in which they wish to remain anonymous.
  - Password-based attacks: attempts are made by intruders to duplicate a valid user password.
- Cyber-crimes: The Internet and smart objects are used to exploit users and data for materialistic gain, such as intellectual property theft, identity theft, brand theft, and fraud

Risk at Social Domain:

In the context of social domain, the entity incudes both human (users) and machine, and the social risks are from both human-to-human and human-to-machine interactions.

- Risk of lacking trust in interactions
  - Human-human interactions: If there is no trust among peoples, their interactions (e.g., exchanging data and information) would have no meaning due to lack of confidence with each other. If the people are not trustworthy, personal interactions do not invoke any response. The unclear decision making or unrealistic situation may be happening from low or broken trust in human relationships.
  - Human-machine interactions: When a human cannot trust a machine (e.g., delivering imprecise data from a machine to a human), meaningful human-machine interactions cannot be established and the potential benefits on system performance will be lost. The human-machine systems have always proved to be unpredictable and fallible, whereas the nature of the system is to function normally. It relies on technological dependency which accentuates risks.
- Risk of attacks in social world[15]: A malicious entity is dishonest and socially uncooperative in nature and can break the basic functionality of a ICT system. This malicious entity can perform the following attacks:
  - Self-promoting attacks: a malicious user can intentionally promote its importance (by providing good recommendations for itself) in order to be selected as the service provider.
  - Whitewashing attacks: a malicious entity can disappear and rejoin the application to wash away its bad reputation.
  - Discriminatory attacks: a malicious entity can discriminatively attack non-friends or nodes without strong social ties (without many common friends) because of human nature or propensity towards friends in social networks.

[14] R. H. Weber, "Internet of Things – New security and privacy challenges," *Computer Law & Security Review*, vol. 26, no. 1, pp. 23-30, 2010.

[15] F. B. J. G. Ing-Ray Chen, "Trust-based Service Management for Social Internet of Things Systems," *IEEE Transactions on Dependable and Secure Computing*, 2015.

  - Bad-mouthing attacks: a malicious entity can ruin the reputation of another well-behaved entity by providing bad recommendations so as to decrease the chance of this good node being selected as a service provider. This is a form of collusion attacks in which collaboration among bad nodes is possible.
  - Ballot-stuffing attacks: a malicious entity can boost the reputation of another bad node by providing good recommendations for it so as to increase the chance of this bad node being selected as a service provider. This is also a form of collusion attacks, i.e., it can collaborate with other bad entities to boost the reputation of each other.

Risks due to the integrity of Physical-Cyber-Social domains

- Mismatch in CPS environments: The CPS cannot be fully operable if there is a mismatch between the physical and cyber world. If the malfunction of a physical system does not notify at the responsible entities in the cyber world, there are some risks to prevent safety in a physical world. An intelligent human in a cyber world can avoid or reduce the risk of failures and minimize the unacceptable situation in a physical world. Time critical convergence applications such as smart grid and intelligent transportation systems require high trust between the cyber world and the physical world. Greater openness, in combination with hiding one’s real identity in a physical world and making a false object in a cyber world, increases the risks that people are becoming victims of deception. They also include identity theft and exposure to inappropriate actions.
- Human errors: Without recognizing a set of rules and external conditions of a physical system, human actions may result in risks or failures. Human errors may be a primary cause or a contributing factor in risks and accidents. Intentional or unintentional human errors may cause serious problems in ICT infrastructures.
- Risk due to the complexity of ICT infrastructures:
  - A numerous number of ICT resources: Risks threaten us to cope with complexity of interactions and mechanisms of ICT infrastructures. The anonymous and/or malicious access of a large number of ICT resources causes irreparable damages and creates unpredictable dangers. It is essential to make ICT resources accessible to all the people with promises but with unknown dangers.
  - Complexity of network operation: There are a lot of algorithms for network resource optimization including efficient routing, congestion avoidance, and guaranteeing Quality of Service (QoS)/Quality of Experience (QoE). When unpredictable situations happen in a network, the out-of-service possibility increases. Natural disaster and distributed denial-of-service (DDoS) attacks are also part of the risks. While network control functions can arrange a by-pass or de-tour route to cope with overflowed traffic, the unexpected side effects like traffic fluctuation and domino effect may bring additional risks. To increase network survivability during network operation, networking protocols and Operations, Administrations, Maintenance, and Provisioning (OAM&P) functions should be re-designed to be trustworthy. Moreover, when a network infrastructure includes a cloud platform with a large volume of storage and processing capabilities, network instability is not only coming from traffic congestion. The operation of the cloud platform and high-level applications are additional harmful sources to increase network risks. The existing security functions including firewall and Deep Packet Inspection (DPI) may be replaced to provide a certain level of trust, through the implementation by a trust gateway system and trust-guaranteed network OAM functions.
  - Complexity of convergence services and applications: ICT based services and applications will continue to be heterogeneous, and this may lead to an increase in the number of convergence services that cover multiple service domains. Within the IoT and CPS environments, people, platforms and devices will be highly inter-connected by a dynamic

network while operating in heterogeneous environments. These kinds of highly connected environments increase the complexity of services and applications (which consume data and information from connected sensors, devices, etc.), and unknown potential risks may be incurred due to complex interactions. As ICT based applications and services will scale over multiple domains and involves multiple stakeholders, methods for assessing trust are needed to enable the users to have confidence to these services and applications.

- Risk in Data, information and knowledge process: Since future ICT infrastructures should provide data, information and knowledge process, the trust provisioning is essential. Data integrity refers to the maintenance and assurance accuracy and consistency in data. The failure of data aggregation is coming from any unintended changes to data as the results of storage, retrieval and processing operation for further information and knowledge. For example, if data stored in a cloud platform are shared by anonymous users, there may be a possibility for undesirable situations to happen. With a certain level of trust, data delivery and cognitive data, information, knowledge and wisdom (DIKW)[16] process may be effective and meaningful.

4.3 Security, privacy and trust in CPS

Traditional information technology (IT) cybersecurity provides information protection (integrity, confidentiality) and readiness for correct services (availability). CPS cybersecurity has the same goals as traditional IT cybersecurity though perhaps with different priorities but should also be focused on how to protect physical components from the results of cyber-attacks.

There are 10 steps to cybersecurity as follows[17].

- Network security: Protect your networks from attack. Defend the network perimeter, filter out unauthorised access and malicious content. Monitor and test security controls.
- User education and awareness: Produce user security policies covering acceptable and secure use of your systems. Include in staff training. Maintain awareness of cyber risks
- Malware prevention: Produce relevant policies and establish anti-malware defences across your organisation
- Removable media controls: Produce a policy to control all access to removable media. Limit media types and use. Scan all media for malware before importing onto the corporate system.
- Secure configuration: Apply security patches and ensure the secure configuration of all systems is maintained. Create a system inventory and define a baseline build for all devices.
- Managing user privileges: Establish effective management processes and limit the number of privileged accounts. Limit user privileges and monitor user activity. Control access to activity and audit logs.
- Incident management: Establish an incident response and disaster recovery capability. Test your incident management plans. Provide specialist training. Report criminal incidents to law enforcement.

[16] DIKW (Data, Information, Knowledge and Wisdom): This refers loosely to a class of models for representing purported structural and/or functional relationships between data, information, knowledge, and wisdom. “Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge”. (Source: https://en.wikipedia.org/wiki/DIKW_Pyramid)

[17] National Cyber Security Centre, www.ncsc.gov.uk.

- Monitoring: Establish a monitoring strategy and produce supporting policies. Continuously monitor all systems and networks. Analyse logs for unusual activity that could indicate an attack.
- Home and mobile working: Develop a mobile working policy and train staff to adhere to it. Apply the secure baseline and build to all devices. Protect data both in transit and at rest.

Two challenges are typical for CPS cybersecurity[18]:

- Detection and prevention of deception attacks (e.g., attacks on sensors that can lead them to input malicious data to the cyber component and, as a result, to provide wrong, or even dangerous, output from the cyber component)
- Detection of compromised cyber components and prevention of incorrect cyber functioning (or failure to function)

These challenges are not unique to CPS; rather, their consequences are potentially more severe because they impact the physical and cyber world. More importantly, the means to prevent these problems include not only cybersecurity controls, but also safety and reliability controls that are not applicable to traditional IT systems. Thus, CPS cybersecurity requirements should be determined in conjunction with safety, reliability, and privacy requirements. In case of system failures or cyberattacks, CPS resilience should provide ways and means to continue not just IT services but also able to provide full CPS recovery. This can be done only through co-design of CPS cybersecurity, including privacy, with safety, reliability, and resilience. As a result, considerations of the traditional tenets of confidentiality, integrity, and availability are no longer the sole focus of cybersecurity for CPS. Neither is providing CPS cybersecurity simply a matter of prioritization and application of existing controls. Rather, it involves the tradeoff of risks. This process of risk management becomes even more critical when considering the potential impact of cybersecurity failures on the ability to deliver capability across the disciplines. In addition, to develop effective CPS cyber protection and mitigation actions, the nature, functions, and interactions of all three types of components of CPS – cyber, analogue, and physical – must be understood. CPS designers and integrators should take into consideration both the intended and unintended effects resulting from the combination of properties where the goals of each property may either contradict or compliment to their counterparts. Trade-off decisions should be considered in light of the system-of-systems objective, if known. This is much more challenging than it sounds.

Trustworthy systems

A main stakeholder goal for a system is that it be trustworthy in respect to the key system characteristics. The importance of each key system characteristic to a given deployment is unique to each system and achieving one can conflict with achieving another. Interactions between the key system characteristics must be understood based on drivers such as regulatory compliance, business process and industry norms, not in isolation[19].

[18] NIST, “Framework for cyber-physical systems” Release 1.0, May 2016.

[19] Industrial Internet consortium, “Industrial Internet of Things Volume G4: Security Framework,” 2016

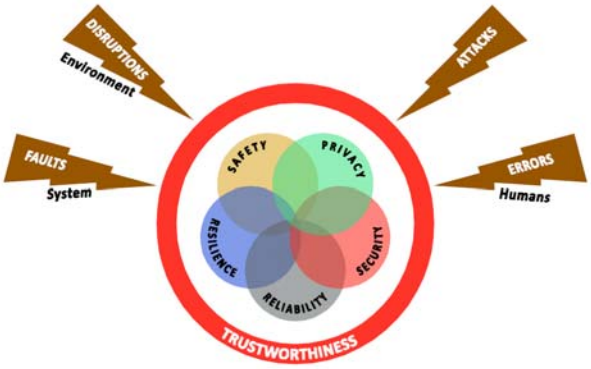

Figure 9. Trustworthiness of an Industrial IoT System.

Trustworthiness is the degree of confidence one has that the system performs as expected in respect to all the key system characteristics in the face of environmental disruptions, human errors, system faults and attacks.

ITU-T developed the first Recommendation (Y.3052) on trust in ICT. Recommendation ITU-T Y.3052[20] provides an overview of trust provisioning in information and communication technology (ICT) infrastructures and services. Recommendation ITU-T Y.3052 introduces necessity of trust to cope with potential risks due to lack of trust. The concept of trust provisioning is explained in the context of trusted ICT infrastructures and services. From the general concept of trust, the key characteristics of trust are described. In addition, a trust relationship model and trust evaluation based on the conceptual model of trust provisioning are introduced. Recommendation ITU-T Y.3052 then describes trust-provisioning processes in ICT infrastructures and services.

Trust provisioning is an integral function of physical, cyber and social trust that provides a valuable method of minimizing risks through identifying the trust characteristics of entities. Using trust provisioning, it is possible to develop trusted ICT infrastructures and services that cooperate with ICT applications in order to support these applications and services for better quality of services and experience by mitigating inherent and extraneous risks.

Figure 10 shows the concept of trusted ICT infrastructures and services. Three types of trust provisioning are classified into: physical trust for physical things (including sensors, actuators and devices); cyber trust for communication, computing and control; and social trust for stakeholders, which are mapped with trust in the physical, cyber and social worlds, respectively. In the trusted ICT world, trust entities may assume DIKW processes to minimize potential risks and to maximize the value of assets.

[20] ITU-T Y.3052, "Overview of trust provisioning in information and communication technology infrastructures and services," March 2017.

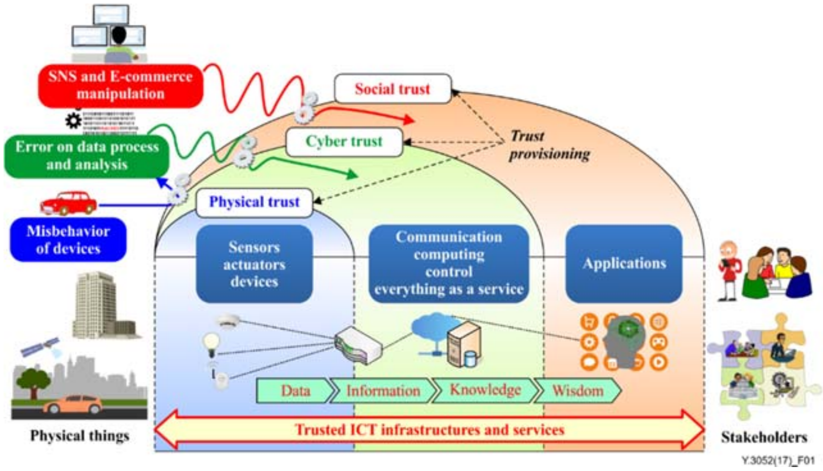
SNS and E-commerce manipulation
Error on data process and analysis
Misbehavior of devices
Social trust
Cyber trust
Physical trust
Trust provisioning
Sensors actuators devices
Communication computing control everything as a service
Applications
Data
Information
Knowledge
Wisdom
Trusted ICT infrastructures and services
Physical things
Stakeholders
Y.3052(17)_F01

Figure 10. The concept of trusted ICT infrastructures and services.

From the perspective of trust provisioning, there are physical, cyber and social worlds. To build an ICT ecosystem, raw data from physical things in the physical world are produced by physical interfaces like sensors and actuators. In the cyber world, there are physical objects and logical objects. Physical objects are those mapping to hardware devices and equipment that have capabilities of data processing, data storage and communication, etc. Logical objects are algorithms, functions and software that work on computing, storage and networking components. In the social world, entities like humans, stakeholders and software agents, which are computer programs that act for a user, produce and consume various data and applications through user interfaces. Physical things, cyber objects and social entities interact to perform trusted ICT applications taking into consideration physical, cyber and social trust, respectively. Figure 11 shows the role of trust provisioning in the ICT world in realizing various trusted ICT applications.

