A

A Securing Routing Scheme for Vehicular Networks with Cognitive Radios

▸ Secure Routing in Cognitive Radio Vehicular Ad Hoc Networks

Access Control

Kan Yang
Department of Computer Science, The University of Memphis, Memphis, TN, USA

Key Applications

Any data or system applications

Definition

In the field of information security, access control (AC) is the selective restriction of access to a resource (e.g., data, system, application, etc.).

Access Control Models

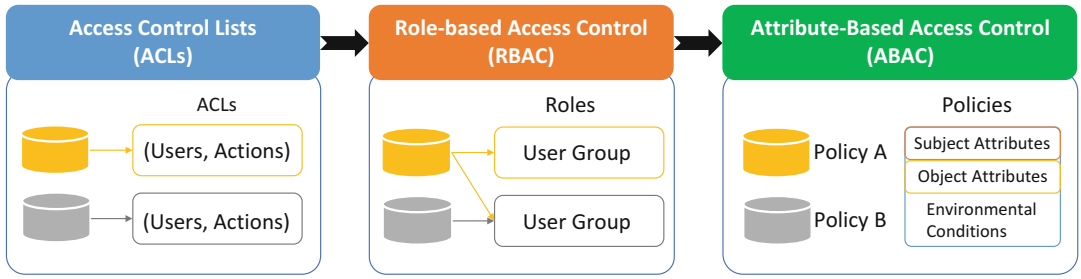
Access control is an effective approach to guarantee confidentiality and privacy of the resource. Figure 1 shows the history of access control models: access control lists, role-based access control, and attribute-based access control.

Access Control Lists

Access control lists (ACLs) are the most basic model of access control, where each data is associated a list of mappings between the set of entities and the set of actions that each entity can take on the data. For example, the home address can be accessed by operators (read only), markets (read only), and utility service providers (read, modify, and delete). Some grid status can only be accessed by power control centers. Whenever a user tries to perform any action (e.g., modify) on the data (e.g., home address), the system or server checks the ACL of this data and determines whether to allow this action for that user. The ACLs can also be applied for a group of data (e.g., account information: name, home address, contact information), which may have the same mappings between actions and entities.

The major limitation of ACL model is that every user is treated as a distinct entity with distinct sets of privileges for each data. That means ACLs have to be defined separately for each data (or group of data), which would be a cumbersome process when many users have different levels of access permissions to a large amount of data. It is time-consuming and error-prone to selectively add, delete, and change ACLs on individual data or even group of data.

X.(S.) Shen et al. (eds.), *Encyclopedia of Wireless Networks*,
https://doi.org/10.1007/978-3-319-78262-1



Access Control, Fig. 1 Access control models (Yang et al. 2016)

Role-Based Access Control

Role-based access control (RBAC) (Ferraiolo et al. 2001; Sandhu et al. 1996) determines whether access will be granted or denied based on user's role or function. Since many peoples may have the same role (e.g., market researcher), RBAC may group them into one category with a particular role, which means that the access control permission on a particular data can be set only once for all members in the group with the same role. Users can also be members of multiple groups which may have different access permissions, where more restrictive permissions override general permissions in RBAC.

Although RBAC has many advantages compared with ACL model, it still suffers from some limitations. One of the most important is that it is difficult to achieve the fine-grained access control for each user when grouping them into categories based on roles. So, how to differentiate individual members of a group is a challenging problem in RBAC.

Attribute-Based Access Control

In order to define specific access permissions for individual users, attribute-based access control (ABAC) is proposed to control data access based on the attributes associated with the user. According to the definition of NIST (Hu et al. 2014), in ABAC, the access policy is defined based on both subject attributes and object attributes under some environmental conditions. ABAC is more flexible in defining access policies than RBAC.

A key advantage of ABAC model is that no specific users are needed to be known in advance when defining access policies. The access can be granted as long as the attributes of users can satisfy access policies. Thus, ABAC is useful for the applications where organizations or data owners allow unanticipated users to be able to access the data if their attributes can satisfy some policies. This feature makes ABAC well suitable for large enterprises. An ABAC system can implement existing role-based access control policies and can support a migration from role-based to a more granular access policy based on many different attributes of individual users. Gartner predicts that "By 2020, 70% of all businesses will use ABAC as the dominant mechanism to protect critical assets, up from 5% today (in 2013)."

Encryption-Based Implementation

The ACLs, RBAC, and ABAC all require the system/server to evaluate access rules/policies and make access decisions. When the server is not fully trusted by data owners, e.g., cloud servers, how to protect data confidentiality and privacy becomes a challenging problem. In order to cope with this challenge, cryptographic techniques are applied to implement the abovementioned three access control models. Data owners encrypt files by using the symmetric encryption approach with content keys and then use every user's public key to encrypt the content keys. However, traditional public key encryption methods may pro-

duce many copies of ciphertexts for multiple users, which is not efficient in large-scale systems, e.g., smart grid. Moreover, the key management is also a complex issue in public key infrastructures (PKI).

To eliminate the need for distributing public keys (or maintaining a certificate directory), identity-based encryption (IBE) (Boneh and Franklin 2001) is a new cryptographic primitive that allows data owners to encrypt their data with identities of users instead of their public keys. To make access policies more flexible and expressive, attribute-based encryption (ABE) (Goyal et al. 2006; Bethencourt et al. 2007) is proposed for access control of encrypted data. There are two complimentary forms of ABE, namely, key-policy ABE (KP-ABE) (Goyal et al. 2006) and ciphertext-policy ABE (CP-ABE) (Bethencourt et al. 2007). In KP-ABE systems, keys are associated with access policies, and ciphertext is associated with a set of attributes, while in CP-ABE systems, keys are associated with a set of attributes, and ciphertext is associated with access policies.

Cross-References

▸ Cloud Computing

References

Bethencourt J, Sahai A, Waters B (2007) Ciphertext-policy attribute-based encryption. In: Proceedings of S&P'07, Washington, DC. IEEE Computer Society, pp 321–334

Boneh D, Franklin MK (2001) Identity-based encryption from the weil pairing. In: Proceedings of CRYPTO'01, London. Springer, pp 213–229

Ferraiolo DF, Sandhu R, Gavrila S, Richard Kuhn D, Chandramouli R (2001) Proposed NIST standard for role-based access control. ACM Trans Inf Syst Secur (TISSEC) 4(3):224–274

Goyal V, Pandey O, Sahai A, Waters B (2006) Attribute-based encryption for fine-grained access control of encrypted data. In: Proceedings of CCS'06, New York. ACM, pp 89–98

Hu VC, Ferraiolo D, Kuhn R, Schnitzer A, Sandlin K, Miller R, Scarfone K (2014) Guide to attribute based access control (abac) definition and considerations. NIST Spec Publ 800:162

Sandhu RS, Coyne EJ, Feinstein HL, Youman CE (1996) Role-based access control models. Computer 29(2):38–47

Yang K, Jia X and Shen X (2016) Privacy-preserving Data Access Control in the Smart Grid. Cyber Security for Industrial Control Systems: from the viewpoint of close-loop. Peng Cheng, Heng Zhang and Jiming Chen (Eds.), CRC

Acoustic Networks

▸ Opportunistic Routing in Underwater Sensor Networks

Acoustic Sensor Networks

▸ Opportunistic Routing in Underwater Sensor Networks

Acoustic Wireless Sensor Networks

▸ Opportunistic Routing in Underwater Sensor Networks

Ad Hoc Networks

▸ Capacity of Wireless Ad Hoc Networks

Adaptive Channel Access Control

▸ Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs

Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs

Qiang Ye[1,2] and Weihua Zhuang[1]
[1]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada
[2]Department of Electrical and Computer Engineering and Technology, Minnesota State University, Mankato, MN, USA

Synonyms

Adaptive channel access control; Contention-Based and Contention-Free Channel Access; Distributed transmission coordination; QoS-aware hybrid medium access control (MAC)

Definitions

Adaptive medium access control (MAC) for Internet-of-Things (IoT)-based ad hoc networking refers to a distributed (infrastructure-less) link-layer mechanism to coordinate channel access for data transmissions from a varying number of IoT devices, generating either delay-sensitive services or data-hungry applications. The adaptive MAC should achieve and maintain high performance for heterogeneous services with differentiated quality-of-service (QoS) requirements (e.g., delay and throughput) in network traffic load dynamics.

Background and Key Applications

Next-generation wireless networks are envisioned to interconnect a proliferation of Internet-of-Things (IoT) devices (e.g., smartphones, smart sensors and actuators, home appliances) to achieve seamless communication interaction and diversified service customization, such as smart cities, industrial automation, and intelligent transportation (Gubbi et al. 2013). To support an increasing number of end devices, current communication infrastructures are required to be boosted extensively with additional installation and operational cost. In the scenarios where the network infrastructures are temporarily not in place or not accessible (e.g., in hotspot areas or postdisaster areas), mobile ad hoc networking is an infrastructure-less and cost-effective networking technology to connect a large number of IoT devices via a device-to-device (D2D) communication mode (Nishiyama et al. 2015). In mobile ad hoc networks (MANETs), nodes are self-organized and initiate transmission requests over a common wireless channel (with 20MHz bandwidth at 2.4GHz or 5.9GHz spectrum frequencies) in a distributed way. To achieve consistently high link-layer performance (low packet transmission delay and high data throughput), an efficient medium access control (MAC) mechanism is required to coordinate channel access from a group of end devices by adapting to the variation of network traffic load (Natkaniec et al. 2013).

Without relying on centralized transmission coordination and device synchronization, the carrier sensing multiple access with collision avoidance (CSMA/CA)-based IEEE 802.11 distributed coordination function (DCF) (Bianchi 2000) is the most commonly used MAC scheme in existing MANETs, where packet transmissions from each node are initiated based on channel contention with an exponential backoff mechanism employed for collision avoidance. The contention-based DCF has the advantage of simplified implementation and high channel utilization in low traffic load conditions (i.e., the number of devices is low). However, its performance degrades substantially when the number of nodes becomes high, since a large portion of channel time is wasted in collided packet transmissions and collision resolution. Distributed time division multiple access (TDMA) schemes (Kanzaki et al. 2003; Wilson et al. 1993) eliminate transmission collisions by allocating time slot(s) exclusively for each device and thus achieve high resource utilization in high network load

conditions. However, the control information exchange among devices for distributed time slot acquisition makes the performance of TDMA inferior to DCF in a low network condition. Due to the performance trade-off between contention-based MAC and reservation-based time slot allocation, hybrid MAC schemes are proposed to combine the advantages of both types of MAC schemes by switching between contention-based and contention-free MAC frame structures, either periodically (Zhang et al. 2010, 2011) or adaptively based on instantaneous network load conditions (Hu et al. 2011). For most existing hybrid MAC schemes, the MAC switching decision is made upon measurement of some MAC parameters (e.g., number of idle TDMA slots (Ahmed 1997), number of lost acknowledgments (ACKs) (Hu et al. 2011), node buffer occupancy (Doerr et al. 2005)) which reflect the current network load condition. However, it is difficult to establish an analytical model between those measurement parameters and certain performance metrics (e.g., network throughput and packet delay), thus making the MAC switching decisions not optimal.

To satisfy differentiated quality-of-service (QoS) requirements from heterogeneous IoT services, the distributed MAC is expected to be both context-aware and service-aware (or QoS-aware) in a changing network environment. For example, delay-sensitive voice communications and remote control applications have stringent delay bound requirements on each transmitted packet so that the packets that are received beyond the delay bound are dropped, whereas the smart sensing applications collect data from a massive number of end devices which are throughput-oriented. Service prioritized contention schemes (e.g., busy tone-based MAC (Wang et al. 2007)) give guaranteed channel access opportunities to voice devices to relieve their contention collisions with data devices, which, on the other hand, suppress the data traffic channel access probability. Moreover, packet collisions among voice devices exist and are accumulated with the increase of voice device number. Existing distributed TDMA mechanisms can alleviate channel contention collisions but are not suitable for supporting heterogeneous services. Since most of the data-hungry sensing applications generate event-driven data traffic, the allocated time slots can be underutilized and need to be adaptive to traffic burstiness. To meet differentiated QoS demands, an adaptive and QoS-aware hybrid MAC scheme is required to differentiate the channel access between delay-sensitive applications and high data rate applications, where time slots are allocated to delay-sensitive traffic and data nodes occupy portions of channel time via contention to exploit traffic multiplexing gain (Zhang et al. 2010, 2011). With a varying heterogeneous network traffic load, how to adaptively allocate time slots to delay-sensitive applications and adjust the channel access parameters among data devices to achieve consistently bounded packet loss rate and maximum data throughput needs investigation.

Adaptive MAC Solutions

As stated precedingly, adaptive and hybrid MAC solutions are desired to coordinate packet transmissions among devices in an IoT-based MANET for achieving consistently maximum network performance, by adapting to a varying number of end devices and providing differentiated QoS guarantee. In the following, the technical steps to develop adaptive MAC solutions are presented for a homogeneous service scenario (i.e., support only high-rate data service) and for a heterogeneous service scenario (i.e., support both voice and data applications), respectively.

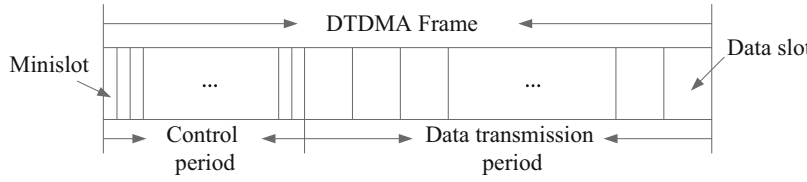
Homogeneous Service Scenario

Consider a fully connected MANET (i.e., each device is within the one-hop transmission range of any other device.) with a single and error-free wireless channel. Note that developing an adaptive MAC solution for a multi-hop MANET is considered in one of our recent works Ye and Zhuang (2017a). There is no central controller in the network, and devices coordinate their packet transmissions in a distributed manner.

Packet transmission failures are due to channel contention collisions. All devices in the network are assumed homogeneous, generating the same type of high-rate data application. Each device randomly chooses its destination among other devices and can send packets to and receive packets from other devices in the half-duplex mode. Every device has a unique device identifier (ID) that is randomly selected and included in its transmitted packet headers. Devices are synchronized in time by receiving an 1PPS signal periodically with a global positioning system (GPS) receiver. The total number of devices in the network is denoted by N, which varies slowly when devices move in or move out the network coverage area. Traffic arrivals at each device are assumed to follow a Poisson process with the rate parameter λ packet/s.

To establish an adaptive MAC framework in the presence of a varying number of devices, a hybrid MAC solution (Ye et al. 2016) is proposed to switch MAC frame structures between the CSMA/CA-based IEEE 802.11 DCF and a dynamic TDMA (DTDMA) scheme (Wilson et al. 1993) under different network load conditions. The DCF with exponential backoff for collision resolution achieves high data throughput when the number of devices is low, but experiences low channel utilization as the transmission collisions are accumulated in high network load conditions. For the DTDMA, time is divided into a sequence of frames, as shown in Fig. 1. Each frame consists of a control period and a data transmission period. The control period has a number of constant-duration minislots used for local information exchange to allocate time slots in the data transmission period to each device. The time slot allocation is conducted in a distributed way to fit the MANET scenario. The number of minislots indicates the maximum number of end devices that can be admitted in the network. The data transmission period is composed of a number of data slots, which equals the current device number in the network, and the duration of each data slot is set the same as one data packet length. We assume that all data packets have an identical and fixed packet length. Since the DTDMA eliminates packet transmission collisions, its performance is superior to the DCF in a high network load condition. Therefore, with the consideration of these two candidate MAC schemes, the proposed adaptive MAC solution uses a separate mediating MAC entity (Doerr et al. 2005) working on top of the MAC candidates maintained at each device to make MAC switching decisions according to the variation of the number of devices, N, in the network.

To maintain consistently maximum network throughput by switching between the MAC candidates, the optimal MAC switching threshold (i.e., the optimal number of devices N^*) needs to be determined. To this end, a unified and closed-form performance analytical framework (Ye et al. 2016) is established for the aggregate network throughput with respect to the number of devices N for both DCF and DTDMA schemes, based on *least-squares curve fitting* and *M/G/1 queueing analysis*. For the throughput analysis, both traffic non-saturation (i.e., the transmission buffer of each device can be empty) and traffic saturation (i.e., each device always has packets to be transmitted) conditions are considered. For a traffic non-saturation condition, closed-form throughput analytical expressions $H_1(N, \lambda)$ and $H_2(N, \lambda)$ are established in a function of N and λ for the DCF and the DTDMA, respectively. With the increase of N, packet service rate for each



Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs, Fig. 1 A DTDMA frame structure

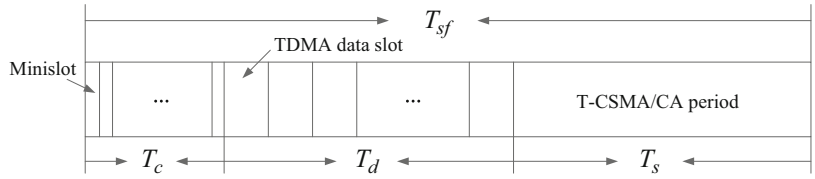
device decreases, making the network operating in DCF or DTDMA enter the traffic saturation state. Thus, the throughput for the DCF and the DTDMA is also analyzed in a closed-form function of N, denoted by $H_3(N)$ and $H_4(N)$.

To calculate the optimal MAC switching threshold, denoted by N^*, throughput comparison between the two MAC candidates is conducted based on the developed closed-form performance analytical models. Since DCF and DTDMA have different network load points to enter the traffic saturation state, the following four combinations of traffic load states for both MAC schemes should be taken into consideration for the performance comparison: (1) the network is in the traffic saturation state for both MAC schemes; (2) the network is in a traffic non-saturation state for both MAC schemes; (3) the network is saturated for the DCF and non-saturated for the DTDMA; and (4) the network is saturated for the DTDMA and non-saturated for the DCF. Then, the optimal MAC switching threshold in terms of the number of devices, N^*, in the network can be determined, which also varies with the traffic arrival statistics λ at each device. Therefore, the adaptive MAC solution consistently makes the MAC switching decision at an optimal point (i.e., the DCF is operated when $N < N^*$ and is switched to the DTDMA when $N \geq N^*$) to achieve maximum aggregate network throughput.

Heterogeneous Service Scenario

To support heterogeneous service types, an adaptive and QoS-aware MAC solution is required. Consider the same system model as in the homogeneous service scenario except that both delay-sensitive voice devices and high-rate data devices are present in the network. The number of voice and data devices are denoted by N_v and N_d, which are slowly varying with time. For the delay-sensitive voice service, packet arrivals at each voice device follow an *on/off* model (Wang et al. 2007), which is a two-state Markov process with the *on* and *off* states being the traffic generation and traffic suppression phases, respectively. The durations each device stays in *on* and *off* states are independent and exponentially distributed with respective average of $\frac{1}{\gamma}$ and $\frac{1}{\delta}$. During the *on* state, packets arrive periodically with the constant rate α packet/s. Each voice packet has a hard delay bound requirement, so that packets received beyond the delay bound are dropped. Therefore, the voice service has a packet loss rate bound requirement denoted by P_l. For the data service, each device is expected to access the channel by exploiting resource multiplexing gain to achieve as high as possible data throughput. All data devices are assumed in the traffic saturation state.

To satisfy the QoS requirements from both voice and data services, a QoS-aware hybrid MAC scheme is proposed to differentiate the channel access among voice and data devices (Ye and Zhuang 2017b), where voice devices are allocated time slots in a distributed way to guarantee a bounded packet delay by avoiding contention collisions and data devices contend for the channel access according to a truncated CSMA/CA (T-CSMA/CA) scheme to exploit high resource multiplexing gain. Time is partitioned into a sequence of fixed-duration superframes, as shown in Fig. 2. The duration of each superframe, denoted by T_{sf}, is set the same as the packet delay bound for voice traffic. Each superframe is composed of a control period, a TDMA period, and a T-CSMA/CA period, the durations of which are denoted by T_c, T_d, and T_s. The control period consists of a number N_{mi} of constant-duration minislots, each with an exclusive minislot sequence number (MN). Each voice device selects a unique minislot to broadcast and exchange its local information among its one-hop neighbors for distributed data time slot scheduling in the following TDMA period. Thus, N_{mi} also indicates the maximum number of voice devices (i.e., voice capacity) that can be admitted in the network. The TDMA period is divided into multiple equal-duration data transmission slots, each of which has a unique data slot sequence number (DN). Every active voice node (i.e., having non-empty transmission buffer) occupies one data slot to send a number of packets (a voice burst) generated during the previous superframe time. The number of

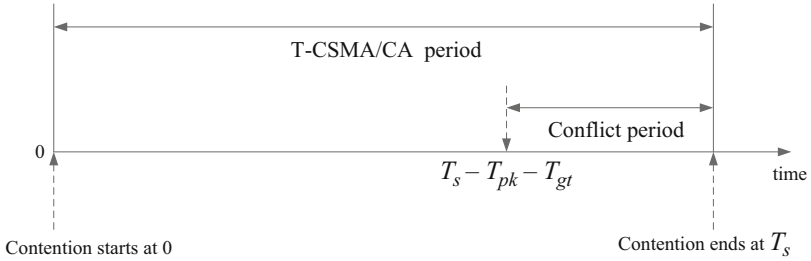


Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs, Fig. 2 A hybrid MAC superframe structure

voice bursts scheduled in each TDMA period is indicated by N_{sv}. To provide voice traffic guaranteed service quality, a maximum fraction ρ (< 1) of channel time in each superframe is allocated to voice traffic, including the control period and the TDMA period. The voice capacity N_{mi} can be determined based on ρ and the packet loss rate bound P_l. In the T-CSMA period, data devices contend the channel according to the CSMA/CA with exponential backoff with periodic interruption of the control period and the TDMA period in each superframe.

Next, a traffic-adaptive time slot allocation mechanism is presented according to instantaneous voice traffic load in the network. Each voice device randomly selects a minislot from the control period of a superframe after the time synchronization and broadcasts a control packet in the selected minislot to its one-hop neighbors. A control packet sent from a tagged device contains the following important information: (1) device IDs from its one-hop neighbors including the tagged device; (2) MN of the minislot occupied by the tagged device; (3) buffer occupancy bit (BO), BO = 1 if the device's transmission buffer is non-empty and 0 otherwise; and (4) previous DN, indicating the DN of the occupied data slot from the tagged device in previous superframe, with DN = 0 if the device was not allocated a data slot in the previous superframe. The selection of one minislot from the tagged device is considered successful if the broadcast control packets at subsequent minislots following the selected minislot contain the tagged device ID, and thus the same minislot will be selected in every subsequent superframe. Otherwise, collision happens for accessing the minislot, and all the devices involved will wait until the next superframe for a new minislot selection. With the consideration of the *on/off* traffic statistics, only active voice devices (i.e., the devices with BO = 1 in broadcast control packets) are allocated data transmission time slots after accessing the minislots, to adapt to instantaneous voice traffic load. Two categories of active voice devices are defined: category I and category II. A category I device is currently active but was not allocated a data slot in previous superframe, whereas a category II device is active in both current and previous superframes. Since the activation of category I device is at some random time instant before its minislot accessing time at current superframe, category I devices are prioritized over category II devices for the time slot scheduling. Specifically, category I devices are scheduled data time slots first by following their minislot accessing sequence to minimize the packet delay bound violation probability, under the condition that each category II device can be scheduled a time slot no later than the same time slot as in previous superframe.

After the TDMA period, the data devices contend for the channel access by employing the T-CSMA/CA scheme. The T-CSMA/CA is similar as the CSMA/CA with exponential backoff, except that the contention-based packet transmissions in one superframe are periodically interrupted by the presence of its subsequent superframe. Therefore, the performance of T-CSMA/CA is different from the CSMA/CA in the following two aspects: (1) the packet waiting time before transmission is enlarged by the control period and the TDMA period of each superframe; and (2) before transmitting one packet at the end of the exponential backoff phase, each device needs to check whether the remaining time in current superframe is



Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs, Fig. 3 An illustration of T-CSMA/CA period in each superframe

enough to transmit at least one packet. If the remaining time is less than the summation of a complete packet transmission time (T_{pk}) and a guard time (T_{gt}), a virtual conflict occurs and ($T_{pk} + T_{gt}$) is the duration of the conflict period, as shown in Fig. 3. Then, all devices involved in a virtual conflict are required to hold on the packet transmissions until the beginning of subsequent T-CSMA/CA period.

The optimal parameters of the proposed hybrid MAC solution are derived, which are adaptive to varying numbers of voice and data devices to achieve bounded voice packet delay and consistently maximum data throughput. For the delay-sensitive voice service, given the maximum fraction of channel time, ρ, allocated to voice traffic in each superframe, the voice capacity N_{mi} supported in the network is analyzed, which is also the number of minislots configured for the control period of each superframe, to guarantee the voice packet loss rate bounded by P_l. Given N_v, the maximum number of voice bursts (data slots), denoted by N_{sm}, that can be scheduled in each superframe to achieve P_l is then determined. Since the actual number of scheduled data slots in each superframe is likely less than N_{sm}, the average number of scheduled data slots $\overline{N_{sv}}$ and the average duration $\overline{T_d}$ of each TDMA period are also calculated. Moreover, N_{sm}, $\overline{N_{sv}}$, and $\overline{T_d}$ are dynamically adapted to the variation of N_v to achieve a consistently bounded voice packet loss rate. After obtaining T_c and $\overline{T_d}$, the average duration $\overline{T_s}$ of the T-CSMA/CA period is also determined, which is employed for channel access by N_d data devices. Therefore, the aggregate data throughput S_d in the T-CSMA/CA period of each superframe is further derived in terms of N_v, N_d, and the packet transmission probability τ_d from each device at a backoff slot. After some algebraic manipulation and certain approximation, the optimal transmission probability τ_d^{opt} and the corresponding optimal contention window size CW^{opt} to achieve maximum aggregate data throughput $S_d^{\max}$ are obtained in closed-form expressions of N_v and N_d. With the derived performance analytical models, the optimal MAC parameters CW^{opt} and τ_d^{opt} can also be dynamically adjusted with variations of N_v and N_d. Therefore, the proposed hybrid MAC solution in supporting heterogeneous services optimizes the MAC configurations and adapts the optimal MAC parameters to heterogeneous traffic load conditions for maintaining consistently maximum network performance.

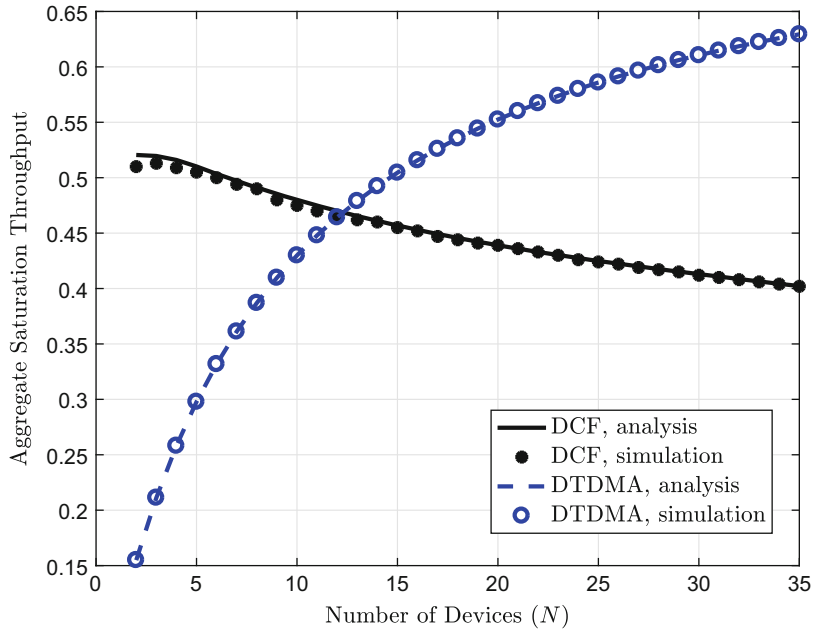
Our proposed adaptive MAC solutions for both homogeneous and heterogeneous service scenarios are useful in an IoT-based network environment, where conventional cellular communication infrastructures are inaccessible and devices are interacted via an ad hoc networking. For example, in a postdisaster area, a large number of smartphones and smart sensors are required to be interconnected for information dissemination. The proposed MAC solutions are efficient in coordinating channel access from the smart devices in an infrastructure-less manner and adapting the MAC performance to the network load fluctuations due to device activation/deactivation and device mobility.

Numerical Results

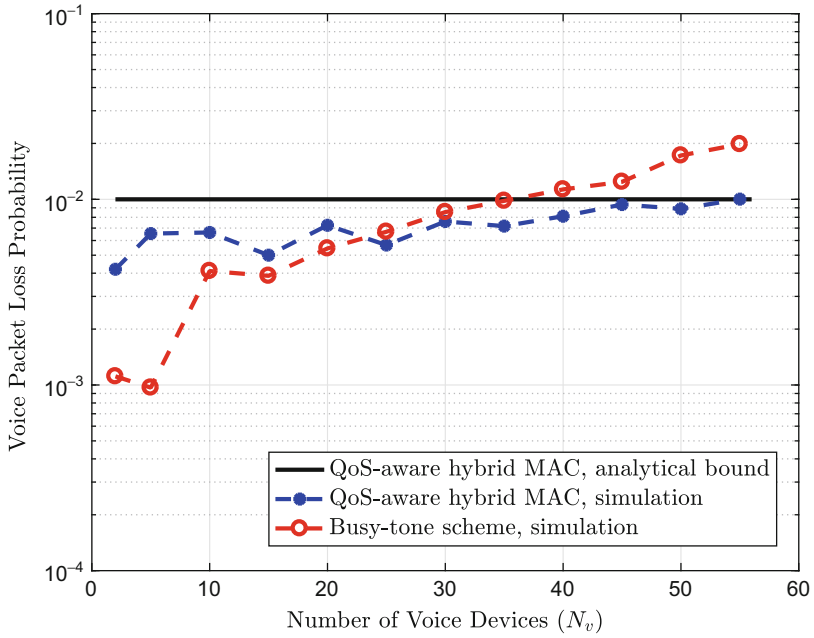
Simulation results are provided to verify the effectiveness of the proposed adaptive MAC solutions. All simulations are carried out by using the network simulator OMNeT++. In the simulation, devices are randomly scattered over a 100 × 100 m square region and are within one-hop communication ranges of other devices. Each source device randomly selects its destination among the other devices. For the homogeneous service scenario, devices are with the same type of high-rate data application. Packet arrivals at each data device are assumed to follow a Poisson process with the average packet arrival rate set as 300 packet/s for the traffic saturation state (results for a traffic non-saturation state are provided in Ye et al. 2016). For the heterogeneous service scenario, there are a mixture of delay-sensitive voice devices and high-rate data devices in the network. Voice traffic is generated periodically during the *on* state with the rate of 50 packet/s. Other system parameters for the simulation are provided in Ye et al. (2016) and Ye and Zhuang (2017b).

Performance comparison between DCF and DTDMA is conducted for the homogeneous service scenario. Figure 4 shows the aggregate saturation throughput for both candidate MAC protocols with the variation of N. It can be seen that the optimal MAC switching threshold, N^*, exists at the network load point when the two MAC candidates achieve the same throughput. Based on the derived closed-form performance analytical models, the optimal MAC switching threshold can be determined in a distributed way with low complexity, upon which the adaptive MAC solution achieves consistently maximum network throughput.

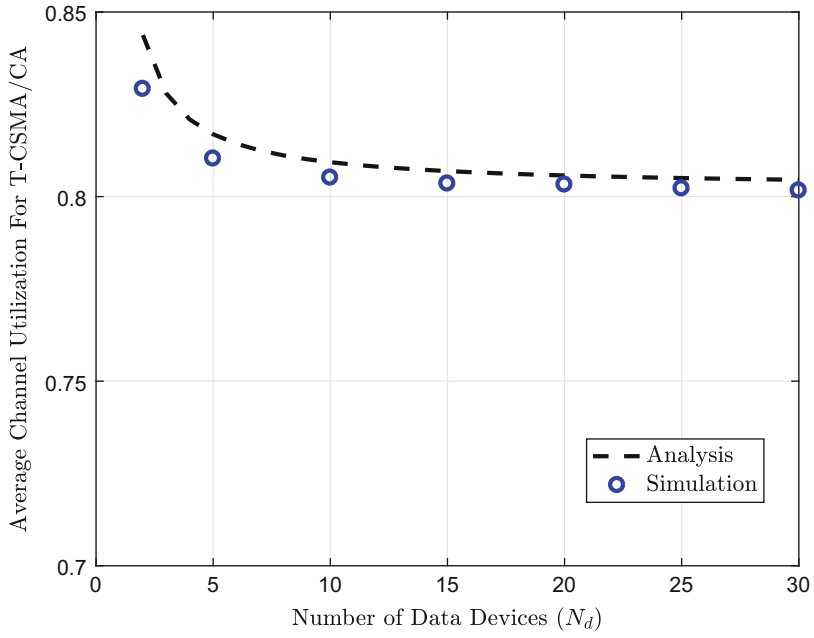
For the proposed QoS-aware hybrid MAC solution, the voice packet loss rate is evaluated in Fig. 5. It is verified through simulations that the proposed distributed and adaptive time slot allocation mechanism always guarantees a packet loss rate below the analytical bound as long as the number of admitted voice devices is within the voice capacity. Although the busy tone-based MAC scheme (Wang et al. 2007) achieves a bounded packet delay in a low network load con-



Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs, Fig. 4 Throughput comparison between DCF and DTDMA



Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs, Fig. 5 Voice packet loss rate (N_d=10, $\rho = 0.5$)



Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs, Fig. 6 Average channel utilization of T-CSMA/CA in each superframe ($N_v = 20$, $\rho = 0.5$)

dition, the transmission collisions among voice devices are accumulated as the network load increases, leading to degraded service performance. The average T-CSMA/CA channel utilization within each hybrid MAC superframe is shown in Fig. 6, which is defined as the ratio of

average time for successful data transmissions inside a T-CSMA/CA period over the length of the T-CSMA/CA period. Since MAC parameters are optimized for the T-CSMA/CA, the channel utilization is consistently maximum with respect to a varying N_d.