Physical trust

Physical trust reflects various trust aspects of physical things, which can be measured by counting on theirs trustworthiness in terms of capability, integrity and cooperation. Its capability means the ability of the physical thing to perform its task with correct functionality. Its integrity means the state of the physical thing, being stable without trouble or breakdown. Its cooperation means that the physical thing works together with other physical things for their common purposes. Physical trust reflects trust propensity that is affected by risks related to the physical world.

Cyber trust

Cyber trust reflects various trust aspects of cyber objects, which can be measured by counting on their trustworthiness in terms of capability, integrity and cooperation. Its capability means that the ability of a cyber object is correct and certain to execute control, computing and communication. Its integrity means that data handled or provided by cyber objects are not accidentally or maliciously altered or destroyed during control, computing and communication. Its cooperation means how well the cyber object works together with other objects. Cyber trust reflects trust propensity that is affected by risks related to the cyber world.

Social trust

Social trust reflects various trust aspects of social entities. Social trust can be measured by considering its trustworthiness in terms of ability, honesty and benevolence. Its ability means human competence in the individual’s activity. Its honesty implies that the social entity treats others honestly. Its benevolence means how nicely the social entity behaves to other social entities or how much the social entity interacts with other entities for their kindness. Social trust reflects trust propensity that is affected by risks in the social world.

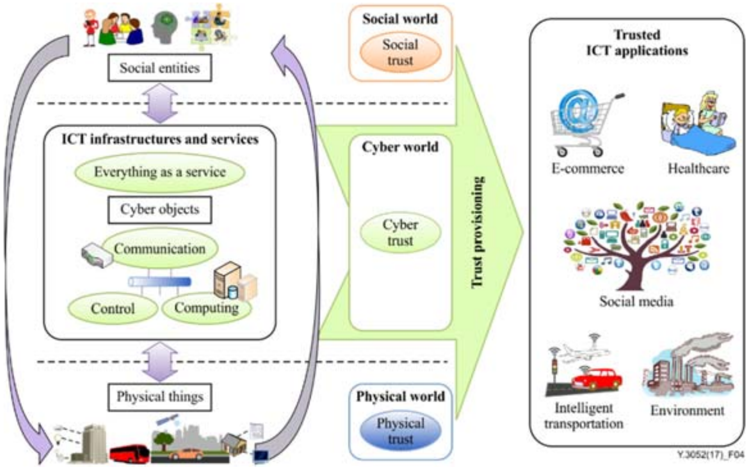

Figure 11. Trust provisioning in the ICT world for trusted ICT applications.

5. Security in IoT and AI

The IoT opens up devices for hostile parties ready to exploit every conceivable vulnerability. Thus, IoT is facing some of the greatest obstacles to its wide-spread adoption and deployment. For its global adoption, the IoT devices and networks must be safeguarded. This report lays out a framework for the adoption of AI and how it can be used to enforce the security of IoT devices and networks. One of the significant challenges of AI based security has been the inadequate resources for implementing AI techniques on resource starved IoT devices since the current approaches rely on the compute power of the Cloud to deploy AI algorithms. This approach is not feasible in practice since the target of attackers are devices and thus, securing these devices using AI techniques requires a completely different approach. This section presents some related work, discusses challenges and proposed areas of further research.

5.1 Introduction

IoT promises to integrate and connect every day object such as sensors, actuators and other physical objects to the Internet providing state-of-the-art intelligent services. The growth of IoT in the last few years has been very rapid with an estimated 50 billion devices to be connected by 2020[21][22]. IoT devices include for instance, internet connected cameras, smartphones, smartwatches and even bracelets that can share our physical activities with your friends and families. Along with the numerous benefits of utilizing IoT, it also comes with risks and security related concerns and issues. In this regards, security in various forms of attacks has been identified as one of the biggest weakness of IoT based platforms. This is due to the heterogeneity nature of these devices, communication protocols, data, as well as the humongous number of devices involved. Security issues such as jamming, spoofing, denial of services, eavesdropping, malwares in the form of viruses, Trojans, worms etc. are a great source of concern when it comes to designing and developing secured IoT systems. They present a variety of potential risks that they could be exploited to harm users or to even bring down an entire system via: (1) unauthorized access and misuse of personal information; (2) attacks facilitation on other systems; (3) risks to personal safety.

One of the major impacts of security issue in the IoT system is that it could undermine consumer confidence. Consumer confidence is necessary for the IoT technology to meet its full potential. The lack of trust and perceived risks of IoT devices may dissuade consumers to fully embrace IoT technology, creating a potential barrier in leveraging the benefits of IoT platforms in city development. These security issues are not new. They have been the major focus of infosec experts in the traditional computers and computer networks for decades[23]. However, in IoT they assume different dimensions since the conventional security mechanisms based on authentication, confidentiality, malware prevention, etc. cannot be directly deployed on IoT devices because of resource scarcity. IoT devices have prohibitively limited resources, battery lifetime, and even network bandwidth to run the traditional compute intensive security mitigation mechanisms. Thus, the lack of effective security measures enables malicious parties to access and misuse personal information, collected and transmitted through the IoT devices and network which is a challenge that needs to be urgently tackled.

A good example can be found within the smart home environment where IoT devices such as smart TV and smart phone are common. They enable consumers to browse the Internet, make purchase such as movies while disclosing sensitive information such as credit card details. This information could be stolen by identity thieves, who may effortlessly exploit the security weakness of IoT devices, in

[21] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT Security Techniques Based on Machine Learning," pp. 1–20, 2018.

[22] C. Tankard, "The security issues of the Internet of Things," *Comput. Fraud Secur.*, vol. 2015, no. 9, pp. 11–14, 2015.

[23] I. Andrea, C. Chrysostomou, and G. Hadjichristofi, "Internet of Things: Security vulnerabilities and challenges," in *Proceedings - IEEE Symposium on Computers and Communications*, 2016.

order to perpetrate fraud. Thus, in such smart home environment, the more the number of devices connected to the network, the more the vulnerabilities a malicious person could exploit to compromise personal information.

Another potential target is the network. Attack on any IoT device can facilitate attacks on the network to which it is connected and with potential to cause attack on several other connected devices. An attacked device can be used to launch denial of service attacks. In addition, considering the large number of IoT devices, the more devices the attackers can access, the more devastating the denial of service attack. Affected devices can also be used to send malicious messages via emails. There have been reports of people devices being hacked and their social media profiles are being accessed to post sensitive information or sometimes to defraud the user's social media friends.

Another important risk is that unauthorised access might be used to exploit the security vulnerability to create risks to physical safety. For example, smart cars can be hacked remotely, tampering its braking system causing physical road hazard.

Since the IoT integrates physical world, communication networks and applications, techniques for addressing security in this new computing infrastructure becomes a critical issue[24].

The goal of AI as a science is to make machines do things that would require intelligence if done by humans. The AI techniques based on machine learning for example, can recognize trends from past experiences and then able to make predictions. Therefore, security solutions based on AI techniques are expected to react more effectively to new threats than the traditional security approaches. Recently, AI based techniques such as machine learning and reinforcement learning, etc. are being suggested as effective solution to address security related issues in IoT platforms. AI based techniques have been explored broadly across the industry and applications as compute power and data availability increase.

In security, available of big data means AI techniques can be exploited to analyse and recognize patterns of security vulnerabilities to prevent such attacks. Thus, the ability of IoT based platform to learn from data to analyse, identify and mitigate security threats is an important feature that every IoT system should incorporate. These techniques are also more accurate in terms of assessment of potential malware threats from large quantity of data. In addition, AI is very suitable to detect and mitigate sophisticated attackers such as advanced persistent threats in which attackers can remain undetected for indefinite period. The rapid development in IoT and the so-called smart attacks have made it imperative to define IoT defence policy and determine various parameters in the security protocols for possible trade-off in the heterogeneous and dynamic networks.

However, one of the greatest challenges of IoT security is the question of how to classify the different types of data for detecting and scanning the IoT traffic that runs various protocols in order to identify patterns that represent security threats and then mitigate such cyber threats.

AI techniques that have been applied to address network security related issue can be broadly classified into three. Machine learning, deep learning and reinforcement learning.

5.2 IoT Security

There is no doubt that for us to push IoT for global penetration and adoption, a key challenge that must be tackled headlong is the issue of security. IoT platforms are expected to connect billions of devices, sensors, actuators and objects through the Internet allowing interactions between these objects, other entities and even humans. To allow this kind of interactions to be meaningful, IoT platforms must be provided with security guarantee to protect individual objects, information, data,

[24] M. Negnevitsky, "Artificial intelligence: a guide to intelligent systems. Pearson Education.," in *Artificial intelligence: a guide to intelligent systems*, 2005.

and services from security threats. Considering the ubiquitous nature of the IoT systems, protecting these systems against attacks is a complex process. This is because anyone can access these systems, anywhere and anytime. Moreover, an access to a single device by a malicious agent is enough to bring down the entire network of IoT systems. Furthermore, the heterogeneous nature of the billions of IoT devices exchanging data and information makes security issue a more difficult problem to address. The smart objects in IoT are connected to the global Internet with capability to communicate with several other objects with high probability of serious security breaches such as authenticity, and integrity problems[25].

With the wide range of facets that IoT system will impact, security issues if not appropriately tackled would have serious consequences, causing dame, disruption of operations and potential to cause loss of lives is quite high. Tankard[26] gave scenarios in which IoT security could come with serious consequences. In autonomous or even smart cars, these cars can be remotely hacked and this could cause massive loss of lives if security issues are not properly addressed.

In smart building, with smart heating, ventilation and air conditioning, lighting, door access control or even video surveillance, smart elevator, etc. all of which are interconnected with each other, any attack by malicious party could lead to loss of lives.

The security challenges in IoT could range from insufficient authentication, authorisation, insecure network services, lack of transport encryption, insecure cloud and edge interfaces, insecure mobile interface, poor security configurability problems, insecure software or firmware and even poor physical security. We should also note that most IoT devices have been developed without taking security into consideration partly because these devices have limited computational resources to execute security mechanisms. One of the key solution is to develop security solutions to the IoT by design. This allows security measure to be built into the IoT devices right from the start.

5.3 IoT Security Attacks

IoT systems including objects or things, networks, services and data are vulnerable to all kinds of attacks. IoT security can be defined as a set of technologies and processes designed to protect IoT devices, IoT networks, data and services from attacks, unauthorised access, change or destruction. In the traditional computing platform, cyber security minimally consists of anti-virus software, firewall and intrusion detection systems[27]. Considering security research in the academic, work on IoT security is still not well established as it is in the traditional computing environment. Most of the body of research consider the adoption of the traditional approaches to addressing the IoT security challenges. However, as said earlier, these approaches cannot be directly deployed on the IoT systems. First, in order to understand specific security issues in IoT, we analyse the peculiar features of this revolutionary computing paradigm that makes it unique and different from the existing computing platforms.

5.4 AI Techniques for IoT Security

The ability to automate the monitoring, management, and control of IoT security will be the driver for securing IoT devices and networks. AI techniques such as supervised and unsupervised machine

[25] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I.S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer, P. Doody, Internet of Things Strategic Research Roadmap, Cluster of European Research Projects on the Internet of Things, CERP-IoT, 2011.

[26] C. Tankard, "The security issues of the Internet of Things," *Comput. Fraud Secur.*, vol. 2015, no. 9, pp. 11–14, 2015.

[27] S. Sicari, A. Rizzardi, L. A. L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Networks*, 2015.

learning, reinforcement learning and even deep learning are being proposed to address security threats in IoT.

The following shows AI and machine learning as cyber tools.

- Behavioural Modelling: AI can be applied to real-time modelling of ALL Network Traffic, Log & Audit Files, Net Nodes, Servers and all “Smart IoT” Devices
- Zero-Day Attacks: AI modelling can mitigate risks of new malware that can no defined “signature”.
- Advanced Persistent Threats (APTs): Adaptive Learning Algorithms can detect the step-by-step penetration of APT malware (Phishing, Trojans, Adware, Botnets, etc.)

6. AI-based Privacy mechanism for Personal Data in the Internet of Things

6.1 Introduction

Privacy is a very broad and diverse notion for which literature offers many definitions and perspectives. With the increasing use and efficiency of electronic data processing, data privacy has become the predominant issue today, especially for the IoT. Data privacy is suitably defined as the appropriate use of data. When companies and merchants use data or information that is provided or entrusted to them, the data should be used according to the agreed purposes. The differences and relations between security and privacy is that security provides protection for all types of information, in any form, so that the information's confidentiality, integrity, and availability are maintained whereas privacy assures that personal information (and sometimes corporate confidential information as well) are collected, processed (used), protected and destroyed legally and fairly.

The evolving nature of the IoT regarding technologies and features and the emerging new ways of interaction with the IoT lead to specific privacy threats and challenges. Generally, data privacy in the IoT is the threefold guarantee to the subject for:

- Awareness of privacy risks imposed by smart things and services surrounding the data subject
- Individual control over the collection and processing of personal information by the surrounding smart things
- Awareness and control of subsequent use and dissemination of personal information by those entities to any entity outside the subject's personal control sphere.

Data privacy in the IoT captures in essence the idea of informational self-determination by enabling the subject (i) to assess its personal privacy risks, (ii) to take appropriate action to protect its privacy, and (iii) to be assured that it is enforced beyond its immediate control sphere. Data Privacy has been a hot research topic in different technology and application areas that are important enablers of the IoT vision.

Despite considerable contributions from research communities, arising privacy issues in the IoT have not been efficiently dealt. This is because the composition of a growing number of technologies and a range of changing features with an explosion in the number of smart things, interactions and inter-communications among users and things in the IoT. These new features of the IoT will aggravate privacy issues and introduce unforeseen threats that pose challenging technical problems.

In the complex IoT environment, privacy problems cannot be optimally solved due to their complexity. In these situations, AI has proven to be extremely useful and well-fitted to solve these problems. Artificial neural networks, evolutionary computation, clustering, fuzzy sets, multi-agent systems, data mining and pattern recognition are just a few examples of AI techniques that can be successfully used to solve some relevant privacy and security problems.

This report identifies the threats and challenges in data privacy in the IoT along with some state-of-the-art AI-based solutions as follows:

- briefly presents a Privacy reference model for the IoT
- introduces major privacy threats and challenges in the IoT
- describes some state-of-the-art approaches based on AI for dealing with such data privacy issues
- concludes the survey as well as propose prospective research directions for AI-based data privacy mechanisms in the IoT.

6.2 Privacy Reference Model for the IoT

6.2.1 Requirements related to IoT environment

Privacy includes the concealment of personal data as well as the ability to control what happens with this data[28]. The right to privacy can be considered as either a basic and inalienable human right, or as a personal right or possession. There are two main approaches for dealing with privacy challenges in the IoT:

- Privacy enhancing technologies (PET): PET refers to specific methods that act in accordance with the laws of data protection. PET is a system of ICT that measures the protection of informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal date[29]. The fulfilment of customer privacy requirements is quite difficult. A number of technologies have been developed in order to achieve privacy goals. PET can be any mechanisms that enhance the privacy[30].
- Legal course of action: Privacy legislation tries to draw boundaries to the evermore data-hungry business models of many Internet enterprises (e.g., data market places, advertising networks and e-commerce sites) and to define mandatory practices and processes for privacy protection.

The European Commission is aware of the security and privacy issues related to the RFID and the IoT. In particular, the Recommendation outlines measures to be taken for the deployment of RFID application to ensure that national legislation is complying with the EU Data Protection Directives 95/46, 99/5 and 2002/58 (No. 2). Member States should ensure that industry in collaboration with relevant civil society stakeholders develops a framework for privacy and data protection impact assessments (PIA; No. 4); this framework should be submitted to the Article 29 Data Protection Working Party within 12 months. The new General Data Protection Regulation (GDPR), adopted in 2016, replaces the EU Data Protection Directive (and the related national acts such as the UK DPA) as it came into force on May 25, 2018.

However, the level of privacy protection offered by legislation is insufficient, as day-to-day data spills and unpunished privacy breaches remain prevalence. The IoT will undoubtedly create new grey areas with ample of space to circumvent legislative boundaries.

6.2.2 Privacy Principles

The concepts of privacy and data protection must not be reduced to protection of data. In fact, the concepts have to be understood more broadly: they address the protection of human beings and their personal rights as well as democratic values of society. Keeping this in mind, privacy and data protection require safeguards concerning specific types of data since data processing may severely threaten informational privacy[31].

[28] Seda F. Gurses/Bettina Berendt/Thomas Santen, Multilateral Security Requirements Analysis for Preserving Privacy in Ubiquitous

Environments, in: Bettina Berendt/Ernestina Menasalvas (eds), Workshop on Ubiquitous Knowledge Discovery for Users (UKDU ’06), at 51–64; for privacy as freedom see Gus Hosein, Privacy as Freedom, in: Rikke Frank Jørgensen (ed.), Human Rights in the Global Information Society, Cambridge/Massachusetts 2006, at 121–147.

[29] https://en.wikipedia.org/wiki/Privacy-enhancing_technologies

[30] Fabian, supra note 6, 61 s; Benjamin Fabian/Oliver Gunther, Security Challenges of the EPCglobal Network, Communications of the ACM, Vol. 52, July 2009, 121–125, at 124 s.

[31] Danezis, George, et al. "Privacy and Data Protection by Design-from policy to engineering." arXiv preprint arXiv:1501.03726 (2015).

Several terms have been introduced to describe types of data that need to be protected. A term very prominently used by industry is "personally identifiable information (PII)", i.e., data that can be related to an individual. Similarly, the European data protection framework centres on "personal data". However, some authors argue that this falls short since also data that is not related to a single individual might still have an impact on the privacy of groups, e.g., an entire group might be discriminated with the help of certain information.

The ISO[32] and the OECD[33] have identified 11 privacy principles from privacy laws and regulations based on the international guidelines that have been defined to protect privacy. Wright and Raab extend that to 20 principles.

They argue that these principles should be considered as new products and services are developed[34]. Some of the principles are particularly applicable to IoT, such as "Right to confidentiality and secrecy of communications", "Consent and choice" and "People should not ... be denied goods or services or offered them on a less preferential basis". It seems that the IoT developers have not taken Wright and Raab's admonition to heart, hence the need for privacy-related IoT privacy-preserving solutions remains largely unfulfilled.

6.2.3 Privacy reference model

The privacy reference model is considered based on the IoT reference model proposed by ITU and IoT European Research Council (IERC) visions[35]. Here in the privacy model, there are 4 main types of entities namely: Smart Things, Subject, Infrastructure and Services with 5 different data flows including Interaction, Collection, Processing, Dissemination and Presentation.

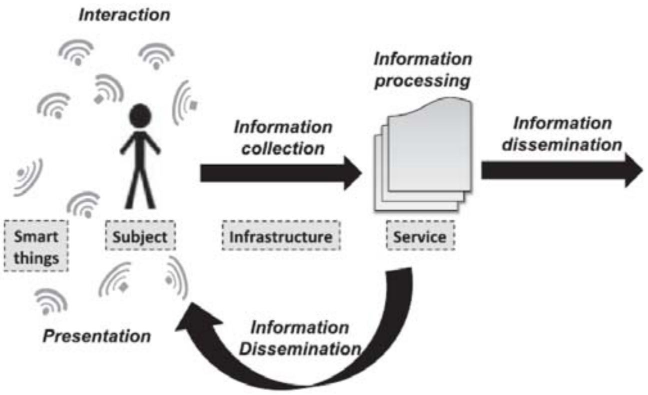

Figure 12. IoT reference model with relevant entities and data flows in a typical IoT application.

[32] International Organization for Standardization. Information technology security tech- niques privacy framework, iso/iec 29100, 2011.

[33] OECD. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. OECD Publishing, 2002

[34] D. Wright and C. Raab. Privacy principles, risks and harms. International Review of Law, Computers & Technology, 28(3):277–298, 2014.

[35] Internet of Things European Research Cluster (IERC). The Internet of Things 2012—New Horizons, 3rd edn.: Halifax, UK, 2012.

This abstract model is well aligned with other IoT models such as the models proposed by the IoT-i consortium[36], existing reference models[37] in Atzori et al. survey and IoT architectures. Considerable progress toward an explicit reference model has been made, for example, by EU FP-7 projects IoT-A[38] and CASAGRAS[39].

6.3 Privacy Threats and Challenges in the Internet of Things

The evolving nature of the IoT regarding technologies and features and the emerging new ways of interaction with the IoT lead to specific privacy threats and challenges.

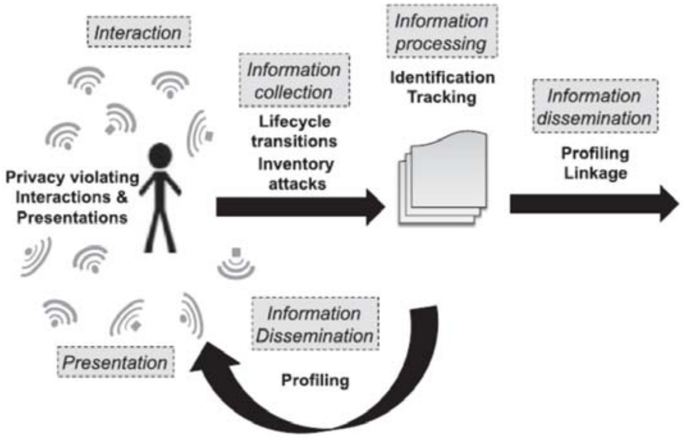

Figure 13. Threats in the reference model.

Figure 13 illustrates seven threat categories (phases) from the privacy reference model in IoT in Figure 12, namely: Identification, Localization and Tracking, Profiling, Privacy-violating Interaction and Presentation, Lifecycle transition, Inventory attacks, and Linkage[40].

6.3.1 Identification

Identification denotes the threat of associating a (persistent) identifier, for example, a name and address or a pseudonym of any kind, with an individual and data about him. The threat thus lies in associating an identity to a specific privacy-violating context, and it also enables and aggravates other threats, for example, profiling and tracking of individuals or combination of different data sources.

[36] Bauer M, Carrez F, Egan R, et al. IOT-I: Internet of Things Initiative: Public Deliverables – D1.2 First Reference Model White Paper, 2011

[37] Atzori L, Iera A, Morabito G. The Internet of Things: a survey. *Computer Networks* 2010; **54**(15): 2787–2805, doi:10.1016/j.comnet.2010.05.010

[38] IOT-A Consortium. Internet of Things architecture, 2011. Available at: http://bit.ly/124jw0M (Accessed 2012-10-12)

[39] Dunkels A, Vasseur J. IP for smart objects. Ipso alliance white paper, 2008

[40] Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. (2014). Privacy in the Internet of Things: threats and challenges. Security and Communication Networks, 7(12), 2728-2742.

The threat of identification is currently most dominant in the information processing phase at the backend services of our reference model, where huge amounts of information are concentrated in a central place outside of the subject's control. In the IoT, also the interaction and collection phase will become relevant because the impact of the evolving technologies and interconnection and interaction features aggravates the threat of identification. Identity protection and, complementary, protection against identification is a predominant topic in RFID privacy, but has also gained much attention in the areas of data anonymization[41], and privacy enhancing identity management. Those approaches (i.e., data anonymization, privacy enhancing identity management) are difficult to fit to the IoT: Most data anonymization techniques can be broken using auxiliary data, which are likely to become available at some point during the IoT evolution. Identity management solutions, besides relying heavily on expensive crypto-operations, are mostly designed for very confined environments, such as enterprise or home networks, and thus difficult to fit to the distributed, diverse, and heterogeneous environment of the IoT.

6.3.2 Localization and Tracking

Localization and tracking is the threat of determining and recording a person's location through time and space. Tracking requires identification of some kind to bind continuous localizations to one individual. Already today, tracking is possible through different means, for example, GPS, internet traffic, or cell phone location. Many concrete privacy violations have been identified related to this threat, for example, GPS stalking[42], disclosure of private information such as an illness[43], or generally the uneasy feeling of being watched[44]. However, localization and tracking of individuals is also an important functionality in many IoT systems.

The location privacy is the protection of location information of user's sensitive information such as residence location, behaviour, health status and other sensitive information. IoT devices have a built-in GPS system for positioning of location information. The user may issue a query to location based services (LBS) for location information. The query may be for a location of interest—for example, the nearest restaurant, hospital, park or other places. The query contains the identity and location of the user. The convenience of using LBS services creates issues of privacy risk. Based on the provided information, an adversary could easily link the identity and location of the user to get more private information. Security and privacy are a critical measure to consider for information gathering and broadcasting. This information and data must be secure from illegal and unauthorized access.

6.3.3 Profiling

Profiling denotes the threat of compiling information dossiers about individuals to infer interests by correlation with other profiles and data. Profiling methods are mostly used for personalization in e-commerce (e.g., in recommender systems, newsletters, and advertisements) and also for internal optimization based on customer demographics and interests. Existing approaches to preserve privacy

[41] Fung BCM, Wang K, Chen R, Yu PS. Privacy preserving data publishing: a survey of recent developments. *ACM Computing Surveys* 2010; 42 (4):14:1–14:53, doi:10.1145/1749603.1749605

[42] Voelcker J. Stalked by satellite—an alarming rise in GPS-enabled harassment. *IEEE Spectrum* 2006; **43**(7): 15–16, doi:10.1109/MSPEC.2006.1652998

[43] Chow CY, Mokbel MF. Privacy in location-based services: a system architecture perspective. *SIGSPATIAL Special* 2009; **1**(2): 23–27, doi:10.1145/1567253. 1567258.

[44] Toch E, Wang Y, Cranor L F. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Modeling and User-Adapted Interaction 2012; 22(1): 203–220, doi:10.1007/s11257-011-9110-z.

include client-side personalization, data perturbation, obfuscation and anonymization, distribution, and working on encrypted data[45][46].

In the IoT, identification and access control technologies provide essential infrastructure to link data between a user's devices with unique identities, and provide seamless and linked up services. At the same time, profiling methods based on linked records can reveal unexpected details about users' identity and private life, which can conflict with privacy rights and lead to economic, social, and other forms of discriminatory treatment. A balance must be struck between identification and access control required for the IoT to function and user rights to privacy and identity. Striking this balance is not an easy task because of weaknesses in cybersecurity and anonymisation techniques. The EU GDPR, set to come into force in May 2018, may provide essential guidance to achieve a fair balance between the interests of IoT providers and users. Through a review of academic and policy literature, this report maps the inherent tension between privacy and identifiability in the IoT. It focuses on four challenges: (1) profiling, inference, and discrimination; (2) control and context-sensitive sharing of identity; (3) consent and uncertainty; and (4) honesty, trust, and transparency. Sandra Wachter et. al examines the extent to which several standards defined in the GDPR will provide meaningful protection for privacy and control over identity for users of IoT[47].

6.3.4 Privacy-violating Interaction and Presentation

This threat refers to conveying private information through a public medium and in the process disclosing it to an unwanted audience. It can be loosely sketched as shoulder surfing but in real-world environments. Many IoT applications, for example, smart retail, transportation, and health care, envision and require heavy interaction with the user. In such systems, it is imaginable that information will be provided to users using smart things in their environment, for example, through advanced lighting installations, speakers, or video screens. Vice versa, users will control systems in new intuitive ways using the things surrounding them, for example, moving, touching, and speaking to smart things. However, many of those interaction and presentation mechanisms are inherently public; that is, people in the vicinity can observe them. This becomes a threat to privacy when private information is exchanged between the system and its user.

6.3.5 Lifecycle transition

Privacy is threatened when smart things disclose private information during changes of control spheres in their lifecycle. The problem has been observed directly with regard to compromising photos and videos that are often found on used cameras or smartphones—in some cases, disturbing data have even been found on ‘new’ devices. Because privacy violations from lifecycle transitions are mainly due to the collected and stored information, this threat relates to the information collection phase of the privacy reference model in IoT (Figure 12).

6.3.6 Inventory attacks

Inventory attacks refer to the unauthorized collection of information about the existence and characteristics of personal things. With the realization of the all-IP and end-to-end vision, smart things become query-able over the Internet. Whereas things can then be queried from anywhere by

[45] Spiekermann S, Cranor L. Engineering privacy. IEEE Transactions on Software Engineering 2009; 35(1): 67–82, doi:10.1109/TSE.2008.88

[46] Kobsa A. Privacy-enhanced web personalization. *The Adaptive Web*. Springer-Verlag: Berlin, Heidelberg, 2007; 628–670

[47] Wachter, Sandra. "Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR." (2017).

legitimate entities (e.g., the owner and authorized users of the system), non-legitimate parties can query and exploit this to compile an inventory list of things at a specific place, for example, of a household, office building, or factory. Even if smart things could distinguish legitimate from illegitimate queries, a fingerprint of their communication speeds, reaction times, and other unique characteristics could potentially be used to determine their type and model. With the predicted proliferation of wireless communication technology, fingerprinting attacks could also be mounted passively, for example, by an eavesdropper in the vicinity of the victim's house. Because inventory attacks are mainly enabled by the increasing communication capabilities of things, the threat arises in the information collection phase of the privacy reference model.