Conclusion

In this entry, the backgrounds of distributed and adaptive MAC schemes for an IoT-enabled MANET are investigated. Novel adaptive MAC solutions are presented for both homogeneous and heterogeneous service scenarios. For the homogeneous service scenario, a hybrid MAC scheme is proposed to switch between the IEEE 802.11 DCF and the DTDMA according to the network load conditions. Based on the throughput analysis of both MAC candidates, an optimal MAC switching threshold is derived in terms of the number of devices in the network. For the heterogeneous service scenario, a QoS-aware hybrid MAC scheme is presented to differentiate the channel access for voice and data devices, where adaptive time slot allocation is conducted for voice traffic and a T-CSMA/CA scheme is employed for data traffic. The MAC parameters of the proposed scheme are optimized and are adaptive to the heterogeneous network traffic load, to achieve bound voice packet delay and consistently maximum aggregate data throughput. Simulation results are presented to verify the advantages of the proposed schemes. The applications of the adaptive MAC solutions are also discussed.

Cross-References

- Media Access Control for Narrowband Internet of Things: A Survey
- Multiple Access Techniques
- QoS-Aware MAC
- Quality of Service in IEEE 802.11 Networks

References

Ahmed R (1997) An adaptive multiple access protocol for broadcast channels. In: Proceedings of IEEE IPCCC'97, pp 371–377

Bianchi G (2000) Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J Select Areas Commun 18(3):535–547

Doerr C, Neufeld M, Fifield J, Weingart T, Sicker DC, Grunwald D (2005) MultiMAC-an adaptive MAC framework for dynamic radio networking. In: Proceedings of IEEE DySPAN'05, pp 548–555

Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of things (IoT): a vision, architectural elements, and future directions. Future Gener Comput Syst 29(7):1645–1660

Hu W, Yousefi'zadeh H, Li X (2011) Load adaptive MAC: a hybrid MAC protocol for MIMO SDR MANETs. IEEE Trans Wireless Commun 10(11):3924–3933

Kanzaki A, Uemukai T, Hara T, Nishio S (2003) Dynamic TDMA slot assignment in ad hoc networks. In: Proceedings of IEEE AINA'03, pp 330–335

Natkaniec M, Kosek-Szott K, Szott S, Bianchi G (2013) A survey of medium access mechanisms for providing QoS in ad-hoc networks. IEEE Commun Surv Tutor 15(2):592–620

Nishiyama H, Ngo T, Oiyama S, Kato N (2015) Relay by smart device: innovative communications for efficient information sharing among vehicles and pedestrians. IEEE Veh Technol Mag 10(4):54–62

OMNeT++ 5.0 [Online]. Available: http://www.omnetpp.org/omnetpp. Accessed on Aug. 2018

Wang P, Jiang H, Zhuang W (2007) Capacity improvement and analysis for voice/data traffic over WLANs. IEEE Trans Wireless Commun 6(4):1530–1541

Wilson N, Ganesh R, Joseph K, Raychaudhuri D (1993) Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN. IEEE J Select Areas Commun 11(6):870–884

Ye Q, Zhuang W (2017a) Token-based adaptive MAC for a two-hop Internet-of-Things enabled MANET. IEEE Internet Things J 4(5):1739–1753

Ye Q, Zhuang W (2017b) Distributed and adaptive medium access control for Internet-of-Things-enabled mobile networks. IEEE Internet Things J 4(2):446–460

Ye Q, Zhuang W, Li L, Vigneron P (2016) Traffic load adaptive medium access control for fully-connected mobile ad hoc networks. IEEE Trans Veh Technol 65(11):9358–9371

Zhang R, Cai L, Pan J (2010) Performance analysis of reservation and contention-based hybrid MAC for wireless networks. In: Proceedings of IEEE ICC'10, pp 1–5

Zhang R, Cai L, Pan J (2011) Performance study of hybrid MAC using soft reservation for wireless networks. In: Proceedings of IEEE ICC'11, pp 1–5

Adaptive Video Streaming

▸ Joint Caching, Computing, and Routing for Video Transcoding in Wireless Networks

Advanced Automated Meter Reading (AMR)

▸ Smart Metering, Specific Challenges, and Solution Approaches

Advanced Metering Infrastructure (AMI)

▸ Smart Metering, Specific Challenges, and Solution Approaches

Advanced Persistent Threats(APT) or Sophisticated Attacks

▸ Security and Privacy in 4G/LTE Network

Advances in Distribution System Monitoring

Omid Ardakanian
Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Synonyms

Wide area monitoring system

Definitions

Distribution system monitoring refers to key technologies for monitoring wide-area power distribution systems, between the substation and customer meters, to facilitate distribution grid planning and operation.

Background

Power distribution systems had a simple design historically. With radial topology, one-way power flow, and predictable demand curves, distribution system planners and operators were only required to evaluate the envelope of design conditions, such as peak loads and fault currents, to ensure reliability and power quality. Thus, there has been little need for costly telemetry beyond the substation, which was remotely monitored through supervisory control and data acquisition (SCADA) at several-second intervals. SCADA is a technology that has been in place for decades to connect sensing and control nodes, mostly in the transmission system, to a control room (Northcote-Green and Wilson 2006) using dedicated telephone lines, cellular, radio, or power line communications.

In recent years, the rapid growth in the deployment of controlled loads and distributed energy resources (DER) has led to unprecedented amounts of variability and uncertainty, which complicate distribution system planning and operation. This has necessitated more comprehensive monitoring of distribution circuits, and a novel planning and operation paradigm centered around pervasive monitoring, real-time analytics, and closed-loop control (Ardakanian et al. 2016).

Foundations

In light of new and complex grid dynamics that span multiple timescales, the operators must closely monitor the voltage and current

waveforms at different locations to detect and characterize certain behaviors, such as harmonics and oscillations, which could not be observed by the traditional SCADA system. This calls for an advanced distribution system monitoring solution which

- has low cost and complexity of deployment,
- covers both primary and secondary distribution networks,
- provides high-resolution measurements at multiple locations,
- uses precise time synchronization so that measurements can be compared across locations,
- utilizes a communication network that provides
  - stable and high bandwidth to allow for transferring high-sample-rate measurements from thousands of end nodes to upstream aggregation nodes and decentralized controllers,
  - low latency which is necessary for real-time applications, such as situational awareness, fault detection, and automated demand response,
  - high degree of reliability, or, equivalently, low link outage and packet loss probabilities,
- incorporates mechanisms to prevent unauthorized access, data manipulation, and denial of access to the sensing and control nodes,
- leverages an extensible data processing pipeline, high-throughput data stores, and real-time event triggers to provide a deeper insight into the operating state of the grid.

Two technologies have emerged in the last couple of years, namely, *smart meters* and *distribution-level phasor measurement units* (D-PMUs), which together can constitute the data acquisition layer of a well-suited distribution system monitoring system that meets the above requirements.

These technologies differ mainly in physical quantities they can measure, the time granularity of data, the location of sensors, and communication requirements. Smart meters are deployed at customer premises, recording their energy usage and voltage level hourly or more frequently (up to every 15 min). The meters send the recorded data at regular intervals to the utility's data center using wireless or wired communications (Gungor et al. 2011). The smart meters, communications networks, and the utility's data management systems are collectively known as *advanced metering infrastructure* (AMI).

The D-PMUs (NASPI Distribution Task Team 2018) offer higher precision and resolution than smart meters. They sample voltage and current waveforms at a high frequency (usually at 120 Hz) and assign a precise time stamp to the measured quantities. The measured quantities are typically voltage and current *phasors* along with frequency, where a phasor represents the magnitude and phase angle of voltage or current. The D-PMUs are installed on distribution circuits supplementing the existing SCADA system by monitoring the network downstream of the substation. The data streams produced by a network of D-PMUs, termed *phasor network*, are aggregated by a small number of phasor data concentrators (PDCs), enabling fast comparison of time-synchronized phasor measurements from multiple locations before they reach the utility's data center. Given the bandwidth requirement of a phasor network, a broadband cellular network technology, such as 4G, is typically used for sending phasor measurements to the utility.

The smart meters and D-PMUs complement each other in the sense that they respectively monitor the end nodes and intermediate nodes in the distribution network. Moreover, in some applications, the availability of smart meter data can compensate for the lack of higher-resolution phasor measurements at some upstream nodes.

Implications for the Grid Planning and Operation

Distribution system operators can utilize the fine-grained measurements along with appropriate analytical tools for many important applications that concern planning and operation of the

Advances in Distribution System Monitoring, Table 1 Applications of distribution-level phasor measurement units

Application	Data streams	Analysis	Minimum resolution
Topology detection	Voltage phasors	On-line	1 cycle
Phase identification	Voltage phase angle	Off-line	1 s
Model parameter estimation	Voltage and current phasors	Off-line	1 s
State estimation	Voltage and current phasors	On-line	1 s
DG characterization	Voltage and current phasors	Off-line	1 min
Event detection/localization	Voltage and current phasors	On-line	1 cycle
Equipment health monitoring	Voltage and current phasors	Off-line	1 min
Outage management	Voltage and current magnitudes	On-line	1–15 min
Phasor-based control	Voltage phasors	On-line	1 cycle

distribution system (Meier et al. 2017). These applications are outlined below:

- *Topology detection* is to determine the set of energized lines and the status of switches to understand the real-time operational structure of the distribution network.
- *Phase identification* is to identify and track the connectivity and loading of the three AC phases throughout the network.
- *Model parameter estimation* is to compute impedances of distribution components, such as line segments and transformers, and update the distribution system model.
- *State estimation* (Abur and Expósito 2004) is to determine unknown state variables, for example, voltage magnitudes and phase angles at specific buses, given a set of known or measured state variables.
- *Distributed Generation (DG) characterization* is to separate the net-metered distributed generation from load.
- *Event detection and localization* is to detect and classify short-term operational and power quality events and pinpoint them to a small part of the distribution network.
- *Equipment health monitoring* is to detect and monitor equipment health issues or early signs of equipment aging to prevent damage to the equipment and support system upgrade decisions.
- *Outage management* is to automatically and reliably detect outages to reduce their duration and the system restoration cost.
- *Phasor-based control* is to incorporate phasor measurements in the feedback control of distribution system components and active end nodes, e.g., solar inverters, battery storage systems, and electric vehicle chargers.

Table 1 describes which physical quantities must be measured and what time granularity is sufficient for each of these applications.

Cross-References

- ▸ Cellular Networks: An Evolution from 1G to 4G
- ▸ Key Technologies in 4G/LTE Network
- ▸ Smart Metering, Specific Challenges, and Solution Approaches
- ▸ Power Line Communications for Grid Discovery and Diagnostics

References

Abur A, Expósito A (2004) Power system state estimation: theory and implementation. Power engineering (Willis). CRC Press, Boca Raton

Ardakanian O, Keshav S, Rosenberg C (2016) Integration of renewable generation and elastic loads into distribution grids. SpringerBriefs in electrical and computer engineering. Springer International Publishing, Cham

Gungor VC, Sahin D, Kocak T, Ergut S, Buccella C, Cecati C, Hancke GP (2011) Smart grid technologies: communication technologies and standards. IEEE Trans Ind Inf 7(4):529–539

Meier A, Stewart E, McEachern A, Andersen M, Mehrmanesh L (2017) Precision micro-synchrophasors for distribution systems: a summary of applications. IEEE Trans Smart Grid 8(6): 2926–2936

NASPI Distribution Task Team (2018) DisTT: synchrophasor monitoring for distribution systems: technical foundations and applications. Technical report, North American Synchrophasor Initiative. https://www.naspi.org/node/688

Northcote-Green J, Wilson RG (2006) Control and automation of electrical power distribution Systems. CRC Press, Boca Raton

Ambient Power Technology

▸ Energy Harvesting Technologies in Wireless Sensor Networks

Analog and Digital Communications

Xin-Lin Huang
Tongji University, Shanghai, China

Synonyms

Analog and digital communications

Definitions

The analog signal can be one of infinite number of values, while the digital signal can only be one of finite number of values. The analog signal transmission from one part to other parts is called analog communications, while the digital signal transmission from one part to other parts is called digital communications.

Historical Background

In practice, most signals that exist in the nature are analog signals, such as voice, image, video, temperature, pressure, and so on. There are infinite numbers of possible values for analog signal, which is continuous with time in most cases. According to the Nyquist theorem, the low-pass and band-pass analog signals can be converted into digital signals through sampling, quantization, and coding processes, which is known as analog-to-digital converting (ADC). During the ADC process, the error called quantization error is caused in the quantization process. After ADC process, the analog signal is converted into digital signal as "0" and "1" binary sequence (Hui and Yeung 2003).

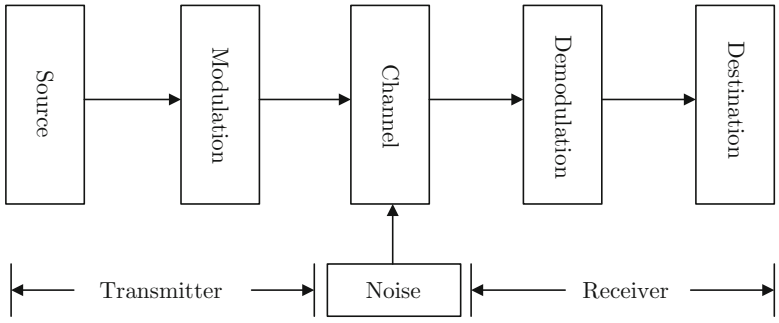
The modulation modes adopted in analog communications include amplitude modulation, frequency modulation, and phase modulation. Digital communication system uses digital amplitude modulation, digital phase modulation, digital frequency modulation, and APK modulation. Since the digital signal can only be one of finite number of values, the main difference of demodulations between analog and digital communications is that a sampling and decision module is used in digital communications. The sampling and decision module maps the demodulated signals into one of the finite number of values. If the noise caused in the digital communication systems is less than the noise tolerance threshold, it can be removed by the sampling and decision unit. However, the quantization noise is irreversible in digital communications.

Analog and digital communications are adopted in mobile communications. The emergence of mobile communications began in the early twentieth century. The mobile communication schemes have experienced from analog communications to digital communications. In 1G to 4G communications, 1G cellular network adopts analog communications, and the others are digital communications (Kumar et al. 2010).

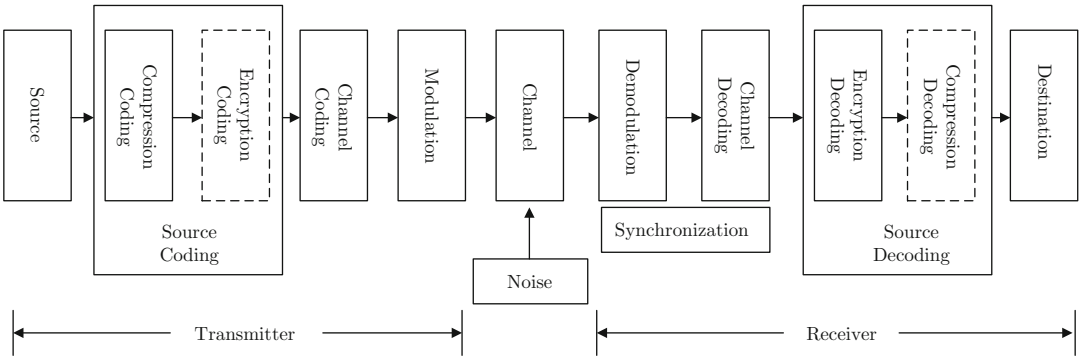
Foundations

Analog Communication Systems

The analog communication systems transmit analog signals, which have infinite number of possible values. Therefore, the receiver cannot com-



Analog and Digital Communications, Fig. 1 Analog communication systems model



Analog and Digital Communications, Fig. 2 Digital communication systems model

pletely eliminate the interference noise caused by the transmission channel.

Due to the amplitude distortion and phase distortion in communication systems, the performance of analog communications is affected. It means that the analog signals sent by the transmitter will not be received by the receiver accurately. The analog communications can be classified into amplitude modulation, phase modulation, and frequency modulation according to which parameter of the carrier is modulated by the transmitting analog signals. The diagram of the general analog communication systems is shown in Fig. 1. The performance of analog communication systems is merited by the signal-to-noise ratio of the reconstructed signals at the receiver.

Digital Communication Systems

The digital communication systems transmit the digital signal which can only be one of finite number of values. The sampling and decision module at the receiver can completely eliminate the interference signal if it is smaller than the noise tolerance threshold and thus greatly improving the quality of digital communications and facilitating long-distance transmission with relays. The general digital communication system is shown in Fig. 2. The performance of digital communication systems is merited by the symbol error probability of the decoded digital sequence at the receiver.

Different from Fig. 1, the source coding and channel coding are adopted at the transmitter, and the corresponding source decoding and channel decoding are adopted at the receiver. Due to the overwhelming bias toward digital communications, there is no effective analog channel coding designed for analog communications.

Discussions

In digital communications, the irreversible quantization error caused by ADC process is introduced into digital transmission. The source coding leads to a serious dependence among the transmission stream and has the drawback of

error propagation during decoding. In the transmission process on time-varying fading channel, the channel quality changes dynamically. When the channel quality is higher, the digital modulation cannot obtain extra transmission quality after correctly demodulating the digital signal.

In analog communications, due to lack of effective analog channel coding scheme, the analog communications are not good at overcoming channel noise. However, analog communications can be well adapted to the dynamics of channel quality, and better signal-to-noise ratio can be obtained when the channel quality is higher. Hence, effective analog channel coding and power allocation schemes are encouraged in future research works.

Key Applications

The analog communications are mainly adopted in telephone communications, radiobroadcasting and television services, and so on. Currently, digital communications are widely used, such as computer communications, mobile communications, and digital television (Hui and Yeung 2003; Tachikawa 2003).

References

Hui SY, Yeung KH (2003) Challenges in the migration to 4G mobile systems. IEEE Commun Mag 41(12):54–59

Kumar A, Liu Y, Sengupta J, Divya JS (2010) Evolution of mobile wireless communication networks 1G to 4G. Int J Electr Commun Technol 1(1):68–72

Tachikawa K (2003) A perspective on the evolution of mobile communications. IEEE Commun Mag 41(10): 66–73

Anomaly

▸ Anomaly Detection for IoT Systems

Anomaly Detection for IoT Systems

Xin-Xue Lin[1], En-Hau Yeh[1], and Phone Lin[2]
[1]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
[2]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, Republic of China

Synonyms

Anomaly; Internet of Things (IoT)

Definitions

Anomaly detection for Internet of Things (IoT) system is to automatically detect whether the IoT devices, components, or systems operate normally or not. Usually there are multiple sensors or external monitors that observe the signals sent from the operating IoT systems. The detection module analyzes the signals to determine whether the system's behavior is normal or abnormal.

Historical Background

The IoT systems link the heterogeneous sensors and IoT servers to provide the IoT applications such as healthcare, industrial automation, environment monitoring, and so on. Because the decade aged IoT systems and new IoT systems may coexist, it is not easy to implement the monitors into the integrated IoT systems. It is usually to treat the integrated IoT system as a black box. Furthermore, because the signals come from one or more types of sensors (i.e., heterogeneous sensors), it is complicated for the monitors to analyze the signals from the IoT systems. One of the solutions to resolve the issue is to use the rules defined by the expert, but it usually takes time and cost. The data-driven technologies can

be applied for anomaly detection in IoT systems. Some of the previous works using these data-driven technologies are summarized as follows:

The works (Xie et al. 2017; Juvonen et al. 2015) built the statistical model based on the normal data. If a data sample does not follow the statistical model, it is judged as an abnormal data point. The work (Zhang et al. 2016) used supervised machine learning algorithms to build a classification model to check whether a data sample is normal or abnormal. It is required to label each data sample as "normal" or "abnormal" in supervised learning algorithms. The works (Valenzuela et al. 2013; Shin et al. 2011) used unsupervised machine learning algorithms to build clustering models. Because it does not rely on the labeled data, it can save lots of labeling effort.

Foundations

Anomaly detection techniques can be applied in many domains such as intrusion detection, system health monitoring, anomalous event detection in sensor networks, and so on. The anomalies could be caused by external factors (e.g., network attacking and human mistake) or internal factors (e.g., hardware failure and resource exhaustion). The anomaly detection techniques can be grouped into the following categories: density-based (Breunig et al. 2000), statistical-based (Xie et al. 2017; Juvonen et al. 2015), clustering-based (Zong et al. 2018), and the machine learning-based (Zhang et al. 2016) techniques. These techniques use historical data to build up models to determine whether the data sample is normal or abnormal. In most cases, the anomalies are rare events (i.e., less data points for the anomalies). It is very hard to directly learn the anomalous patterns from less data samples. To resolve the issue, we may only use the data samples of normal events to build up the model for anomaly detection. If a sample data does not follow the model (e.g., does not fall into the distribution of the data points of normal event), it is determined as an abnormal event.

Key Applications

In the following, we elaborate on some IoT applications (Porkodi and Bhuvaneswari 2014) for which anomaly detection can be applied:

Industrial Environment/System

In Industry 4.0 (Stojanovic et al. 2016) for automation and data exchange in the manufacturing industry, wireless sensor networks are deployed to monitor the physical environment to ensure that the manufacturing systems operate continuously and safely. With the help of the monitoring sensors, the status of the monitored object is represented by several metrics, with which the anomaly detection can determine whether the status of the monitored object is normal or abnormal.

Smart Home

In the smart home application (Fahad and Rajarajan 2015), the electrical devices, such as air-condition, television, and so on, are connected with the IoT network, through which users can interact with the devices. The anomaly detection for smart home can be applied for home security, where it can detect whether someone without authority breaks into, or check whether the door is locked or not.

Health Monitoring

With the wearable devices connected to the IoT network, the human's physiological status can be monitored remotely in real time, which makes elderly care (Shin et al. 2011) or healthcare more functional. The anomaly detection can be applied for checking whether the status deviates from distribution of the historical data, and determines whether the owner of the devices is in trouble (i.e., an emergency event occurs). It can also reduce the response time for an emergency event. Applying anomaly detection helps health monitoring system be able to react more quickly.

Connected Cars

A connected car is a vehicle with capability to connect to other vehicles, devices, and networks through wireless local area networks which assists users to drive (Kwak et al. 2016). One of the applications of the anomaly detection on the connected car is to check the user's driving behavior (e.g., it can determine whether the user has dangerous driving behavior), or check if the functionality of the car is in good status.

Smart City

In the smart city (Difallah et al. 2013), the sensors and activators (e.g., traffic light) are connected to IoT network. Examples of the smart city applications include traffic flow management and environmental monitoring. The anomaly detection in smart city can be applied to detect the car accident (i.e., abnormal traffic flow).

IoT Network Security

In the IoT network security (Hodo et al. 2016), the traditional firewall determines whether a packet can go into the Intranet by referencing the predefined security rules. However, it is hard to identify new types of attacks by using the predefined rules. The unsupervised anomaly detection mechanism can train a behavior model by using the normal Internet traffic. The model can be used to determine abnormal network traffics, e.g., attack patterns, and it can reduce the effort for human expert to find out the attack patterns.

Cross-References

- ▸ Data-Driven Security
- ▸ Data-Driven Smart City
- ▸ Industrial Big Data

References

Breunig MM, Kriegel H-P, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. ACM SIGMOD Rec 29(2):93–104

Difallah DE, Cudre-Mauroux P, McKenna SA (2013) Scalable anomaly detection for smart city infrastructure networks. IEEE Internet Comput 17(6):39–47

Fahad LG, Rajarajan M (2015) Anomalies detection in smart-home activities. In: Proceedings of IEEE international conference on machine learning and applications

Hodo E, Bellekens X, Hamilton A, Dubouilh PL, Iorkyase E, Tachtatzis C, Atkinson R (2016) Threat analysis of IoT networks using artificial neural network intrusion detection system. In: Proceedings of IEEE international symposium on networks, computers and communications

Juvonen A, Sipola T, Hämäläinen T (2015) Online anomaly detection using dimensionality reduction techniques for http log analysis. Comput Netw Int J Comput Telecommun Netw 91:46–56

Kwak BI, Woo J, Kim HK (2016) Know your master: driver profiling-based anti-theft method. In: Proceedings of IEEE annual conference on privacy, security and trust

Porkodi R, Bhuvaneswari V (2014) The internet of things applications and communication enabling technology standards: an overview. In: Proceedings of the international conference on intelligent computing applications

Shin JK, Lee B, Park KS (2011) Detection of abnormal living patterns for elderly living alone using support vector data description. IEEE Trans Inf Technol Biomed 15(3):438–448

Stojanovic L, Dinic M, Stojanovic N, Stojadinovic A (2016) Big-data-driven anomaly detection in industry (4.0): an approach and a case study. In: Proceedings of IEEE international conference on big data

Valenzuela J, Wang J, Bissinger N (2013) Real-time intrusion detection in power system operations. IEEE Trans Power Syst 28(2):1052–1062

Xie M, Hu J, Guo S, Zomaya AY (2017) Distributed segment-based anomaly detection with Kullback–Leibler divergence in wireless sensor networks. IEEE Trans Inf Forensics Secur 12(1):101–110

Zhang Z, Wang X, Lin S (2016) Mobile payment anomaly detection mechanism based on information entropy. IET Netw 5(1):1–7

Zong B, Song Q, Min MR, Cheng W, Lumezanu C, Cho D, Chen H (2018) Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: Proceedings of international conference on learning representations

Anti-jamming

- ▸ Wireless Jamming Attack

Any-Path Routing

- ▸ Opportunistic Routing (OR)

AODV Protocol with Trust Based Mechanism

▸ Trust-Based Routing in Software-Defined Vehicular Ad Hoc Networks

Application of Big Data on Service Provisioning

▸ Data-Driven Service Provisioning

Application of Machine Learning in Wireless Sensor Network

Vaidehi Vijayakumar
School of Computing Science and Engineering, VIT University, Chennai, India

Synonyms

Clustering; Data aggregation; Machine learning; Routing; Wireless sensor Networks

Definition

A Wireless Sensor Network (WSN) consists of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions. A WSN system incorporates a gateway that provides wireless connectivity back to the wired world and distributed nodes. Machine learning (ML) is the science of getting computers to learn and act like humans do and improve their learning over time in autonomous fashion, by feeding those data and information in the form of observations and real-world interactions.

Introduction

Wireless Sensor Networks (WSN) is a vital component in Internet of Things (IoT). The small sized, low powered sensors are capable of monitoring and collecting data from environment. Most of the recent research works have not concentrated to provide a solution for analyzing the potentially huge amount of data generated by these sensor nodes. Thus, there is a need for machine learning (ML) algorithms in WSNs. When volume, velocity, and variety of data generated by WSN is very high, then data analytics tools are needed for data aggregation and clustering. ML tools are used in several applications such as intrusion detection, target tracking, healthcare, home automation, smart city. The main aim of this entry is to provide a basic knowledge in machine learning and its applications in WSN.

Background

Generally, the designers of sensor network symbolize machine learning as a branch of artificial intelligence and it is a collection of algorithms that is capable of creating prediction models. On the other hand, ML experts characterize it as a field, which is having huge amount of patterns and themes useful in sensor network applications (Alsheikh et al. 2014).

Supervised Learning

Supervised learning is merely a formulation of the concept of learning from examples. Supervised learning approach is used to resolve various issues for WSNs such as event detection, objects targeting, localization (Shareef et al. 2008), processing of query, Medium Access Control (MAC), intrusion detection, security, data integrity, and QoS.

Unsupervised Learning

No output vectors or labels are provided in unsupervised learning. Data sets are classified by finding the similarities among them. Unsupervised learning algorithm determines the concealed rela-

tionships and it can be used for solving the problems in WSN, where relationship between the variable is complex. These types of algorithms are mainly used for clustering and collection of data in WSN.

Reinforcement Learning (RL)

It allows the agent to learn from the environment by interacting with it. Here, sensor nodes learn to capture the best measurement in order to maximize the advantage. The most famous reinforcement learning algorithm is Q-learning, in which every node tries to extract measurements which are expected to increase the rewards. Sensor nodes regularly update the bonuses it has achieved derived from the actions taken at a given state. Total rewards of future can be computed by the equation:

$$Q\left(s_t+1, \alpha_t+1\right) = Q\left(s_t, \alpha_t\right) + \gamma\left(\mathrm{r}\left(s_t, \alpha_t\right) - Q\left(s_t, \alpha_t\right)\right) \quad (1)$$

(,) indicates the incentive for taking the action a_t at state s_t and γ is the rate of learning, which decides how frequently the learning happens (between the values 0 and 1).

Machine Learning for Localization

A sensor node recognizes the position either by using GPS device or by any specific localization methods. One of such method is to collect the knowledge (e.g., connectivity, pair-wise distance measure) on the subject of the whole network into single place, where the composed information is managed in a centralized manner and identify the nodes' locations using mathematical algorithms.

Security and localization are the two main applications of Support Vector Machine in WSN. The localization method (Tran and Nguyen 2008) for mobile node uses SVM and knowledge about connectivity capabilities. By using RSSI metric (Received signal strength indicator), it detects the node location. Despite the fact that Localized SVM is capable for distributed localization in a speedy manner, it is still susceptible to outliers in training data set.

A localization system based on Self-Organizing Map (SOM) (Paladina and Paone 2007) provides a few artificial intelligence (AI) features to sensor nodes. Without any supervision, SOM can able to learn how to classify. Self-Organizing Map is used on each sensor nodes in order to evaluate the position of the node. Eight anchor nodes spatial coordinates surrounding the unspecified node form the input layer. Output layer gives the special coordinates in 2D space of the unspecified node after the training. Demerit of this method is that nodes should be organized consistently throughout the monitoring area.

An improved localization using Support Vector Regression (SVR) (Wang et al. 2016) discusses about a novel extraction method of training data for localization model. It improves the accuracy of localization method without affecting the hardware cost. The location of the unknown node is found accurately using SVM-based learning with minimum number of anchor nodes (Mary Livinsa and Jayashri 2015). Accuracy is provided by finite size of grid cells. For larger scale networks, SVM-based DV-hop is suitable.

Machine Learning for Clustering and Data Aggregation

It is really difficult to transfer entire data to sink instantly in a large scale energy-constrained sensor network. An effective solution for this problem is to send the data to an aggregator node, which is also called as cluster node. The node aggregates information from all members of the cluster and sends to the sink or base station, thus reduces the energy consumption. Numerous algorithms are actually suggested for the election of Cluster Head to increase the energy efficiency. The machine learning dependent strategies can develop the advantages of clustering and aggregation of data between nodes in WSNs through several procedures.

- ML algorithm is used to identify and remove the nonfunctional nodes from routing schemes

and compress data at CH using dimensionality reduction method.
- Machine learning algorithms are executed for electing the Cluster Head efficiently, where election of suitable cluster head will considerably decrease the usage of energy and augment the network's lifespan.

The CH election problem can be solved by the application of Decision Tree algorithm (Ahmed et al. 2010) (DT). By considering prominent features as battery, distance, and mobility, DT is electing suitable cluster head. By estimating these features, DT can provide a proficient method for identifying link reliability in Wireless Sensor Networks. Role-Free Clustering (Forster and Murphy 2010) is a clustering method with Q-learning for WSNs. The ability of the cluster head is evaluated by Q-learning algorithm with active network parameters.

A distributed spectral clustering method to group sensor nodes on their location is proposed (Muniraju et al. 2017) to avoid data congestion. This is the combination of two algorithms, distributed eigen vector computation and distributed K-Means clustering. Eigen vector of graph Laplacian is calculated by distributed power iteration technique. Clustering on eigen vector is done by distributed K-Means algorithm. In order to establish the network topology, only the location information of sensor nodes is used and this information is not exchanged in the network. In (Park et al. 2013), energy efficiency is maximized by finding cluster head with k means algorithm.

The Self-Organizing Map (SOM) is unsupervised algorithm method that learns to map from high to low dimensional space. In Cluster-based self-Organizing Data Aggregation (CODA) architecture (Lee and Chung 2006), the nodes classify the aggregated data j^* as the winning neuron, which is having a weight W(t) nearest to the value of input vector X(t). With CODA scheme, quality of data is increased and energy consumption and network traffic are reduced.

Principal Component Analysis (PCA) algorithm is used for dimensionality reduction and famous in the field of data compression. Principle components are a set of orthogonal variables. Main aim is to select the significant information from data concerning the principal components. Both data compression and dimensionality reduction are multivariate methods. This can reduce the amount of data transmission between the nodes in WSN. By selecting only the momentous principal components and tossing out the lower order components, it solves the big data problem into small data. For improving the process of data aggregation, the combination of following two significant algorithms is used along with Principal Component Analysis (PCA) in WSNs (Morell et al. 2016).

- Compressive Sensing (CS): The conventional scheme of "sample then compress" is recently replaced by CS. Compressive Sensing uses sparsity property of signal.
- Expectation-Maximization (EM): This algorithm comprises of two stages, an expectation stage and maximization stage. While performing its expectation (E) stage, EM evaluates the cost function by saving the current expectation of parameters. In the maximization (M) stage, it re-computes parameters that can minimize the estimation error.

Adaptive Learning Vector Quantization (ALVQ) is used to extract compressed model of information from the nodes accurately (Lin et al. 2009). ALVQ uses the LVQ learning algorithm with previous training data for predicting the code-book. This method reduces the needed bandwidth while transmitting data and improves the accuracy of correct reading restoration from the compressed information.

Machine Learning for Routing

Considering features such as memory and computational requirements, communication costs, wireless ad-hoc nature, restricted energy, mobility and topology changes, different ML algorithms are used for energy-efficient routing in WSNs.

In fuzzy logic-based routing method (Arabi 2010), the entire network is grouped into different clusters and the election of suitable cluster head

is based on fuzzy variables. The routing in WSNs depends on the combination of hybrid routing methods and fuzzy logic for improved energy savings and improved network life span. Energy and Delay Efficient Routing protocol for sensor network (EDEAR) (Sharma and Shukla 2012) is adaptive routing method (Kumar and Kumar 2010) using Reinforcement Learning (RL). It finds the best path with minimum energy consumption and end to end delay. RL renews routing tables and thus considers all the dynamic parameters that characterize the traffic. The adaptation of routing depends on the varying traffic conditions and hence minimizes data transfer time.

Energy-aware QoS Routing algorithm using Reinforcement Learning (EQR-RL) (Jafarzadeh and Moghaddam 2014) is able to balance QoS requirements and improve network life time. Using a random load balancing algorithm, next hop neighbor is chosen. QoS parameters such as number of hops, latency, and geographical distance are considered. Dynamics of the network is learned using Q-Learning technique and routing decision is made accordingly; however, the network state information is not maintained. Sensor Intelligence Routing (SIR) using Self-Organizing Map (Barbancho et al. 2008) detects optimal routing path. The combination of SOM and Dijkstra's algorithm model offers QoS guarantees like throughput, latency, duty cycle, and packet error rate.