6.3.7 Linkage

This threat consists in linking different previously separated systems such that the combination of data sources reveals (truthful or erroneous) information that the subject did not disclose to the previously isolated sources and, most importantly, also did not want to reveal. Users fear poor judgment and loss of context when data that were gathered from different parties under different contexts and permissions are combined[48]. Privacy violations can also arise from bypassing privacy protection mechanisms, as the risks of unauthorized access and leaks of private information increases when systems collaborate to combine data sources. A third example of privacy violations through linkage of data sources and systems is the increased risk of re-identification of anonymized data. A common approach toward protecting privacy is working on anonymized data only, but the act of combining different sets of anonymous data can often enable re-identification through unforeseen effects[49][50].

6.4 AI-based Privacy Techniques and Mechanisms

Successful security success is about having the right combination of people, process, policy and technology. This can be achieved by developing a network management systems capable of intellectual reasoning, dynamic real time decision making, and self-adaptation and improvement based on experiences. The design of such efficient, dynamic and automated social network management framework requires support from the field of AI. Dealing with uncertainty and inconsistency has been a part of AI since its origins.

6.4.1 Traditional Privacy Preserving Approaches

In order to address the privacy concerns of end-users and privacy considerations of service providers, several approaches have been proposed by the research community:

a) Cryptographic techniques and information manipulation: Although researchers have spent many years proposing novel privacy-preserving schemes, cryptography is still the most dominant solution. However, due to limited storage and computation resources, cryptography often cannot offer adequate security protocols to safeguard end-users' data.

[48] Spiekermann S, Cranor L. Engineering privacy. *IEEE Transactions on Software Engineering* 2009; **35**(1): 67–82, doi:10.1109/TSE.2008.88

[49] Narayanan A, Shmatikov V. Myths and fallacies of "personally identifiable information". *Communications of the ACM* 2010; **53**: 24–26, doi:10.1145/ 1743546.1743558

[50] El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. *PLoS ONE* 2011; **6**(12) , doi:10.1371/journal.pone.0028071

b) Privacy awareness or context awareness: The solutions for the lack of privacy awareness have been primarily focused on relying individual applications to provide a basic privacy terms and conditions to the end-users. This practice is common among devices such as smart TVs, wearable fitness devices, and health monitor systems. For instance, in a recent research, a framework called SeCoMan was proposed to act as a trusted third party for the users as applications might not be reliable enough with the location information that they manage[51].

c) Access control: Access control is one of the viable solutions to be used in addition to encryption and raising privacy awareness. This gives users the power to manage their own data. An example of this approach is CapBAC, as proposed by Skarmeta, Hernandez, and Moreno[52]. It is essentially a distributed approach in which smart things themselves are able to make fine-grained authorization decisions.

d) Data minimization: The principle of "data minimization" means that the IoT service providers should limit the collection of personal information to what is directly relevant. They should also retain the data only for as long as it is necessary to fulfill the purpose of the services. In other words, they should collect only the personal data they really need and should keep it only for as long as they need it.

There are other proposed solutions that do not fall into the previous four categories, such as hitchhiking. This is a new approach to ensure the anonymity of users who provide their locations. Hitchhiking applications handle locations as the entity of interest. Because the knowledge of who is at a particular location is unnecessary, the fidelity tradeoff is removed[53]. Another example is the introspection technique that proactively protects users' personal information by examining the activities of the virtual machine (VM). It gathers and analyzes the CPU state of every VM, the memory contents, file I/O activity, network information that is delivered via hypervisor and detects malicious software on the VM. However, if IoT device loses integrity due to any malicious attack, it creates risks to the users' privacy[54].

6.4.2 Prospective AI-based Privacy Preserving

Capabilities of AI can be leveraged to deal with privacy challenges in the IoT.

AI-based Identification Management

- AI - is it the answer for identity management[55]?
- Could AI improve identity management and security[56]?

Identity and Access Management (IAM) is already a key weapon in the security arsenal of many organisations as a way to mitigate against data breaches and manage the additional risks that come with remote working and Bring Your Own Device (BYOD). And the take up of IAM solutions is set to gain even more momentum. IAM solutions enable a network or system to authenticate the identity of a user against a set of pre-prescribed credentials. Depending on the system being accessed, these

[51] A. Huertas Celdran, G. Clemente, J. Felix, M. Gil Perez, and G. Martinez Perez. Secoman: A semantic-aware policy framework for developing privacy-preserving and context-aware smart applications. IEEE Systems Journal, 99:1–14, 2013.

[52] A. F. Skarmeta, J. L. Hernandez-Ramos, and M. Moreno. A decentralized approach for security and privacy challenges in the internet ofthings. In Internet of Things (WF-IoT), 2014 IEEE World Forum on,pages 67–72. IEEE, 2014.

[53] K. P. Tang, P. Keyani, J. Fogarty, and J. I. Hong. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 93–102. ACM, 2006.

[54] C. Kang, F. Abbas, and H. Oh. Protection scheme for IoT devices using introspection. In Network of the Future (NOF), 2015 6th International Conference on the, pages 1–5. IEEE, 2015

[55] https://www.scmagazineuk.com/artificial-intelligence--is-it-the-answer-for-identity-management/article/531630/

[56] https://www.digitalcatapultcentre.org.uk/could-ai-improve-identity-management-and-security/

can range from a simple username and password to digital certificates, physical tokens, biometric passwords (such as fingerprints, iris scans, or facial recognition), or a combination of these features.

Traditionally, the strength of the authentication required depends on the sensitivity of the material being accessed, as well as the impact should these resources fall into unauthorised hands. Public information might require little or no authentication, while proprietary or classified data or accounts with administrative privileges will require stronger authentication, preferably using multiple factors.

While the above still holds true, recent thinking around best practice in IAM has moved on. The focus has shifted from authenticating identity to controlling access based on the principle of least privilege access. In practice, what that means is that every user – whether an individual, a device, a programme or a process – is given access only to the resources needed to fulfil their role.

Least privilege is an approach that acknowledges how serious the insider threat is to businesses and the fact that just because someone has established their identity as an employee with the right credentials should not mean unfettered access to company systems.

While this is sound in principle, least privilege and deciding who should have access to what and when can be difficult for organisations. One issue with least privilege in IAM is that users are usually given access privileges based on their role in an organisation, but employees rarely fit neatly into a single role. It is common that they may need special one-time access or each person fulfilling the same role may need slightly different types of access. Another challenge is that some organisations fail to extend the concept of least privilege access right across the organisation, monitoring those classified as privileged users, such as systems administrators.

But how might AI help? So often with data breaches it's not the management of the identity that causes the breach, but the transfer of credentials to some unknown party. While least privilege access control does afford some protection here, there are clearly insufficient. Identity management and access control have always been two sides of a coin, but in the future AI will be the glue to bind them together to much greater effect.

Moving on to biometric passwords, it's not difficult to conceive that AI could identify a user by using sight and sound. Rather than checking pre-defined credentials, a machine would be able to identify whether a person using visual and aural clues, granting access to this person accordingly.

AI also offers the potential for intelligent, real-time security by implementing fine-grained access control. Just because a user proved who they were at log on two minutes ago, should the system continue to believe they are who they say they are? Visual images and voice could obviously still play a part here, constantly monitoring users as they move around the network. However, in addition to behavioural factors and real-time, risk analysis can also come into play.

Working within a user's access permissions, AI systems could monitor in real-time whether a user is accessing or trying to access a part of the system they never normally would or suddenly downloading more documents than they generally would. The rhythm of a user's keyboard and mouse movements could be observed to identify irregular or unusual patterns. Taking this a step further it's not inconceivable that insights from an individual's online identity and activity – their social profile, groups they are part of, people they follow, websites they visit – could be used to determine a risk score. Drawing this data together, actions taken by the AI system could range from an alert being triggered, to specific areas of a corporate system being switched off for a user, to access being instantly revoked.

In the future, a truly intelligent system will know, understand, monitor and act drawing whatever clues it requires on a user. Identity and credentials will not be separate elements. An individual's identity will become their credentials. That should be the ultimate goal of any AI system.

AI and block-chain helping streamline data and identity management[57]

As we push further into the digital age, we unconsciously change how we perceive and expect services from government. This is particularly true of digital natives who, by birth right, have the "ask only once" expectation.

Digitization addresses two broad government objectives—to increase transparency and to improve speed of service for citizens. The two elements necessary to achieve this are data and identity. When combined effectively, they provide potent fuel to provide public services. The challenges of yesterday are the opportunities today and large data is easier to manage and share with citizens or agencies in order to effectively provide services. The key is to "un-duplicate" the work for citizens and public servants using disruptive technologies like AI, robotics, and block-chain.

With all the tools (networks) and fuel (data) available, government should aim for the "ask only once" point. Most data and identity is with the agencies or available to governments and same goes with citizens, this situation leaning towards transparency. To put it simply: citizens and governments know each other much better than they did a generation ago; we need ask fewer questions of one another and don't have to provide KYC at each stage of requesting a service.

The Aadhar card is one example of Indian government trying to seamlessly integrate services into citizens' everyday life. High school grades are now available in a government database and linked to each student's Aadhar number. Students should be allowed to send in applications without transcripts. The college can check everything with the Aadhar number.

Humans have historically spent a lot of time and resources producing things or providing services—from manpower to produce food to fossil fuels to ignite the industrial revolution. While we have succeeded in making the production of goods and food produce more efficient, we now have to better manage the huge amounts of data we generate and become leaner and more effective with data management. The future of data management means producing relevant data but not producing too much to manage. With tools like AI and block-chain we can avoid duplication of work and streamline identity management, while with biometrics we largely resolve issues relating to individual identity. The next step is to identify what data governments should manage in order to reach the "ask only once" point.

Although we seem to have everything to make "ask only once" a reality, we must still address how agencies secure this data. This is a challenge in democratic societies where privacy issues are respected; a recent Supreme Court judgment on privacy in India being a case in point. Citizens must have confidence in their agencies and this confidence must stem from legislation, much like it already does for the protection of physical property. The Estonian Tax and Customs Board's strong focus on transparency for example, has helped it become one of the most trusted organizations in the country.

The challenge is that unlike physical assets, our digital data is easy to access and potentially to tamper with. If government is to provide services seamlessly, it must first adequately secure our data with digital locks. This can be difficult insofar as much of it resides in the cloud and must be regulated by international treaties. Facebook and Google, for example, are often ordered to share information by national courts. Things become complicated when companies fall back on national data protection laws or rely on the unenforceability of national legislation.

Finally, most citizens today already have a digital footprint. It is in fact difficult to imagine that in a few more years anybody with citizenship will not have one. We are past the stage of asking who needs to be in the system. We are in the system.

A blockchain is a distributed immutable database consisting of a continuous growing list of blocks used to record transactions between peers in a network. The blockchain is then synchronized and

[57] https://www.capgemini.com/2017/09/ai-and-blockchain-helping-streamline-identity-management/

distributed across the network, playing as a distributed ledger. By nature, blockchain is inherently resistant to data modification because data in any given block cannot be altered retroactively as this would invalidate all hashes in the previous blocks in a blockchain; and break the consensus agreed among nodes in the network.

In Smart cities context, a GDPR-compliant data management solution for a Smart City platform leveraging Blockchain and Smart Contracts is envisioned. From citizens' point of view, personal data usage control should be strengthened by a trusted and transparent solution that enables citizens to verify if the data was accessed, processed, and transferred without violating consents, and to withdraw the consent arbitrarily whenever needed. From Smart City operators' perspective, the solution should provide a legal contract that the operators have received the consent from citizens for managing their personal data, and the agreement on data usage obligations from service providers.

In Smart Cities, operators and agencies are supposed to be trustworthy to manage all data from citizens and control the data usage from third-party service providers. Therefore, a permissioned blockchain is more suitable for a GDPR-compliant data management model rather than a public one, as many advantageous simplifications can be achieved such as it is not required to implement a cryptographic token for incentivizing mining and for determining financial stake. Transactions in a permissioned network can be considered immediately final, as the possibility of having to resolve a fork can be eliminated. In this regard, Smart City operators are in charge of restricting stakeholders to join the network, to read the ledgers, to propose transactions, and to engage in consensus process. In addition to high-speed transactions and finality, permissioned blockchain also benefits from improved privacy, control, and scalability. The adoption of the per-missioned blockchain has impact on privacy, anonymity, performance, and scalability.

<u>AI-based Localization and Tracking</u>

Some traditional AI-based approaches have been proposed for privacy preserving of location & tracking:

- Techniques based on anonymization - K-anonymity is one of the basic techniques for protection of privacy proposed for the first time by Sweeney[58]. The k-anonymity model addresses the re-identification problem during broadcasting sensitive information for the research objective. Gedik and Liu presented a new architecture for the protection of location privacy from several threats due to unrestrained practice of LBS[59]. This strategy contains a personalized k-anonymity prototype and a suite of algorithms based on anxiety to protect privacy. The distinctive feature of this design is the elastic personalization privacy to sustain k-anonymity for wide-ranging mobile clients. The prototype is designed to be on a trusted platform of an anonymization server. Wang et al. presented mobile user location privacy in active and varied scenarios, reinforcing it to articulate the location awareness and location privacy protection (L2P2) problem. The problem is additionally distributed into basic and enhanced problems, and a distinct algorithm offered for each problem[60]. AI can be used for improving the effectiveness of such techniques.
- Randomized noise-based techniques - Such techniques which are based on random noise added to the original location. This random noise changes or blurs the original location in such way that the adversary cannot acquire the actual location of the user. A location privacy preserving mechanism (LPPM) must contemplate three fundamental features: user privacy requirements, knowledge and abilities of adversary, and tolerated service quality. Shokri et al. introduced an

[58] Sweeny, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557–570.

[59] Gedik, B., & Liu, L. (2008). Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, *7*(1), 1–18.

[60] Wang, Y., Li, F., & Xu, B. (2012). L2P2: Location-aware location privacy protection for location-based services. In *INFOCOM, 2012 proceedings IEEE* (pp. 1996–2004).

optimum LPPM for LBS which gives users a service quality constraint against an adversary optimal inference algorithm[61]. The authors formalize mutual optimization with location privacy versus correctness of localization by using Stackelberg Bayesian games. It reports that an adversary could not observe that the location has been disturbed by the user.

AI-based Profiling

There are at least three possible ways of monitoring and profiling that offer grounds for discrimination in IoT systems: (a) data collection that leads to inferences about the person (e.g. Internet browsing behaviour); (b) profiling at large through linking IoT datasets (sometimes called 'sensor fusion'); and (c) profiling that occurs when data is shared with third parties that combine data with other datasets (e.g. employers, insurers).

Weaknesses of anonymisation to prevent profiling and resulting discrimination cause further problems. According to Gudgel, "there is special concern that if data is not anonymized then it could potentially be used to track specific individuals, linked to information in other databases, and possibly used to predict future behaviour."[62] Tracking data of the type that many IoT devices generate is notorious to open to re-identification and reverse engineering of identity. Following the assumption that data cannot be permanently anonymised without destroying its analytical value, non-technical methods may be necessary to prevent profiling and discrimination in the IoT. One potential solution is to treat all IoT generated data that refers to a user as personal data under data protection law, as it will always be possible in principle to link the data back to a person. This approach would ensure that the user would be able to exercise his/her rights granted under data protection law over all information that IoT devices create and manage. This concept would not prevent profiling as a result; but rather, extend the scope of existing user rights against privacy risks to cover all data, including inferences and profiles.

These problems will be exacerbated by the proliferation of machine learning in the IoT. Machine learning will lead to even less predictable inferences, while the complexity and opaqueness of machine learning algorithms can inadvertently hide discriminatory treatment from users. Systems operating as 'black boxes', for which the inputs, internal logic, and outputs may be unavailable or incomprehensible to individual users do not facilitate systematic observation, identification of harmful effects, or investigation of their causes. Machine learning can inadvertently and unknowingly reinforce existing biases and prejudices as a result[63].

European legislators have addressed the risks of profiling and discrimination, albeit often lacking detailed recommendations. European regulators have raised "major concerns in declarations on profiling," acknowledging that profiling offers grounds for discrimination, especially when datasets are combined. The European Commission has called for the creation of a set of guiding principles to govern IoT regulation, urging that "always being connected to the things around us has the potential to lead to more surveillance or more profiling by public authorities and private entities." Similarly, the European Data Protection Supervisor has voiced concerns that RFID tags used in IoT systems could lead to profiling by linking users to specific devices and usage records.

Similar concerns are reflected in the GDPR, especially in Article 21 (Right to object) and Article 22 (Automated individual decision-making, including profiling)[64].

AI-based Privacy-violating Interaction and Presentation

[61] Shokri, R., Theodorakopoulos, G., & Troncoso, C. Protecting location privacy: Optimal strategy against localization attacks. In *Proceedings of the 2012 ACM conference on computer and communications security* (pp. 617–627). ACM.

[62] John Gudgel, 'Objects of Concern? Risks, Rewards and Regulation in the Internet of Things"' (Social Science Research Network 2014) SSRN Scholarly Paper ID 2430780 12, July 2017.

[63] Brent Mittelstadt, 'Auditing for Transparency in Content Personalization Systems' (2016) 10 International Journal of Communication 12.

[64] Wachter, Sandra. "Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR." (2017).

Because such advanced IoT services are still in the future, privacy-violating interactions have not received much attention from research. Interaction mechanisms are, however, crucial to usable IoT systems, and privacy threats must consequently be addressed.

We identify two specific challenges that will have to be solved: First, we need means for automatic detection of privacy-sensitive content. It is easily imaginable that the provisioning of content and rendering it for the user are handled in two steps by two different systems: For example, company A generates recommendations for customers of a store, which are then delivered to the customer by company B's system either by special lighting and the use of speakers or through a push to his smartphone. How to choose between those two interactions mechanisms, one public one private? Should company A mark privacy-sensitive content or should company B detect it? How can company B (committed to privacy) protect itself from A's lax privacy attitude? Automatic detection of privacy-sensitive content can help to decide these questions. Second, with the previous point in mind, scoping will be necessary; that is, how can we scope public presentation medium to a specific subgroup of recipients or a specific physical area? This approach would prove useful to support users, which have no smartphone (or any other device providing a private channel for interactions and presentations). However, it will be difficult to accurately determine the captive audience of a particular presentation medium, separate the intended target group, and adjust the scope accordingly. For example, what if the target user is in the midst of a group of people? Applications for privacy-preserving pervasive interaction mechanisms are, for example, smart stores and malls, smart cities, and healthcare applications. Here, it would certainly be an achievement to provide similar levels of privacy as people would expect in the contexts of their everyday conversations, that is, interactions with their peers.

AI-based Linkage

The purpose of a privacy preserving technique for a privacy threat is to link data across organisations such that besides the linked records (the ones classified to refer to the same entities) no information about the sensitive source data can be learned by any party involved in the linking, or any external party. Challenges with this technique is that the linkage unit needs access to personal details (metadata might also reveal sensitive information). Also, collusion between parties, and internal and external attacks, make these data vulnerable.

Privacy-preserving record linkage (PPRL) is a solution that aims to overcome these drawbacks. No un-encoded data ever leave a data source; only details about matched records are revealed. However, provable security against different attacks PPRL is challenging (employs techniques from cryptography, databases, etc.)[65].

[65] Verykios VS, Karakasidis A and Mitrogiannis VK: Privacy preserving record linkage approaches. International Journal of Data Mining, Modelling and Management, 2009.

7. AI-based Trust mechanisms in the Internet of Things

7.1 Introduction

This section explains the latest advances and applications of trust solutions based on AI techniques for IoT through the various type of references including scientific publications, text books, and online articles. The scope of the survey includes, but is not limited to, titles in the areas of trust solutions based on AI techniques like supervised-unsupervised machine learning, reputation systems, reinforcement learning, deep learning, and multi agent systems.

First, in order to consider trust issues, it is required to understand basic concepts of trust in Section 4.3. Then, it is important to clarify that trust is neither a property of a trustor (e.g., trustor's preferences) nor a property of a trustee (e.g., trustee's trustworthiness and trustee's reputation). It is a relationship between the trustor and the trustee that is subjective and asymmetric which is derived from the triad of trustee's trustworthiness, trustor's propensity and environment's characteristics. Based on the clarification of the trust concept, a conceptual trust model in the IoT is proposed as illustrated in Figure 14. Then, a more specific trust definition in the IoT associated with the conceptual trust model is proposed as follows:

"Trust is the perception of a trustor on trustee's trustworthiness under a particular environment (within a period of time) so-called perceived trustworthiness."

According to the model illustrated in Figure 14, trust will be obtained by harmonizing the trustor's propensity and environment conditions into the trustee's trustworthiness. The harmonization is accomplished by aggregating both the observation of a trustor toward a trustee and the interactions between the two. It is worth to note that the environment conditions are reflected as risks taken during the observations and interactions. The trustor's propensity includes both requirements for the trust goal and the trustor's preferences about the trustee's trustworthiness whereas the environment conditions are the considerations for some factors such as vulnerabilities, threats and risks. The trust goal requirements with the environmental factors helps determining the set of TAs for deriving the perceived trustworthiness whereas the trustor's preferences is to help combining these TAs to obtain an overall trust value for making a decision. For example, trustor's preferences could be represented in forms of weights of TAs, indicate the levels of importance of the TAs when constructing trust. Trust as perceived trustworthiness is as an instance of trustee's trustworthiness respecting to a particular trustor and an environment, thus, even same a trustee and same an environment, different trustors might have different propensities of the trustee's trustworthiness. This illustrates the subjective characteristic of trust. Another important characteristic of trust is the context-dependence that can also be illustrated using this conceptual model as follows: with the same trustor and trustee, different environments might result in different TAs and different trustor's propensities.

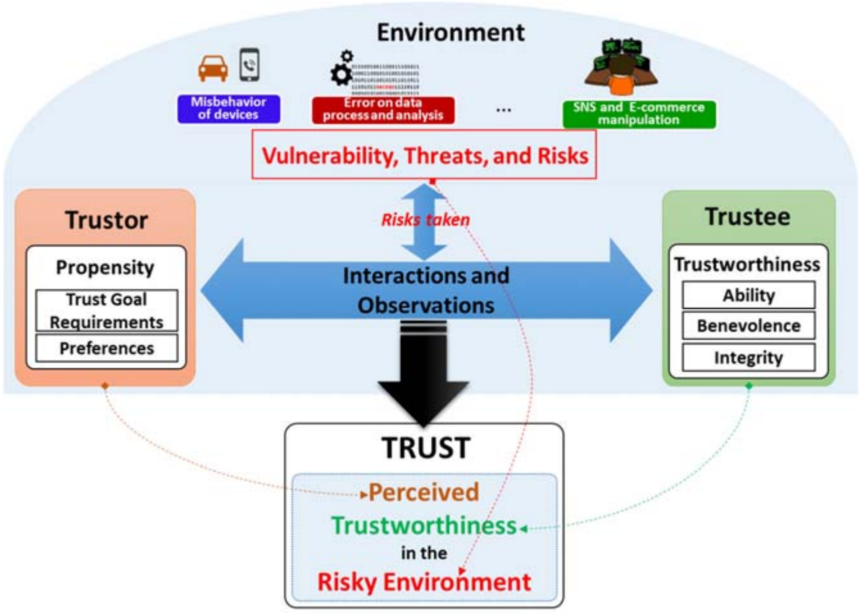

Figure 14. Conceptual Trust Model in the IoT environment.

Based on the conceptual model, the goal of any trust model is two-fold: (i) to specify and evaluate TAs of the trustworthiness of a trustee respecting to the trustor's propensity and the environment conditions; (ii) to combine the TAs to finalize the perceived trustworthiness as the trust value. From now on in this article, the term "trust" is referred to this conceptual model and it is interchangeably used with the term "perceived trustworthiness".

According to the proposed conceptual trust model, in order to quantify trust, it is necessary to investigate trustee's trustworthiness by specifying TAs associated with it. As mentioned above, trustworthiness is as a composite of a variety of TAs that illustrate different characteristics of the trustee. Despite a large number of TAs have been figured out in trust-related literature, TAs are mostly fallen into three categories as the three main dimensions of trustworthiness: Ability, Benevolence and Integrity. This classification is appropriate for consideration of trustworthiness in the IoT environment.

- Ability is a dimension of trustworthiness showing the capability of a trustee to accomplish a trust goal. An entity may be high benevolent and integrity for fulfilling a trust goal but the results may not be satisfactory if it is not capable. This term incorporates some other terms that have been used as TAs in many trust-related literature such as competence, expertness, and credibility.

- Benevolence is a dimension of trustworthiness showing to what extent a trustee is willing to do good things or not harm the trustor. Benevolence ensures that the trustee will have good intentions toward the trustor. This term incorporates some TAs such as credibility, relevance, and assurance as TAs.

- Integrity is a dimension of trustworthiness showing the trustee adheres to a set of principles that helps the trustor believe that the trustee is not harmful and not betray what it has committed to do. These principles can come from various sources such as fairness, or morality. This term incorporates some TAs such as honesty, completeness, and consistency.

7.2 The Challenge

Trust management technologies have been widely investigated in many fields including economics, sociology, and computer science [1-3].

For example, trust management systems established on policies are presented in [4], and [5], based on reputation systems are explained in [6], [7],and [8], authors in [9] and [10] have developed a trust model corresponding to their owner's social behaviour, [11], [12], [13] and [14] present a computational model for trust based on concepts like a community of interest, friendship, followers, similarity, information reliability, and social opinions.

However, the influence of a trust attribute on trust in above methods are determined by weighting factors. But the assessment of a proper weightage is a complex task since the trust is a varying quantity which depends on many factors, e.g. expectations of a trustor, time, context, etc. Thus, schemes that are more intelligent are required to find these weighting factors and a threshold that defines a trustworthy boundary. Further, existing technologies have investigated limited number of trust features, like packet forwarding ratio, QoS, privacy, reputation, feedbacks and some social attributes. Further, existing technologies heavily rely on readily available knowledge for trust assessment and dynamic nature of the trust is almost ignored in most systems.

7.3 AI for Trust Solutions

Complex characteristics in Trust make it challenging for traditional analysis, but ideal for the application of AI, machine learning techniques and big data analytics. The objective of AI is to investigate the very large volumes of data produced by various components in the IoT ecosystem, and transform the data into meaningful outputs such as trust based decision making, fault detection, service composition and generate ultimate wisdom.

In this regard, authors in [15] and [16] outline the requirements for robust probabilistic trust assessment using supervised learning and apply a selection of estimators to a real-world dataset, in order to show the effectiveness of supervised methods.

Another interesting work that applies machine learning techniques is found in [17]; in this work, they propose to use neural networks in order to provide a global reputation model using the distributed reputation evaluations.

The global reputation is determined by the neural network's output unit, a two class classification in this case. A trust model with a broader scope, not only considering reputation was introduced in [18]; in this work, the authors propose the use of a Bayesian-Network trust model to properly interact with trustworthy peers. Furthermore, authors in [4], [19] and [20] investigate more innovative models and solutions for privacy, security and data integrity based on statistical and deep learning concepts. Moreover, authors in [21] and [22] propose a regression based model which compares the variation of trustworthiness with respect to trust features in mobile ad-hoc networks (MANET) and WSN. Recently, authors in [23], [24] and [25] present several trust management frameworks based on reinforcement learning and multiclass classification techniques.

On the other hand, authors in [26, 27] discuss a generic machine learning framework called MetaTrust for identifying relevant features to determine trust. Specifically, a trustor uses its own previous transactions (with other agents) to build a knowledge base, and utilizes this to assess trustworthiness of a potential transaction based on associated features, which are capable of distinguishing successful transactions from unsuccessful ones.

From a different perspective, Tang et al. [28] addressed the issue of initial trust assessment using the homophily effect. Homophily suggests that similar users are more likely to establish trust relations.

For instance, people with similar tastes about fiction movies tend to trust each other. This work employs low-rank matrix factorization to study trust relations.

In [29], a trust model for multi-agent systems is developed to help the agent make optimal trust decisions over time in a dynamic environment. The target agent's behaviour is predicted according to the Hidden Markov Model (HMM) trust estimation module following the Q-learning greedy policy. ElSalamouny et al. [30] modelled the real dynamic behaviour of an agent by HMMs. They further justified the consistency of the model by measuring the difference between real and estimated predictive probability distributions using relative entropy. The works [31] and [32] demonstrate how HMM-based trust models are applied to distinct application scenarios: routing protocol design in mobile and ad-hoc networks (MANET) and Web service providers selection.

Many trust schemes for multi-agent e-marketplaces have been proposed to deal with the unfair rating problem. The Beta Reputation System (BRS) [33] calculates seller reputation using a probabilistic model based on the beta probability density function, which can be used to represent probability distributions of binary events. The personalized approach proposed by authors in [34] combines buyer's personal experience and the public knowledge held by the system. Several other approaches have also been proposed to deal with unfair ratings. Dellarocas [35] proposed a clustering-based algorithm to separate the advisor's ratings into two clusters (one cluster including lower ratings and another cluster including higher ratings). The ratings in the higher cluster are considered as unfairly high ratings and are discarded. However, this approach cannot effectively handle unfairly low ratings. The iCLUB approach [36] adopts a clustering technique (DBSCAN) to filter out dishonest advisors based on local and global information.