Example Application Scenarios

Regression-Based Adaptive Incremental Algorithm for Health Abnormality Prediction (RBAIL)

This method employs a regression based incremental algorithm for providing the learning features (Srinivasan et al. 2013). The incremental learning system is wirelessly connected to the patient and receives the stream of input parameters from patient for a fixed time of interval. RBAIL algorithm performs regression on the important health parameters for predicting the indiscretion of patient. System uses history of the patient to check whether previous anomalies were there in order to get the updates and feedbacks. Main features here are aggregation, learning, and prediction. Correctness of the parameter is verified during aggregation. If previous data and current input data are valid and parameter value is greater than a threshold value, then abnormality is detected. If the difference of current and predicted value becomes greater than a threshold, then doctor provides feedback to correct the learning algorithm.

Object Detection and Tracking in Wireless Sensor Networks

Prediction logic is used to predict the exact location of the sink node using current location (Vaidehi et al. 2011). The estimated position is sent to Cluster Head to wake up the node which is in sleep mode. The combination of sleep wake scheduling, clustering, tracking, prediction logic, and shortest path routing minimize the energy consumption in sensor networks. Sink nodes awake cluster head that helps to reach target. Further complex events processing engine is used for detecting abnormal events in a multisensor scenario (Bhargavi and Vaidehi 2013; Gao et al. 2012).

Semantic Intrusion Detection System by Means of Pattern Matching and State Transition Analysis (SIDS)

Semantics Intrusion Detection System (SIDS) (Sri Ganseh et al. 2011) combines pattern matching, state transition, and data mining for increasing the accuracy of intrusion detection. Multiple sensors are deployed in the sensor area. The events generated by sensors are correlated in time spatial domain. The outputs from the sensors are symbolized as patterns and states. When the patterns generated by sensors violate the rule, it is detected as an intrusion. Semantics rules are developed using Another Tool for Language Recognition (ANTLR).

Online Incremental Learning Algorithm for Anomaly Detection and Prediction in Health Care

An Online Incremental Learning Algorithm (OILA) is proposed (Kirthana and Bhargavi 2014) for processing the data in online. It uses the combination of regression and feedback mechanism in order to decrease the prediction error and hence improves accuracy. The vital health parameters are received from the body sensors of a patient. Online Incremental Algorithm evaluates several parameters based on the received data and checks whether any anomalies are found. An alert is set to the doctor, if any anomalies are detected. Regression-based method is used to predict next instance. Prediction of each patient is personalized according to his/her health parameters. This algorithm calculates overall trend by *longvalue* and recent trend by *shortvalue* in health parameters. The parameters *maxthresh* and *minthresh* captures maximum and minimum threshold value of tolerance. Difference between *maxthresh* and *minthresh* is captured by a parameter *diffthresh*. Patient *sensitivity range* can be defined through sensitivity range parameter by the doctor. *History factor* is a parameter that defines number of times a patient affected to abnormalities. After reading every new instance, these parameters are updated, error is adjusted, and according to that prediction is made. The algorithm predicts the abnormality using updated parameters and triggers alert.

A Genetic Approach for Personalized Healthcare

Genetic Algorithm-based Personalized Healthcare System (GAPHS) (Vaidehi et al. 2015), uses a sensor integrated wearable chest strap for the non-invasive monitoring of physiological parameters and body parameters. Wrist wear wireless Blood Pressure (BP) sensor is used for monitoring blood pressure. A fingertip wearable oxygen saturation level (SPO2) sensor is used to detect blood oxygen saturation level. The abnormality levels of the vital parameters are classified into very low (VL), low (L), medium (M), high (H), and very high (VH) and encoded into a 5-bit representation to determine the severity level of the patient. Using fitting function, the best chromosome that represents the personalized vital parameter of the patient is obtained. The proposed GAPHS provides an intelligent, personalized, and efficient healthcare system to serve the needy patient in right time by the doctor.

Dynamic Higher Level Learning Radial Basis Function for Healthcare Application

Traditional Radial Basis Function (RBF) has issues with using complete training set and large number of neurons. Due to these issues, computation time and complexity are increased. Dynamic Higher Level Learning RBF (DHLRBF) (Chandraskar et al. 2014) is applied to health parameters to find normal and abnormal category. The DHLRBF uses both cognitive and higher level learning component for effective classification with less complexity.

Sensor Based Decision Making Inference System for Remote Health Monitoring

Most of the existing methods have difficulty to differentiate between original and fall like patterns. Intelligent Modeling technique, Adaptive Neuro-Fuzzy Inference System (ANFIS) (Dhivya Poorani et al. 2012) is used for detecting the fall automatically with higher accuracy and less complexity. The data received form 3 axis accelerometer is categorized into five states (sit, stand, walk, lie, and fall) using ANFIS model. Mean, median, and standard deviation are selected for training the neural network. When the state is detected as fall, it examines ECG and heart rate of patient to check the abnormal condition and raise alarm.

Future Direction

The new revolution, Internet of Things is rapidly gathering momentum driven by advances in Sensor Networks and cloud technologies. The cloud has to provide security for sensed data. The storage for sensed data needs to be minimized by compression schemes. Multisensor data fusion is used for situation, object and threat refinement. The complex events sensed by sensor network

can be analyzed with data analytics tools to find the trends, patterns, and gain new insights and knowledge for variety of applications.

Cross-References

- Data Aggregation
- Fingerprinting Localization
- Routing

References

Ahmed G, Khan NM, Khalid Z, Ramer R (2010) Cluster head selection using decision trees for wireless sensor networks. In: IEEE international conference on intelligent sensors, sensor networks and information processing, Sydney

Alsheikh MA, Lin S, Niyato D, Tan HP (2014) Machine learning in wireless sensor networks: algorithms, strategies, and applications. IEEE Commun Surv Tutor 16(4):1996–2018

Arabi Z (2010) HERF: a hybrid energy efficient routing using a fuzzy method in wireless sensor networks. In: International Conference on Intelligent and Advanced Systems (ICIAS), Manila

Barbancho J, León C, Molina F, Barbancho A (2008) A new QoS routing algorithm based on self-organizing maps for wireless sensor networks. Telecommun Syst 36:73–83

Bhargavi R, Vaidehi V (2013) Semantic intrusion detection with multisensor data fusion using complex event processing. Sadhana 38(2):169–185

Chandraskar JB, Ganapathy K, Vaidehi V (2014) Dynamic higher level learning radial basis function for healthcare application. In: International conference on recent trends in information technology, Chennai

Dhivya Poorani V, Ganapathy K, Vaidehi V (2012) Sensor based decision making inference system for remote health monitoring. In: International conference on recent trends in information technology, Chennai

Forster A, Murphy AL (2010) CLIQUE: role-free clustering with Q-learning for wireless sensor networks. In: 29th IEEE international conference on distributed computing systems, Montreal

Jafarzadeh SZ, Moghaddam MHY (2014) Design of energy-aware QoS routing algorithm in wireless sensor networks using reinforcement learning. In: 4th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad

Kirthana R, Bhargavi VV (2014) Online incremental learning algorithm for anomaly detection and prediction in health care. In: International Conference on Recent Trends in Information Technology (ICRTIT), Chennai

Kumar N, Kumar M (2010) Neural network based energy efficient clustering and routing in wireless sensor networks. In: First international conference on networks & communications, Chennai

Lee SH, Chung TC (2006) Data aggregation for wireless sensor networks using self-organizing map. In: International conference on AI, simulation and planning in high autonomy systems, Berlin

Lin S, Kalogeraki V et al (2009) Online information compression in sensor networks. In: IEEE international conference on communications, Istanbul

Mary Livinsa Z, Jayashri S (2015) Localization with beacon based support vector machine in wireless sensor networks. In: International conference on robotics, automation, control and embedded systems (RACE), Chennai

Mingyan Gao, Ramesh Jain et al (2012) Eventshop: from heterogeneous web streams to personalized situation detection and control. In: Proceedings of the 4th annual ACM web science conference, pp 105–108

Morell A, Correa A et al (2016) Data aggregation and principal component analysis in WSNs. IEEE Trans Wirel Commun 15(6):3908–3919

Muniraju G, Zhang S, Tepedelenlio C (2017) Location based distributed spectral clustering for wireless sensor networks. In: Sensor Signal Processing for Defense Conference (SSPD), London

Paladina L, Paone M (2007) Self-organizing maps for distributed localization in wireless sensor networks. In: 12th IEEE symposium on computers and communications, Las Vegas

Park GY, Kim H, Jeong HW, Youn HY (2013) A novel cluster head selection method based on K-means algorithm for energy efficient wireless sensor network. In: WAINA'14 proceedings of the 27th international conference on advanced information networking and applications, Barcelona

Shareef A, Zhu Y, Musavi M (2008) Localization using neural networks in wireless sensor networks. In: Proceedings of the 1st international conference on mobile wireless middleware, operating systems, and applications, Turkey

Sharma VK, Shukla SSP (2012) A tailored Q-learning/or routing in wireless sensor networks. In: 2nd IEEE international conference on parallel, distributed and grid computing, Solan

Sri Ganseh K, Shekhar R, Vaidehi V (2011) Semantic intrusion detection system using pattern matching and state transition analysis. In: IEEE-International Conference on Recent Trends in Information Technology, ICRTIT, Chennai

Srinivasan S, Bhargavi R, Ramkumar K, Vaidehi V (2013) An incremental algorithm technique for health abnormality prediction. In: EEE International Conference on Recent Trends in Information Technology, ICRTIT, Chennai

Tran D, Nguyen T (2008) Localization in wireless sensor networks based on support vector machines. IEEE Trans Parallel Distrib Syst 19(7): 981–994

Vaidehi V, Sandhya M, Karthika J (2011) Power optimization for object detection and tracking in wireless sensor networks. In: IEEE-International Conference on Recent Trends in Information Technology, ICRTIT, Chennai

Vaidehi V, Ganapathy K, Raghuraman V (2015) A genetic approach for personalized healthcare. In: IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), Halifax

Wang M, Wen-xin F, Ya-dong L, Heng-wei L (2016) An improved localization for wireless sensor network using support vector regression. In: IEEE International Conference on Computational Electromagnetics (ICCEM), Guangzhou

Applications and Architectures of Social Internet of Things

Chen Zhang[1,2,3] and Yawei Wang[2]
[1]Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada
[2]The George Washington University, Washington, DC, USA
[3]Southeast University, Nanjing, China

Synonyms

Social internet of things (SIoT)

Historical Background

Internet of Things (IoT) has been an integral component of our lives and thriving with new products and services. This term was first introduced by Kevin Ashton at Procter & Gamble (*P&G*) in 1999 (Ashton et al. 2009). He presented the idea that computers could be empowered to observe, identify, and understand the physical world with RFID (radio-frequency identification) and sensor technology. By now, this definition has evolved continually with the development of embedded systems, wireless sensor networks, machine learning, etc. As defined in ITU (2012), IoT is the network of interconnecting physical and virtual things which have embedded information and communication technologies that allow these things to interact and exchange data. In addition to the personal devices such as smartphones, tablets, and laptops, IoT objects have been widely used in others fields which include, but not restricted to, agriculture (Mekala and Viswanathan 2017), smart city (Arasteh et al. 2016), education (Gul et al. 2017), manufacturing (Yang et al. 2018), medical and healthcare (Joyia et al. 2017), etc.

Existing technologies have enhanced the computing capacity of IoT objects, but most of them embedded in specific and necessary hardware features are only able to finish a particular task. In Guinard et al. (2010), the authors propose a platform that enables an individual to share the services offered by his smart objects with his friends or their objects through social networks of humans. Using interconnected objects to extend the Internet and consequently leading to efficient conversations is addressed in the study (Mendes 2011). In Atzori et al. (2011), the concept of *Social Internet of Things (SIoT)*, which is a social network among the objects of IoT, is formalized. Generally, a social network for human beings is a platform (like Facebook, Twitter, Instagram, etc.) which people and organizations can share user-generated content such as videos, photos, comments, posts, location, and so on. Members of this kind of platform prefer to interact and build social relationships via the Internet with those who share similar backgrounds, interests, or friends. For objects, the SIoT is a platform for global interconnected IoT objects to obtain social-like capabilities, which allow these objects to autonomously establish social relationships, request or provide services, and collaborate with others to achieve a common goal in an autonomous way with proper authorization from users.

As a new paradigm of IoT, SIoT introduces the social relationship concept among IoT objects. This does not mean that a common architectural model of social network services for human beings can be directly applied to SIoT. The differences between main components of social network platforms for humans (Boyd and Ellison 2007) and IoT objects are presented in Atzori

et al. (2011). And the explanations of main components of social network platforms for IoT objects are organized as follows:

- *ID management (ID)*: assigns a unique ID to each IoT object and manage existing objects.
- *Object profiling (OP)*: contains static and dynamic information about the object.
- *Owner control (OC)*: sets up rules to carry out access control, relationship control, and any other operations.
- *Relationship management (RM)*: is an important component based on the user-defined rules to establish, update, and terminate relationships among objects in a network.
- *Service discovery (SD)*: finds objects that can provide required services from the network, a very similar way to searching for friends and information from an online social network.
- *Service composition (SC)*: enables interaction between objects.
- *Trustworthiness management (TM)*: provides trust evaluation among objects for relationship management and information sharing.

In a SIoT network, several kinds of relationships can be defined among objects. The types and characteristics of social relations among IoT objects are systematically summarized in Atzori et al. (2012):

- *Parental object relationship (POR)*: established when objects are created in the same period by the same manufacturer.
- *Co-location object relationship (CLOR)*: established when objects are deployed in the same location.
- *Co-work object relationship (CWOR)*: established if the objects work on the same task.
- *Ownership object relationship (OOR)*: established when objects have the same owner.
- *Social object relationship (SOR)*: established when the owners get in touch and the objects are active.

These relationships among objects in SIoT should be established and managed *without human intervention*. Several SIoT architectures based on the above components and relationships are introduced in the next section.

SIoT Architectures

A three-layered architecture model proposed in Zheng et al. (2011) for IoT should be introduced before the SIoT architectures, since several existing SIoT architectures are built upon this IoT architecture. In this architecture model, the sensor data are acquired by sensing devices in the *Sensing Layer*. Then these data are safely and reliably transmitted to the *Application Layer* through the *Network Layer*. The *Network Layer* is devoted to transferring data across different networks such as cellular networks, WLAN (wireless local area network), satellite communication networks, and so on. By applying distributed computing technologies to massive data, the *Application Layer* performs data processing and analysis intelligently to support intelligent control in the IoT.

The following are SIoT architectures summarized from existing research:

- An architecture consisting of the Server Side and Client Side is proposed in Atzori et al. (2011). This architecture has three layers on both sides. On the Client Side, the Objects Layer consists of different physical objects. The Object Abstraction Layer is used to harmonize communications among physical objects. The top layer on the Client Side consists of the Social Agent element and the Service Management element. The Social Agent element is designed for friendships and object profile updates, service discovery, and service request from the network. Management element provides interface for humans to control the object behaviors. On the Server Side, the Base Layer provides data storage and data management. The Component Layer includes the main components of SIoT (OP, DC SD, TM, ID, RM, and SC). Application Layer consists of interfaces to human, objects, and third-party services.

- A three-layer SIoT architecture is proposed in Atzori et al. (2012). This proposed system consists of three subsystems: the SIoT Server, the Gateway, and the Object. The SIoT Server has a Network Layer and an Application Layer. The Application Layer includes three sub-layers. The database of data storage, data management, and relevant descriptors is in the Base Sub-layer. The main components of the SIoT system discussed in the previous section are located in the Component Sub-layer; they provide main capabilities for the SIoT system. The Interface Sub-layer consists of the third-party interfaces for objects, humans, and services. The Gateway and the Object both have Sensing Layer, Network Layer, and Application Layer. The characteristics of devices decide the combination of these layers. But no matter what the combination is, the Application Layer includes SIoT Applications, the Social Agent, and the Service Management. The Social Agent communicates with the SIoT server to update its profile and friendship and also provides service discovery and service request. When it is necessary, it is devoted to direct communications between objects which are close to each other in the physical world as well. The Service Management Agent provides interfaces for humans to control the behaviors of the objects.
- An architecture with Actors, the Intelligent system, the Interface, and the Internet is introduced in Ortiz et al. (2014). The Internet is the foundation of this architecture; Actors in the architecture refer to humans and things who have equal rights to take part in managing data, executing queries, and receiving services and commands. The Intelligent system is devoted to coordinating the interactions between Actors; it also has capabilities of recommendation, data and context management, service discovery and search, and service and application management. The Interface provides ports for inputs and outputs of the interactions with the system.
- A distributed sensor storage for SIoT is presented in Wu et al. (2015). The Sensing Layer is responsible for collecting data from the physical world through sensors and other objects. The Communication Layer transfers the sensing data collected from Sensing Layers to the users via the Internet, mobile network, etc. The Application Layer consists of consumer applications, enterprise applications, and government applications. Different services can provide data for different users. In this architecture, the IoT, consisting of the Internet and sensor network, can be organized into SIoT based on the social network. In SIoT scenarios, the sensors with large storage can help normal sensors which can only sense the environment to store the sensing data. The storage should be able to repair the lost fragment and to protect data secrecy.
- A four-layered SIoT architecture is introduced in Gulati and Kaur (2019). In this architecture, the Base Layer called Object Layer is where the IoT devices and sensors are deployed. Communication and collaboration among these objects are through local sensor networks. Communication technologies and protocols used by SIoT are defined in the Communication Layer. The SIoT Management Layer defines the SIoT platforms to manage SIoT services. The main components of SIoT that were introduced in the previous section are in this layer. The top layer is the Application Layer which provides APIs (application programming interfaces) for SIoT applications.

Key Applications

Following the theoretical development of SIoT, several projects have been proposed to integrate IoT objects into a social network. Existing SIoT platforms such as *Toyota Friend Network, Nike+, Xively*, and *Paraimpu* are summarized in Atzori et al. (2014). Toyota Friend Network is a private social network that allows Toyota customers to create a virtual community among them and interact with their own cars. For example, customers can get a maintenance alert from their cars and can connect to the dealerships to get service information through this social network. Nike+ provides a platform for customers to share their fitness data in a social network. Xively

and Paraimpu are two similar platforms which enable objects to be connected to each other or to be linked to existing social networks so that these objects can respond to their owners' requests. Furthermore, Paraimpu allows users to share the data generated by objects with their friends. A protocol for traffic management in *social Internet of Vehicles (SIoV)* (an extension to the concept of SIoT) is presented in Jain et al. (2018). This protocol can improve the performance of data transmission over vehicular social networks. Other cases that use SIoT are illustrated in Afzal et al. (2019), for instance, smart shopping mart retailing, healthcare and telemedicine, traffic surveillance and road safety, and so on. In Fu et al. (2017), an intelligent painting device intended for children is designed based on SIoT. This device aims at creating a social network consisting of humans, objects, and emotion to facilitate the children's intellectual development and personality development.

Security and Trust in SIoT

As IoT devices are becoming more ubiquitous, IoT security has been a growing concern. Several IoT security incidents make people realize the importance of IoT security. In 2016, the largest DDoS attack was launched by an IoT botnet (Goodin 2016), a malware called *Mirai* that brought people's attention to the seriousness of IoT security. Beyond that, there are even more serious attacks related to daily life. For example, the vulnerable web camera at home can be used to pry into the owner's daily life by the attacker (Whittaker 2015). A Jeep over the Internet can be hijacked, and the attacker can gain access to many functions that could be used to cause a traffic accident (Rouse 2015). In 2017, the Food and Drug Administration (FDA) warned that cybersecurity vulnerabilities are identified in several medical devices (FDA 2017).

In order to address the security threats in IoT, a series of studies are carried out from different aspects including encryption, authentication, access control, network security, and application security (Zhou et al. 2018; Alaba et al. 2017; Diro and Chilamkurti 2018; Jia et al. 2018; Liang et al. 2018; Zheng et al. 2018). In addition, object diversity and complex relationships among objects make security communication and trustworthiness management become more important and serious for SIoT (Nitti et al. 2014). Furthermore, related policies and laws should also be formulated to regulate the ethical use of SIoT (Berman and Cerf 2017).

Conclusion

As a new paradigm of the Internet of Things (IoT), the Social Internet of Things (SIoT) will create a brand-new way of communication and interaction for IoT objects. This article systematically introduces the definition and components of SIoT as well as the existing architectures and applications. Furthermore, security and trust issues which could bring great challenges to the development of SIoT are discussed at the end of this article.

Cross-References

- Internet of Medical Things
- Social IoT Crowd-Sourcing on Disaster Reduction

References

Afzal B, Umair M, Shah GA, Ahmed E (2019) Enabling IoT platforms for social IoT applications: vision, feature mapping, and challenges. Futur Gener Comput Syst 92:718–731

Alaba FA, Othman M, Hashem IAT, Alotaibi F (2017) Internet of things security: a survey. J Netw Comput Appl 88:10–28

Arasteh H, Hosseinnezhad V, Loia V, Tommasetti A, Troisi O, Shafie-Khah M, Siano P (2016) IoT-based smart cities: a survey. In: 2016 IEEE 16th international conference on environment and electrical engineering (EEEIC). IEEE, pp 1–6

Ashton K et al (2009) That "internet of things" thing. RFID J 22(7):97–114

Atzori L, Iera A, Morabito G (2011) Siot: giving a social structure to the internet of things. IEEE Commun Lett 15(11):1193–1195

Atzori L, Iera A, Morabito G, Nitti M (2012) The social internet of things (siot)–when social networks meet the internet of things: concept, architecture and network characterization. Comput Netw 56(16):3594–3608
Atzori L, Iera A, Morabito G (2014) From "smart objects" to "social objects": the next evolutionary step of the internet of things. IEEE Commun Mag 52(1): 97–105
Berman F, Cerf VG (2017) Social and ethical behavior in the internet of things. Commun ACM 60(2):6–7
Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput-Mediat Commun 13(1):210–230
Diro AA, Chilamkurti N (2018) Distributed attack detection scheme using deep learning approach for internet of things. Futur Gener Comput Syst 82:761–768
FDA (2017) Cybersecurity vulnerabilities identified in st. jude medical's implantable cardiac devices and merlin@home transmitter: FDA safety communication. https://www.fda.gov/medical-devices/safety-communications/cybersecurity-vulnerabilities-identified-st-jude-medicals-implantable-cardiac-devices-and-merlinhome
Fu Z, Lin J, Li Z, Du W, Zhang J, Ye S (2017) Intelligent painting based on social internet of things. In: International conference on distributed, ambient, and pervasive interactions. Springer, pp 335–346
Goodin D (2016) Record-breaking DDoS reportedly delivered by >145k hacked cameras. Ars Technica 28. https://arstechnica.com/?post_type=post&p=966459 Accessed 10 February 2019
Guinard D, Fischer M, Trifa V (2010) Sharing using social networks in a composable web of things. In: PerCom workshops, pp 702–707
Gul S, Asif M, Ahmad S, Yasir M, Majid M, Malik M, Arshad S (2017) A survey on role of internet of things in education. Int J Comput Sci Netw Secur 17(5): 159–165
Gulati N, Kaur PD (2019) When things become friends: a semantic perspective on the social internet of things. In: Smart innovations in communication and computational sciences. Springer, pp 149–159
ITU (2012) Overview of the internet of things. http://handle.itu.int/11.1002/1000/11559
Jain B, Brar G, Malhotra J, Rani S, Ahmed SH (2018) A cross layer protocol for traffic management in social internet of vehicles. Futur Gener Comput Syst 82: 707–714
Jia Y, Xiao Y, Yu J, Cheng X, Liang Z, Wan Z (2018) A novel graph-based mechanism for identifying traffic vulnerabilities in smart home IoT. In: IEEE INFOCOM 2018-IEEE conference on computer communications. IEEE, pp 1493–1501
Joyia GJ, Liaqat RM, Farooq A, Rehman S (2017) Internet of medical things (ioMT): applications, benefits and future challenges in healthcare domain. J Commun 12:240–247
Liang Y, Cai Z, Yu J, Han Q, Li Y (2018) Deep learning based inference of private information using embedded sensors in smart devices. IEEE Netw 32(4): 8–14
Mekala MS, Viswanathan P (2017) A survey: smart agriculture IoT with cloud computing. In: 2017 international conference on microelectronic devices, circuits and systems (ICMDCS). IEEE, pp 1–7
Mendes P (2011) Social-driven internet of connected objects. IAB workshop on interconnecting smart objects with the internet
Nitti M, Girau R, Atzori L (2014) Trustworthiness management in the social internet of things. IEEE Trans Knowl Data Eng 26(5):1253–1266
Ortiz AM, Hussein D, Park S, Han SN, Crespi N (2014) The cluster between internet of things and social networks: review and research challenges. IEEE Internet Things J 1(3):206–215
Rouse M (2015) What is car hacking? Definition from whatis.com. https://internetofthingsagenda.techtarget.com/definition/car-hacking
Whittaker Z (2015) New security flaws found in popular IoT baby monitors. https://www.zdnet.com/article/security-vulnerability-flaw-internet-things-baby-monitors
Wu J, Dong M, Ota K, Liang L, Zhou Z (2015) Securing distributed storage for social internet of things using regenerating code and blom key agreement. Peer-to-Peer Netw Appl 8(6):1133–1142
Yang C, Shen W, Wang X (2018) The internet of things in manufacturing: key issues and potential applications. IEEE Syst Man Cybern Mag 4(1):6–15
Zheng L, Zhang H, Han W, Zhou X, He J, Zhang Z, Gu Y, Wang J et al (2011) Technologies, applications, and governance in the internet of things. Internet of things-global technological and societal trends From smart environments and spaces to green ICT
Zheng X, Cai Z, Li Y (2018) Data linkage in smart internet of things systems: a consideration from a privacy perspective. IEEE Commun Mag 56(9):55–61
Zhou R, Zhang X, Du X, Wang X, Yang G, Guizani M (2018) File-centric multi-key aggregate keyword searchable encryption for industrial internet of things. IEEE Trans Ind Inform 14(8):3648–3658

Applications of Molecular Communication Systems

Tadashi Nakano, Yutaka Okaie, and Takahiro Hara
Osaka University, Suita, Japan

Synonyms

Molecular, biological, and multiscale communications; Nano-networks

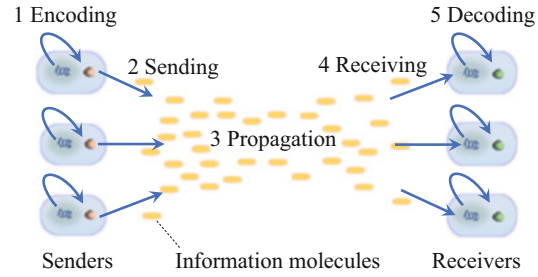
Definitions

A molecular communication system is defined as a system of bio-nanomachines that transmit and receive information using chemical signals or molecules. A bio-nanomachine that constitutes a molecular communication system is made of biomaterials with or without non-biological materials, approximately 1–100 μm in size, and capable of processing molecules. Examples of molecular communication systems are naturally occurring biological systems such as bacterial populations, epithelial sheets, and immune systems where biological cells represent bio-nanomachines. Examples of molecular communication systems also include artificial or synthetic biological systems designed for specific applications such as biomolecular sensing and targeted drug delivery.

Historical Background

Molecular communication was proposed as an unexplored research area at the intersection of communications engineering and biology (Nakano 2017). The basic concept of molecular communication is simple (Fig. 1); sender and receiver bio-nanomachines communicate using chemical signals or molecules. To transmit information, sender bio-nanomachines generate "information molecules" that represent information (i.e., encoding in Fig. 1) and transmit the information molecules into the environment (sending). Information molecules propagate in the environment (propagation) and react to receiver bio-nanomachines (receiving and decoding). Namely, receiver bio-nanomachines receive information from information molecules.

Molecular communication research can be divided into its first phase (2005–2009) and second phase (2010–present). The first phase of research (2005–2009) focuses on experimentally recreating functionalities and components of naturally occurring biological systems. This includes the first prototype implementation of molecule transport systems using motor proteins



Applications of Molecular Communication Systems, Fig. 1 Molecular communication (Nakano 2017)

and DNA molecules (Hiyama et al. 2008a). This also includes experimental demonstration of signal amplification mechanisms through calcium signaling (Nakano et al. 2009), self-organized microtubule networks (Enomoto et al. 2006), and encapsulation of information molecules and transport of information molecules from liposomes to biological cells (Moritani et al. 2010).

The second phase of molecular communication research (2010–present) has seen a wider variety of research topics concerning molecular communication. This phase of research includes the use of communication theory and mathematical tools to understand physical properties of molecular communication (Mahfuz et al. 2010; Farsad et al. 2012; Pierobon and Akyildiz 2013). It also includes design, analysis, and computer simulations of higher-layer mechanisms of molecular communication (Lio and Balasubramaniam 2012; Felicetti et al. 2014). It further includes standardization efforts to establish molecular communication as a standard framework and to develop simulation tools (Bush et al. 2015).

Key Applications

Functional applications of molecular communication systems are anticipated in a variety of domains (Nakano et al. 2012).

- **Biomolecular sensing**: Specific molecules in the human body serve as biomarkers

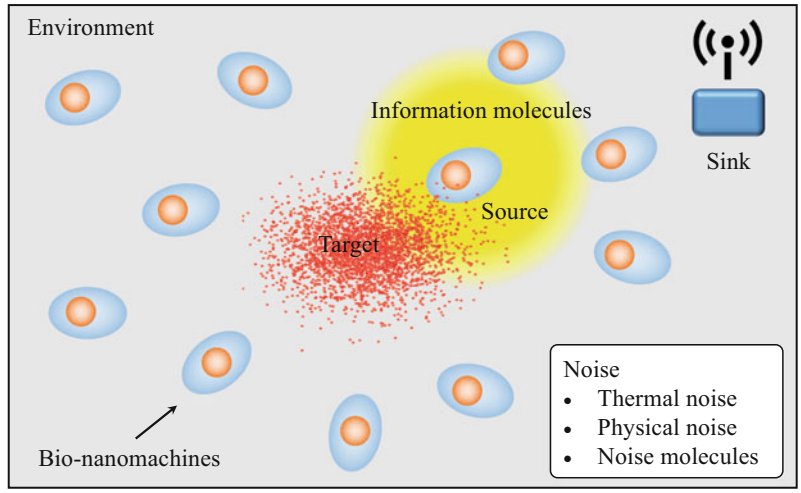
for certain diseases or medical conditions. More detailed information such as the spatial distribution of molecules is potentially useful for in-depth diagnosis. For biomolecular sensing, bio-nanomachines may sense their environment, communicate with others, and collectively determine whether specific molecules exist in their environment. Alternatively, bio-nanomachines may transmit sensed information to external devices or control units for diagnosis of their environment (Rogers and shung Koh 2016; Abdi et al. 2017).

- **Targeted drug delivery**: The delivery of drug molecules to target sites in the human body is expected in nanomedicine; it maximizes the efficacy of drug molecules and at the same time reduces potential side effects at nontarget sites. For drug delivery, bio-nanomachines may be embedded with drug molecules and either injected directly into target sites or intravenously to propagate to target sites. Bio-nanomachines may use molecular communication to collaborate to search for target sites, aggregate at the target sites, and release embedded drug molecules (Okaie et al. 2016; Wei et al. 2013).
- **Tissue regeneration**: In tissue development, biological cells communicate through synthesizing growth factor molecules and transmitting them into the environment. Growth factor molecules propagate in the environment and bind to cell surface receptors of the target cells. The concentrations and types of growth factor molecules modulate migration, proliferation, and differentiation of target cells, leading to the formation of a tissue structure. For the repair or construction of a tissue structure, bio-nanomachines made of living cells may be deployed in the human body. Bio-nanomachines use molecules to communicate with tissue-forming cells, while they divide and grow to help the tissue structure formation.
- **Internet of Bio-NanoThings**: Molecular communication systems may also be interfaced to and integrated with existing communication systems. Future mobile phones or wearable devices may be integrated with bio-nanomachines capable of molecular communication for on-chip analysis of biochemical signals (e.g., molecules in blood or from sweat) (Hiyama et al. 2008b). Further, such devices and molecular communication systems may be integrated into the Internet to form the Internet of NanoThings (Akyildiz and Jornet 2010) or the Internet of Bio-NanoThings (Akyildiz et al. 2015).

Molecular Communication System Model

Figure 2 shows a model of molecular communication systems.

- **Bio-nanomachines** are sensors and actuators to form molecular communication systems. A bio-nanomachine is characterized with the three criteria: material, size, and functionality (Nakano et al. 2012). A bio-nanomachine is composed of biomaterials with or without non-biological materials. The size of a bio-nanomachine ranges from the size of a macromolecule to that of a biological cell (i.e., up to tens of μm). A bio-nanomachine implements a set of biochemical functionalities such as sensing a certain type of molecule to achieve application-dependent goals (e.g., biomolecular sensing). Examples of bio-nanomachines are genetically engineered biological cells.
- The **environment** is a three-dimensional space where bio-nanomachines are deployed. It is typically an aqueous environment (e.g., an internal environment of the human body), containing molecules and energy sources for bio-nanomachines to operate. The environment may also contain flow by which the operation of bio-nanomachines can be disturbed or alternatively enhanced (Kadloor et al. 2012; Tavakkoli et al. 2017). The environment may generate events or information of interest, or contain targets that generate such information. For example, in biomolecular sensing applications, a change of the human body into an abnormal state



Applications of Molecular Communication Systems, Fig. 2 Molecular communication system model

represents such an event, and in targeted drug delivery applications, cancer cells represent a target.