On the other hand, Wang et al. [37] describe a super-agent-based framework for Web service selection, where service clients with more capabilities act as super-agents. These super-agents maintain reputation information of the service providers and share such information with other service clients that have less capabilities than the super-agents. Also, super-agents maintain communities and build a community-based reputation for a service provider based on the opinions from all community members (service clients in a community) that have similar interests and judgement criteria as the super-agents or the other community members. A reward mechanism is also introduced to create incentives for super-agents to contribute their resources (to maintain reputation and form communities) and provide truthful reputation information.

While most of the works on trust evaluation in service oriented computing (SOC) have focused on accurately predicting trust scores, Conner et al. [38] present a trust model that allows each service client (with different trust requirements) to use different scoring functions over the same feedback data for customized evaluations. Rather than assuming a single global trust metric as with many existing reputation systems, they allow each service client to use its own trust metrics to meet its local trust requirements. They also propose a novel scheme to cache the calculated trust values based on recent client activity.

There is a different perspective to consider, that is, trust for AI. As a classical security application of machine learning, intrusion detection followed various learning-based approaches, in particular, anomaly detection [39, 40], rule inference [41-43] and supervised learning [44, 45]. Although most of the proposed methods performed well in controlled experiments, the practical intrusion detection systems, such as Snort [46] and Bro [47], are still rooted in the more conservative signature-based approach. Sommer and Paxson discussed several practical difficulties faced by learning-based intrusion detection systems [48]. Due to the extreme versatility of web applications, it is next to impossible to devise signatures for specific attack patterns.

7.4 References

[1] F. Huang, "Building social trust: A human-capital approach," Journal of Institutional and Theoretical Economics JITE, vol. 163, no. 4, pp. 552-573, 2007.

[2] G. Möllering, "The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension," Sociology, vol. 35, no. 2, pp. 403-420, 2001.

[3] S. P. Marsh, "Formalising trust as a computational concept," Ph.D. dissertation Ph.D. dissertation, Dept. Computing Science and Mathematics, University of Stirling, Stirling, Scotland, UK., 1994.

[4] F. Fei, S. Li, H. Dai, C. Hu, W. Dou, and Q. Ni, "A K-Anonymity Based Schema for Location Privacy Preservation," IEEE Transactions on Sustainable Computing, vol. PP, no. 99, pp. 1-1, 2017.

[5] T. Jim, "SD3: a trust management system with certified evaluation," presented at the Proceedings 2001 IEEE Symposium on Security and Privacy S&P, 2001, 2001.

[6] M. A. Azer, S. M. El-Kassas, A. W. F. Hassan, and M. S. El-Soudani, "A survey on trust and reputation schemes in ad hoc networks," presented at the Third International Conference on Availability, Reliability and Security ARES 08, 2008.

[7] U. Jayasinghe, N. B. Truong, G. M. Lee, and T.-W. Um, "RpR: A Trust Computation Model for Social Internet of Things," presented at the Smart World Congress , Intl IEEE Conferences on Ubiquitous Intelligence & Computing, 2016.

[8] Y. Zhang, H. Chen, and Z. Wu, "A Social Network-Based Trust Model for the Semantic Web," presented at the International Conference on Autonomic and Trusted Computing, Berlin, Heidelberg, 2006. Available: http://dx.doi.org/10.1007/11839569_18

[9] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl, and H.-W. Gellersen, "Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts," presented at the Proceedings of the 3rd international conference on Ubiquitous Computing, Atlanta, Georgia, USA, 2001.

[10] L. Atzori, A. Iera, and G. Morabito, "From "smart objects" to "social objects": The next evolutionary step of the internet of things," IEEE Communications Magazine, vol. 52, no. 1, pp. 97-105, 2014.

[11] Y. Hu, D. Wang, H. Zhong, and F. Wu, "SocialTrust: Enabling long-term social cooperation in peer-to-peer services," Springer Peer-to-Peer Networking and Applications, journal article vol. 7, no. 4, pp. 525-538, 2014.

[12] M. Nitti, R. Girau, L. Atzori, A. Lera, and G. Morabito, "A Subjective Model for Trustworthiness Evaluation in the Social Internet of Things," presented at the IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC, Australia, 2013.

[13] J. Zhan and X. Fang, "A novel trust computing system for social networks," presented at the IEEE Third International Conference onPrivacy, Security, Risk and Trust (PASSAT) and IEEE Third Inernational Conference on Social Computing (SocialCom), 2011.

[14] G. Yin, F. Jiang, S. Cheng, X. Li, and X. He, "Autrust: A practical trust measurement for adjacent users in social networks," presented at the Second International Conference on Cloud and Green Computing (CGC), 2012.

[15] S. Hauke, S. Biedermann, M. Mühlhäuser, and D. Heider, "On the Application of Supervised Machine Learning to Trustworthiness Assessment," in 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 2013, pp. 525-534.

[16] K. Zhao and L. Pan, "A Machine Learning Based Trust Evaluation Framework for Online Social Networks," in 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, 2014, pp. 69-74.

[17] S. Weihua and V. V. Phoha, "Neural network-based reputation model in a distributed system," in Proceedings. IEEE International Conference on e-Commerce Technology, 2004. CEC 2004., 2004, pp. 321-324.

[18] Y. Wang and J. Vassileva, "Bayesian network-based trust model," in Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), 2003, pp. 372-378.

[19] F. Jiang et al., "Deep Learning based Multi-channel intelligent attack detection for Data Security," IEEE Transactions on Sustainable Computing, vol. PP, no. 99, pp. 1-1, 2018.

[20] J. Shen, D. Liu, D. He, X. Huang, and Y. Xiang, "Algebraic Signatures-based Data Integrity Auditing for Efficient Data Dynamics in Cloud Computing," IEEE Transactions on Sustainable Computing, 2017.

[21] Y. Wang, Y.-C. Lu, I.-R. Chen, J.-H. Cho, A. Swami, and C.-T. Lu, "LogitTrust: A Logit Regression-based Trust Model for Mobile Ad Hoc Networks," presented at the Proceedings of the 6th ASE International Conference on Privacy, Security, Risk and Trust Cambridge, MA, 2014.

[22] Z. Li, X. Li, V. Narasimhan, A. Nayak, and I. Stojmenovic, "Autoregression Models for Trust Management in Wireless Ad Hoc Networks," presented at the IEEE Global Telecommunications Conference (GLOBECOM), Kathmandu, Nepal, 5-9 Dec. 2011, 2011.

[23] F. Boustanifar and Z. Movahedi, "A Trust-Based Offloading for Mobile M2M Communications," presented at the Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Toulouse, France, 2016.

[24] W. Li, W. Meng, L.-F. Kwok, and H. Horace, "Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model," Journal of Network and Computer Applications, vol. 77, pp. 135-145, 2017.

[25] A. Bolster and A. Marshall, "Analytical metric weight generation for multi-domain trust in autonomous underwater MANETs," presented at the IEEE Third Underwater Communications and Networking Conference (UComms), Lerici, Italy, 2016.

[26] L. Xin, G. Tredan, and A. Datta, "Metatrust: Discriminant analysis of local information for global trust assessment," in The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3, 2011, pp. 1071-1072: International Foundation for Autonomous Agents and Multiagent Systems.

[27] X. Liu, G. Tredan, and A. Datta, "A GENERIC TRUST FRAMEWORK FOR LARGE-SCALE OPEN SYSTEMS USING MACHINE LEARNING," Computational Intelligence, vol. 30, no. 4, pp. 700-721, 2014.

[28] J. Tang, H. Gao, X. Hu, and H. Liu, "Exploiting homophily effect for trust prediction," in Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 53-62: ACM.

[29] M. E. G. Moe, M. Tavakolifard, and S. J. Knapskog, "Learning trust in dynamic multiagent environments using HMMs," in Proceedings of the 13th Nordic Workshop on Secure IT Systems (NordSec 2008), 2008.

[30] E. ElSalamouny, V. Sassone, and M. Nielsen, "HMM-based trust model," in International Workshop on Formal Aspects in Security and Trust, 2009, pp. 21-35: Springer.

[31] Z. Malik, I. Akbar, and A. Bouguettaya, "Web services reputation assessment using a hidden markov model," in Service-Oriented Computing: Springer, 2009, pp. 576-591.

[32] M. E. Moe, B. E. Helvik, and S. J. Knapskog, "TSR: Trust-based secure MANET routing using HMMs," in Proceedings of the 4th ACM symposium on QoS and security for wireless and mobile networks, 2008, pp. 83-90: ACM.

[33] A. Josang and R. Ismail, "The beta reputation system," in Proceedings of the 15th bled electronic commerce conference, 2002, vol. 5, pp. 2502-2511.

[34] J. Zhang and R. Cohen, "Evaluating the trustworthiness of advice about seller agents in e-marketplaces: A personalized approach," Electronic Commerce Research and Applications, vol. 7, no. 3, pp. 330-340, 2008.

[35] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in Proceedings of the 2nd ACM conference on Electronic commerce, 2000, pp. 150-157: ACM.

[36] S. Liu, J. Zhang, C. Miao, Y.-L. Theng, and A. C. Kot, "iCLUB: An integrated clustering-based approach to improve the robustness of reputation systems," in The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3, 2011, pp. 1151-1152: International Foundation for Autonomous Agents and Multiagent Systems.

[37] Y. Wang, J. Zhang, and J. Vassileva, "A SUPER-AGENT-BASED FRAMEWORK FOR REPUTATION MANAGEMENT AND COMMUNITY FORMATION IN DECENTRALIZED SYSTEMS," Computational Intelligence, vol. 30, no. 4, pp. 722-751, 2014.

[38] W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou, and K. Nahrstedt, "A trust management framework for service-oriented environments," in Proceedings of the 18th international conference on World wide web, 2009, pp. 891-900: ACM.

[39] P. Laskov, C. Schäfer, I. Kotenko, and K.-R. Müller, "Intrusion detection in unlabeled data with quarter-sphere support vector machines," Praxis der Informationsverarbeitung und Kommunikation, vol. 27, no. 4, pp. 228-236, 2004.

[40] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection," in Proceedings of the 2003 SIAM International Conference on Data Mining, 2003, pp. 25-36: SIAM.

[41] W. Lee and S. J. Stolfo, "Data Mining Approaches for Intrusion Detection," in USENIX Security Symposium, 1998, pp. 79-93: San Antonio, TX.

[42] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," ACM transactions on Information and system security (TiSSEC), vol. 3, no. 4, pp. 227-261, 2000.

[43] M. V. Mahoney and P. K. Chan, "Learning nonstationary models of normal network traffic for detecting novel attacks," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 376-385: ACM.

[44] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on, 2002, vol. 2, pp. 1702-1707: IEEE.

[45] S. Mukkamala, A. H. Sung, and A. Abraham, "Intrusion detection using ensemble of soft computing paradigms," in Intelligent systems design and applications: Springer, 2003, pp. 239-248.

[46] M. Roesch, "Snort: Lightweight intrusion detection for networks," in Lisa, 1999, vol. 99, no. 1, pp. 229-238.

[47] V. Paxson, "Bro: a system for detecting network intruders in real-time," Computer networks, vol. 31, no. 23-24, pp. 2435-2463, 1999.

[48] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in Security and Privacy (SP), 2010 IEEE Symposium on, 2010, pp. 305-316: IEEE.

8. Case Study – Securing IoT based Applications

This section identifies various use cases for securing IoT based applications and investigate related case study or good practices from industries and academia for AI enabled applications.

The secure cyber-physical transportation system (CPTS) is an important case on the security of cyber physical systems, which is for secure IoT environment, and it is important to avoid traffic congestion and traffic accidents. Figure 15 show an example of

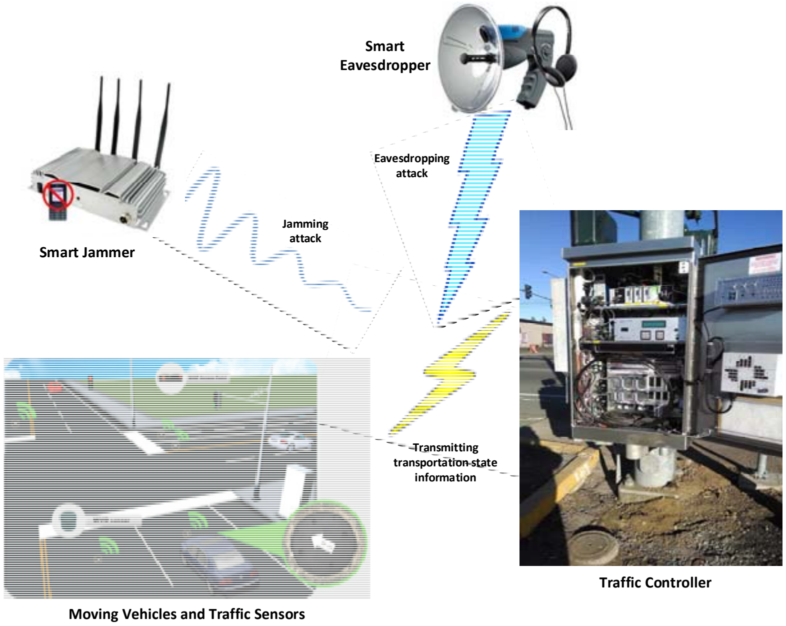

Figure 15. An example of cyber-physical transportation system.

CPTS combine the capabilities of sensing, control, communication and computing together, and in such systems, the sensors send the sensing information to traffic controllers through the open and wireless communication media. The open property of wireless communications makes it easy to be under the threat of security attacks, e.g., eavesdropping and jamming, as shown in Figure 15, and these security attacks will cause serious traffic accidents to result in deaths and economic loss. More importantly, the eavesdroppers and jammers are smart. They can adjust the power strategy to maximize the negative effects on the wireless communications, according to the feature of information transmission power. It is a challenge to actively extract and learn the features of malicious attacks, e.g., the above-mentioned feature: power strategy, which is dynamically changed with the change of the information transmission power. The wireless communications between sensors and controllers are open-mode. The open property makes the wireless communications easy to be under the threat of eavesdropping and jamming attacks. Moreover, the eavesdropper and jammer are smart in adjusting their power strategy to maximize the negative effects on the wireless communications.

Business continuity management (BCM) ensures continuation of an organization's business processes by utilizing data collected from the organization's IT and operation technology (OT) systems. The objective of BCM is to provide an advanced risk assessment processed from collected IT and OT data and to implement necessary measures to mitigate the impact to the organization's business processes, see Figure 16.