- **Sources** are entities in the environment that provide useful information. A source may be a bio-nanomachine or a group of bio-nanomachines that detects an event or target in the environment. A bio-nanomachine functioning as a source may detect an event or target biochemically and transmit the detected information by propagating information molecules to other bio-nanomachines or sinks.
- **Sinks** are entities that receive and collect information from bio-nanomachines or sources. A sink may be a bio-nanomachine or a group of bio-nanomachines that processes information and performs application-dependent functionalities such as releasing drug molecules. A sink may also be a conventional device capable of traditional communication (e.g., wireless communication) (Nakano et al. 2014). A sink may be made from materials that are not compatible with the environment and may be orders of magnitude larger than bio-nanomachines. A sink may function as a gateway that interconnects molecular communication systems with external systems. Examples of sink devices include implantable medical devices (Kiourti et al. 2014).
- **Noise** exists in molecular communication systems in various forms. The first type of noise is thermal noise. Due to thermal noise, molecular communication among bio-nanomachines and the operation of bio-nanomachines become stochastic. The second type of noise is physical noise. The high viscosity of the environment and fluid in the environment generate physical force, making it difficult for molecules to propagate and for bio-nanomachines to move. The third type of noise is caused by molecules existing in the environment or noise molecules. Due to noise molecules, molecular communication among bio-nanomachines, and the operation bio-nanomachines can be disturbed.

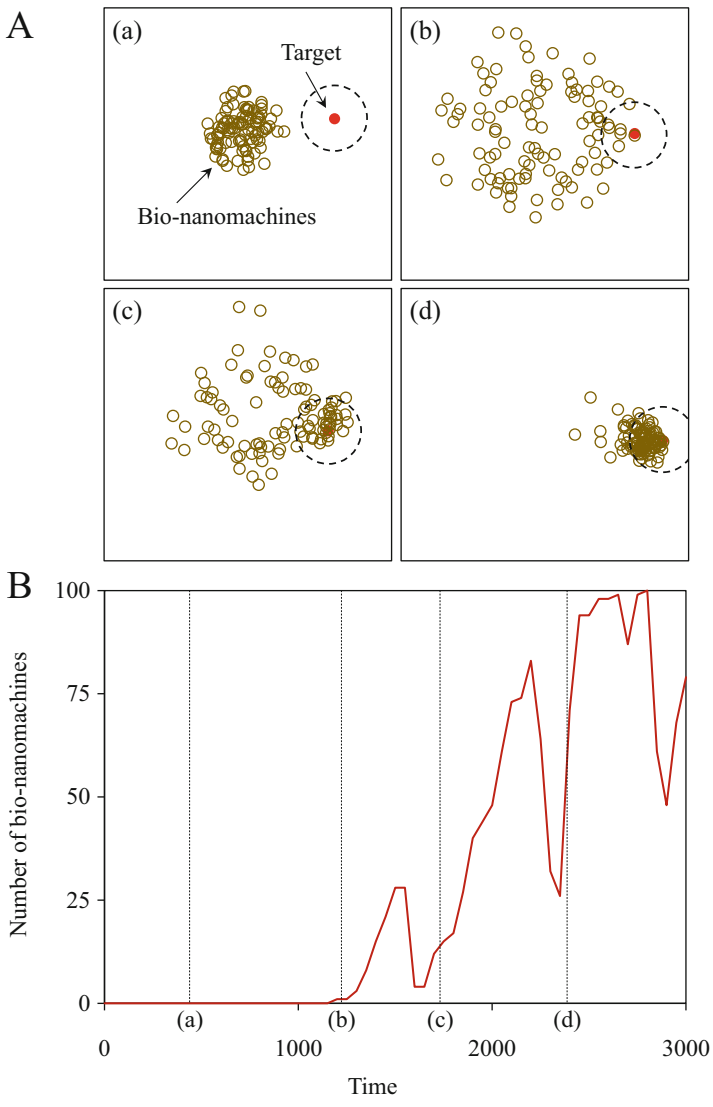
Target Detection and Tracking

Okaie et al. (2016) describes key functionalities of molecular communication systems: target detection and tracking. Target detection is a functionality of molecular communication systems to detect a target in a given environment, while target tracking is aimed at detecting and tracking targets as they move. In nanomedical applications, targets can be disease sites, pathogens, infectious microorganisms, or biochemical weapons that represent a potential threat to the environment;

the timely detection of targets and tracking of targets are important to provide immediate treatments or further analysis of the environment.

The mechanisms for target detection and tracking proposed in Okaie et al. (2016) use two types of molecule: repellents and attractants. In search of a target, bio-nanomachines release repellents to quickly spread in the environment; the released repellents form the concentration gradient in the environment, bio-nanomachines move toward lower concentrations of repellents, and therefore they move away from each other to help their search process. Upon detecting a target, they release attractants to recruit other bio-nanomachines in the environment toward the target location; the released attractants also form the concentration gradient in the environment, and bio-nanomachines move toward higher concentrations of attractants, namely, toward the target location.

Figure 3A illustrates target detection and tracking processes. Here (a) a group of bio-nanomachines is placed in an environment where a single target exists; (b) the group of bio-nanomachines first spreads in the environment using repellents, and as a result one of the



Applications of Molecular Communication Systems, Fig. 3 Target detection and tracking processes (Okaie et al. 2016). (**A**) A group of 100 bio-nanomachines is placed in an environment containing a single moving target. The bio-nanomachines communicate to distribute in the environment and aggregates around the target location. (**B**) A performance measure is given by the number of bio-nanomachines in the proximity of the target

bio-nanomachines detects the target; (c) this bio-nanomachine, upon detecting the target, starts releasing attractants, and nearby bio-nanomachines are thus attracted to the location of the bio-nanomachine (i.e., near the target); and (d) as the target moves, the group of bio-nanomachines uses repellents and attractants in the same manner to move and track the target. Figure 3B shows how the number of bio-nanomachines in the proximity of the target (namely, the number of bio-nanomachines within the circle in Fig. 3B) changes with time. The up-and-down behavior of the graph indicates target detection and tracking processes that bio-nanomachines perform.

Future Directions

The current molecular communication research focuses more on theoretical work than experimental one; wet laboratory experiments form an importance piece of future work. Wet laboratory experiments help identify practical issues and gain insight into biologically implementable designs of molecular communication systems. An objective would be to show the collective behavior of bio-nanomachines for a specific application such as target detection and tracking.

Wet laboratory experiments are however challenging for communication engineers, since experimental facilities are often not available to them. To solve this problem, tabletop molecular communication platforms are developed in (Farsad et al. 2013). Tabletop molecular communication platforms are built using macroscale devices such as electronic sprays and sensors, yet communication is performed by propagating molecules between the macroscale devices. Tabletop MC platforms help communication engineers gain experience in molecular communication and develop realistic models to analyze the unique features and characteristics of molecular communication systems.

Nonetheless, theoretical work remains important in future work. Biologically realistic modelling and computer simulations will help reduce the cost and time required to carry out wet laboratory experiments. This will also help us understand the detailed dynamics of molecular communication, such as spatiotemporal concentrations of molecules, which is difficult to observe in wet laboratory experiments. Future work should therefore use an integrated approach of experimental and theoretical investigations in order to design and develop practical applications of molecular communication systems.

Cross-References

- Connectivity via Molecular Signaling
- Drug Delivery via Nanomachines
- Modeling Approaches for Simulating Molecular Communications
- Molecular Bit Detection
- Molecular Event Detection
- Modulation in Molecular Signaling
- Nanonetworks
- Neuronal Communication Channels
- Receiver Mechanisms for Synthetic Molecular Communication Systems with Diffusion
- Reaction-Diffusion Channels

References

Abdi A, Einolghozati A, Fekri F (2017) Quantization in molecular signal sensing via biological agents. IEEE Trans Mol Biol Multi-Scale Commun 3(2): 106–117

Akyildiz IF, Jornet JM (2010) The Internet of nano-things. IEEE Wirel Commun 17(6):58–63

Akyildiz IF, Pierobon M, Balasubramaniam S, Koucheryavy Y (2015) The Internet of bio-nano things. IEEE Commun Mag 53:32–40

Bush SF, Paluh JL, Piro G, Rao V, Prasad RV, Eckford A (2015) Defining communication at the bottom. IEEE Trans Mol Biol Multi-Scale Commun 1(1):90–96

Enomoto A, Moore M, Nakano T, Egashira R, Suda T, Kayasuga A, Kojima H, Sakakibara H, Oiwa K (2006) A molecular communication system using a network of cytoskeletal filaments. In: Proceedings of 2006 NSTI nanotechnology conference and trade show, vol 1, pp 725–728

Farsad N, Eckford AW, Hiyama S, Moritani Y (2012) On-chip molecular communication: analysis and design. IEEE Trans Nanobiosci 11(3):304–314

Farsad N, Guo W, Eckford AW (2013) Tabletop molecular communication: text messages through chemical signals. PLOS ONE 8(12):e82935
Felicetti L, Femminella M, Reali G, Nakano T, Vasilakos AV (2014) TCP-like molecular communications. IEEE J Sel Areas Commun (JSAC) 32(12):2354–2367
Hiyama S, Inoue T, Shima T, Moritani Y, Suda T, Sutoh K (2008a) Autonomous loading, transport, and unloading of specified cargoes by using DNA hybridization and biological motor-based motility. Small 4(4): 410–415
Hiyama S, Moritani Y, Suda T (2008b) Molecular transport system in molecular communication. NTT DOCOMO Tech J 10(3):49–53
Kadloor S, Adve RS, Eckford AW (2012) Molecular communication using Brownian motion with drift. IEEE Trans Nanobiosci 11(2):89–99
Kiourti A, Psathas KA, Nikita KS (2014) Implantable and ingestible medical devices with wireless telemetry functionalities: a review of current status and challenges. Bioelectromagnetics 35(1):1–15
Lio P, Balasubramaniam S (2012) Opportunistic routing through conjugation in bacteria communication nanonetwork. Nano Commun Netw 3(1):36–45
Mahfuz MU, Makrakis D, Mouftah HT (2010) On the characterization of binary concentration-encoded molecular communication in nanonetworks. Nano Commun Netw 1(4):289–300
Moritani Y, Nomura S-iM, Morita I, Akiyoshi K (2010) Direct integration of cell-free-synthesized connexin-43 into liposomes and hemichannel formation. FEBS J 277:3343–3352
Nakano T (2017) Molecular communication: a 10 year retrospective. IEEE Trans Mole Biol Multi-Scale Commun 3(2):71–78
Nakano T, Koujin T, Suda T, Hiraoka Y, Haraguchi T (2009) A locally induced increase in intracellular Ca^{2+} propagates cell-to-cell in the presence of plasma membrane ATPase inhibitors in non-excitable cells. FEBS Lett 583(22):3593–3599
Nakano T, Moore M, Wei F, Vasilakos AV, Shuai JW (2012) Molecular communication and networking: opportunities and challenges. IEEE Trans NanoBiosci 11(2):135–148
Nakano T, Kobayashi S, Suda T, Okaie Y, Hiraoka Y, Haraguchi T (2014) Externally controllable molecular communication. IEEE J Sel Areas Commun (JSAC) 32(12):2417–2431
Okaie Y, Nakano T, Hara T, Nishio S (2016) Target detection and tracking by bionanosensor networks. SpringerBriefs in computer science. Springer, Singapore
Pierobon M, Akyildiz IF (2013) Capacity of a diffusion-based molecular communication system with channel memory and molecular noise. IEEE Trans Inf Theory 59(2):942–954
Rogers U, shung Koh M (2016) Parallel molecular distributed detection with brownian motion. IEEE Trans NanoBiosci 15(8):871–880
Tavakkoli N, Azmi P, Mokari N (2017) Performance evaluation and optimal detection of relay-assisted diffusion-based molecular communication with drift. IEEE Trans NanoBiosci 16(1):34–42
Wei G, Bogdan P, Marculescu R (2013) Bumpy rides: modeling the dynamics of chemotactic interacting bacteria. IEEE J Sel Areas Commun (JSAC) 31(12): 879–890

A

Architecture and Data Management for Smart Community Information Platform

Hiroaki Nishi
Department of System Design Engineering, Keio University, Yokohama, Japan

Synonyms

Data Privacy; Edge Computing; Fog Computing; Information Platform; Internet of Things (IoT); Smart City; Smart Community; Smart Community Services; Smart Infrastructure; Social Data Management; Vender and Consumer Relationship Management

Definition

Smart Infrastructure: A new style of infrastructure composed of conventional infrastructure and information and communication technologies (ICTs). Smart infrastructure refers to an efficient and highly functional infrastructure accomplished by ICT-based monitoring, control, and management.

Smart Community: An integration of smart infrastructures implemented in a certain region.

Introduction

A smart community is an integration of smart infrastructures implemented in a certain region. Each smart infrastructure may provide dedicated merits. However, the essential merit of a smart

community is community big data, which covers all infrastructures. To pursue the effective use of the community, big data is the key to the success of smart community projects. Therefore, data infrastructure is a necessity for a smart community. The information platform considering the safe data exchange and privacy preservation is described as a case study in Saitama City, Japan.

Smart Community

A smart city is an integration of smart infrastructures that are composed of conventional infrastructure and information and communication technologies (ICTs). Smart infrastructure refers to an efficient and highly functional infrastructure accomplished by ICT-based monitoring, control, and management (Nishi 2018). For example, a smart grid is composed of the interaction between the power grid and ICTs, and it accomplishes high-functioning grid operation and effective electricity usage. Smart transportation achieves automated drive and fare systems in logistics, cars, and road systems, which is also referred to as an intelligent transportation system (ITS). Smart agriculture improves the product value by controlling the growth of crops and the total efficiency of farm work. A smart government provides administrative services using the Internet to improve usability and operational efficiency. A smart city is an implementation of several smart infrastructures in a city. Similarly, a smart town and a smart island are implemented in a town and on an island, respectively. A smart community refers to a similar concept and is unrelated to the target region. Moreover, it is meaningless when these smart infrastructures are independently implemented in the target region. A smart community provides new services by integrating and linking information of different smart infrastructures. An example of this information integration is the efficient charge/discharge management for electric vehicles, combining data from an electric vehicle and the electric power system with traffic information. The intensive introduction of smart infrastructures and strong information linkage enable the provision of more advanced services to communities. The penetration of Internet of Things (IoT) has created a large amount of data, and the success of the smart community project depends on the effective use of the data. The data management for a smart community is indispensable to provide attractive smart community services.

A smart community is described in the Smart Communities Guidebook (1997) by the State University of San Diego as "a geographical area ranging in size from neighborhood to a multi-county region whose residents, organizations, and governing institutions are using information technology to transform their region in significant ways. Co-operation among government, industry, educators, and the citizenry, instead of individual groups acting in isolation, is preferred. The technological enhancements undertaken as part of this effort should result in fundamental, rather than incremental, changes." The ICT created a paradigm shift in infrastructures, and a significant amount of data processing and network transactions was generated. The data processing is primarily achieved in cloud services. However, the location of cloud causes obstacles to some services. The communication latency may cause a serious problem to hard real-time control applications. Open Fog (OpenFog Consortium 2018; Bonomi et al. 2012; Yi et al. 2015) proposed the placement of services closer to the terminal devices than the cloud for improving the efficiency of providing the services. Therefore, Fog would improve the service latency and improve the service distribution in the networks. Edge computing is used as a similar technology to Open Fog. Edge computing provides a computing environment at the edge of the Internet. This also means that the locations of the computing resources are closer to the IoT nodes than to the cloud. Data freshness is indispensable for some services, such as the ancillary service of power grids, and the automated car drive service. The study regarding Fog and Edge computing has become active as the discussions of new smart community services become popular. For data management,

FIWARE (https://www.fiware.org) (Ferreira et al. 2017) provides several types of data management API for smart community services. FIWARE provides various types of APIs to manage the smart community data; this shows the importance of data management, especially the multiple perspectives for the data management of smart city services. Herein, the smart community information platform from the perspective of its data management is focused and described.

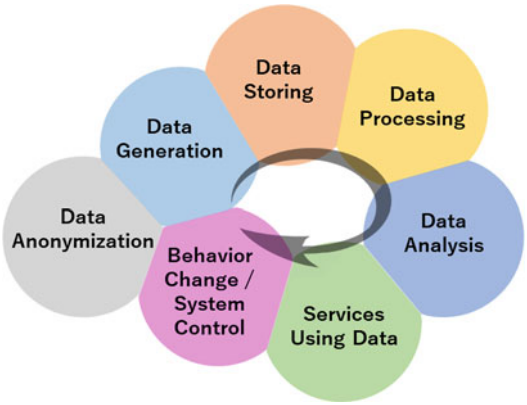
Smart Community Data Specifications and Requirements

The requirements of smart city services are diverse. When providing smart city services, data specifications are required to provide the service. Several points for data specifications and requirements are used and described as follows.

Immediate data handling: Regarding time constraints, for example, ancillary services in a smart grid do not legally allow a delay of over 1 Hz. The IEC 61000-4 standard determines that the control delay has to be smaller than 10-ms intervals. Moreover, the IEC 61850-9-2 standard determines the tolerance of the delay to be less than 1 μs. Thus, it is almost impossible to provide ancillary services as cloud services. However, it is necessary to accommodate new services that handle hard real-time system requirements, such as haptic communication for telemedicine and tactile sensing, which are sensitive to delay because they are required to transmit action–reaction force feedbacks to convey tactile sensations. This application also permits a 10-ms delay for maintaining stabilization in the control of its feedback control system (Anderson and Spong 1989). Cloud services cannot meet this need owing to their comparatively large communication delay.

Flexible data handling: Many protocols are proposed for communicating with IoT nodes. When focusing only on smart community protocols, especially the application layer, several examples can be provided: IEEE1888, IEEE1451, HTTP, MQTT, SEP2.0, Bluetooth (such as health thermometer profile, Bluetooth Low Energy (BLE)), etc. In the application layer protocols, the primary purpose is to define the data formats. Therefore, server applications are required to support different protocols for receiving data from IoT devices, which differ in the types of application layer protocol. Moreover, new protocols and new data expression rules are successively developed to support the emerging smart community services. However, it is better to continually use conventional IoT devices that do not support new protocols because the replacement of all old IoT devices is costly. Moreover, it is typical that IoT terminals, which have similar sensors and functions, are installed redundantly at the same place owing to different protocols required to handle the data, and this poses a significant problem in the implementation of smart community services. It is important to design a smart community information infrastructure to address this problem.

Data cycle for improving QoL and QoS: For providing smart community services, it is essential to generate, store, process, and analyze the data, as well as foster the data to useful services, use the data through the services, change someone's behavior or control something using the data, and anonymize the data for its secondary use. Finally, the result of the behavior or control is measured by the sensors, and it creates the data again. This data cycle is shown in Fig. 1. This cycle enables the human behavior and machine

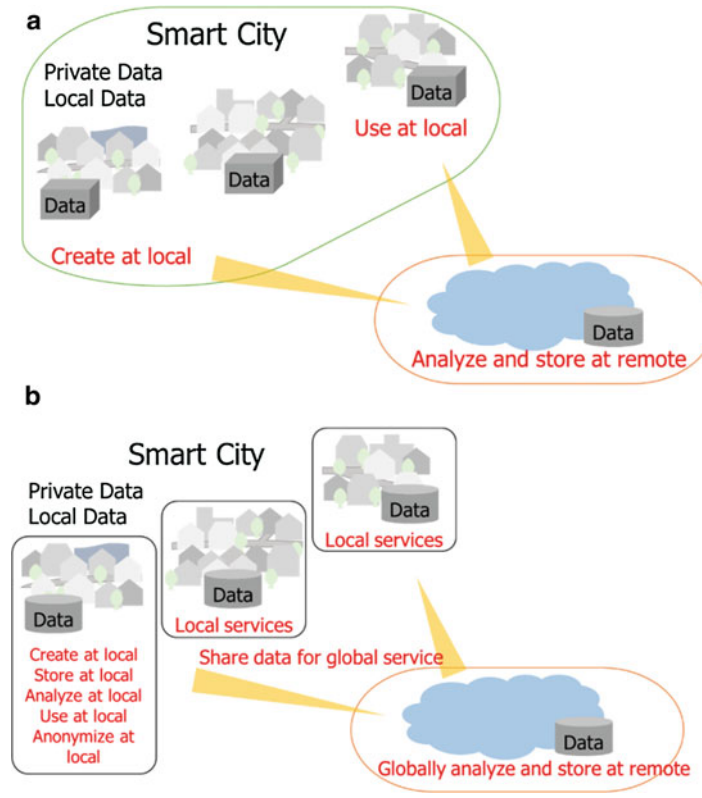


Architecture and Data Management for Smart Community Information Platform, Fig. 1 Smart community data cycle

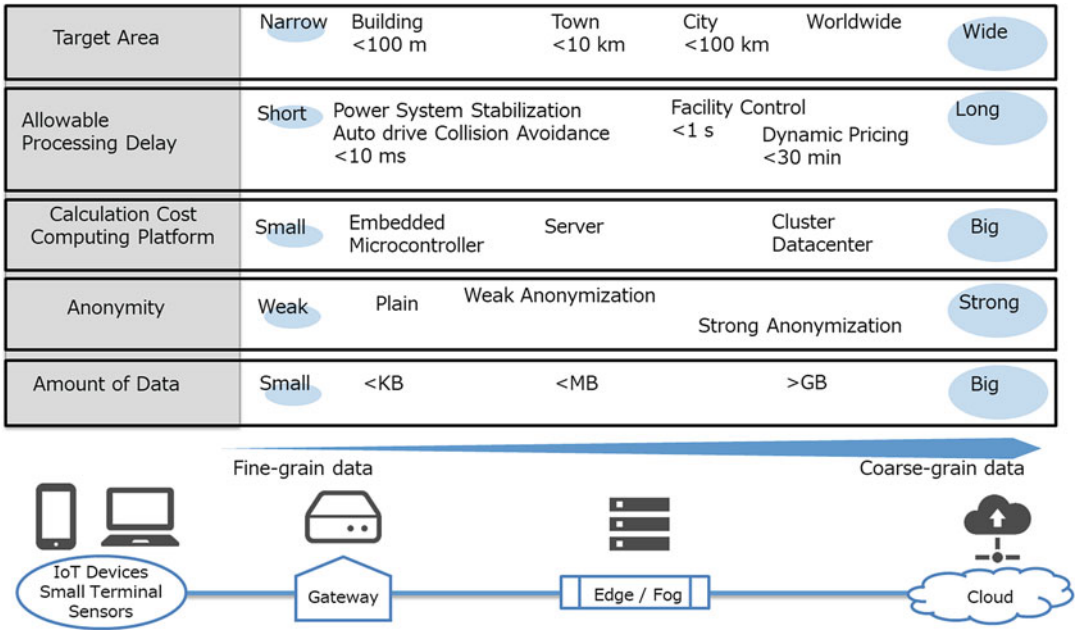
control to be improved. Therefore, it improves the quality of life (QoL), such as the improvement in comfort or wellbeing, and the quality of service (QoS), such as the system efficiency or functionality. Thus, the smart city data infrastructure has to maintain the effective data cycle. The similar cycle is also discussed in ambient intelligence from a different perspective.

Data encapsulation: In the series of data processes, most parts of the process are achieved in the cloud, as shown in Fig. 2a. The use of cloud has the merit of low management cost. However, to use the cloud means to use a centralized system. Therefore, cloud services may cause a single point of failure and suffer from an advanced persistent threat. What types of countermeasures can be provided for it? It is strategically correct to prohibit gatherings and distribute people when the danger of terrorism is predicted. However, despite the repeated incidents of cyber terrorism, private data continue to be centrally managed in the cloud. It is also natural for people to resist the revelation of personal information by moving these data from the cloud to their residential area. However, the service providers only consider their points of view and would force users to observe their service provision policy: "personal information must be managed in the cloud for providing services." This situation is not desirable in the provision of smart community services. Private data should be encapsulated locally, and this encapsulation is effective from the perspective of protecting several types of attacks or threats, as shown in Fig. 2b. The cloud should provide a global service using abstract and unified information.

Data hierarchy: IoT devices send measured data to the cloud. On their way to the cloud, the data may pass a gateway at a smart house or smart building, or switches and routers at the internet providers, and finally, it will arrive at the service application programs in the cloud. As explained in data encapsulation, the data should be processed at an appropriate location when required



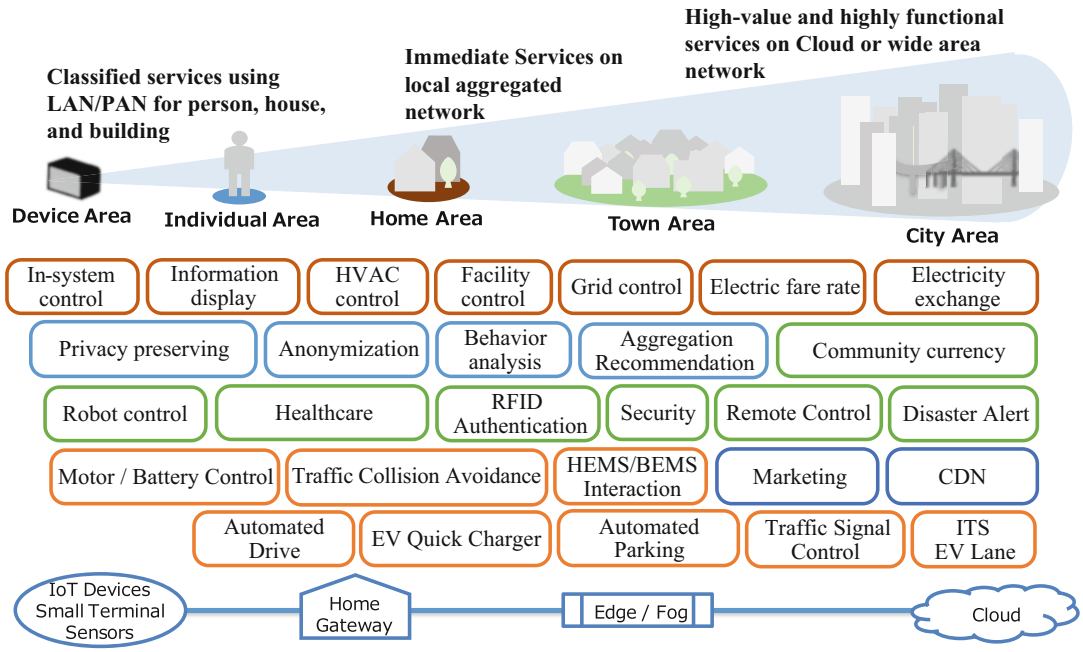
Architecture and Data Management for Smart Community Information Platform, Fig. 2 (**a**) Conventional smart city data services. (**b**) Encapsulated smart city data services



Architecture and Data Management for Smart Community Information Platform, Fig. 3 Smart community service hierarchy

by the service. As shown in Fig. 3, the switches and routers as well as the gateways can become computational resources referred as edge and fog computing. The selection of the appropriate computational resources requires appropriate indexes, and the typical indexes are shown in the figure. When the services are provided in a gateway, it focuses on narrow space service provisioning and short-range communications, e.g., an indoor environment, and the amount of calculation cost and computing platform can be small because the required data amount is small. Anonymity is a new and important index, as given in data encapsulation. The gateway is often connected with IoT devices and handles the raw data generated by the devices. Contrastingly, the services provided in the cloud can use anonymized data, and vice versa, as explained in data encapsulation. The Edge or Fog computing environment provides a balanced environment between the gateway and cloud. By using these indexes and the differentiated environment, data can be stored at an appropriate network location and processed at an appropriate network location. Moreover, the needs of data processing chains on the Internet are increasing. This chain of data processing is also discussed as service function chaining in network function virtualization (NFV) (Trajkovska et al. 2017). This data chain is important in providing services using the data. Figure 4 shows the mapping of promising smart community services according to the indexes. For example, HVAC control is provided in a building and may use private information. Electricity exchange service is provided in the city area, and the price can be defined using generalized information. Additionally, the service applications have to select the best location.

IoT data security: The risk of IoT terminals being hacked has increased in the recent years. Hence, the maintenance of security levels must be ensured. However, most IoT terminals have low computing and power capacities. Therefore, it is difficult to facilitate better resistance to cyberattacks using complicated protocols to add new functionalities. Moreover, introducing additional security software, or new protocols, is undesirable from the viewpoint of power consumption and system cost. Thus, it is necessary to design an information infrastructure to address these

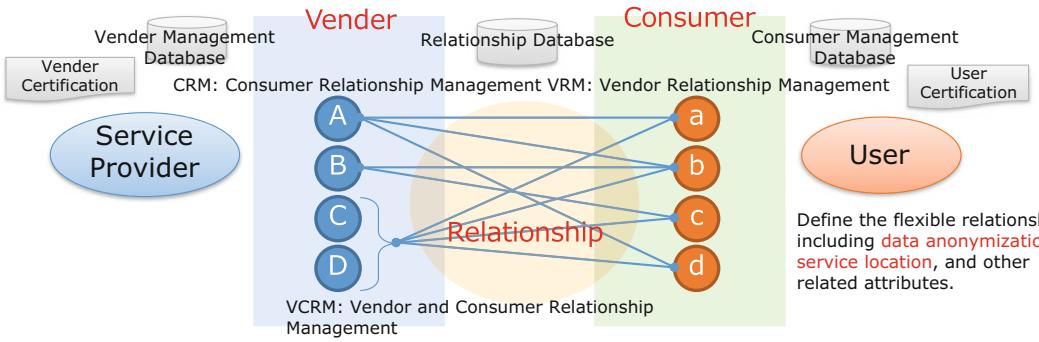


Architecture and Data Management for Smart Community Information Platform, Fig. 4 Mapping of smart community services considering service locations (Nishi 2018)

problems. Namely, the network system should provide new security services for IoT devices and locally stored data.

Private data handling: The protection of private information is important in smart community services because the data generated by residents for receiving smart community services could have privacy constraints. One of the protection methods is the aforementioned local data encapsulation. However, it is desirable to share the data for secondary use. This secondary use of data is the most promising service in the future smart community. Anonymization technology is key for the safe sharing of private data. If the data are perfectly anonymized and any type of private data is not extracted from the anonymized data, it is safe but is not useful for smart community services. The balance between difficulty in revealing private information and the usefulness in providing services is important. For achieving the balance at the highest level, the appropriate anonymization method must be designed and selected for each data service. Medical records, smart meters, locations, and other data will have their respective suitable anonymization methods. The development of an anonymization method is indispensable. Moreover, watermarking technology for anonymized data is indispensable (Nakamura et al. 2017). In data service, the information of user rights must be managed, such as who generates the data, who uses the data, and what the purpose is. Watermarking technology can include the information in the anonymized data. The watermark in anonymized data can prevent data leakage by tracing the data leakage and protect the user rights holders.

Data ownership: Who possesses the data? Although this is a simple question, it is sometimes difficult to answer because the ownership and stakeholders of data are easy to entangle and handover. Moreover, the national speculations have worsened this situation. The General Data Protection Regulation (GDPR) was issued, and it provided a new stage of data management from the perspective of data privacy. In consideration of these situations, flexible data ownership management is required. Regarding the style of ser-



Architecture and Data Management for Smart Community Information Platform, Fig. 5 Vendor and consumer relationship management

vice provision, vendors have to manage the consumers; vendors draft contracts with consumers for providing services and provide reasonable prices for the services, in which the vendors aim to maximize both their benefit and the consumer satisfaction. In this contract, the vendors request the disclosure of private data. This is called consumer relationship management (CRM). Contrastingly, new trend requires an alternative relationship management between vendors and consumers such that the consumer can manage the vendors' services. This is called vender relationship management (VRM). A user may think that a certain service is useful and worth providing the user's private data to the service vendor for a more efficient service. Meanwhile, the user may think that the service is enough to use as a trial and not worth providing the private information. In this case, the data management system will anonymize the data to preserve privacy. For attaining the flexible service relationship management between vendors and users, vendor and consumer relationship management (VCRM) is proposed (Niwa and Nishi 2017). Figure 5 shows the structure of VCRM. Both the vendor and consumer manage the database, which stores the information of the service provider or who the data user. A relationship is established when a contract of providing and receiving a service is engaged. The relationship is managed by its dedicated database, and the database has various information, such as the method and level of data anonymization, target area of service, location to be processed, service provider, service user, expiration date, constraints for providing services, required computational power, required latency, and amount of data. VCRM manages these data and verifies the integrity with real use.

Smart Community Information Platform (SCIP)

The fusion of Fog/Edge, gateway, and cloud are appropriate as a desired future direction. However, their current shortfalls must be compensated. Therefore, an information mechanism called authorized stream contents analysis (ASCA) was proposed (Nishi 2018). ASCA is a mixed mechanism of software and hardware for supporting service provision. It supplies a method to process the information flowing through a communication network on intermediate communication devices. ASCA reconstructs the TCP stream on the device, decodes the SSL using the key shared with the cloud service, decodes Chunk and Gzip, executes string matching and the extraction function using regular expression rules, and subsequently sends the analyzed result of the stream to the dedicated service process. ASCA can modify the stream contents directly, under the constraint of the buffer size. When using 128 G of memory, it can analyze more than two million TCP streams simultaneously without the limitation of the TCP stream length by employing a dedicated context-switch technology. ASCA on a general enterprise server can provide 20

Gbps of processing performance using hardware accelerators, such as DPDK for the accelerating network throughput using userland zero-copy communication, HyperScan on the Intel Xeon Processor for accelerating the string matching function, and Intel QuickAssist Technology for accelerating the throughput of the encryption, decryption, compression, and decompression.

A service application using ASCA on a Fog/Edge terminal can process the network stream and provide dedicated services at an intermediate location on the route to the target in the cloud. This does not require any modification of the IoT terminals; the destination IP address of the data stream can maintain its original IP address of the target in the cloud. Moreover, ASCA provides the functionality to modify the stream contents, enabling the direct removal or anonymization of private information at the intermediate nodes. NEGI (Takagiwa and Nishi 2015) is an original library created for ASCA. It is designed with no acceleration and can therefore be utilized in any Linux-based environment, including an ARM-based embedded platform. The Intel Xeon platform can achieve 1 Gbps of streaming. DooR is a hardware accelerated library of the ASCA, and it achieves a 20-Gbps stream process. The proposed smart community information platform (SCIP) uses NEGI or DooR as basic libraries for stream processing. On these ASCA basic libraries, the user and service provider can design their services as a Docker container (Miura et al. 2017). Docker is a virtualized environment for executing application programs, and it reduces the cost of launching and terminating application software, compared with the conventional virtual machine environment. Because Docker provides a virtualized environment isolated from the host machine, the zero-copy architecture including the DPDK is not available for the application software as a Docker container. Therefore, Docker's shared memory option is used to communicate with a DPDK-supported NIC via shared memory to cope with this problem.

ASCA can handle the aforementioned smart community data specifications and requirements. ASCA provides a transparent environment as network nodes. Because ASCA can monitor or modify a network stream at any point, it enables immediate data handling using zero-copy communication and hardware accelerators. It can also offer flexible data handling because it can provide services at the intermediate nodes on the Internet and change a communication protocol transparently. When ASCA is used as a basic library in the smart community data platform, it can be the key device for rotating the data cycle for improving the QoL and QoS. All the given functions in Fig. 1 can be achieved or measured using service applications with ASCA. Transparent data encapsulation can be achieved by designing the data firewall of the data anonymization application on the ASCA. Moreover, the hierarchical design of the ASCA-based encapsulation enables the data hierarchy. The firewall function and antivirus function on ASCA-supported devices can provide IoT data security. By implementing the appropriate applications on the ASCA, private data can be handled and the data can be anonymized. VCRM can be an application of ASCA. Therefore, either the user or the service provider can define and modify the relationship at any time. This feature adopts the opt-in and opt-out of data registration. The default value is given as the opt-in. The on-demand modification of the relationship achieves the opt-out.

Another important point is the mobility of the service applications. The initial allocation to an appropriate location and the subsequent execution of a service should ideally occur automatically, and be migrated as necessary, without the need for an explicit migration request. Executing this effectively requires managing several resources and activities, including memory, storage, CPU, communication, task allocation, and distribution, to manage the Docker containers with service applications. Thus, a mechanism for resource management that performs functions similar to the basic functions of an operating system is necessary. This resource management system is called the SCIP OS. Some basic functions of the SCIP OS resemble those of the orchestrator for Docker containers. However, the orchestrator only considers the management of the Docker containers, whereas the SCIP OS focuses on

the service applications more broadly from the perspectives of service feasibility, IoT feasibility, and future feasibility. Moreover, ASCA enabled wireless station can be a center of local data manager in the wireless environment. It gathers data from sensor nodes and offers a variety of benefits to local services in the reachable range of its wireless signals.

System Demonstration

The data anonymization and watermark insertion application of UDCMi, a smart town project in Misono Town, Saitama City, Japan, was demonstrated at the Global City Team Challenge EXPO (https://pages.nist.gov/GCTC/). This application uses the smart metering of smart houses in UDCMi. As a smart community service, a lifestyle recommendation service for eco-life is provided as a nudge service. In this demonstration scenario, the smart meter sends the data to the cloud, and the nudge report was automatically generated using machine-learning technology. En route to the cloud, the data are also captured, anonymized, and watermarked at the gateway by ASCA. The extension of the function of IoT devices was proven, and the effectiveness of the nudge report using anonymized data was compared with the other report using raw data.

Conclusion

Data-oriented smart community services are indispensable for the sustainable advancement of smart cities. The proposed smart community information platform is a system considering the given data specification and requirements. The complexity of data handling at the network infrastructure will be increased according to the development of a society structure. However, an information platform maximizing the flexible data management can solve various regional problems, including urbanization problems. The improvement of QoL and QoS, i.e., the ultimate goal of a smart community, is not achieved by a single metric but by an interoperable data approach including the residents' behavior change and machinery control that enables hopeful societies to be established. Further developments of the architecture and data management of the smart community information platform are expected.

Cross-References

- Advances in Distribution System Monitoring
- Cyber-physical Security in Smart Grid Communications
- Game Theory Meeting Vehicle-to-Grid Regulation: Past, Present, and Future
- Privacy-Preserving Data Aggregation for Smart Grids
- Smart Metering, Specific Challenges, and Solution Approaches
- Ultra-reliable and Low-Latency Communications for the Smart Grid

Acknowledgments This work was partially supported by MEXT/JSPS KAKENHI Grant (B) Numbers JP17H01739, and also by the Technology Foundation of the R&D project, "Design of Information and Communication Platform for Future Smart Community Services" by the Ministry of Internal Affairs and Communications of Japan.

References

Anderson RJ, Spong MW (1989) Bilateral control of teleoperators with time delay. IEEE Trans Autom Control 34(5):494–501. https://doi.org/10.1109/ICSMC.1988.754257

Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on mobile cloud computing. ACM, New York, pp 13–16. https://doi.org/10.1145/2342509.2342513

Ferreira D, Corista P, Gião J, Ghimire S, Sarraipa J, Jardim-Gonçalves R (2017) Towards smart agriculture using FIWARE enablers. In: International conference on engineering, technology and innovation (ICE/ITMC), Madeira Island, pp 1544–1551. https://doi.org/10.1109/ICE.2017.8280066

Miura T, Wijekoon JL, Prageeth S, Nishi H (2017) Novel infrastructure with common API using docker for scaling the degree of platforms for smart community services. In: International conference on industrial informatics, Emden, pp 474–479. https://doi.org/10.1109/INDIN.2017.8104818

Nakamura Y, Nakatsuka Y, Nishi H (2017) Novel Method to Watermark Anonymized Data for Data Publishing. IEICE Trans Inf Syst E100-D(8):1671–1679

Nishi H (2018) Information and communication platform for providing smart community services system implementation and use case in Saitama City In: Proceedings of the 2018 IEEE international conference on industrial technology (ICIT), Lyon, pp 1375–1380. ISBN: 978-1-5386-4053-1/18/. https://doi.org/10.1109/ICIT.2018.8352380

Niwa A, Nishi H (2017) An information platform for smart communities realizing data usage authentication and secure data sharing. In: Fifth international symposium on computing and networking, Aomori, pp 119–125. https://doi.org/10.1109/CANDAR.2017.73

OpenFog Consortium. https://www.openfogconsortium.org/#fog-computing. Accessed 7 July 2018

Smart Communities Guidebook: Building smart communities, how California's communities can thrive in the digital age. International Center for Communications, College of Professional Studies and Fine Arts, San Diego State University, San Diego, 1997

Takagiwa K, Nishi H (2015) Local trend detection from network traffic using topic model and network router. In: ICOMP'15 – The 2015 international conference on internet computing and big data, Las Vegas, pp 53–59

Trajkovska I, Kourtis M-A, Sakkas C, Baudinot D, Silva J, Harsh P, Xylouris G, Bohnert TM, Koumaras H (2017) SDN-based service function chaining mechanism and service prototype umplementation in NFV scenario. J Comput Stand Interfaces Arche 54(P4):247–265 . Elsevier Science Publishers. https://doi.org/10.1016/j.csi.2017.01.002

Yi S, Li C, Li Q (2015) A survey of fog computing: concepts, applications and issues. In: Proceedings of the 2015 workshop on mobile big data. ACM, New York, pp 37–42. https://doi.org/10.1145/2757384.2757397

Architectures, Key Techniques, and Future Trends of Heterogeneous Cellular Networks

Mugen Peng and Yaohua Sun
Key Laboratory of Universal Wireless Communications (Ministry of Education), Beijing University of Posts and Telecommunications, Beijing, China

Synonyms

Multilayer wireless networks

Definitions

HetNets are seen as new network paradigm evolutions to the fifth-generation wireless systems, which can cost-efficiently improve system coverage, capacity, latency, and so on.

Historical Background

With the explosive increase in mobile data traffic, operators have to continuously improve network performance. As a promising solution, the HetNet is a new technology that can cost-efficiently improve system coverage and capacity (Peng et al. 2015a). Communication nodes with high transmit power, such as MBSs, are deployed for large coverage and high capacity, while LPNs, such as SCAPs, help to supply service in the coverage of some hotspots. By deploying additional LPNs within the local-area range and bringing LPNs closer to end-UEs, HetNets can potentially improve spatial resource reuse and extend the coverage, thus allowing future cellular systems to achieve higher data rates while retaining the uninterrupted connectivity and seamless mobility of cellular networks.

The LPN is identified as one of the key components to increase the capacity of cellular networks in dense areas with high traffic demands. When traffic is clustered in hotspots, such LPNs can be combined with HPN to form a HetNet. HetNets have advantages of serving hotspot customers with high bit rates through deploying dense LPNs, providing ubiquitous coverage, and delivering the overall control signallings to all UEs through the powerful HPNs. Actuarially, too dense LPNs will incur severe interference, which restricts performance gains and commercial developments of HetNets. Therefore, it is critical to control interference through advanced signal processing techniques to fully unleash the potential gains of HetNets. The CoMP transmission and reception is presented as one of the most promising techniques in 4G systems. Unfortunately, CoMP has some disadvantages in real networks because its performance gain depends

heavily on the backhaul constraints and even degrades with increasing density of LPNs. Further, it was reported that the average SE performance gains from the uplink CoMP in downtown Dresden field trials were only about 20% with non-ideal backhaul and distributed cooperation processing located on the base station.

To overcome the SE performance degradations and decrease the energy consumption in dense HetNets, a new paradigm for improving both SE and EE through suppressing inter-tier interference and enhancing the cooperative processing capabilities is needed in the practical evolution of HetNets. Meanwhile, cloud computing technology has emerged as a promising solution for providing high energy efficiency together with gigabit data rates across software-defined wireless communication networks, in which the virtualization of communication hardware and software elements place stress on communication networks and protocols.

To achieve these goals, the C-RAN has been proposed as a combination of emerging technologies from both the wireless and the information technology industries by incorporating cloud computing into RANs (Peng et al. 2016a). C-RANs have come with their own challenges in the constrained fronthaul and centralized BBU pool. A prerequisite requirement for the centralized processing in the BBU pool is an interconnection fronthaul with high bandwidth and low latency. Unfortunately, the practical fronthaul in C-RANs is often capacity and time-delay constrained, which has a significant decrease on SE and EE gains.

Consequently, H-CRANs are proposed in Peng et al. (2014) as cost-effective potential solutions to alleviating inter-tier interference and improving cooperative processing gains in HetNets through combination with cloud computing. The motivation of H-CRANs is to enhance the capabilities of HPNs with massive MIMO and simplify LPNs through connecting to a "signal processing cloud" with high-speed optical fibers. As such, the baseband datapath processing and the radio resource control for LPNs are moved to the cloud server so as to take advantage of cloud computing capabilities.

Unfortunately, H-CRANs are still challenging in practice. First, since the location-based social applications become more and more popular, the traffic data over the fronthaul between RRHs and the centralized BBU pool surges a lot of redundant information, which worsens the fronthaul constraints. Besides, H-CRANs do not take full advantage of processing and storage capabilities in edge devices, such as RRHs and "smart" UEs, which is a promising approach to successfully alleviating the burden of the fronthaul and BBU pool. Moreover, operators need to deploy a huge number of fixed RRHs and HPNs in H-CRANs to meet the requirements of peak capacity, which makes a serious waste when the volume of delivery traffic is not sufficiently large.

To solve such challenges in H-CRANs, revolutionary approaches involving new RAN architectures and advanced technologies need to be explored. Fog computing is a term for an alternative to cloud computing that puts a substantial amount of storage, communication, control, configuration, measurement, and management at the edge of a network, rather than establishing channels for the centralized cloud storage and utilization (Bonomi et al. 2012). Inspired by the advanced characteristics of fog computing and to alleviate the existing challenges of H-CRANs and take full advantages of local caching, the F-RAN architecture has been proposed in Peng et al. (2016b).

In this chapter, we are motivated to make an effort to offer a comprehensive survey on system architectures, key techniques, and future trends in heterogeneous cellular networks, including HetNets, H-CRANs, and F-RANs 'in addition, all the abbreviations are listed (Table 1).

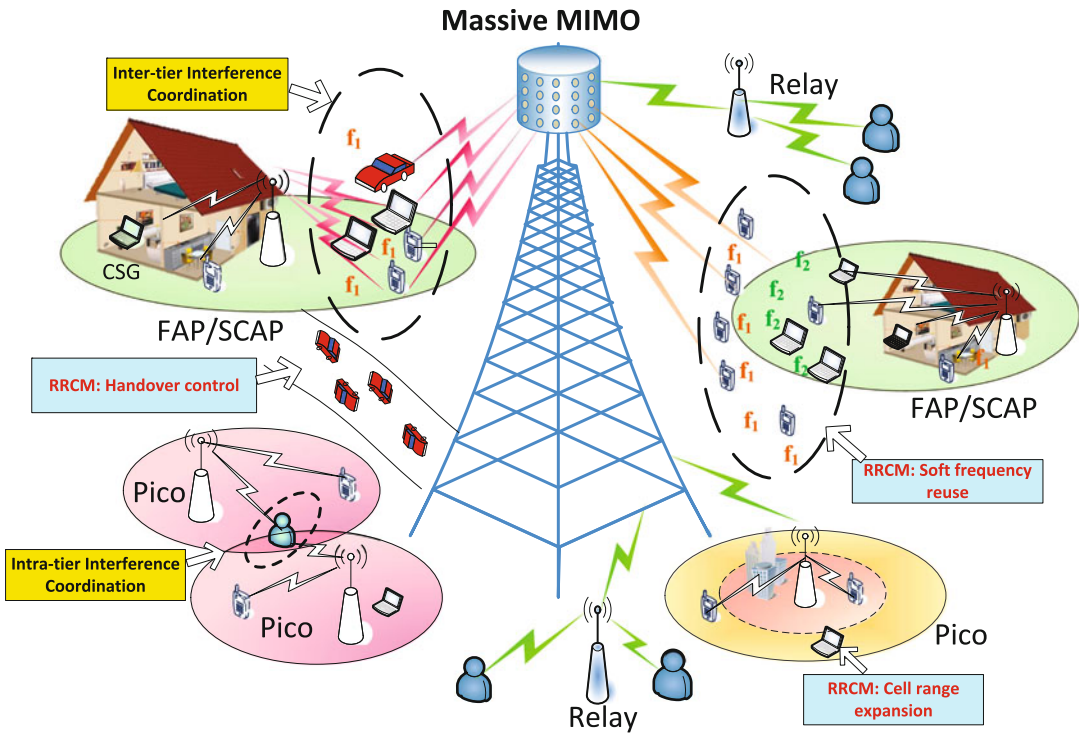
Foundations

The Architecture of Traditional HetNets

To provide a high-speed performance for cell-edge users, a hierarchical HetNet architecture Fig. 1 has been proposed for converging multiple LPNs and supporting unicast and multicast services simultaneously, where the coverage areas are divided into three layers: hierarchical

Architectures, Key Techniques, and Future Trends of Heterogeneous Cellular Networks, Table 1 Abbreviations in this chapter

3GPP	Third generation partnership project	JR	Joint reception
4G	Fourth generation	JT	Joint transmission
BBU	Baseband unit	LPN	Low power node
CNN	Convolutional neural network	LTE	Long-term evolution
CN	Core network	MBS	Macro base station
CoMP	Coordinated multi-point	MIMO	Multiple-input-multiple-output
CS/CB	Coordinated scheduling/coordinated beamforming	PHY layer	Physical layer
CRSP	Collaborative radio signal processing	QoE	Quality of experience
CRRM	Collaborative radio resource management	QoS	Quality of service
C-RAN	Cloud radio access network	RAN	Radio access network
DL	Downlink	RF	Radio frequency
DRL	Deep reinforcement learning	RNN	Recurrent neural network
EE	Energy efficiency	SCAP	Small cell access point
F-AP	Fog access point	SE	Spectral efficiency
F-RAN	Fog computing-based radio access network	SON	Self-organizing network
F-UE	Fog user equipment	UE	User equipment
H-CRAN	Heterogeneous cloud radio access network	UL	Uplink
HetNet	Heterogeneous network	WLAN	Wireless local area network
HPN	High power node		



Architectures, Key Techniques, and Future Trends of Heterogeneous Cellular Networks, Fig. 1 HetNet architecture for E-UTRAN systems

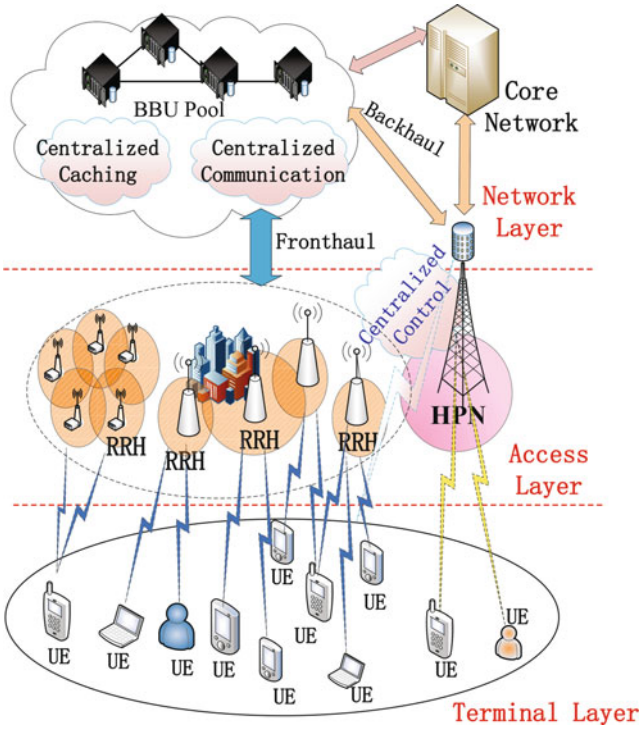
cooperative basic layer, homogeneous cooperative enhanced layer, and heterogeneous cooperative extended layer. For the hierarchical cooperative basic layer, UEs may be located near the BS; high-speed service is supported due to the utilizations of high-order modulation and coding schemes for the unicast services and hierarchical modulation schemes with unequal error protection space-time code for the multicast service. In the homogeneous cooperative enhanced layer, cooperative homogeneous diversity gain can be achieved, where UEs may be located near a cell boundary. For the heterogeneous cooperative extended layer, heterogeneous cooperative diversity gain guarantees the convergence and interworking of multiple RANs.

The Architecture of H-CRANs

Similarly with the traditional C-RAN, As shown in Fig. 2, a huge number of RRHs with low energy consumptions in H-CRANs cooperate with each other in the centralized BBU pool to achieve high cooperative gains. Only the front RF and simple symbol processing functionalities are implemented in RRHs, while the other important baseband physical processing and the procedures of the upper layers are executed jointly in the BBU pool. Sequently, only partial functionalities in the PHY layer are incorporated in RRHs, and the model with these partial functionalities is denoted as PHY_RF in Fig. 2.

However, different from C-RANs, the BBU pool in H-CRANs is interfaced to HPNs for mitigating the cross-tier interference between RRHs and HPNs through the centralized cloud computing-based cooperative processing techniques. Further, the data and control interfaces between the BBU pool and HPNs are added and denoted by S1 and X2, respectively, whose definitions are inherited from the standardization definitions of 3GPP. Since the voice service can be provided efficiently through the packet switch mode in 4G systems, the proposed H-CRAN can support both voice and data services simultaneously, and the voice service is preferred to be administrated by HPNs, while the high-data packet traffic is mainly served by RRHs.

Compared with the traditional C-RAN architecture, the proposed H-CRAN alleviates the fronthaul requirements with the participation



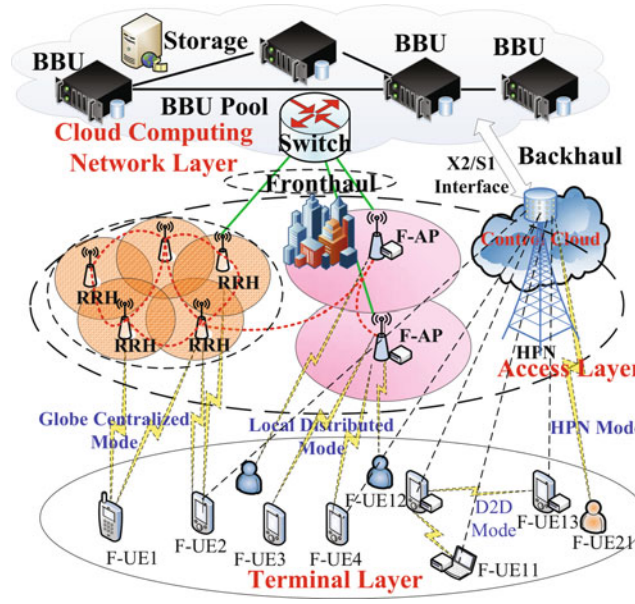
Architectures, Key Techniques, and Future Trends of Heterogeneous Cellular Networks, Fig. 2 System architecture of H-CRANs

of HPNs. Owing to the incorporation of HPNs, the control signallings and data symbols are decoupled in H-CRANs. All control signallings and system broadcasting information are delivered by HPNs to UEs, which simplifies the capacity and time-delay constraints of the fronthaul links between RRHs and the BBU pool, and can make RRHs active or sleep efficiently to save the energy consumption. Further, some burst traffic or instant messaging service with a small amount of data can be efficiently supported by HPNs. The adaptive signalling/control mechanism between connection-oriented and connectionless is supported in H-CRANs, which can achieve significant overhead savings in the radio connection/release by moving away from a pure connection-oriented mechanism. For RRHs, different transmission technologies in the PHY layer can be utilized to improve transmission bit rates, such as IEEE 802.11 ac/ad, millimeter wave, and even optical light. For HPNs, the massive MIMO is one potential approach to extend the coverage and enrich the capacity.

The Architecture of F-RANs

In the proposed F-RAN architecture shown in Fig. 3, the F-APs and F-UEs in the terminal layer and network access layer constitute the fog computing layer, which are capable of conducting local caching, signal processing, and radio resource management. In the terminal layer, adjacent F-UEs can communicate with each other through the D2D mode or the F-UE-based relay mode. For example, F-UE13 and F-UE11 can communicate with each other with F-UE12 helping relay the signal, while F-UE11 and F-UE12 can operate in D2D mode to transmit data directly. In the network access layer consisting of F-APs and HPNs, HPNs are responsible for delivering control signalling to F-UEs, and F-APs are used to forward and process the received data. In addition, the data between F-APs and the BBU pool in the cloud computing layer is transmitted over fronthaul links, while HPNs are interfaced to the BBU pool with backhaul links for coordinations. The traditional X2/S1 interface for the backhaul link is backward compatible with that defined in 3GPP standards for LTE and LTE-Advanced systems.

Different from H-CRANs, a large number of CRSP and CRRM functions originally located in the cloud computing layer are shifted to F-APs and F-UEs, which alleviates the burden



Architectures, Key Techniques, and Future Trends of Heterogeneous Cellular Networks, Fig. 3 System model for implementing F-RANs

on the fronthaul and BBU pool. Meanwhile, the limited caching in F-APs and F-UEs can make some requests served locally. The benefit of this enhancement is significant, considering that the location-based social applications are becoming popular and much redundant information over the fronthaul links will worsen the fronthaul constraints. Furthermore, local caching, local signal processing, and local radio resource management can better satisfy the low latency requirement of some applications in 5G era.

Cooperative Spatial Processing

To boost the SE and EE performance of HetNets, the inter-tier interference should be properly handled. In the PHY layer, this interference can be mitigated by inter-tier CoMP transmission. The CoMP technique is aimed at enhancing the performance of cell-edge UEs. By the coordination between transmissions of adjacent cells, the CoMP can effectively suppress the adjacent cells' interferences when applied in UL or DL. Specifically, all access nodes can jointly decode transmitted signals from UEs in UL via JR while allocate transmission power to different UEs in a network MIMO manner with optimal precoding matrices, i.e., JT and CS/CB (Access 2010). It should be noted that the main difference between JT and CS/CB is that the data of each UE is transmitted by only one access node in CS/CB, while multiple cooperative nodes serve each UE simultaneously in JT. In Lee et al. (2012), it is demonstrated that JT and CS/CB can improve the performance of cell-edge UEs by 100% and 30%, respectively, compared to multiuser MIMO.

However, the backhaul transmission will incur latency for coordination in practice, and meanwhile the number of quantization bits for the exchanged information is limited, which will both degrade the performance of CoMP. The impacts of non-ideal backhaul are investigated in Xia et al. (2013), and it is concluded that the backhaul latency will affect the coverage and normalized throughput. Moreover, it is shown that the coverage and throughput achieved by CoMP zero-forcing beamforming decrease almost linearly with the average latency growing from zero. Even worse, when the latency exceeds 60% of the fading coherence time, the scheme does not have any benefit. Furthermore, the large number of cooperative cells does not essentially lead to better performance. On the contrary, it is shown that the number of coordinated cells should be kept fairly small in HetNets.

Radio Resource Management

To maximize the SE and EE performances of HetNets, the multidimensional radio resources should be optimized, including power allocation, subchannel allocation, and user association. However, because of the involvement of integer variables like subchannel allocation indicator and the existence of inter-tier and intra-tier interference, multidimensional radio resource allocation in HetNets is always non-convex, making the optimization problem hard to solve. To deal with the non-convexity, there are generally two kinds of ways. One is using heuristic algorithms which are suitable for finding the near-optimal solutions to NP-hard problems. The other one is to transform these non-convex problems into convex problems by different methods, and then the primal problems can be solved by settling the corresponding convex problems. In Peng et al. (2015b), the author focuses on maximizing EE by jointly allocating the resource block and transmit power for a heterogeneous cloud radio access network. Due to the non-convexity of the primal problem, the author tackles the primal problem by solving its dual problem aiming to minimize the Lagrange function of the primal problem which is always convex by definition. Moreover, since the primal problem satisfies the time-sharing condition, zero duality gap holds. Thus, the non-convex primal problem can be equivalently transformed into a convex optimization problem, i.e., its dual problem.

Furthermore, to achieve better global performance, it is very important to consider both the conventional physical layer performance metrics and the traffic delay in the upper layer. As a simple framework to deal with delay-aware RRM optimization problems, the Lyapunov optimization approach has been commonly adopted in HetNets. In Urgaonkar and Neely (2012), the Lyapunov optimization approach is utilized to

solve the problem of opportunistic cooperation in a two-tier HetNet. The access nodes would like to maximize their own throughput under average power constraints by optimizing access admission, cooperation decision, and power control. The obtained online control algorithm can stabilize the traffic queue without requiring any knowledge of the traffic arrival rates. While in Chen and Lau (2013), a two-stage queue-aware cross-layer resource management algorithm is proposed by minimizing drift-plus-utility. The cross-layer RRM consists of a queue-aware limited CSI feedback filtering optimization stage over a long timescale and an SINR-based optimal user scheduling stage over a short timescale. The utility is defined as the average feedback cost, and a large parameter V reduces the average CSI feedback at the cost of a larger average queue length.

Self-Organizing HetNets

Due to the ever-increasing number of radio resource management functionalities, it is impractical to manually adjust resource management parameters in a multidimensional parameter scenario. Faced with this issue, SON techniques have attracted a lot of attentions from both industry and academia. Generally speaking, the self-organizing capability of HetNets can be divided into three components: self-configuration, self-optimization, and self-healing. By using SON, the network configuration, installation, coverage, capacity, spectrum, and quality can be automatically managed and optimized taking into account target QoS performance, interference level, signal strength, traffic pattern, and so on.

To take full advantages of both centralized and distributed SON architectures, the author in Peng et al. (2013) proposes a hybrid SON architecture, and differences between traditional homogeneous networks and HetNets are discussed in terms of self-configuration like prediction-based radio resource configuration as well as self-optimization including mobility robustness optimization and energy saving. Moreover, it should be noted that time-varying traffic load makes the deployment of completely self-organized HetNets nontrivial. Hence, the features of SON coordination in HetNets should be emphasized. Specifically, the graph-based decision framework proposed in Gelabert et al. (2014) can be adopted to enable efficient interaction and coordination of SON mechanisms, where the interaction and conflict relationship between multiple SON mechanisms are described by the metric event and action graph. Based on this proposed framework, the strategy-based SON coordination for HetNets can be easily implemented for a given event or a combination of events.

Key Applications

HetNets are the key to satisfying the diverse and stringent performance requirements in future wireless networks.

Future Directions

Access Slicing in HetNets

As a cost-efficient solution to support diverse use cases in the 5G era, network slicing enables the provision of networks in an as-a-service fashion. However, current studies mainly concentrate on CN-based slicing, and the impact of the characteristics of RANs is not well considered. To get a more effective network slicing solution, the author in Xiang et al. (2017) proposes an enhanced network slicing approach in F-RANs, termed as access slicing. The proposed access slicing architecture is featured with a centralized orchestration layer, a slice instance layer, and the information awareness capability to guarantee diverse QoS and QoE requirements. Although the above work makes a big step for network slicing, several challenges exist for its implementation. First, as stated in Xiang et al. (2017), the information awareness allows the access slice orchestration layer makes intelligent decisions on slice instance creation and resource management. Nevertheless, more specifications should be investigated like the types of information needed, the frequency for collecting information, and the way

of slice instance configuration based on processed information. Second, resource allocation between multiple slices should be investigated to fully utilize the resources of HetNets to meet diverse performance requirements. Such a problem can be very challenging, since it is basically a multi-objective optimization problem. Meanwhile, besides allocating radio resource, some slices also need caching resource and computation resource, which incurs an extra challenge.

HetNets Driven by Deep Learning

As an important technology for enabling artificial intelligence, deep learning has been successfully applied in many areas, including computer vision and speech recognition, and has drawn a lot of attentions of researchers in the wireless communication area. In Cao et al. (2017), a learning framework is proposed in which a CNN and an RNN are used to extract features in spatial domain and time domain from raw information collected by wireless networks, respectively, and this manner avoids identifying features manually. Taking the derived features as state input, DRL is adopted to control wireless networks intelligently, and the superiority of the proposal is verified by applying it to mobility management in WLAN. However, to apply deep learning in HetNets, more studies should be conducted. For example, more theoretical analysis should be done to provide guidelines on the architecture design of neural networks and the selection of hyper-parameters. Furthermore, the infrastructure of HetNets needs to be upgraded to support the implementation of deep learning algorithms, especially at the network edge to facilitate real-time network optimization and control.

Cross-References

- Heterogeneous Small Cell Networks
- Resource Allocation in SDN/NFV-Enabled 5G Networks
- Resource Allocation
- Wireless Edge Caching

References

Access EUTR (2010) Further advancements for E-UTRA physical layer aspects. 3GPP technical specification TR 36:V2

Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on mobile cloud computing. ACM, Helsinki, Finland pp 13–16

Cao G, Lu Z, Wen X, Lei T, Hu Z (2017) Aif: an artificial intelligence framework for smart wireless network management. IEEE Commun Lett 22:400–403

Chen J, Lau VK (2013) Large deviation delay analysis of queue-aware multi-user MIMO systems with two-timescale mobile-driven feedback. IEEE Trans Signal Process 61(16):4067–4076

Gelabert X, Sayrac B, Jemaa SB (2014) A heuristic coordination framework for self-optimizing mechanisms in LTE HetNets. IEEE Trans Veh Technol 63(3):1320–1334

Lee D, Seo H, Clerckx B, Hardouin E, Mazzarese D, Nagata S, Sayana K (2012) Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges. IEEE Commun Mag 50(2):148–155

Peng M, Liang D, Wei Y, Li J, Chen HH (2013) Self-configuration and self-optimization in LTE-advanced heterogeneous networks. IEEE Commun Mag 51(5):36–45

Peng M, Li Y, Jiang J, Li J, Wang C (2014) Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. IEEE Wirel Commun 21(6):126–135

Peng M, Wang C, Li J, Xiang H, Lau V (2015a) Recent advances in underlay heterogeneous networks: interference control, resource allocation, and self-organization. IEEE Commun Surv Tutorials 17(2):700–729

Peng M, Zhang K, Jiang J, Wang J, Wang W (2015b) Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks. IEEE Trans Veh Technol 64(11):5275–5287

Peng M, Sun Y, Li X, Mao Z, Wang C (2016a) Recent advances in cloud radio access networks: system architectures, key techniques, and open issues. IEEE Commun Surv Tutorials 18(3):2282–2308

Peng M, Yan S, Zhang K, Wang C (2016b) Fog-computing-based radio access networks: issues and challenges. IEEE Netw 30(4):46–53

Urgaonkar R, Neely MJ (2012) Opportunistic cooperation in cognitive femtocell networks. IEEE J Sel Areas Commun 30(3):607–616

Xia P, Liu CH, Andrews JG (2013) Downlink coordinated multi-point with overhead modeling in heterogeneous cellular networks. IEEE Trans Wirel Commun 12(8):4025–4037

Xiang H, Zhou W, Daneshmand M, Peng M (2017) Network slicing in fog radio access networks: issues and challenges. IEEE Commun Mag 55(12):110–116

Area Spectral Efficiency

▸ Area Spectral Efficiency of Ultradense Networks

Area Spectral Efficiency of Ultradense Networks

Guoqiang Mao
University of Technology Sydney, Sydney, NSW, Australia

Synonyms

Area spectral efficiency; Ultradense networks

Definition

Area spectral efficiency refers to the data rate that can be achieved per unit bandwidth and in a unit area of the wireless network. It has the unit of b/s/Hz/m^2 or b/s/Hz/km^2.

Historical Background

Network densification has been the main driver of wireless network capacity increase in the past and will play an even more crucial role in the development of the next generation mobile communication systems (5G). Network densification refers to the deployment of more base stations (BSs) and wireless access points per unit area and the associated technological advances to support such densification. There are three primary ways of increasing the wireless network capacity: (1) adding more spectrum, (2) enhancing spectral efficiency through advanced communication techniques, and (3) enhancing spatial reuse of frequency spectrum through network densification. The area spectral efficiency (ASE) is a major metric to measure the efficiency in the spatial reuse of frequency spectrum. According to a study by Webb (Henderson 2007), among the 1-million-fold increase in wireless network capacity achieved in 50 years from 1950 to 2000, 15× improvement was achieved from a wider spectrum, 5× improvement from better Medium Access Control (MAC) and modulation schemes, 5× improvement by designing better coding techniques, and an astounding 2700× gain through network densification and reduced cell sizes. As we move to the next generation mobile communication systems and seek further capacity increases, network densification, manifested through heterogeneous and ultradense networks, will play an even more important role. First, the combined amount of spectrum available for licensed mobile broadband and unlicensed Wi-Fi is scarce. Spectrum approaching millimeter wavelengths is more abundant but is yet to be proven for cellular and Wi-Fi use due to difficulty in penetration and supporting significant mobility (Andrews et al. 2016a; Kutty and Sen 2016). Second, the scope for enhanced spectral efficiency may be limited as many current wireless systems are already running at a spectral efficiency close to the performance limit prescribed by the Shannon. Therefore, mobile operators have increasingly turned to network densification and small cell technology to meet the 1000× capacity increase expected on 5G. Small cells are now deployed in massive numbers (Small Cell Forum 2016), which drives the paradigm shift into ultradense networks. Ultradense networks feature a very dense deployment of BSs and wireless access points.