The BCM IoT platform will gather incident information from various security systems (i.e. IT systems) as well as planned and actual production data from production control systems (i.e. OT systems) utilizing sensor fusion technology. The BCM IoT platform will import threat intelligence information from other organizations to acquire insights into the situations of other interdependent systems and cohesive knowledge of the current and future attacks. The platform will analyse the incident information and perform risk analysis of the incident. It will also create security measures such as risk mitigation plans that will minimize the effect to the production activities. The BCM IoT platform will implement security measures, such as isolation of the affected subsystems or interruption of production lines. The IoT platform will analyse the production data to create an optimal production plan in response to affected capabilities of each production site.

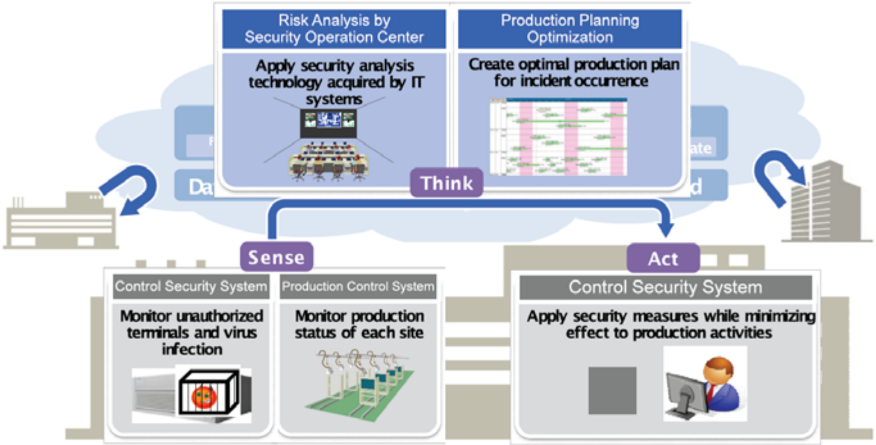
Risk Analysis by Security Operation Center
Apply security analysis technology acquired by IT systems
Production Planning Optimization
Create optimal production plan for incident occurrence
Think
Sense
Act
Control Security System
Monitor unauthorized terminals and virus infection
Production Control System
Monitor production status of each site
Control Security System
Apply security measures while minimizing effect to production activities

Figure 16. Industrial domain use case[66].

Smart City solutions bring together a number of heterogeneous IoT platforms that collectively contain a wide variety of sensors and data sources. These include temperature, humidity, noise, gas, and motion sensors, cameras, mobile devices, network sniffers, smart meters, water meters and a plethora of others devices that collectively monitor the dynamics of a city and optimize city operations while also enhancing citizen services. The Smart City IoT platform will transform the multi-modal sensing information from these various IoT platforms into cross-domain and real-time information mashups, using semantic interoperability. Advanced data mining and machine learning techniques will easily access a variety of different platforms and their operating environments to provide applications for residents and multiple agencies and enable intelligent actions. The Smart City platform of platforms will include real-time applications to enhance public safety, improve city mobility, optimize utility usage and enhance the plethora of citizen services that involve physical objects. Smart City platforms and the platform of platforms will rely heavily on smart and secure sensing to optimize services. They will use a combination of data from the public, private and personal sectors (anonymized as necessary) to seamlessly gain a more holistic view of the Smart City environment. They will use cross-domain communication techniques to bring together the disparate IoT systems deployed by individual agencies in a geopolitical entity and enable cross-domain cooperation and optimization, see Figure 17.

The Smart City platforms and platform of platforms will rely heavily on semantic disambiguation and contextualization of information to support advanced data processing and next-generation analytics that are developed for, and focused exclusively on, optimization of citizen services. These platforms

[66] IEC, "IoT 2020: smart and secure IoT platform," White Paper 2016.

will leverage advanced connectivity such as 5G and in-memory databases to move and process the vast amounts of data generated by the plethora of devices within their geopolitical boundary. The platforms will support edge-aware stream processing to handle processing topologies on parallel-networked systems. The Smart City platform of platforms and individual platforms will use next-generation technology to create smarter and more secure citizen quality of life environments.

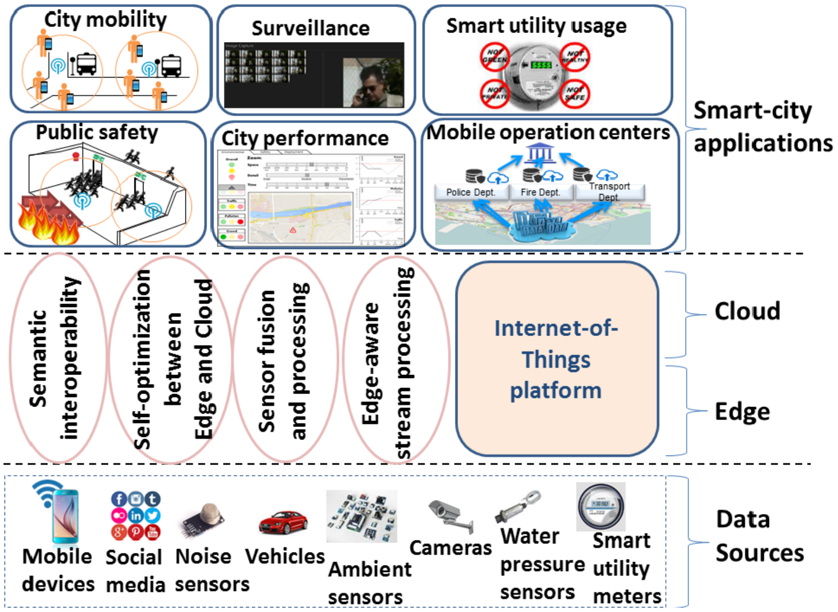

Figure 17. Smart city optimization[67].

[67] IEC, "IoT 2020: smart and secure IoT platform," White Paper 2016.

9. Challenges for Global Standardization

This section identifies the necessity of global standardization considering big technical trends towards the use of AI in various IoT applications while ensuring security and privacy. Then, it clearly identifies important standardization items and challenges to stimulate related activities in related standardization bodies.

For future standardization, it's necessary to investigate potential topics. Autonomous IoT with big data and AI and cyber security including trust with bloc-chain are very important topics. For IoT standardization, global interoperability for cross-border applications will be always important in support of various standardization bodies and fora in different sectors for vertical application domains and telecommunications. Furthermore, open source projects and alliances will be of interest. The most important technological trends is technology convergence among 5G for network connectivity, computing resources and platforms for Big data and AI while supporting security, privacy and trust including governance. It will ultimately target trustworthy autonomous IoT applications for supporting increasing intelligence like human brain.

As driving force for changes, there are three keyword: Data, Network and AI – i.e., DNA as shown in Figure 18. The important thing is to link between data and AI considering data from IoT and diffusion of AI. There are very important roles to support data-driven networking and services. Ultimately, standardization of trustworthy data-driven ICT combining data and AI. In fact, there are common features to support these technical requirements. However, rather than unlimited number of solutions, it is required to standard critical technologies globally with a fermentation and assembling approach.

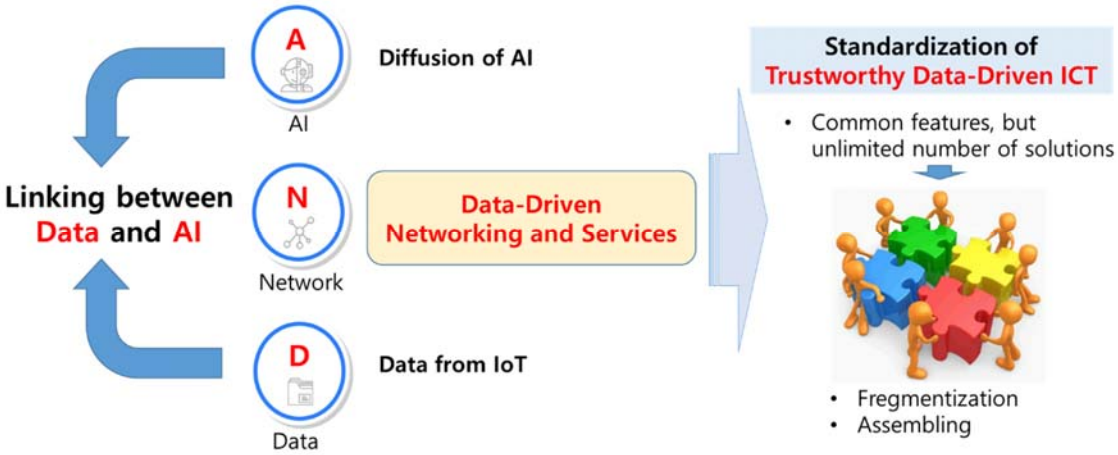

Figure 18. Future standardization towards trustworthy data-driven ICT.

Considering AI and IoT with security, privacy and trust, there are many potential items for future standardization as follows.

- DNA core technologies
  - A new networking paradigm – Data-driven networking
  - DNA platform – Technology convergence (IoT+Big Data+Cloud+AI)
  - Data-Information-Knowledge-Wisdom (DIKW) process
- Use of AI in ICT infrastructures and services (trustworthy autonomous ICT)

- Automotive control and management in networking and services
- Operational efficiency in Things + Processing + Communications + Storage
- Data-driven applications with AI (linking between data and AI)

- Security, privacy and trust including regulatory issues
  - Trust in DNA, particularly human-technology interface including social aspects

For trust technology, ITU-T Q16/13 made a preliminary version of roadmap for future trust standardization as shown in Figure 19. Based on the Data-Information-Knowledge concept, at the 1st stage, basic issues and key features on trust have been focused. Then, core technical solutions for trust provisioning from ICT infrastructures and services perspectives at the 2nd stage should be standardized. Finally, technology deployment as well as new services and business aspects on trust-based networks and eco-platforms are necessary. For this, there will be many related issues with other groups in ITU-T.

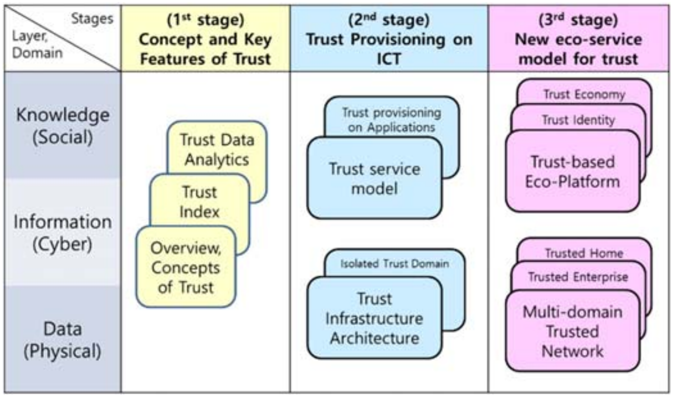
Stages
Layer, Domain
(1st stage) Concept and Key Features of Trust
(2nd stage) Trust Provisioning on ICT
(3rd stage) New eco-service model for trust
Knowledge (Social)
Information (Cyber)
Data (Physical)
Trust Data Analytics
Trust Index
Overview, Concepts of Trust
Trust provisioning on Applications
Trust service model
Isolated Trust Domain
Trust Infrastructure Architecture
Trust Economy
Trust Identity
Trust-based Eco-Platform
Trusted Home
Trusted Enterprise
Multi-domain Trusted Network

Figure 19. Roadmap for future trust standardization.

The social and economic value of data is mainly reaped during two moments: first when data is transformed into knowledge (gaining insights) and then when it is used for decision making (taking action). The knowledge is accumulated by individuals or systems through data analytics over time. So far data processing, management and interpretation for awareness and understanding have been considered as fundamental processes for obtaining the knowledge. As shown in Figure 20, trust is positioned throughout the whole process including belief between knowledge (i.e., awareness and understanding) and action.

- Trusted data collection and aggregation - Since an ICT convergence ecosystem is widely deployed and the number of data sources and types are dramatically increased, the trustworthiness of data itself comes to the fore. Data collection and aggregation should be trustworthy. However, as the vulnerability of the data sources, it is important to detect wrong data caused by various reasons. Since collection and aggregation of false data will cause damage and waste of system resources, trust metrics and models, which are able to use as criteria for checking trustworthiness, should be well managed to achieve trusted data collection and aggregation.

- Trusted data process and analysis

  Data processing means data manipulation such as filtering, fusion or mining. When the huge amounts of data are collected to a system, these data should be processed and analyzed in trustworthy ways. Data process and analysis mainly occurs in cyber domain (for example, utilizing cloud computing for Big-data analysis), however, it also can be done in physical domain as well as social domain. Each domain has their own intelligence to process incoming data to create new useful information. This information is usually propagated to different entities and domains, so there are some ways to check whether given data process and analysis mechanism is trustworthy or not. Measurable trust value should be defined to analyze trust of entities, and it is also important to find appropriate trust evaluation mechanisms for analyzing trust values for a specific domain. Trust bootstrapping is also an important issue. When a new entity is joined to an existing system, the system is able to score trust values for a newcomer.

- Trustworthy decision-making, action and data dissemination - ach entity makes their decisions, takes necessary actions and disseminates data based on data process and analysis. Also, each domain has different criteria for making a decision and taking an action, so different trust mechanism should be considered on each step. The result of decision-making or action can be disseminated among physical, cyber and social domains. In the dissemination stage, the feature of data should be considered; for example, some data can be related with user privacy, or they can have urgent information. These kinds of data characteristics should be considered for trustworthy data dissemination.

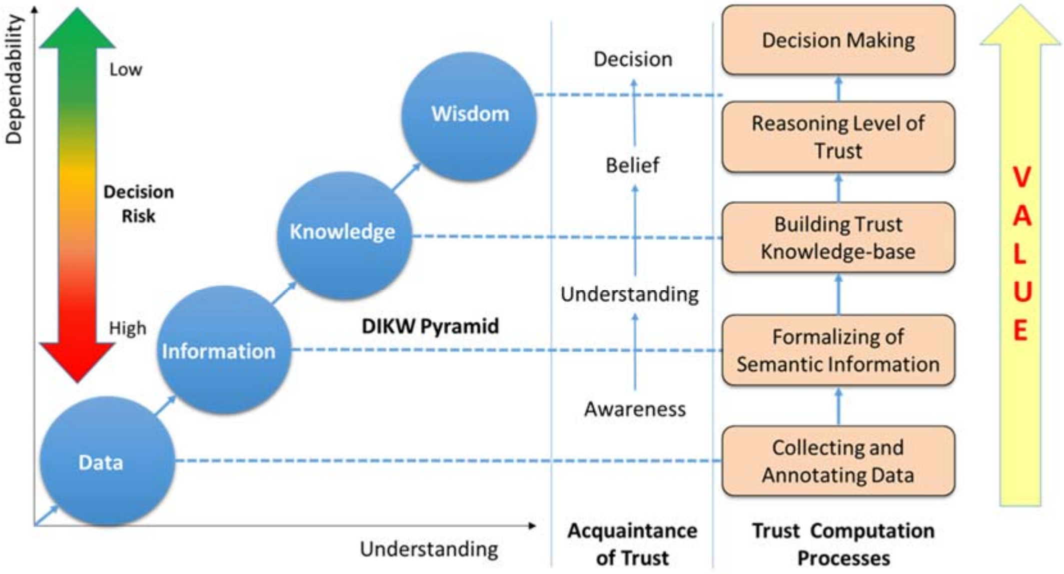

Figure 20. Trusted computing from raw data to actionable knowledge.