Foundations

The ASE and wireless network capacity can be used interchangeably in the sense that the total capacity achieved by a network deployed in a given geographical area and using a specific amount of bandwidth is equal to the ASE times the size of the area and the amount of bandwidth. It is of crucial interest to investigate how the ASE

varies as more and more BSs are deployed and the network becomes denser and denser.

The SINR Invariance Principle for Low-to-Medium BS Density

Until recently, there was widely held belief that the ASE, or equivalently the network capacity, may increase indefinitely as the network densifies (Haenggi et al. 2009; Andrews et al. 2011). This view has been underpinned by the so-called SINR (signal-to-interference-plus-noise ratio) invariance principle, which is valid for a low-to-medium BS density. We use the following example to illustrate the SINR invariance principle. Consider a set of BSs deployed on an infinite plane, and number these BSs according to their distances to the origin such that the k-th nearest BS has a distance of l_k. Further assume that all BSs transmit at the same fixed power P_t and the wireless signal experiences standard power-law attenuation such that the received power at a distance d from the transmitter is $P_r(d) = P_t L d^{-\alpha}$, where L is a reference path loss at unit distance and α is the path loss exponent. For a "typical" user located at the origin and associated with its nearest BS, its SINR can be expressed as

$$\text{SINR} = \frac{P_t L l_1^{-\alpha}}{\sum_{k=2}^{\infty} P_t L l_k^{-\alpha} + \sigma^2}, \tag{1}$$

where σ^2 represents the noise power. Now consider scaling the distances between all BSs by a factor of t. The density of BSs will increase by a factor of $\frac{1}{t^2}$. Further assume that noise power is negligible, which is valid in an interference-limited regime where most wireless networks now operate in. The SINR then becomes

$$\begin{aligned} \text{SINR} &= \frac{P_t L (t l_1)^{-\alpha}}{\sum_{k=2}^{\infty} P_t L (t l_k)^{-\alpha} + \sigma^2} \\ &\simeq \frac{P_t L (t l_1)^{-\alpha}}{\sum_{k=2}^{\infty} P_t L (t l_k)^{-\alpha}} \\ &= \frac{l_1^{-\alpha}}{\sum_{k=2}^{\infty} (l_k)^{-\alpha}}, \end{aligned} \tag{2}$$

i.e., the SINR is invariant with the increase of BS density.

The SINR invariance principle implies that as the network densifies and the distances between transmitters and receivers reduce, the increase in interference will be counterbalanced by the increase in the desired signal. Consequently, the SINR will stay approximately the same. The principle suggests that other things being equal, the spectral efficiency, or equivalently the capacity, per BS cell is invariant as network densifies. Therefore, from the network perspective, the ASE, or equivalently the overall network capacity, will linearly increase with the number of cells per unit area; from the user perspective, as each user maintains the same SINR but shares its BS with an ever-smaller number of other users, each user can achieve approximately linear growth in its achievable data rate as BSs are added, until the limit of one user per cell is reached. The SINR invariance principle is not affected by the BS layout, transmit power, shadowing and fading distributions, and path loss exponent (Andrews et al. 2016b).

Area Spectral Efficiency for Ultradense BS Regime

Recent research suggested however that the SINR invariance principle may no longer apply when the BS density becomes very large and that there may exist a limit in network densification, beyond which further densification will not bring the expected linear increase in capacity and may even reduce the capacity (Ding et al. 2015, 2016a,b; Ge et al. 2016; Liu et al. 2016a,b; Andrews et al. 2016b; Zhang and Andrews 2015). Specifically, by incorporating those effects that have negligible impacts on the ASE when the BS density is small or moderate but become dominant factors determining the ASE when the BS density is very large, it was shown that the ASE may either exhibit a sublinear increase with the BS density, or reduce at certain region of the BS density, or even monotonically reduce to zero beyond a certain BS density threshold.

In Ding et al. (2015, 2016a,b), and Ge et al. (2016), researchers considered the impact of non-line-of-sight (NLoS) and line-of-sight

(LoS) transmissions on the ASE. NLoS and LoS transmissions are ubiquitous in wireless communications. Other things being equal, as the distance between a transmitter and a receiver increases, the probability that their direct LoS path gets blocked increases and the converse. NLoS transmissions will generally suffer much higher path loss than LoS transmissions. Specifically, by employing a 3GPP-endorsed LoS/NLoS probability model given below

$$\Pr^{\mathrm{L}}(x) = \begin{cases} 1 - \frac{x}{d_1}, & 0 < x \leq d_1 \\ 0, & x > d_1 \end{cases}, \quad (3)$$

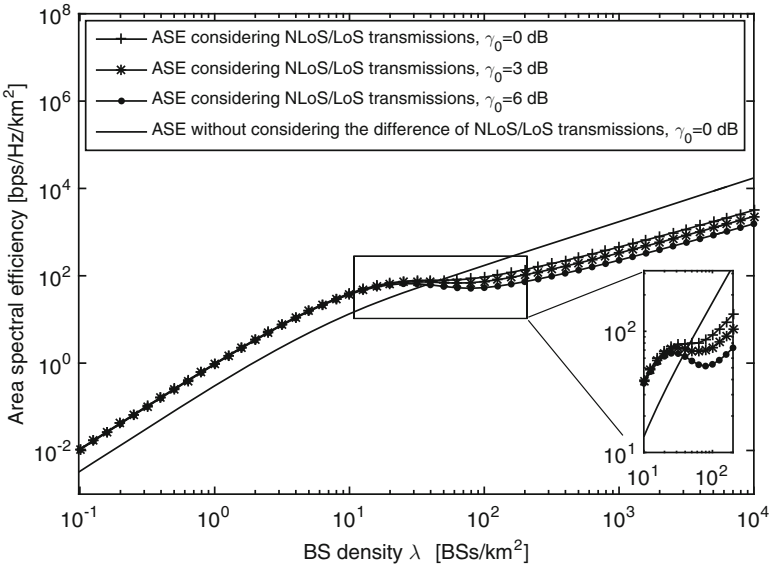
where $\Pr^{\mathrm{L}}(x)$ is the probability that a transmitter and a receiver separated by a distance x experience LoS transmission, $1 - \Pr^{\mathrm{L}}(x)$ is the probability that the same pair of transmitter and receiver experiences NLoS transmission, and d_1 is a parameter determining the decreasing slope of the linear function $\Pr^{\mathrm{L}}(r)$. Ding et al. (2016a) investigated the variation of the ASE with the BS density, shown in Fig. 1. The ASE is determined using the following equation

$$ASE = \lambda \Pr(\mathrm{SINR} \geq \gamma_0) \int_{\gamma_0}^{\infty} \log_2(1+x)\, f_{\mathrm{SINR}}(x)\, dx \quad (4)$$

where λ is the BS density, γ_0 is the SINR threshold required for establishing a connection, $f_{\mathrm{SINR}}(x)$ is the probability density function of the SINR, and the term $\log_2(1+x)$ comes from the Shannon capacity formula. In the literature, the ASE has also been calculated using the following formula $ASE = \lambda \Pr(\mathrm{SINR} \geq \gamma_0) \log_2(1+\gamma_0)$, which reflects the fact that some wireless systems may not be able to explore the extra SINR above the SINR threshold γ_0 to boost the data rate.

Figure 1 shows that when LoS/NLoS transmissions are considered, the ASE exhibits not only quantitatively but also qualitatively different trends with the BS density, compared with the case that does not consider the distinction of LoS and NLoS transmissions. Specifically,

1. when the BS density is small, the ASE quickly increases with the BS density because the network is generally noise-limited and adding more BSs immensely benefits the ASE by reducing the transmission distance between the BS and the mobile user (MU). Most transmissions in this regime are NLoS transmissions; however the transmission between a MU and its desired BS has a higher probability of being LoS transmission, and as the BS density further increases, this probability increases to a non-negligible value. Therefore, in this region, the ASE considering LoS/NLoS transmissions is higher than that without considering LoS/NLoS transmissions, but their trend is the same;
2. when the network is dense enough, i.e., greater than $20\,\mathrm{BS/km^2}$ in Fig. 1, the growth of the ASE with the BS density, when considering LoS/NLoS transmissions, becomes flat or even exhibits a decrease (for $\gamma_0 = 3\,\mathrm{dB}$ and $\gamma_0 = 6\,\mathrm{dB}$), which is distinctly different from that predicted without considering LoS/NLoS transmissions. This can be explained by the so-called NLoS-LoS transition effect. Particularly, in this region, the transmissions between MSs and their desired BS are already dominated by LoS transmissions but in the beginning signals from interfering BSs, are mainly NLoS transmissions suffering higher loss. As BS density further increases and the distances between a MU and BSs, including both the desired BS and interfering BSs, further reduce, signals from interfering BSs start to transit from NLoS transmissions to LoS transmissions. Consequently, as the BS density increases, interference experiences a larger increase compared with the desired signal, causing the SINR to reduce sharply in this region. Therefore, the ASE remains largely flat with the increase in BS density or may even reduce;
3. when the network becomes very dense, i.e., greater than $100\,\mathrm{BS/km^2}$ in Fig. 1, the dominant interfering BSs have completed their NLoS-LoS transitions. Both the desired signal and the dominant interference are now



Area Spectral Efficiency of Ultradense Networks, Fig. 1 The variation of the ASE with the BS density at various SINR thresholds. Both the case considering the coexistence of NLoS and LoS transmissions and the case without considering the coexistence of NLoS and LoS transmissions are shown. The case without considering the coexistence is plotted using the result in Andrews et al. (2011) and assuming $\alpha = 3.75$, $P_t = 24\,\text{dBm}$, $\sigma^2 = -95\,\text{dBm}$, and $L = 10^{-14.54}$. The case considering the coexistence of NLoS and LoS transmissions is obtained assuming $d_1 = 300\,\text{m}$, $P_t = 24\,\text{dBm}$, $\sigma^2 = -95\,\text{dBm}$, $\alpha = 2.09$, and $L = 10^{-10.38}$ for LoS transmissions and $\alpha = 3.75$ and $L = 10^{-14.54}$ for NLoS transmissions. These parameters are chosen following 3GPP recommendations

LoS transmissions. However, non-dominant interfering BSs further away from the MS may still experience the NLoS-LoS transition effect. Therefore, in this region, the SINR may still reduce modestly with the increase in BS density. This modest decrease in the SINR combined with the increase in the BS density causes the ASE to increase but at a rate below that predicted without considering LoS/NLoS transmissions.

Open data at Ofcom, an independent regulator and competition authority for the UK (http://sitefinder.ofcom.org.uk/search), shows that in some dense urban regions in the UK, the BS density already exceeds $10\,\text{BS/km}^2$. As we move into the regime of ultradense networks, we are bound to pass through the aforementioned second region. Therefore, it is important to consider the effect of LoS and NLoS transmissions when analyzing the ASE of ultradense networks.

In 2016b, Liu et al. investigated the ASE from a different perspective by considering the transition from far-field to near-field radio propagation which may be triggered as the BS density increases to a very large value. Specifically, it is well known that wireless signal attenuates faster at larger distances. In a region close to the transmitter, the signal may suffer little attenuation, while this attenuation sharply increases as the distance from the transmitter becomes large. Based on some field measurements, they proposed the following multi-slope BPM model to better capture the signal attenuation

$$g_N^B\left(\{\alpha_n\}_{n=0}^{N-1}; x\right) = \eta_n\left(1 + x^{\alpha_n}\right)^{-1}, \quad R_n < x \leq R_{n+1}, \tag{5}$$

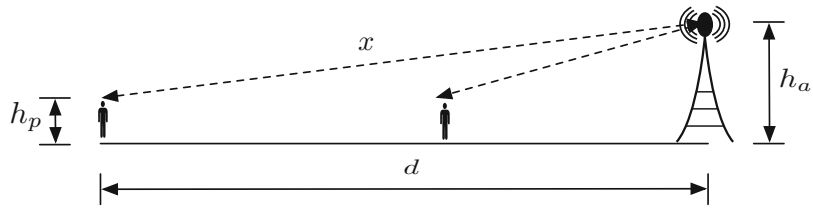
where x denotes the distance from the transmitter to the receiver, R_n separates the attenuation func-

tion into several distinct regions, and α_n denotes the path loss exponent for $R_n < x \leq R_{n+1}$, $\eta_0 = 1$, and $\eta_n = \prod_{i=1}^{n} \frac{1+R_i^{\alpha_1}}{1+R_i^{\alpha_{i-1}}}$. As signals are attenuated faster at larger distances, $\alpha_n < \alpha_{n+1}$ holds. Using the model, they predicted that the ASE may monotonically decrease to zero when the BS density increases beyond a certain critical threshold. When the SINR threshold $\gamma_0 = 20\,\text{dB}$, this critical BS density lies between $5000 - -25{,}000\,\text{BS/km}^2$ depending on the values of α_is; when $\gamma_0 = 0\,\text{dB}$, this critical BS density lies between $2 \times 10^5 - 3 \times 10^5\,\text{BS/km}^2$. In a separate work (Zhang and Andrews 2015), Zhang and Andrews used a dual-slope (power-law) path loss function to model the different radio propagations in a region close to the transmitter and in a region far away:

$$l_2(\alpha_1, \alpha_2; x) = \begin{cases} x^{-\alpha_0} & x \leq R_c \\ \eta x^{-\alpha_0} & x > R_c \end{cases}, \quad (6)$$

where $\eta = R_c^{\alpha_1 - \alpha_0}$, $R_c > 0$ is the critical distance, and α_0 and α_1 are the near- and far-field path loss exponents with $0 \leq \alpha_0 \leq \alpha_1$. Using the model, they predicted that the ASE grows linearly with the BS density λ if $\alpha_0 > 2$, scales sublinearly with rate $\lambda^{2-\frac{2}{\alpha_0}}$ if $1 < \alpha_0 < 2$, and decays to zero if $\alpha_0 < 1$. It is worth noting that compared with extensive studies on far-field propagation models, surprisingly little is known about near-field radio propagation. Therefore, the models in (5) and (6) are plausible. However, a common insight revealed in their work (Zhang and Andrews 2015; Liu et al. 2016b) is that the ASE exhibits quantitatively and qualitatively different trends with the BS density λ when the different propagation conditions in the near-field and in the far-field are considered.

In Ding and Perez (2016) and Gruber (2016), researchers considered the impact of some geometric conditions on the ASE. Specifically, in Ding and Perez (2016), Ding and David Lopez investigated the impact of antenna height on the ASE by considering a scenario where the BS antenna height is kept constant when its density increases. They showed that when the height of the BS relative to the MU is maintained at a constant value of 8.5 m, the ASE may peak when the BS density increases to $\sim 15\text{BS/km}^2$ and then starts to monotonically decrease to zero as the BS density further increases. In comparison, when the impact of this height is ignored, the ASE may increase monotonically toward infinity. Their result can be intuitively explained using Fig. 2, which shows that while the impact of antenna height may be small at small BS density, its impact may play a dominant impact on the ASE in an ultradense network with a large BS density. In 2016, Gruber considered the geometric constraints limiting the BS deployment locations. Specifically, by considering that BSs can only be deployed along the two sides of the street, some distinctly different results on the ASE were obtained compared with that assuming BSs can be deployed randomly and anywhere in the street. It is worth noting that while some assumptions



Area Spectral Efficiency of Ultradense Networks, Fig. 2 An illustration of the different impacts of the antenna height as the distance between MU and BS changes. When the distance d between the MU and BS is large, the impact of the relative height between BS antenna, h_a, and MU, h_p, on the propagation distance x is small and negligible. When the BS density increases to a very large value and hence the distance d reduces to a small value, the impact of $h_a - h_p$ on the propagation distance $x$is no longer negligible and may even be the dominant factor determining x

used in these studies are quite plausible, e.g., as the density of BS increases, one would expect that the BS size and height also reduce instead of being constant; the insight revealed in these studies holds: as we move into the ultradense regime, such geometric restrictions as physical dimensions of transmitters and receivers and deployment restrictions on BSs that previously have small or negligible impact on the ASE may now become dominant factors in the consideration.

In summary, distinct from conventional low-to-medium density networks, in ultradense networks, impacts of LoS/NLoS transmissions, the different radio propagation conditions in the near-field and far-field of BSs, and network geometric constraints may play an important role in determining the ASE.

Key Applications

The ASE is a fundamental performance metric of ultradense networks and more general wireless networks. The ASE determines the capacity that can be achieved by a wireless network and measures the efficiency in the spatial reuse of frequency spectrum. Knowledge of the ASE helps to guide the design and deployment of wireless networks.

Cross-References

- ▸ Area Spectral Efficiency
- ▸ Cognitive Heterogeneous Networks
- ▸ Network Capacity

References

Andrews JG, Baccelli F, Ganti RK (2011) A tractable approach to coverage and rate in cellular networks. IEEE Trans Commun 59(11):3122–3134

Andrews JG, Gupta AK, Dhillon HS (2016a) A primer on cellular network analysis using stochastic geometry. eprint arXiv:1604.03183

Andrews JG, Zhang X, Durgin GD, Gupta AK (2016b) Are we approaching the fundamental limits of wireless network densification? IEEE Commun Mag 54(10):184–190

Ding M, Lopez-Perez D, Mao G, Wang P, Lin Z (2015) Will the area spectral efficiency monotonically grow as small cells go dense? In: IEEE GLOBECOM, pp 1–7

Ding M, Perez DL (2016) Please lower small cell antenna heights in 5G. In: IEEE Globecom, pp 1–6

Ding M, Wang P, López-Pérez D, Mao G, Lin Z (2016a) Performance impact of LoS and NLoS transmissions in dense cellular networks. IEEE Trans Wirel Commun 15(3):2365–2380

Ding T, Ding M, Mao G, Lin Z, López-Pérez D (2016b) Uplink performance analysis of dense cellular networks with LoS and NLoS transmissions. In: IEEE ICC, pp 1–6

Ge X, Tu S, Mao G, Wang CX, Han T (2016) 5G ultra-dense cellular networks. IEEE Wirel Commun 23(1):72–79

Gruber M (2016) Scalability study of ultra-dense networks with access point placement restrictions. In: IEEE ICC workshops, pp 650–655

Haenggi M, Andrews JG, Baccelli F, Dousse O, Franceschetti M (2009) Stochastic geometry and random graphs for the analysis and design of wireless networks. IEEE J Sel Areas Commun 27(7):1029–1046

Henderson W (2007) Webb report, ofcom.

Kutty S, Sen D (2016) Beamforming for millimeter wave communications: an inclusive survey. IEEE Commun Surv Tutor 18(2):949–973

Liu J, Sheng M, Liu L, Li J (2016a) Effect of densification on cellular network performance with bounded pathloss model. IEEE Commun Lett 21(2):346–349

Liu J, Sheng M, Liu L, Li J (2016b) How dense is ultra-dense for wireless networks: from far- to near-field communications. eprint arXiv:1606.04749

Small Cell Forum (2016) Small cell market status report, May 2016

Zhang X, Andrews JG (2015) Downlink cellular network analysis with multi-slope path loss models. IEEE Trans Commun 63(5):1881–1894

Artificial Intelligence

- ▸ Machine Learning in Wireless Sensor Networks for the Internet of Things

Artificial Jamming schemes

- ▸ Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks

Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks

Yi-Liang Liu
Department of Electronics Information Engineering, Harbin Institute of Technology, Harbin, China

Synonyms

Artificial jamming schemes; Physical layer security

Definition

Cellular communications and networks are particularly vulnerable to eavesdropping attacks due to the opened and broadcasting nature of wireless channels. Due to the rapid development of cellular networks and wireless business, the security issue has attracted a lot of attention. Physical layer security takes the advantages of channel randomness nature of transmission media to achieve communication confidentiality, which is the most important and interesting topic of privacy communication technologies. Artificial noise (AN) or jamming signals with MIMO technologies are supposed to implement physical layer security and enhance secrecy capacities in cellular networks, where AN signals along with confidential data confuse potential eavesdroppers via utilizing orthogonal spaces provided by transmit antennas. In these AN-based schemes, message streams were sent in a multiplexing mode via all eigen-subchannels (positive eigenvalue channels) at desired directions, and the AN signals can be transmitted to the null spaces (zero eigenvalue channels) of desired directions, so that they do not affect the desired user, while eavesdropper channels are degraded with a high probability. This chapter introduces a physical layer security review, covering the information theory and physical layer security schemes based on MIMO technologies, and provides an overview on the state-of-the-art works on the AN-based scheme along with conclusions and future research directions.

Historical Background

The origin of physical layer security research can be traced back to Wyner's definition of an information-theoretic secrecy capacity (Wyner 1975), which is a maximum message transmission rate in confidential communications. Compared to conventional cryptography that works to ensure all involved entities to load proper and authenticated cryptographic information, physical layer security technologies perform confidentiality functions without considering about how those security protocols are executed. In other words, it does not require to implement any extra security schemes or algorithms on other layers above the physical layer. In addition, the physical layer has the same confidentiality level with the one-time pad, which ensures its security performance. The concept of physical layer security has become more popular with the help of MIMO technologies, because these emerging technologies can improve the secrecy capacity massively via utilizing extra orthogonal spaces provided by multiple antennas.

Information Theory

Wyner implied that the intrinsic and unpredictable elements of physical channels, such as noises, interferences, and wireless fading, play central roles in protecting secure messages (Wyner 1975) and claimed that there exists a randomized codebook maximizing transmission rates that can achieve prefect secrecy where any eavesdroppers cannot get any information. The maximizing secrecy rate is defined as a secrecy capacity C_s as

$$C_s = \max_{V \to X \to YZ} I(V;Y) - I(V;Z), \quad (1)$$

where $I(\cdot;\cdot)$ denotes mutual information between V and Y and V is an auxiliary input

variable with a joint auxiliary input distribution $p(v)p(x|v)$. Given a discrete memoryless channel $P_{YZ|X}$, wiretap codes achieve the secrecy capacity via maximization over the choices of the joint distribution $P_{YZ|X}$, such that the Markov chain $V \rightarrow X \rightarrow YZ$ holds, which has been expected to completely address the underlying problem of that traditional cryptographic algorithms are vulnerable to quantum attacks.

In the last two decades, researchers have developed a significant amount of mathematical theories, such as secrecy capacity characterization, wiretap coding designs, wireless fading managements, etc. And the advancements in cellular technologies have improved physical layer security significantly, by exploring spatial diversities and multiplexing gains with the help of multi-antennas technologies.

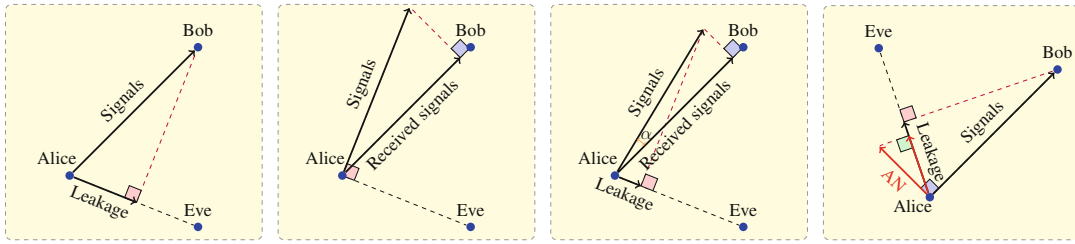
Physical Layer Security Schemes Based on MIMO Technologies

From the traditional communication viewpoints, improving the link quality of the main channel is one of the most feasible approaches to improve the secrecy capacities with existing installations of cellular networks, because most of the cellular wireless communication technologies can improve the channel spectrum utilization, and if the eavesdropping channel spectrum utilization is unchanged, these traditional wireless communication technologies will undoubtedly improve the secrecy capacities. However, this viewpoint has a great security risk, i.e., the secrecy capacities are small if the eavesdropper has a better channel quality than the desired users. Therefore, current research insists on another viewpoint that jointly uses more degree of freedom of main channels and artificial randomness to protect messages while reduces the quality of eavesdropping channels. Prospectively, the randomness of multipath fading of a MIMO channel is stronger than a single channel. Multiplexing technologies of MIMO are promising ways to enhance the degree of freedom of main channels. And AN signals can interfere with eavesdroppers while they are sent into the null spaces of desired directions.

We briefly summarize MIMO-based physical layer security techniques of recent research in four categories, as in Fig. 1, including secure unitary beamforming, zero-forcing (ZF) beamforming, convex optimization-based precoding, and AN-based precoding. It is emphasized that a MIMO-based physical layer security scheme of cellular networks can combine multiple techniques.

As in Fig. 1a, unitary beamforming techniques use unitary matrices, semi-unitary matrices, or unitary vectors as transmit precoding, which is similar with traditional beamforming technologies of cellular networks. Unitary beamforming techniques send a message stream via multi-antennas to make it as close to the main channel direction as possible. ZF beamforming is shown in Fig. 1b, which is a secure beamforming technology based on ZF precoding, where the message stream is transmitted to a desired receiver via a shifted beamforming direction, which is orthogonal to the eavesdropper's channel. The illustrative diagram of CVX-based precoding can be seen in Fig. 1c. The problem of optimizing secrecy capacities is usually nonconvex in MIMO systems, but Newton methods or Lagrangian dual transformations can be used to trap it in a local maximum problem and address the transmit covariance matrix optimization based on convex optimization tools. AN signals can be transmitted to the null spaces of desired directions, so that they do not affect the desired user, while an eavesdropper's channel is degraded with a high probability, as in Fig. 1d. This AN-based precoding exploits a fraction of the transmit power to send artificially generated noise signals, so the power for transmitting messages will be reduced.

AN-based schemes have been seen as promising methods because this type of techniques has two advantages. Firstly, there are no requirements on the condition of better main channels. Secondly, when the number of transmit antennas is larger the number of the eavesdroppers, the main channel state information (CSI) and precoding matrices of security schemes can be broadcasted to both legitimate receivers and eavesdroppers. We do not need to worry about the



Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks, Fig. 1 Illustrative diagrams of four basic secure multi-antenna technologies. For simplicity, this sets two-dimensional diagrams and a single message stream. (**a**) Secure unitary beamforming, (**b**) Zero-forcing precoding, (**c**) CVX-based precoding, (**d**) AN-based precoding (Liu et al. 2017a)

leakage of key precoding messages. More details of the second advantage can be seen in Liu et al. (2017b).

Artificial Noise Schemes

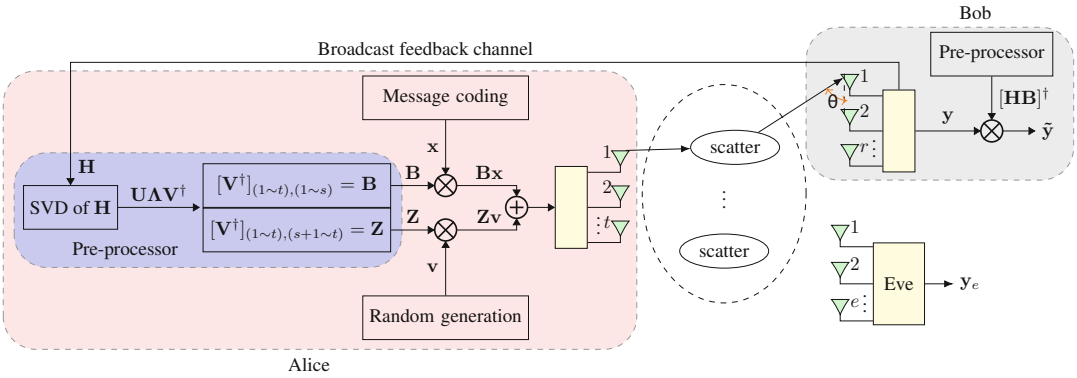
The section will introduce a general AN-based model, which is first provided in Liu et al. (2017b). Assuming t is the number of transmit antennas, messages are encoded in s (which is a variable) strongest eigen-subchannels based on ordered eigenvalues of Wishart matrices ($\mathbf{H}\mathbf{H}^\dagger$ or $\mathbf{H}^\dagger\mathbf{H}$, here, $\mathbf{H}$ is the main channel CSI matrix), while AN signals are generated in remaining $t - s$ eigen-subchannels. This scheme treats the number of eigen-subchannels for message streams, i.e., s, as an optimization objective that can be leveraged to optimize secrecy capacities.

General AN-Based Model

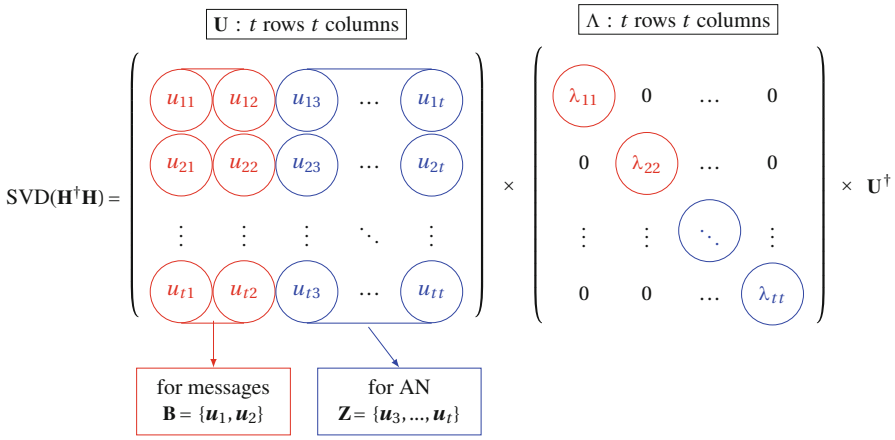
Let us consider a MIMO communication system in the presence of correlated Rayleigh fading at receiver and eavesdropper sides. The system consists of a transmitter (Alice) with t transmit antennas, a legitimate receiver (Bob) with r receive antennas, and an eavesdropper (Eve) with e receive antennas, as shown in Fig. 2. Define $m = \max(t, r)$ and $n = \min(t, r)$. In general, the main channel between Alice and Bob and the wiretap channel between Alice and Eve are defined by receiver-side correlated complex Gaussian matrices $\mathbf{H} \in \mathbb{C}^{r\times t}$ and $\mathbf{H}_e \in \mathbb{C}^{e\times t}$, whose elements obey distribution $\mathcal{CN}(0, 1)$. $\mathbf{R}_r \in \mathbb{C}^{r\times r}$ and $\mathbf{R}_e \in \mathbb{C}^{e\times e}$ are the receiver-side correlated channel matrices from Bob and Eve, respectively. Also assume that Alice knows full CSI of Bob via a broadcast feedback channel from Bob, including $\mathbf{H}$ and $\mathbf{R}_r$, but only knows the channel distribution information (CDI) of Eve and $\mathbf{R}_e$. Eve knows the CSI of all channels, including $\mathbf{H}$, $\mathbf{H}_e$, $\mathbf{R}_r$, and $\mathbf{R}_e$, where it is assumed that $t > e$.

In this AN-based scheme, there are s ($s \le t$) message-sending eigen-subchannels, which are selected by Alice based on the CSI feedback from Bob. More specifically, Alice performs SVD of $\mathbf{H}^\dagger\mathbf{H} \in \mathbb{C}^{t\times t}$ in a preprocessor, whose output is a unitary matrix $\mathbf{U} \in \mathbb{C}^{t\times t}$, its Hermitian transpose form $\mathbf{U}^\dagger \in \mathbb{C}^{t\times t}$, and a diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{t\times t}$, which consists of positive and zero eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$. Then, Alice generates a message precoding matrix $\mathbf{B} \in \mathbb{C}^{t\times s}$, whose columns are the eigenvectors corresponding to the first to the sth largest eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$, and an AN precoding matrix $\mathbf{Z} \in \mathbb{C}^{t\times d}$ ($s + d = t$), whose columns are the eigenvectors of the remaining eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$.

Alice transmits $\mathbf{B}\mathbf{w} + \mathbf{Z}\mathbf{v}$ via t transmit antennas. It means that each antenna transmits a combination of message components and AN components, but the AN components can be eliminated by the preprocessor at Bob. In this way, we create a capacity difference between the main channels and wiretap channels. Note that $\mathbf{B}$ and $\mathbf{Z}$ are fixed semi-unitary matrices derived from $\mathbf{H}$. Hence, we have $\mathbf{B}^\dagger\mathbf{B} = \mathbf{I}_s$ and



Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks, Fig. 2 Illustration of an artificial noisy MIMO wiretap channel model with receiver-side correlated Rayleigh fading. Alice has t transmit antennas, Bob has r receive antennas, and Eve has e receive antennas. Assumed that Alice, Bob, and Eve have uniformly linear array antennas with d_u antenna spacing. θ represents AoA between a scattered path and the antenna array, which can be viewed as a random variable with enough scatters, and θ follows a Gaussian distribution



1) The $r \times t$ matrix $\mathbf{H}$ is a channel matrix with t transmit antennas and r receive antennas.
2) Two antenna arrays $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ corresponding to two maximum eigenvalues λ_{11} and λ_{22} are allocated for message transmission.
3) Remained antenna array $\boldsymbol{u}_3, ..., \boldsymbol{u}_t$ corresponding to the remained eigenvalues is allocated for AN signal transmission.
4) By the precoding $\mathbf{B}$ and $\mathbf{Z}$, the information signal $\boldsymbol{w}$ and AN signal $\boldsymbol{s}$ are encoded as $\mathbf{B}\boldsymbol{w}$ and $\mathbf{Z}\boldsymbol{s}$.

Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks, Fig. 3 An example of this AN precoding scheme with $s = 2$

$\mathbf{Z}^\dagger\mathbf{Z} = \mathbf{I}_d$. An example of AN precoding scheme with $s = 2$ is illustrated in Fig. 3.

Theorem 1 *The AN signal can be eliminated at Bob if and only if* $[\mathbf{HB}]^\dagger\mathbf{HZ} = \mathbf{0}$ *(Liu et al. 2017b).*

Theorem 2 *The AN signal cannot be eliminated at Eve if and only if* $t > e$, $[\mathbf{HB}]^\dagger\mathbf{H}_e\mathbf{Z} \neq \mathbf{0}$, *and* $[\mathbf{H}_e\mathbf{B}]^\dagger\mathbf{H}_e\mathbf{Z} \neq \mathbf{0}$ *(Liu et al. 2017b).*

According to CSI matrix $\mathbf{H}$ and the preprocessing method, the received signals at Bob and

Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks, Table 1 Secrecy metrics with their conditions

Measures	Symbols	Conditions
Instantaneous secrecy capacity	C_s	$\mathbf{H}_e$ is available and optimal input distribution.
Average secrecy capacity	$\tilde{C}_s$	$\mathbf{H}_e$ is unavailable and optimal input distribution
Instantaneous secrecy rate	R_s	$\mathbf{H}_e$ is available and Gaussian distribution input
Average secrecy rate	$\tilde{R}_s$	$\mathbf{H}_e$ is unavailable and Gaussian distribution input

Eve can be expressed as

$$\mathbf{y} = \mathbf{HBw} + \mathbf{HZv} + \mathbf{n}, \tag{2}$$

$$\mathbf{y}_e = \mathbf{H}_e\mathbf{Bw} + \mathbf{H}_e\mathbf{Zv} + \mathbf{n}_e, \tag{3}$$

respectively. Here, $\mathbf{w}$ is a transmitted signal of the desired user, and $\mathbf{v}$ is a random AN signal. Both $\mathbf{w}$ and $\mathbf{v}$ are circularly symmetric complex Gaussian vectors with zero means and its covariance matrices $P/t\mathbf{I}_s$ and $P/t\mathbf{I}_d$, respectively, where P is the average transmit power constraint. For simplification, we distribute total power to each antenna evenly. $\mathbf{n}$ and $\mathbf{n}_e$ are additive white Gaussian noise (AWGN) vectors with their covariance matrices $\mathbf{I}_r$ and $\mathbf{I}_e$, respectively.

Bob can eliminate the AN signal $\mathbf{v}$ by preprocessing ($[\mathbf{HB}]^\dagger\mathbf{HZ} = \mathbf{0}$) the received signal $\mathbf{y}$ as

$$\tilde{\mathbf{y}} = \left[\mathbf{HB}\right]^\dagger\mathbf{y} = \boldsymbol{\Lambda}_s\mathbf{w} + \tilde{\mathbf{n}}, \tag{4}$$

where $\tilde{\mathbf{n}} = [\mathbf{HB}]^\dagger\mathbf{n} \in \mathbb{C}^{s\times 1}$ is an AWGN vector with its distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Lambda}_s)$. $\boldsymbol{\Lambda}_s \in \mathbb{R}^{s\times s}$ is a diagonal matrix formed by the first to the sth eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$. In the elimination process, the received signal multiplied by a fixed matrix will not change its capacity. Even if we consider the worst case that Eve has the knowledge of $\mathbf{H}$, $\mathbf{H}_e$, $\mathbf{B}$, and $\mathbf{Z}$, the AN signal still degrades Eve's channel because Eve cannot eliminate the AN signal because $t > e$, $[\mathbf{HB}]^\dagger\mathbf{H}_e\mathbf{Z} \neq \mathbf{0}$, and $[\mathbf{H}_e\mathbf{B}]^\dagger\mathbf{H}_e\mathbf{Z} \neq \mathbf{0}$, as in **Theorem 1**.

Secrecy Metric

The secrecy capacity and secrecy rate are important measures of AN-based schemes. Here, four expressions are provided to represent the instantaneous secrecy capacity, the average secrecy capacity, the instantaneous secrecy rate, and the average secrecy rate, respectively. These measures are fit in with different conditions as in Table 1

Instantaneous Secrecy Capacity

In a MIMO wiretap channel model, while Bob has the knowledge of $\mathbf{H}$, $\mathbf{H}_e$, $\mathbf{R}_r$, and $\mathbf{R}_e$, the instantaneous secrecy capacity is

$$C_s = \max_{p(\mathbf{w}),p(\mathbf{v})} \left\{I(\mathbf{w};\mathbf{y}) - I(\mathbf{w};\mathbf{y}_e)\right\}, \tag{5}$$

where $I(\mathbf{w};\mathbf{y})$ is the mutual information between information variables $\mathbf{w}$ and received vector $\mathbf{y}$ at Bob and $I(\mathbf{w};\mathbf{y}_e)$ is the mutual information between information variables $\mathbf{w}$ and received vector $\mathbf{y}_e$ at Eve. And the maximization is taken over all possible input distributions of $p(\mathbf{w})$ and $p(\mathbf{v})$. However, it is hard to begin the optimization process when $\mathbf{H}_e$ is unavailable.

Average Secrecy Capacity

Assuming that the communication lasts longer enough to experience all channel states, to average out the randomness of C_s when only CDI of $\mathbf{H}_e$ is available, we remark that C_s is a function of $\mathbf{H}_e$, and then, the average secrecy capacity is

$$\tilde{C}_s = \max_{p(\mathbf{w}),p(\mathbf{v})} \left\{I(\mathbf{w};\mathbf{y}) - I(\mathbf{w};\mathbf{y}_e|\mathbf{H}_e)\right\}, \tag{6}$$

where $I(\mathbf{w};\mathbf{y}|\mathbf{H}_e) = \mathrm{E}_{\mathbf{H}_e}[I(\mathbf{w};\mathbf{y}_e)]$ is the expected value of the conditional mutual information between $\mathbf{w}$ and $\mathbf{y}_e$ for given $\mathbf{H}_e$. And the maximization is taken over all possible input distributions of $p(\mathbf{w})$ and $p(\mathbf{v})$.

Instantaneous Secrecy Rate

Either in the instantaneous expression or in the average expression, it is hard to find optimal distributions of $p(\mathbf{w})$ and $p(\mathbf{v})$ to maximize the

secrecy capacity. Here follow the convention in Liu et al. (2017b, 2015), and use the secrecy rate instead of the secrecy capacity with Gaussian input alphabets and Gaussian AN, i.e., both $\mathbf{w}$ and $\mathbf{v}$ are circularly symmetric complex Gaussian vectors. In this case, the instantaneous secrecy rate can be expressed as

$$R_s = [C_m - C_w]^+, \tag{7}$$

where $[x]^+ = \max(x, 0)$. C_m and C_w are

$$C_m = \log_2 \det(\mathbf{I}_r + (P/t)\mathbf{H}_1\mathbf{H}_1^\dagger), \tag{8}$$

$$C_w = \log_2 \det\left(\mathbf{I}_e + \frac{(P/t)\mathbf{H}_2\mathbf{H}_2^\dagger}{(P/t)\mathbf{H}_3\mathbf{H}_3^\dagger + \mathbf{I}_e}\right),$$

respectively. Here, $\mathbf{H}_1 = \mathbf{HB} \in \mathbb{C}^{r\times s}$, $\mathbf{H}_2 = \mathbf{H}_e\mathbf{B} \in \mathbb{C}^{e\times s}$, and $\mathbf{H}_3 = \mathbf{H}_e\mathbf{Z} \in \mathbb{C}^{e\times d}$. However, the instantaneous secrecy capacity is hard to be calculated by Eq. (7) in the absence of Eve's CSI.

Average Secrecy Rate

Assuming the CDI of $\mathbf{H}_e$, the average secrecy rate can be expressed as

$$\begin{aligned}\tilde{R}_s &= \mathrm{E}_{\mathbf{H}_e}[C_m - C_w]^+ \\ &\geq C_m - \mathrm{E}_{\mathbf{H}_e}[C_w].\end{aligned} \tag{9}$$

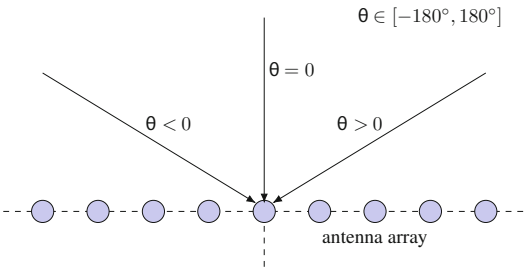
Equation (9) gets an equality if and only if the secrecy rates are always positive over all channel states. Since AN signals are used to jam the wiretap channels, the average secrecy rate is always nonnegative. From **Lemma 1** in the Appendix, $\mathbf{H}_2$, $\mathbf{H}_3$, and $\mathbf{H}_4$ are complex Gaussian matrices with their distributions

$$\mathbf{H}_2 \sim \mathscr{CN}_{e,s}(\mathbf{0}, \mathbf{R}_e \otimes \mathbf{I}_s), \tag{10}$$

and

$$\mathbf{H}_3 \sim \mathscr{CN}_{e,d}(\mathbf{0}, \mathbf{R}_e \otimes \mathbf{I}_d), \tag{11}$$

respectively. Hence, the expected value of average secrecy rates can be calculated by Monte Carlo simulations with Eve's CDI.



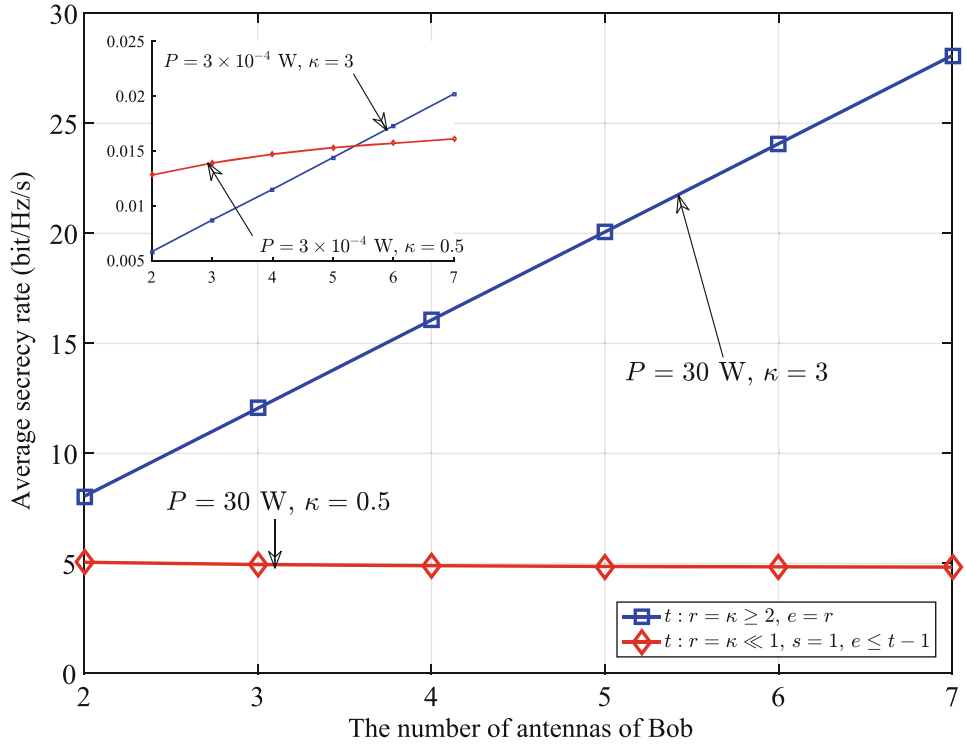
Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks, Fig. 4 AoA model with respect to a uniform antenna array

Correlated Matrices

The correlated matrices $\mathbf{R}_r$ and $\mathbf{R}_e$ can affect all the secrecy metrics of an AN-based system. As in Bolcskei et al. (2003), a correlated matrix $\mathbf{R}$ (generalized for $\mathbf{R}_r$ and $\mathbf{R}_e$) is a function of AoA (defined by θ) distribution. Here consider an AoA model with respect to a uniform antenna array based on 3GPP 3GP (2003), as shown in Fig. 4, where the AoA model belongs to a Gaussian distribution, the mean AoA of θ is $\bar{\theta}$, and the RAS (variance) of θ is δ. Then, we get the following simple model for the correlated coefficient $[\mathbf{R}]_{u,v}$[5], which is

$$\begin{aligned}[\mathbf{R}]_{u,v} &= \exp\left\{-j2\pi(u-v)d_u\cos\bar{\theta}\right\} \\ &\times \exp\left\{-\frac{1}{2}\left(2\pi\delta(u-v)d_u\sin\bar{\theta}\right)^2\right\},\end{aligned} \tag{12}$$

where $u \in \{1, \ldots, r\}$ and $v \in \{1, \ldots, r\}$ are the receive antenna index numbers. According to the parameters from 3GPP standard 3GP (2003), $\bar{\theta}$ is set in the range $[0°, 100°]$ and δ in the range $[5°, 35°]$, respectively. Assume that all antennas form a uniformly linear antenna array with $d_u = d_{\min}/\omega$, where $d_{\min}$ is the normalized minimum distance and ω is the wavelength. In the wiretap channel model, $\mathbf{R}_r$ and $\mathbf{R}_e$ have the same structure as $\mathbf{R}$. Here define the mean AoA at Bob and Eve as $\bar{\theta}_r$ and $\bar{\theta}_e$, respectively, and define the RAS at Bob and Eve as δ_r and δ_e, respectively.



Artificial Noise Schemes Based on MIMO Technology in Secure Cellular Networks, Fig. 5 Average secrecy rates in terms of the number of Bob's antennas

Simulations

Simulation results are provided to investigate joint impacts of the number of antennas and the number of selected message-sending eigen-subchannels on the average secrecy rates.

Figure 5 shows the relationship between the average secrecy rates and the number of Bob's antennas. In the case of adequate transmit antennas, i.e., $t : r = \kappa \geq 2$ and $e = r$, the average secrecy rates increase with an increasing number of Bob's antennas in both high ($P = 30\,\mathrm{W}$) and low SNR ($P = 3 \times 10^{-4}\,\mathrm{W}$) regions. In the case with a small number of transmit antennas and adequate receive antennas (here we set $t : r = \leq 1$, $s = 1$, and $e = t - 1$), we find that the average secrecy rates converge to a deterministic constant when e becomes large. It means that an increasing number of Bob's antennas bring in no benefit when Alice uses less antennas than Bob and selects $s = 1$, i.e., uses a transmit diversity scheme.

Conclusion and Further Work

Physical layer security will play a critical role in the future security architecture of cellular communications. This research should keep competing with cryptography and quantum communications, which are believed to be the three major security architectures for cellular communication systems. AN-based schemes are seen as promising methods and now have produced a raft of fascinating and important results. However, these AN-based schemes are optimal only under the condition of that the number of transmit antennas is larger than eavesdropper antennas. When the number of transmitted antennas is constrained or even smaller than that of eavesdropper

antennas, AN-based schemes cannot get positive secrecy capacities or rates. This motivates us to design a better AN-based scheme. In addition, the secrecy capacity (rate) quantization is a great challenge in AN-based schemes. Without perfect CSI of eavesdroppers, it is hard for encoders to calculate instantaneous secrecy capacities or rates. And the influence of fading does not allow us to use a simple AWGN or Rayleigh channel model to calculate a secrecy rate. It seems that we need to investigate main and wiretap fading channels that belong to other models of cellular communications. At last, power allocation in cellular networks should be investigated further, because power allocation problems are usually non-convex ones with a lot of cellular users.

Appendix

Lemma 1 (Proved in Gupta and Nagar 1999) *Define an* $r \times t$ *matrix* $\mathbf{A} \sim \mathscr{CN}_{r,t}(\mathbf{0}, \mathbf{R} \otimes \mathbf{I}_t)$ *as a receiver-side correlated central complex Gaussian matrix, and establish an independent* $(t \times s)$ *unitary matrix* $\mathbf{B}$. *We have*

$$\mathbf{AB} \sim \mathscr{CN}_{r,s}(\mathbf{0}, \mathbf{R} \otimes \mathbf{I}_s), \tag{13}$$

where $s \in \mathbb{R}$ *and* $t \geq s$.

Key Applications

Artificial noise (AN) schemes based on MIMO technologies can be applied to various cellular scenarios. For example, Massive MIMO systems have an enormous number of antennas, which offer more degrees of freedom for wireless channels, and a more secure performance in terms of secrecy capacities and the number of AN beamforming. The small call base stations deployed as cooperative jammers in cellular networks can be used to provide well-designed AN signals. In addition, the long-term evolution advanced (LTE-A) system supports device to device (D2D) communications, which is defined as the direct communications between two mobile users via shared radio resources with cellular users. D2D interference caused by the shared radio resources can be seen as AN signals to interfere with the illegitimate eavesdropper.

Cross-References

- IoT Security
- Security and Privacy in 4G/LTE Network
- Vehicle Ad-Hoc Networks: Services, Security, and Privacy
- Wireless Data Security
- Wireless Jamming Attack
- Wireless Key Establishment
- Wireless Sensor Network Security

References

3GP (2003) Spatial channel model for multiple input multiple output (MIMO) simulations. Rev. 6

Bolcskei H, Borgmann M, Paulraj AJ (2003) Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM. IEEE J Sel Areas Commun 21(3):427–439. https://doi.org/10.1109/JSAC.2003.809723

Gupta AK, Nagar DK (1999) Matrix variate distributions, vol 104. CRC Press, London

Liu S, Hong Y, Viterbo E (2015) Artificial noise revisited. IEEE Trans Inf Theory 61(7):3901–3911. https://doi.org/10.1109/TIT.2015.2437882

Liu Y, Chen HH, Wang L (2017a) Physical layer security for next generation wireless networks: theories, technologies, and challenges. IEEE Commun Surv Tutorials 19(1):347–376. https://doi.org/10.1109/COMST.2016.2598968

Liu Y, Chen HH, Wang L (2017b) Secrecy capacity analysis of artificial noisy MIMO channels–an approach based on ordered eigenvalues of Wishart matrices. IEEE Trans Inf Forensics Secur 12(3):617–630. https://doi.org/10.1109/TIFS.2016.2627219

Wyner AD (1975) The wire-tap channel. Bell System Tech J 54(8):1355–1367

Asynchronous Coordination Scheme/Protocol

- Asynchronous Coordination Techniques

Asynchronous Coordination Techniques

Ruoyu Su[1], Dengyin Zhang[1], Ramachandran Venkatesan[2], Cheng Li[2], Zijun Gong[3], and Fan Jiang[3]
[1]School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, China
[2]Faculty of Engineering and Applied Science, Memorial University, St. John's, NL, Canada
[3]Department of Electrical and Computer Engineering, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada

Synonyms

Asynchronous coordination scheme/protocol; Asynchronous wake-up coordination scheme/protocol; Asynchronous wake-up scheme/protocol

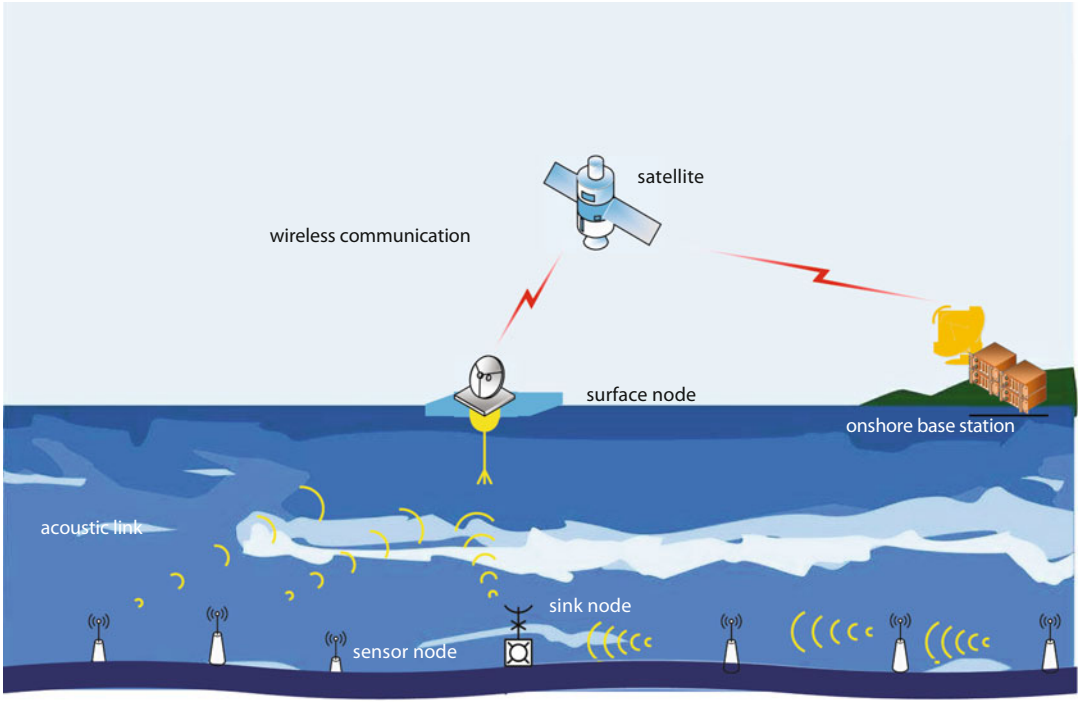
Definition

Asynchronous coordination technique is a communication approach that sensor nodes use different predetermined wake-up and sleep patterns for each cycle of network operations in order to achieve high energy efficiency and to guarantee network connectivity based on certain properties in linear algebra and optimization. No time synchronization is required in the network operations.

Historical Background

Underwater acoustic sensor networks (UWSNs) have attracted much research interest in recent years due to their wide range of potential applications such as marine environmental monitoring, undersea resources exploration, disaster monitoring and prevention, assisted location and navigation, and security monitoring (Akyildiz et al. 2015). Early generation of UWSNs include the Acoustic Local Area Network (ALAN), which was deployed by the Woods Hole Oceanographic Institution (WHOI) in 1994 of the coast of Monterey, California (Catipovic et al. 1993). ALAN, aiming at long-term data acquisition and ocean monitoring, consists of a number of sensor nodes placed on the seafloor and one surface node attached to a buoy. The underwater sensor nodes located on the seafloor transmit data to the surface node via acoustic links. Data was transmitted to the onshore base station via a radio frequency link. Another example of UWSNs is Seaweb 2000 (Rice et al. 2000), which was one of the first multi-hop underwater acoustic networks. This project advanced the underwater communication technology to achieve near-realtime, synoptic observation of the underwater environment. In Seaweb 2000, 14 sensor nodes (to sense, collect, and relay data) were placed on the seafloor and three gateway nodes with buoys (to collect data from sensor nodes and transmit data the onshore station) were deployed on the sea surface (Rice et al. 2002). A fundamental architecture of UWSNs is presented in Fig. 1. As shown in Fig. 1, underwater acoustic sensor nodes either directly report data to the surface node or transmit data to the sink node via multi-hop. The sink node aggregates data from different sensor nodes and then transmits to the surface node. The onshore base station will receive data from the satellite after the surface node reports collected data to the satellite.

Underwater acoustic communications consume more energy especially for long-range transmission (Zhang et al. 2015) compared with terrestrial communications. Furthermore, energy availability is limited in underwater acoustic communications because of the unchangeable and nonchargeable batteries in the sensor nodes. Frequently replacing sensor nodes generates high deployment cost, which is not realistic in long-term marine monitoring applications (Heidemann et al. 2012; Zhang et al. 2015). Therefore, unlike the high requirements of network throughput and packet delay, energy efficiency is a fundamental



Asynchronous Coordination Techniques, Fig. 1 The fundamental architecture of UWSNs

and significant issue when designing underwater network protocols.

An effective approach to energy saving is the wake-up coordination scheme. In the wake-up coordination scheme, each sensor node works periodically and has two modes of operations in each period: wake-up mode and sleep mode. In the wake-up mode, the transceiver is turned on for neighboring node discovery and data exchange (once the communication between a source node and destination node is established), whereas it remains off during the sleep mode so that energy consumption can be reduced. In general, the wake-up coordination scheme can be categorized into on-demand, synchronous, and asynchronous wake-up coordination schemes.

On-demand Wake-up Coordination Schemes

On-demand wake-up coordination scheme is a communication approach that each sensor node is equipped with a low power radio module to wake up the node for communication. The wake-up time can be predetermined according to application requirements, network data traffic, and so on. For example, cooperative underwater multichannel media access control (MAC) protocol (CUMAC) is a typical protocol based on the on-demand wake-up coordination mechanism, which adopts two acoustic transceivers to be equipped for one sensor node (Zhou et al. 2012). The low-priced acoustic transceiver is used to detect the channel by sending and receiving tone signal. The other acoustic transceiver is for data transmission. The tone pulse sent by the low-price acoustic transceiver at a predetermined time arrives at different sensor nodes at different times. Sensor nodes can catch the tone pulse by the low-price acoustic transceiver at the right time and then wake up the acoustic transceiver to exchange data packets at proper time. Similar mechanisms can be referred in Syed et al. (2008) and Jurdak et al. (2010).

Synchronous Wake-up Coordination Schemes

Synchronous wake-up coordination scheme is a communication approach that all sensor nodes are time synchronized before data transmission so that each sensor node can wake up at the proper time to send or receive data. Also, sensor nodes can turn off their transceivers to conserve energy when there is no data to transmit or receive. For example, underwater wireless acoustic networks MAC protocol (UWAN-MAC) is a distributed scalable energy-efficient scheme based on synchronous wake-up coordination mechanism (Min and Rodoplu 2008). In the UWAN-MAC protocol, before data transmission, each sensor node broadcasts signaling packets (i.e., time stamp) which includes information about the amount of time in sleep mode during one cycle and records information transmitted by its neighboring nodes to achieve time synchronization among all sensor nodes. The time stamps guarantee that neighboring nodes would wake up at the correct time in order to receive data packets without the knowledge of the propagation delay. Similar protocols can be found in van Dam and Langendoen (2003) and Ye et al. (2004).

Asynchronous Wake-up Coordination Schemes

Unlike the on-demand and synchronous wake-up coordination schemes, asynchronous wake-up coordination schemes utilize different predetermined wake-up and sleep patterns for each cycle of network operations to achieve high energy efficiency and to guarantee network connectivity without time synchronization during the network operations. By using asynchronous wake-up coordination schemes, two sensor nodes are able to communicate with each other at least once in one cycle if they are in the communication range. Meanwhile, sensor nodes can sleep as long as possible to reduce energy consumption when they do not have data to transmit and to forward. In general, each sensor node works periodically. In one period, the operations of each sensor node alternate between the wake-up mode and sleep mode. In the wake-up mode, the transceiver of the sensor node is turned on for neighboring node discovery and data exchange (once the communication between a source node and destination node is established). In the sleep mode, the transceiver of the sensor node remains off so that energy consumption can be reduced. Both the wake-up mode and the sleep mode may contain different numbers of time slots. A time slot is called an active slot when the sensor node operates in the wake-up mode and is called an inactive slot when the sensor node stays in the sleep mode. In this case, the operation status of a sensor node alternates between a predetermined number of active slots and inactive slots. To understand the diverse asynchronous wake-up coordination schemes, the following definitions and theorems are presented.

Preliminaries

Definition 1 (Quorum system) (Jiang et al. 2005) Let U be a universal set and $U = \{0,1,2,\ldots, n-1\}$. A quorum system Q is defined as a collection of non-empty subsets of U (e.g., G and H), which satisfies

$$\forall G, H \in Q, G \cap H \neq \emptyset. \tag{1}$$

For example, $U = \{0,1,2\}$, $Q = \{\{0,1\}, \{0,2\}, \{1,2\}\}$ is a quorum system under U.

Definition 2 (Rotational closure property) (Jiang et al. 2005) A quorum system has the rotation closure property when

$$\begin{aligned} &\forall G, H \in Q, i \in \{0, 1, 2, \ldots, n-1\}, \\ &\quad G \cap rotate\,(H, i) = \emptyset, \end{aligned} \tag{2}$$

where i is a nonnegative integer and *rotate* $(H, i) = \{(j + i) \bmod n | j \epsilon H\}$.

For example, the quorum system $Q = \{\{0,1\}, \{0,2\}, \{1,2\}\}$ under $\{0,1,2\}$ has the rotation closure property. However, the quorum system $Q' = \{\{0,1\}, \{0,2\}, \{0,3\}, \{1,2,3\}\}$ under $U = \{0,1,2,3\}$ has no rotation closure property

because $\{0,1\} \cap rotate\ (\{0,3\}, 3) = \emptyset$ (Jiang et al. 2005).

Definition 3 $(Z_v, +)$ (Stinson 2004) $(Z_v, +)$ is a finite group of order v, where is v is a positive integer.

Definition 4 (Cyclic difference set) (Stinson 2004; Choi and Shen 2011) Given a set D with v elements and k subset elements, $D : a_1, a_2, \cdots, a_k (\text{mod } v)$ is called a (v,k,λ) – cyclic difference set in $(Z_v, +)$ if for every $d \neq 0 \ (\text{mod } v)$ there are exactly λ ordered pairs (a_i, a_j), such that $a_i - a_j = d (\text{mod } v)$.

For example, $D = \{1,2,4\}$ is called a (7,3,1)–cyclic difference set in $(Z_7, +)$.

Theorem 1

*(Singer difference set) (*Stinson 2004 *;* Choi and Shen 2011 *) Let q be a prime power. Then there exists a* $(q^2 + q + 1, q + 1, 1)$*–cyclic difference set in* $\left(Z_{q^2+q+1}, +\right)$.

The cyclic difference sets can be constructed according to Theorem 1. For instance, (7,3,1)–cyclic difference set when $q = 2$, (13,4,1)–cyclic difference set when $q = 3$, (21,5,1)–cyclic difference set when $q = 4$. More cyclic difference sets can be found in Appendix A of Stinson (2004).

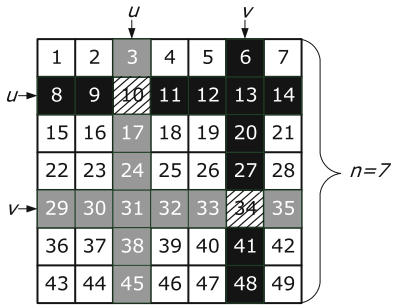
Theorem 2

(Multiplier theorem) (Stinson 2004) *Suppose there exists a* (v,k,λ)*–difference set D in an Abelian group* $(Z_v, +)$ *of order v. p is a multiplier of D if the following four conditions are satisfied: (1) p is prime; (2) gcd(p, v) = 1, gcd is greatest common divisor; (3)* $k - \lambda = 0 (\text{mod } p)$*; and (4)* $p > \lambda$.

According to Theorems 1 and 2, $p = 2$ is a multiplier of (21,5,1)–cyclic difference set because 2 satisfies the four conditions in Theorem 2. Based on the multiplier, it is easy to construct two (7,3,1)–cyclic difference sets: {1,2,4} and {3,5,6}.

Quorum-Based Energy Conserving (QEC) Protocol

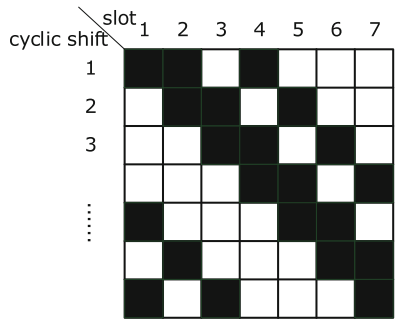
Figure 2 presents the basic idea of network operations in QEC. The white cube represents an inactive slot and the black and gray cube are an active slot. When node u and node v each arbitrarily choose a row and a column from a square of size n, respectively, their active intervals would overlap for at least two active slots, e.g., the 10-*th* and the 34-*th* slots shown in this figure ($n = 7$). Furthermore, there exists at least one overlapped active slot if two sensor nodes adopt two different grid sizes (Chao et al. 2006). It is obvious that a large value of n can lead to better energy conservation (larger number of time slots in one period associated with larger number of inactive slots) but it would also increase the latency. The QEC (Chao et al. 2006) accommodates different traffic loads by varying the value of n and thus achieve a balance between energy efficiency and packet delay. The simulation results show that QEC can outperform IEEE 802.11 in terms of energy consumption.



Asynchronous Coordination Techniques, Fig. 2 The fundamental idea of QEC

Cyclic Difference Set (CDS)-Based Protocol

The CDS-based protocol determines the number and positions of active and inactive slots in one cycle using as Singer difference set of a (v, k, λ)–cyclic difference set, where v, k, and λ represent the total number of slots, the number of active slots, and the minimum number of overlapping active slots in one cycle, respectively (Zheng et al. 2003, 2006). The CDS-based protocol guarantees at least one active overlapping active slot between any two sensor nodes for any cyclic shift.



Asynchronous Coordination Techniques, Fig. 3 An example of the network operations in the CDS-based protocol followed (7,3,1)–cyclic different set

Figure 3 presents an asynchronous wake-up coordination schemes followed (7,3,1)–cyclic different set. In one period, a sensor node has seven time slots including three active slots and four inactive slots. The positions of active slots (i.e., black cubes) and inactive slots (i.e., white cubes) are determined by the (7,3,1)–cyclic different set. There exists at least one active overlapping slot between any two cycle shift for the cases of perfect and no alignment of time slot boundaries Zheng et al. (2006). That is, the network connectivity can be guaranteed and the energy consumption on idle listening is reduced as well. Furthermore, the CDS-based protocol conserves more energy on idle listening than that of the QEC protocol due to the minimum number of overlapping active slots is one between any cycle shift.

Extension of CDS-Based Protocols

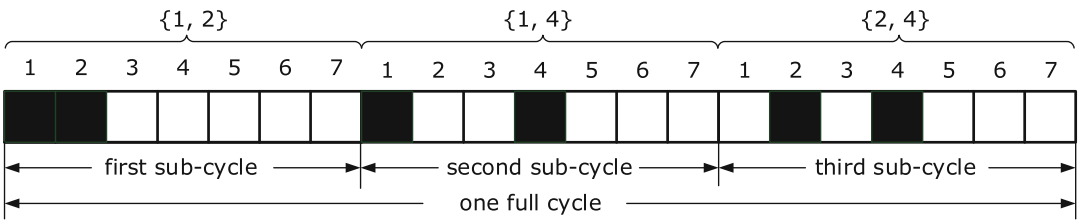
The extension of CDS-based protocols aims to reduce more energy consumption by leveraging a longer cycle period with less active slots without sacrificing network connectivity.

Exponential adaptive CDS-based protocol (EACDS) and multiplicative adaptive CDS-based protocol (MACDS) are the two extensions of the CDS-based protocols (Choi and Shen 2011). The key idea of these two schemes is the use of hierarchical arrangements of sets with the Kronecker product (Anderson 1998; Bernstein 2008). The EACDS-based scheme utilized a difference set scaled by another difference set, which is called the exponential set. The scaling was done by the Kronecker product and guaranteed that there was at least one overlapping slot between any two EACDSs with different duty cycles. For example, a difference set (v_I, k_I, λ_I), called an initial set, is scaled by another difference set (v_E, k_E, λ_E), called an exponential set. Then, the higher level hierarchical set can be obtained by the Kronecker product, i.e., $(v_I, k_I, \lambda_I) \otimes (v_E, k_E, \lambda_E)$, where operator $\otimes$ denotes the Kronecker product. The MACDS-based scheme used a multiplier set instead of the exponential set. All multiplier sets were selected from the relaxed difference sets (Luk and Wong 1997). The performance evaluation revealed that the MACDS-based scheme conserves more energy, whereas the EACDS is more suitable for scenarios where many different duty cycles are required.