Recently decentralization is one of the emerging technologies that implements proof of trust using block-chains but applying it to overall holistic network security is yet to be studied thoroughly. Assuming a trustworthy model is in place, several functional benefits become apparent. For example, domain-based trust can be represented as a function of time and credibility score from number of other domains. This is computed by verification of correctness and that data remained unimpaired while transiting a domain. New routing paradigms (and corresponding business incentives) could emerge as now domains may choose to transit through systems that are better in trust-scores. A side effect of such approach will be finding a balance of cost and security within the network[68]. As one of

[68] Huawei, “Internet 2030, Towards a new Internet for the Year 2020 and beyond,” 2018

tools to support trust, block-chain should be considered as an important standardization item taking into account the 4th industrial revolution.

10. Considerations for promoting safe use of IoT based application with AI

This section presents required actions to promote safe use of IoT based application with AI. It presents considerations for data protection, ensuring privacy, ethical, policy and regulatory issues.

10.1 Data protection, privacy and ethical considerations

The key characteristics of big data analytics still represent a step change in the processing of personal data. The analysis of big data using techniques made possible by AI creates implications for data protection, and it can be more challenging to apply the data protection principles when using personal data in a big data context. These implications arise not only from the volume of the data but from the ways in which it is generated, the propensity to find new uses for it, the complexity of the processing and the possibility of unexpected consequences for individuals.

Article 8 of the EU's Charter of Fundamental Rights enshrines data protection as a fundamental right, which reinforces the necessity to ensure the implementation of high standards for data protection, privacy and information security.

ITU-T Focus Group on Data Processing and Management (FG-DPM) is developing deliverables on data security, privacy and trust issues including governance. The deliverable on framework of security and privacy presents key considerations of data protection[69]. Data protection within the context of smart cities can be seen from different points of view. First of all, one has to bear in mind that when we look at smart cities they are subject to different legislations in many parts of the world and therefore their regulation is not uniform.

If we take into account the European situation, starting from an user-centred point of view, and taking into account the GDPR, we can derive important principles of significant importance for smart cities[70]. The GDPR enshrines a set of fundamental principles and norms that are always to be taken into account in the context of smart cities. Among them: lawfulness; fairness; transparency; purpose limitation; Data minimisation; accuracy; storage limitation; integrity and accountability. The GDPR also provides detailed norms for the collection of consent. The GDPR is more prescriptive when it comes to the conditions for consent, however the new rules transpose into law what was already required by certain supervisory authorities. According to article 4(11) of the GDPR consent means any "freely given, specific, informed, and unambiguous indication of the data subject's wishes by which he or she, by statement or by clear affirmative action, signifies agreement to the processing of personal data relating to him or her". The GDPR also details the requirement for the processing of personal data of underage persons and processing of special categories of data. The GDPR sets out obligations provides for obligations towards the facilitation of the exercise of the data subject's right to information such as access to personal data, rectification and erasure, right to data portability. The legal provisions also enable the data subject to restrict processing of his data under certain circumstances and detail processes for objection and seeks to protect the individual vis-à-vis automated decision making mechanisms.

The GDPR will require i) the use of Privacy by design (PbD), privacy by default and the use of the Privacy Impact Assessment (PIA) in the design and management of ICT solutions using personal data, and ii) the appointment of data protection officers (DPO).

Many researchers have argued that, by way of an "intelligence explosion" sometime in the 21st century, a self-improving AI could become so vastly more powerful than humans that we would not be able to stop it from achieving its goals. General super-intelligence would be capable of independent initiative and of making its own plans, and may therefore be more appropriately

[69] ITU-T FG-DPM, "Draft Technical Report – Framework of security and privacy in DPM," DPM-O-067.

[70] For a general scenario of legal framework see [Create-IoT-17].

thought of as an autonomous agent. Since artificial intellects need not share our human motivational tendencies, it would be up to the designers of the super-intelligence to specify its original motivations. In theory, a super-intelligent AI would be able to bring about almost any possible outcome and to thwart any attempt to prevent the implementation of its top goal, many uncontrolled unintended consequences could arise. It could kill off all other agents, persuade them to change their behaviour, or block their attempts at interference[71].

10.2 Technical considerations and challenges

There are several tools and approaches, including anonymization, PIAs and privacy by design, that can help organisations to ensure their processing complies with data protection legislation and minimises the impact on privacy. It needs the trend towards organisations developing their own ethical principles and building relationships of trust with the public, because putting this into practice will assist compliance with data protection requirements. Recent moves towards setting up 'councils of ethics', within organisations and nationally, are a positive development that should also support this.

10.2.1 Privacy by Design and Privacy by Default

Usually it is not the technology that increases the risks for privacy, data protection and security, but the way it is developed and applied. The negative consequences of this practice for privacy, data protection and security will significantly increase, if applied to IoT systems. Therefore mechanisms are needed to ensure that no unwanted processing of personal data takes place and that individuals are informed of the processing, its purposes, the identity of the processor and of how to exercise their rights. At the same time processors need to comply with the data protection principles as data minimisation, purpose limitation etc., which might be especially challenging in an IoT environment, where automatic communication without human intervention between objects and between objects and persons is at the core of the system.

Definition - Privacy-by-Default

The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons (art. 25.2 in GDPR).

Definition - Privacy-by-Design

Institutionalisation of the concepts of privacy in organisations and integration of these concepts in the design, and life cycle of systems.

The GDPR defines it at a methodology according to which: "Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymization, which are designed to implement data-protection principles, such as data minimisation in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and to protect the rights of data subjects (art. 25.1 in GDPR).

[71] https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence

On a technical level an appropriate process to protect personal data should be defined. For example, privacy policies (at the user level) that can be pushed/built inside the objects (and translated according to the technical specificities of each object) and finally be enforced by the object with appropriate mechanisms to ensure data protection. The technical challenge is however, to enable objects with limited processing power and / or memory to receive and respect such policies.

The possibilities for individuals to exercise their data subject rights need to be enhanced. It needs to be ensured that clear, easily understandable information on the data processing of IoT systems, their objects, functions and purposes, is provided to individuals. Mechanisms need to be found to make individuals aware of the processing and to provide information on the processor, the purpose of the processing and possibilities to exercise data subject rights, as most IoT applications are expected to operate in the background, invisible to and unrecognised by the individual.

Information on how to build privacy-friendly applications needs to be provided to IT engineers, system designers and standardisation bodies, to ensure that the concepts of Privacy by design and Privacy by default settings get implemented in practice. Data protection officers might have an important role here, provided that defined minimum requirements ensure that they themselves are sufficiently trained.

Challenges addressed with this option include i) design considerations for IoT technologies, ii) the risks are context-aware and situational, iii) traceability / profiling / unlawful processing, iv) exercising data protection rights for individuals and compliance with DP legislation for organisations, and v) loss of user control / difficulty in making decisions.

10.2.2 Human/technology interface considering ethics

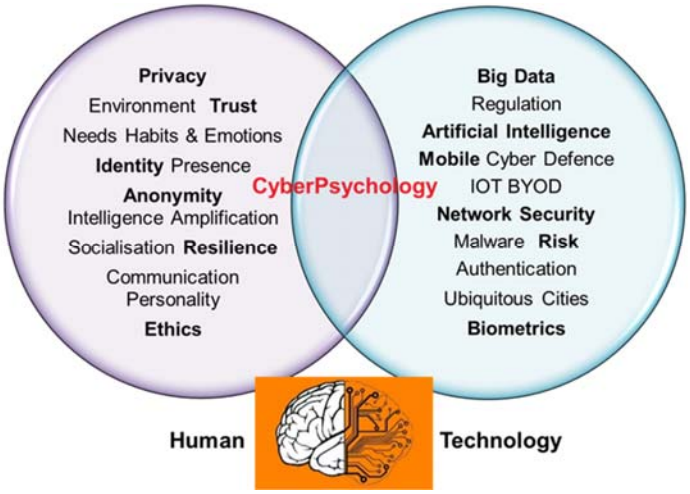

Figure 21. Deliberating insight at the human/technology interface[72].

Beyond conventional security solutions, it's very important to consider transparency, data protection, privacy preserving, policy and regulatory issues including ethics.

The biggest threat to network security is company employees and firms need to adapt management structures to deal with leaderless environments says a leader in sphere of what is now known as cyber-

[72] Source: Mary Aiken, "The CyberPsychology of Cyber Security" RCSI

psychology (see Figure 21), a world with no authority for authority, leaving it very difficult to establish robust controls. Worse still, this was a world where people assume that they do not need leaders and everything should be done by consent. And one in which technology was evolving much fast than the required corporate psychology required to cope best with this change, meaning that there was tension between traditional and leaderless corporations.

An ethics point of view, the sheer complexity of human value systems makes it very difficult to make AI's motivations human-friendly. Unless moral philosophy provides us with a flawless ethical theory, an AI's utility function could allow for many potentially harmful scenarios that conform to a given ethical framework but not "common sense". It needs an AI design that avoids several types of unintended AI behaviour including self-delusion, unintended instrumental actions, and corruption of the reward generator.

10.2.3 Policy and regulatory considerations

Main goals of any regulatory approach should therefore be i) to ensure full compliance of IoT technology and applications with the Charter of Fundamental Rights and ii) to minimise potential barriers for the adoption of IoT technology in order to benefit of it's full economic potential.

Lack of trust makes consumers hesitate to buy online and adopt new services, including public e-government services. If not addressed, this lack of confidence will continue to slow down the development of innovative uses of new technologies, to act as an obstacle to economic growth and to block the public sector from reaping the potential benefits of digitisation of its services, e.g. in more efficient and less resource intensive provisions of services. This is why data protection plays a central role in the Digital Agenda for Europe, and more generally in the Europe 2020 Strategy.

Standards could provide presumption of conformity with the legal requirements and could be used for certification; in addition, they could provide definitions for clear information to individuals on how to exercise their data protection rights, mindful that clear information is a prerequisite for informed consent.

On the other hand standards are voluntary and non-binding. Therefore the tool might possibly be too “weak” for the intended outcome. Other regulatory measures which are more binding might also be needed. A recommendation could be envisaged and comitology procedures could be launched to build a European privacy risk management tool (methodology + best practices).

According to the GDPR, which is probably the most stringent legislation at the global level, many cities will need to comply with legal accountability obligations to European Data protection law. As data controllers, cities will be required to implement appropriate technical and organizational measures to ensure and be able to demonstrate that data processing is performed in accordance with the GDPR, and review and update those measures when necessary. In each case cities will be called to evaluate which measures will be appropriate. This will depend on the nature, scope, context and purpose of the processing and also the risks for rights and freedoms of individuals. Regulators around the word are also embracing the concept of accountability as a key principle in data process management (privacy regulators in Canada, Hong Kong, Australia have issued “Accountability Guides” or “Privacy Governance Frameworks” in order to assist the private sector).

Basic frameworks to deal with IoT solutions, also in the context of smart cities, have been recently elaborated. Hyper-connectivity in fact entails a great deal of risks and calls for the development of a proper legal framework. The regulatory ecosystem of course will vary according to the different kind of activities in place. For what may concern data protection and privacy the Article 29 Working party has specifically raised the issue of privacy and security issues raised by the IoT.

10.2.4 Risk management

Data processing and management in Smart Cities and IoT generate uncertainty and may constitute a threat for the stability of the system, the efficiency of the operations, or the wealth, the health or the dignity of the persons. Thus, it is the responsibility of the ecosystem to measure the risks linked to data, and to manage it.

Different methodologies/frameworks already exist to manage the risks. We can cite for example Coso Enterprise Risk Management[73] (Committee of Sponsoring Organizations of the Treadway Commission). We can also cite the Operational Risk Framework as sponsored by the Basel Committee on Banking Supervision[74]. Even if Coso emphasizes on Governance and Communication, while Basel standard emphasizes on calculation accuracy and the losses modelization, both methodologies have the same canvas to manage the risks linked to data processing and management in Smart Cities and IoT[75].

[73] https://www.coso.org/Documents/2017-COSO-ERM-Integrating-with-Strategy-and-Performance-Executive-Summary.pdf

[74] https://www.bis.org/publ/bcbs195.pdf

[75] ITU-T FG-DPM, “Draft technical specification on risk management in DPM for IoT and smart cities,” FG-DPM-O-64.

11.Conclusion

This report has examined the relevance of AI in the current and future development of IoT and how security should be addressed, including data protection and privacy as follows.

- It has provided key technical trends for digital transformation from literature and highlight the important of AI and IoT technology. It has also addressed potential risks and threats while the number of new technologies continues to grow.
- It has provided basic concepts and background of AI and Machine Learning including general application areas of AI.
- It has discussed IoT and security issues, including a security framework in CPS in order to understand the overall features and technical issues for secure IoT environment.
- It has surveyed security in IoT and AI, AI-based privacy mechanism for personal data, and AI-based trust mechanisms.
- It has identified the necessity of global standardization considering big technical trends towards the use of AI in various IoT applications while ensuring security and privacy. Then, it has also identified important standardization items and challenges to stimulate related activities in related standardization bodies.
- It has presented required actions to promote safe use of IoT based application with AI. It presents considerations for data protection, ensuring privacy, ethical, policy and regulatory issues.

In the area of security, privacy and trust research in IoT, AI plays an important role. Recently, the increasing amount of data as well as the rich metadata brought by large-scale Web applications (e.g., social media, e-commerce, recommender systems) has led to a new trend of applying formerly unutilized machine learning methodologies, such as deep learning, to more precisely model trust. In this report, we have discussed potential benefit of integrating machine learning and other AI technologies with trust and security concepts and its crucial role in data-driven applications, service composition, social networking, recommendation systems and security aspects. Further, it is observed that there is an emerging trend in scientific community in developing many research activities at the junction of trust and AI techniques. To understand the prevailing issues related to trust, and provide more smart solutions, it is essential to bring researchers in these two communities closer to each other.

In the context of the IoT, standards should be elaborated specifically on the aspects of good application design, user application interfaces and on tools for individuals to play their part in security, data protection and privacy for the IoT for people. In order to make Privacy by design and Privacy by default a reality, the consideration of data protection requirements should become a mandatory design goal in standardisation, as standards can serve as a multiplier for privacy friendly application design.