Another extension of CDS-based protocol is to utilize permutation and combination to generate a longer cycle period for each sensor node (Su et al. 2016). Based on the Singer difference set, there exist $q+1$ $\left(C_{q+1}^{q}\right)$ possible position combinations for q active slots in one subcycle. For example, when $q = 2$, we get a (7,3,1)–cyclic different set: $D = \{1,2,4\}$. There are three possible position combinations which has two active slots in one subcycle, and they are {1,2}, {1,4}, and {2,4}, as shown in Fig. 4.

One full operating cycle of a sensor node contains concatenation of all possible subcycles, each corresponding to a distinct position combination of any q active slots. Therefore, an operating cycle is comprised of $(q+1)$ subcycles, which corresponds to $(q+1)\cdot(q^2+q+1)$ slots, $(q+1)\cdot q$ of which are active slots. Any ordering of subcycles can be used for concatenation. One working cycle period of a sensor node contains $(q+1)\cdot(q^2+q+1)$ active slots. Thus, the duty cycle is $q/(q^2+q+1)$. Figure 4 shows an example when $q = 2$.

Table 1 summarizes duty cycles used in Su et al. (2016) and the CDS-based protocol (Zheng et al. 2003). For a given value of q, the proposed scheme in Su et al. (2016) can generate a smaller duty cycle than the CDS-based protocol. It is obvious that a larger value of q can achieve a



Asynchronous Coordination Techniques, Fig. 4 Extension of the CDS-based protocol by permutation and combination

Asynchronous Coordination Techniques, Table 1 Duty cycles used in Su et al. (2016) and the CDS-based protocol (Zheng et al. 2003)

q	Duty cycles (%) (Su et al. 2016)	Duty cycles (%) (Zheng et al. 2003)
2	28.57	42.85
3	23.08	30.77
4	19.05	23.81
5	16.13	19.35
7	12.28	14.04
8	10.96	12.33
9	9.89	10.99
11	8.27	9.02

Asynchronous Coordination Techniques, Table 2 Duty cycles used in Su et al. (2016) with Kronecker product and EACDS

n	Duty cycles (%) (Su et al. 2016)	Duty cycles in EACDS (%) (Choi and Shen 2011)
1	19.04	28.57
2	12.70	19.04
3	8.47	12.70
4	5.64	8.47
5	3.76	5.64
6	2.51	3.76

Asynchronous Coordination Techniques, Table 3 Duty cycles used in Su et al. (2016) with Kronecker product and MACDS

(V_M, k_M, λ_M)	Duty cycles (%) (Su et al. 2016)	Duty cycles in MACDS (%) (Choi and Shen 2011)
(4,3,1)	21.43	32.14
(5,3,1)	17.14	25.71
(6,3,1)	14.29	21.42
(12,4,1)	9.52	14.29
(24,6,1)	7.14	10.71
(48,8,1)	4.76	7.14

smaller duty cycle and thus reduce more energy consumption for idle listening. However, normally, for a given smaller duty cycle, a sensor node requires more active slots to connect to a next-hop sensor node.

Tables 2 and 3 compare the proposed scheme in Su et al. (2016) that is extended using Kronecker product to the EACDS and MACDS proposed in Choi and Shen (2011), respectively. For EACDS, $(v_E, k_E, \lambda_E) = (3,2,1)$. For MACDS, (v_M, k_M, λ_M) is selected from the rotational set group as used in Luk and Wong (1997). For the scheme proposed in Su et al. (2016) and the EACDS and MACDS, $q = 2$ is chosen as an example. From Tables 2 and 3, for given values of q, n, and (v_M, k_M, λ_M), the smaller duty cycles can be generated by the scheme proposed in Su et al. (2016) using the Kronecker product rather than EACDS and MACDS.

Heterogenous CDS-Based Schemes

The length of a cycle period and the corresponding number of active slots determines the energy consumption of underwater sensor nodes when data exchanges do not frequently occur in long-term monitoring applications. When some nodes need to change the cycle length dynamically to satisfy the requirements of energy and packet delay and the other sensor nodes keep unchanged, the network connectivity will be lost when the wake-up coordination scheme do not guarantee that any two sensor nodes can communicate at least once in one cycle period. The heterogenous CDS-based schemes can keep the network connectivity when different CDS-based wake-up

coordination schemes are adopted by different sensor nodes.

Cyclic quorum system pair (CQS-pair) is a heterogenous asynchronous wake-up coordination scheme based on the CDS-based protocol, which is derived by using the Singer cyclic difference set (Theorem 1) and the multiplier theorem (Theorem 2) (Lai et al. 2010). Moreover, a verification matrix is used to verify that any two heterogenous difference sets have the rotation closure property. To be more specific, sensor node a adopts a difference set $\{a_1, a_2, a_3, \ldots, a_k\}$ in $(Z_N, +)$ and sensor node b adopts another difference set $\{b_1, b_2, b_3, \ldots, b_l\}$ in $(Z_M, +)$, where $N \leq M$ and $p = \lceil M/N \rceil$. The proposed verification matrix (Lai et al. 2010) is defined as a $pk \times l$ matrix $M_{l \times pk}$ whose element $m_{i,j}$ in the matrix is calculated by $\left(b_i - a_j^p\right) \pmod M$ and $a_j^p \in A^p$, as presented as follows.

$$M_{l \times pk} = \begin{bmatrix} b_1 - a_1^p & \cdots & b_1 - a_{pk}^p \\ \vdots & b_i - a_j^p & \vdots \\ b_l - a_1^p & \cdots & b_l - a_{pk}^p \end{bmatrix} \quad (3)$$

Sensor node a and b have the rotation closure property if $M_{l \times pk}$ contains all elements from 0 to $M - 1$, which means these two heterogeneous difference sets can keep network connectivity when sensor nodes utilize them respectively.

For example, according to Theorems 1 and 2, there are two (7,3,1)–cyclic difference sets: {1,2,4} and {3,5,6} in $(Z_7, +)$ when a cycle period contains 7 slots. Similarly, there are two (21,5,1)–cyclic difference sets: {3,6,7,12,14} and {7,9,14,15,18} in $(Z_{21}, +)$ when a cycle period contains 21 slots. By using Eq. (3), sensor nodes can switch between a (7,3,1)–cyclic difference set with {1,2,4} and a (21,5,1)–cyclic difference set with {7,9,14,15,18} without sacrificing the network connectivity. Another pair of wake-up coordination schemes is a (7,3,1)–cyclic difference set with {3,5,6} and a (21,5,1)–cyclic difference set with {3,6,7,12,14}. Unlike the proposed schemes in Chao et al. (2006), Zheng et al. (2003), Choi and Shen (2011), and Su et al. (2016), the CQS-pair (Lai et al. 2010) provides more choices for sensor nodes to adapt dynamic changes of network requirements without further improving the energy consumption for long-term monitoring applications.

Key Applications

Asynchronous wake-up coordination scheme has been involved in different scenarios, such as long-term marine environmental monitoring, undersea resource explorations, assisted location and navigation, delay-tolerant data transmission, and some energy-limited applications. With the development of internet of underwater things (IoUT), the communications among multiple underwater autonomous vehicles (AUV) and the cooperation between AUV and glider can utilize asynchronous wake-up coordination scheme to prolong the network lifetime and extend the monitoring area.

Future Directions

Asynchronous wake-up coordination scheme brings energy efficiency for UWSNs. However, it also prolongs the data packet delay compared with on-demand and synchronous wake-up coordination schemes. Some challenges are presented below.

- Balancing between energy consumption and packet delay

  It is obvious that the longer cycle period leads to low duty cycles for sensor nodes so as to significantly improve the energy efficiency. However, the packet delay is prolonged because not all sensor nodes are always active and the neighboring nodes discovery requires more time (or time slots) with a low duty cycle. Consequently, to achieve a balance between packet delay and energy consumption is a challenge for UWSNs when considering the nontrivial propagation delay by adopting asynchronous wake-up coordination scheme. On the other aspect, according to the data traffic or historic connection information (e.g., which active slots always enable the

communication connection), how to select a proper duty cycle (i.e., network operation pattern such as (7,3,1)–cyclic difference set-s: {1,2,4}) achieves a balance between energy consumption and packet delay, which can be further investigated.

- Mobility of sensor nodes

  In traditional UWSNs, all sensor nodes are supposed to be fixed on the seafloor in UWSNs. In the internet of underwater things (IoUT), AUVs serve as mobile sensor nodes, which can gather data in a large area for long-term monitoring applications. Directly utilizing aforementioned asynchronous wake-up coordination schemes may lead to non-trivial energy consumption and packet delay caused by the neighboring node discovery when a mobile sensor node is out of communication range of its intended nodes. How to adjust the asynchronous wake-up coordination scheme to accommodate the varying network topology in an energy-efficient manner is still an open issue in this area.

Cross-References

- ▸ Asynchronous Coordination Scheme/Protocol
- ▸ Asynchronous Wake-Up Coordination Scheme/Protocol
- ▸ Asynchronous Wake-Up Scheme/Protocol

Acknowledgment This work was supported in part by the National Natural Science Foundation of China under Grant 61571241.

References

Akyildiz F, Wang P, Sun Z (2015) Realizing underwater communication through magnetic induction. IEEE Commun Mag 53(11):42–48

Anderson I (1998) Combinatorial designs and tournaments. Oxford University Press, Oxford

Bernstein D (2008) Matrix mathematics: theory, facts, and formulas. Princeton University Press, Princeton

Catipovic J, Brady D, Etchemendy S (1993) Development of underwater acoustic modems and networks. Oceanography 6(3):112–119

Chao C, Sheu J, Chou I (2006) An adaptive quorum-based energy conserving protocol for IEEE 802.11 ad hoc networks. IEEE Trans Mob Comput 5(5):560–570

Choi B, Shen X (2011) Adaptive asynchronous sleep scheduling protocols for delay tolerant networks. IEEE Trans Mob Comput 10(9):1283–1296

Heidemann J, Stojanovic M, Zorzi M (2012) Underwater sensor networks: applications, advances and challenges. Philos Trans 370(1958):158

Jiang J, Tseng Y, Hsu C, Lai T (2005) Quorum-based asynchronous power-saving protocols for IEEE 802.11 ad hoc networks. Mob Netw Appl 10(1–2):169–181

Jurdak R, Ruzzelli A, O'Hare G (2010) Radio sleep mode optimization in wireless sensor networks. IEEE Trans Mob Comput 9(7):955–968

Lai S, Ravindran B, Cho H (2010) Heterogenous quorum-based wake-up scheduling in wireless sensor networks. IEEE Trans Comput 59(11):1562–1575

Luk W, Wong T (1997) Two new quorum based algorithms for distributed mutual exclusion. In: International conference on distributed computing systems, Baltimore, pp 100–106

Min KP, Rodoplu V (2008) UWAN-MAC: an energy-efficient MAC protocol for underwater acoustic wireless sensor networks. IEEE J Ocean Eng 32(3):710–720

Rice J, Creber B, Fletcher C, Baxley P, Rogers K, McDonald K, Rees D, Wolf M, Merriam S, Mehio R, Proakis J, Scussel K, Porta D, Baker J, Hardiman J, Green D (2000) Evolution of seaweb underwater acoustic networking. In: Proceedings of the MTS/IEEE conference and exhibition for ocean engineering, science and technology (OCEANS), Providence, vol 3, pp 2007–2017

Rice J, Amundsen K, Scussel K (2002) Seaweb 2002 experiment quick-look report. SPAWAR Systems Center, San Diego

Stinson D (2004) Combinatorial designs: constructions and analysis. Springer, New York

Su R, Venkatesan R, Li C (2016) An energy-efficient asynchronous wake-up scheme for underwater acoustic sensor networks. Wirel Commun Mob Comput 16(9):1158–1172

Syed A, Ye W, Heidemann J (2008) Comparison and evaluation of the T-Lohi MAC for underwater acoustic sensor networks. IEEE J Sel Areas Commun 26(9):1731–1743

van Dam T, Langendoen K (2003) An adaptive energy-efficient MAC protocol for wireless sensor networks. In: Proceedings of the ACM conference on embedded networked sensor systems (SenSys), Los Angeles, pp 171–180

Ye W, Heidemann J, Estrin D (2004) Medium access control with coordinated adaptive sleeping for wireless sensor networks. IEEE/ACM Trans Networking 12(3):493–506

Zhang X, Cui J, Das S, Gerla M (2015) Underwater wireless communications and networks: theory and application: part 1 [guest editorial]. IEEE Commun Mag 53(11):40–41

Zheng R, Hou C, Sha L (2003) Asynchronous wakeup for ad hoc networks. In: Proceedings of the ACM

international symposium on mobile ad hoc networking & computing (MobiHoc), Annapolis, pp 35–45
Zheng R, Hou C, Sha L (2006) Optimal block design for asynchronous wake-up schedules and its applications in multihop wireless networks. IEEE Trans Mob Comput 5(9):1228–1241
Zhou Z, Peng Z, Cui J, Jiang Z (2012) Handling triple hidden terminal problems for multichannel MAC in long-delay underwater sensor networks. IEEE Trans Mob Comput 11(1):139–154

Asynchronous Wake-Up Coordination Scheme/Protocol

▸ Asynchronous Coordination Techniques

Asynchronous Wake-Up Scheme/Protocol

▸ Asynchronous Coordination Techniques

Auction

Yanjiao Chen[1] and Qian Zhang[2]
[1]School of Computer Science, Wuhan University, Wuhan, People's Republic of China
[2]Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, People's Republic of China

Synonyms

Bidding; Transaction

Definitions

Auction is a concept from microeconomics, referring to the practice of buying and selling goods through the process of bidding. Auction participants include buyers, sellers, and the auctioneer, and the role of the auctioneer may be assumed by the seller, especially if there is a single seller. An auction mechanism is a series of rules specifying how an auction is conducted. The three most important components of an auction mechanism are the determination of the winners, the allocation of items, and the payment.

Historical Background

Auction is deemed as an effective way to assign items to buyers who value them the most. According to different numbers of participants, auctions can be categorized into three settings.

- *Forward auction*. One seller, who also acts as the auctioneer, and multiple buyers.
- *Reverse auction*. One buyer, who also acts as the auctioneer, and multiple sellers.
- *Double auction*. Multiple buyers, multiple sellers, and a third-party auctioneer.

In the forward auction, if there is a single piece of item, there are four standard auction mechanisms.

- *English auctions* or open ascending-bid auctions. Buyers openly bid against each other by iteratively raising their bids to be higher than the current highest bid. The auction terminates when no buyer offers a higher bid, and the buyer with the highest bid wins the item and pays the corresponding price.
- *Dutch auctions* or open descending-bid auctions. The auctioneer starts with a high bid and then lowers the bid until some buyer accepts and pays the corresponding price.
- *First-price sealed-bid auctions*. Buyers bid simultaneously without knowing each other's bids, i.e., sealed bids. The buyer with the highest bid wins the item and pays the corresponding price. Different from the English auction, buyers can only submit their bids once and cannot adjust the bids according to other buyer's bids.

- *Second-price sealed-bid auctions*. This type of auction is identical to the first-price sealed-bid auctions, with the exception that the winner pays the second-highest bid.

These four types of auctions can be easily extended to reverse auctions and to accommodate multiunit items, and there have been many variants and extensions of these standard auction mechanisms. Besides, there are several other more advanced auction mechanisms.

- *Combinatorial auctions*. Combinatorial auctions allow buyers to bid on combinations of items, such that the bid on a combination may not equal the sum of bids on individual items in the combination. For example, a buyer bids \$1 on item A and \$2 on item B, but may bid \$2.5 on the combination $\{A, B\}$ if A and B are substitutes, or bid \$4 on the combination $\{A, B\}$ if A and B are complements to each other.
- *All-pay auctions*. Normally, nonwinning buyers or sellers will not pay in auctions. All-pay auctions exert a compulsory payment on every participant and may be used to model entry fees or other costs in the process of auctions.
- *VCG auctions* (Vickrey 1961; Clarke 1971; Groves 1973). Vickrey-Clarke-Groves (VCG) auction mechanism tries to maximize social welfare with feasible allocations that satisfy the constraints of the auction (e.g., total number of auctioned items). Take forward auction as an example. The auctioneer finds one optimal feasible allocation A^* that maximizes the total bids of all winning buyers (usually through brute force). The payment of a winning buyer i with bid b_i is calculated as follows. The auctioneer removes buyer i and finds the optimal feasible allocation $\widetilde{A}^*$ among the rest of the buyers. Then, buyer i will be charged the price of $\sum_{j\neq i} b_j(\widetilde{A}^*) - \sum_{j\neq i} b_j(A^*)$, i.e., the difference of the utility of other buyers when buyer i participates in the auction or not. The VCG mechanism can achieve truthfulness and social welfare maximization, but its major drawback is high computational complexity, since it is often NP-hard to find the optimal feasible allocation.
- *McAfee auctions* (McAfee 1992). The McAfee auction mechanism is developed for single-unit or multiunit double auctions. Take the single-unit auction, for instance. Based on their bids, the auctioneer sorts the sellers in a non-descending order and the buyers in a non-ascending order. The auctioneer finds k, the last profitable pair of seller and buyer, i.e., the bid of the k-th buyer is higher than that of the k-th seller, but the bid of the $(k+1)$-th buyer is lower than that of the $(k+1)$-th seller. The first $(k-1)$ sellers win, and each receives the payment of the bid of the k-th seller; the first $(k-1)$ buyer win and each pays the bid of the k-th buyer.

Since auction participants are selfish and rational individuals who will adopt the optimal bidding strategy to gain the highest utility, game theory can be applied to analyze auctions. An auction is economic robust if the following three properties are satisfied.

- *Individual rationality*. A buyer or a seller is individually rational in the sense that they will not participate in the auction if the obtained utility is less than the utility of no participation. An auction mechanism is individually rational, if all sellers and buyers achieve nonnegative utility. This usually means that any seller is paid more than its bid, and any buyer pays less than its bid.
- *Truthfulness or strategy-proofness*. The buyers and sellers have the motivation to manipulate their bids to maximize their own utilities. Being truthful means that a seller or a buyer will submit a bid that equals their true valuation for the item. A truthful auction mechanism guarantees that a buyer or a seller cannot get a higher payoff by misreporting their true valuations; thus they will have no incentive to be untruthful.
- *Budget balance*. Budget balance is often considered in the double auction. It means that the auctioneer maintains nonnegative budget. In

other words, the payment that the auctioneer receives from all buyers is no less than the payment given to all sellers. For regulators, budget balance is often enough to motivate them to host spectrum auctions. The profit-oriented auctioneers, however, may aim at revenue maximization.

Apart from economic robustness, there are several other concerns in the design of auction mechanisms.

- *Social welfare*. Generally speaking, social welfare is the utility of all auction participants, including the buyers, the sellers, and the auctioneer. As the payment merely transfers utility among auction participants, it does not contribute to the social welfare. Therefore, the social welfare can be calculated as the difference between the total true valuation of winning buyers and the total true valuation of winning sellers. In a truthful auction, the social welfare is simply the total bid of winning buyers minus the total bid of winning sellers.
- *Collusions*. Auction participants may collude to rig the auction and gain a higher utility. For example, buyers may form collusions to bid against other buyers but not the buyers in the collusion, resulting in a favorable allocation and a lower payment. The seller may collude with the auctioneer to insert dummy bids to gain a higher payment. The practice of collusion in auctions has been revealed by many empirical studies, calling for effective countermeasures.
- *Privacy-preserving*. The bid of a buyer may be sensitive and private information that should be shielded from the non-trustworthy auctioneer and other rival bidders. For example, in a truthful auction, a buyer's bid is its true valuation, which may be closely related to its profit of winning the spectrum. The disclosure of sensitive information will render unbalanced advantage to the informed entities and cause economic damage to those whose information is divulged. Therefore, privacy-preserving auction mechanism design is gaining more and more attention.

Key Applications

The most successful application of auctions in wireless networks is dynamic spectrum allocation. Spectrum is an indispensable resource for wireless communications. To address the under-utilization problem resulted from static spectrum allocation, it is proposed to dynamically redistribute spectrum through auctions. Different from traditional goods, spectrum features interference-constrained spatial reuse, i.e., the same channel can be reused by multiple non-interfering buyers whose transmission ranges do not overlap. Therefore, the objective of spectrum auctions is to realize spectrum reusability while achieving economic robustness or social welfare maximization. The interference relationship of buyers is usually represented by the interference graph, an undirected graph constructed based on the transmission range of the spectrum and geographic information of the buyers. A typical interference graph can be denoted as $G = (V, E)$, in which the set of nodes V represents the set of buyers and the set of edges E represents interference relationship. If two buyers interfere with each other, there is an edge between them; otherwise, there is no edge between them.

The following are some typical spectrum auction mechanisms.

- *Forward spectrum auction* (Zhou et al. 2008; Jia et al. 2009). The single spectrum owner with M channels acts as the auctioneer, and each buyer can demand more than one channels. After sorting the buyers according to their bids in a non-ascending order, the auctioneer will sequentially check each buyer. If the buyer's demand d_i is fewer than the available channels $M - e_i$, in which e_i is the number of channels allocated to the buyer's interfering neighbors, the buyer will become a winner. A winning buyer will pay according to the critical value, which is the lowest value that the buyer has to bid in order to win.

- *Double spectrum auction* (Zhou and Zheng 2009). Assume that each buyer demands one channel and each seller owns one channel. The third-party auctioneer first divides buyers into groups, each of which only contains non-interfering buyers. The group bid is determined by the group size and the minimum bid in the group. Regarding each buyer group as a virtual buyer, the auctioneer can execute the McAfee auction mechanism to determine the spectrum allocation and the payment.
- *Heterogeneous spectrum auction* (Feng et al. 2012; Chen et al. 2014). The difference between homogeneous and heterogeneous spectrum auctions is that heterogeneous spectrum auction groups non-interfering buyers on a per-channel basis, since the interference relationship of buyers on different channels is different.
- *Combinatorial spectrum auction* (Zheng et al. 2015). Motivated by the fact that channels of contiguous frequency are easier to operate, combinatorial spectrum auction allows buyers to bid on combinations of channels. To realize spatial reuse, the same channel in the combinations of two non-interfering buyers is represented by two different virtual channels. Then, buyers are sorted in the non-ascending order according to the ratio of the bid to the combination size. The auctioneer then sequentially checks each buyer and selects winners as those whose demanded combination does not contain any channel that has already been allocated. A winning buyer will also pay its critical value, as in the forward spectrum auction.
- *Online spectrum auction* (Wang et al. 2010). Online spectrum auctions involve the temporal dynamics of spectrum demand and supply and require buyers to specify their requested time slots. At each time slot, the auctioneer collects bids from the arriving buyers and examines the availability of channels. Given the current buyers, the auctioneer will first conduct a screening process to remove the buyers whose bids are lower than the estimated future value of the channel. Then, the auctioneer can determine spectrum allocation among the remaining buyers using the double spectrum auction mechanism.

Apart from spectrum distribution, auction has been applied to other scenarios in wireless networks.

- *Power allocation*. Auctions can enable wireless users to cooperate in distributed power allocation to improve the overall network performance (Liu et al. 2013).
- *Crowdsensing*. Mobile crowdsensing platforms can adopt auctions to incentivize the participation and high-quality works from crowd workers (Luo et al. 2017).
- *Mobile data offloading*. Wireless service providers can use auctions to employ Wi-Fi or femtocell access points for data offloading (Iosifidis et al. 2015).
- *Video streaming*. Auctions can be used to motivate nearby wireless users to cooperate in video downloading and sharing, thus improving wireless video streaming performance (Han et al. 2009).

Cross-References

▸ Collusion
▸ Matching for Cooperative Spectrum Sharing

References

Chen Y, Zhang J, Wu K, Zhang Q (2014) TAMES: a truthful double auction for multi-demand heterogeneous spectrums. IEEE Trans Parallel Distrib Syst 25(11):3012–3024

Clarke EH (1971) Multipart pricing of public goods. Public Choice 11(1):17–33

Feng X, Chen Y, Zhang J, Zhang Q, Li B (2012) TAHES: a truthful double auction mechanism for heterogeneous spectrums. IEEE Trans Wirel Commun 11(11): 4038–4047

Groves T (1973) Incentives in teams. Econ J Econ Soc 41(4):617–631

Han Z, Su GM, Wang H, Ci S, Su W (2009) Auction-based resource allocation for cooperative video transmission protocols over wireless networks. EURASIP J Adv Sig Process 2009:4

Iosifidis G, Gao L, Huang J, Tassiulas L (2015) A double-auction mechanism for mobile data-offloading markets. IEEE/ACM Trans Netw (TON) 23(5):1634–1647
Jia J, Zhang Q, Zhang Q, Liu M (2009) Revenue generation for truthful spectrum auction in dynamic spectrum access. In: The 10th international symposium on mobile ad hoc networking and computing. ACM, pp 3–12
Liu Y, Tao M, Huang J (2013) An auction approach to distributed power allocation for multiuser cooperative networks. IEEE Trans Wirel Commun 12(1):237–247
Luo T, Kanhere SS, Huang J, Das SK, Wu F (2017) Sustainable incentives for mobile crowdsensing: auctions, lotteries, and trust and reputation systems. IEEE Commun Mag 55(3):68–74
McAfee RP (1992) A dominant strategy double auction. J Econ Theory 56(2):434–450
Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. J Financ 16(1):8–37
Wang S, Xu P, Xu X, Tang S, Li X, Liu X (2010) TODA: truthful online double auction for spectrum allocation in wireless networks. In: IEEE symposium on new frontiers in dynamic spectrum. IEEE, pp 1–10
Zheng Z, Wu F, Chen G (2015) A strategy-proof combinatorial heterogeneous channel auction framework in noncooperative wireless networks. IEEE Trans Mob Comput 14(6):1123–1137
Zhou X, Zheng H (2009) TRUST: a general framework for truthful double spectrum auctions. In: International conference on computer communications. IEEE, pp 999–1007
Zhou X, Gandhi S, Suri S, Zheng H (2008) eBay in the sky: strategy-proof wireless spectrum auctions. In: The 14th annual international conference on mobile computing and networking. ACM, pp 2–13

Auction Mechanism

▸ Incentive Mechanism for Crowdsourcing-Based Spectrum Measurement

Authenticated Query Processing

▸ Verifiable Cloud Computing

Authentication

Si Chen
West Chester University, West Chester, PA, USA

Synonyms

IEEE 802.11; IEEE 802.1x; Wireless credentials; WLAN authentication

Definitions

Authentication methods in wireless networks allow the wireless access point (AP) or broadband wireless router to authorize individuals and devices, and only those who get authorized can access the network. For maximum security, client devices should also authenticate to the wireless network using pre-shared key (PSK) or Extensible Authentication Protocol (EAP) authentication. The use of mutual authentication is essential in a wireless network. This will guard against many security issues, such as man-in-the-middle (MITM) attacks. With authentication, the wireless client and the wireless AP must prove their identity to each other.

Historical Background

When setting up a wireless network, the network administrator has to have some way to ensure that wireless clients, and the devices they are using, are trusted. A typical solution to this question is to use a standard kind of authentication, which in today's wireless network area means IEEE 802.11 and 802.1x authentication.

The first wireless network standard, 802.11, permits a network to be made relatively secure if the user establishes their wireless network using the Wired Equivalent Privacy (WEP) protocol (IEEE 1999). Since the data is transmitted over the air, it is a simple task for a malicious device to sniff it and grab sensitive information. To avoid this type of attack, WEP consists of a

passphrase that uses secret and shared encryption keys that are generated by the wireless host and then passed to the wireless client. These keys are used for encrypting all the data that travels across the airwave and thus thwarting anyone trying to sniff the network. In WEP, there are two authentication mechanisms: (1) open system authentication (OSA) and (2) pre-shared key authentication (PSK). However, it is well known that the IEEE 802.11 WEP protocol is vulnerable to two types of attacks: (1) message privacy and message integrity attacks (Johnson and Maltz 1996) and (2) probabilistic cipher key recovery attacks, such as the Fluhrer-Mantin-Shamir attack (Fluhrer et al. 2001; Stubblefield et al. 2002). In 2003, the Wi-Fi Alliance announced that WEP had been superseded by Wi-Fi Protected Access (WPA) protocol. WPA protocol has the addition of the Temporal Key Integrity Protocol (TKIP). TKIP avoids the problems of WEP by implementing per-packet key mixing with a rekeying system and message integrity check (MIC) mechanism. In 2004, as an updated version of WPA, WPA2 was released along with the IEEE 802.11i standard. Both WPA and WPA2 support the PSK authentication key distribution mechanism and use a stronger encryption than WEP. It is designed for home and small office networks and doesn't require an authentication server.

Another wireless network standard, the IEEE 802.1x, is widely adopted for large wireless area networks such as a hotel or big company. It takes the predecessor IEEE 802.11 one step further and includes computer and user identification, dynamic key creation, and centralized authentication. IEEE 802.1x supports the Internet Authentication Service (IAS) which uses the Remote Authentication Dial-In Service (RADIUS) protocol. In addition, 802.1x ties to the Extensible Authentication Protocol (EAP) which supports multiple authentication methods such as Kerberos, public key authentication, and certificates.

Many security issues have been found in these wireless authentication protocols. In 2008, an attack on TKIP (Beck and Tews 2009) was released which allows an adversary exploiting WPA's MIC. In 2017, security researchers have discovered a significant vulnerability, named KRACKs (Vanhoef and Piessens 2017), in WPA2. It uses a group of vulnerabilities that exist in the implementation of the authentication protocol to launch key reinstallation attack and eventually allows the attacker to intercept and steal data. In January 2018, Wi-Fi Alliance announced the release of WPA3 with several security improvements over WPA2.

Foundations

There are three main methods of authentication that are used on wireless networks.

- **Open System Authentication:** Open system authentication allows any wireless device to connect to the network as long as the end device knows the service set identifier (SSID) used on the network. The significant advantage of this type of authentication is its simplicity: any client can directly connect without complicated configuration. The problem with this authentication scheme is that the SSID is typically broadcast so any device within the wireless coverage range can join the network. Moreover, the SSID can be obtained from the association request frames even if the access point is set not to broadcast the SSID.
- **Pre-Shared Key Authentication:** The pre-shared key authentication method is commonly used on individual and small business wireless networks. It uses a pre-shared key that is distributed to the client device before establishing the connection. The PSK allows any wireless client who has the key to join the network. The client also has the option to use the key to encrypt the data. This can protect sensitive data traveling across the wireless network from becoming captured. The original 802.11 WEP protocol utilizes RC4 for encryption and has been depreciated due to several vulnerabilities. The improved WPA protocol ties with the TKIP which utilizes dynamic keys that were not supported with WEP for encryption. However, a similar vulnerability has been found since TKIP uses the same procedures that WEP does. WPA2

replaced TKIP with Counter Mode Cipher Block Chaining Message Authentication Code Protocol (CCMP) which is primarily based on Advanced Encryption Standard (AES), and it is proven to provide strong encryption.

A PSK is relatively easy to configure though it requires some configuration at the client side. However, a wireless network based on a PSK does not scale well. With a large number of clients, it becomes more difficult to periodically update the PSK.

- **EAP (Extensible Authentication Protocol) Authentication:** EAP authentication is an arbitrary authentication mechanism that is used in the wireless network, and it is part of the 802.1x standard. EAP allows for several choices of authentication methods including MD5, EAP-TLS, EAP-TTLS, and PEAP. In order to use EAP to authenticate users when they connect to the wireless network, a third-party authentication server is required at time of the client association, such as RADIUS. The wireless host opens the client's port for other types of traffic based on access rights stored in the authentication server.

Key Applications

Authentication is one of the most important parts of wireless protocol which is used for establishing user identities and for securing the wireless communication.

Cross-References

- ▸ Access Control
- ▸ Adaptive Medium Access Control for Internet-of-Things-Enabled MANETs
- ▸ Blockchain: Enabling Future Internet with Intrinsic Security

References

Fluhrer S, Mantin I, Shamir A (2001) Weaknesses in the key scheduling algorithm of rc4. In: International workshop on selected areas in cryptography. Springer, Berlin, Heidelberg, pp 1–24

IEEE (1999) Wireless lan medium access control (mac) and physical layer (phy) specifications. IEEE Std 80211

Johnson DB, Maltz DA (1996) Dynamic source routing in ad hoc wireless networks. In: Mobile computing. Springer, Boston, MA, pp 153–181

Stubblefield A, Ioannidis J, Rubin AD (2002, February) Using the Fluhrer, Mantin, and Shamir Attack to Break WEP. In: Proceedings of the Network and Distributed System Security Symposium (NDSS), San Diego, California, USA

Tews E, Beck M (2009) Practical attacks against WEP and WPA. In: Proceedings of the second ACM conference on Wireless network security (WiSec '09). ACM, New York, NY, USA, 79–86. https://doi.org/10.1145/1514274.1514286

Vanhoef M, Piessens F (2017) Key reinstallation attacks: Forcing nonce reuse in wpa2. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. ACM, Dallas, Texas, USA, pp 1313–1328

Authentication in Delay-/Disruption-Tolerant Networks

▸ Security in Delay-Tolerant Networks

Automotive Cybersecurity

▸ Automotive Security

Automotive Security

Nevrus Kaja, Ahmad Nasser, Di Ma, and Adnan Shaout
University of Michigan, Dearborn, MI, USA

Synonyms

Automotive cybersecurity; In-vehicle security; Vehicle cybersecurity

Definitions

Automotive Security is the field of assessing cyber-threats against automotive systems and developing security countermeasures for detecting, preventing the corresponding attacks, and reacting to a security breach. As modern vehicles become connected, the threat of attack becomes increasingly real. Coupled with advances in driver assistance and autonomous solutions, the impact on security is further expanded. Thus, automotive security aims at ensuring automotive systems can achieve their intended function without the danger of being maliciously manipulated by an attacker. A few definitions used in vehicle security are provided below:

- Cyber-physical system: A system of collaborating computational elements controlling physical entities
- Cybersecurity: An attribute of a cyber-physical system that relates to avoiding unreasonable risk due to an attack (Committee SVESS et al. 2016)
- Attack: Exploitation of vulnerabilities to obtain unauthorized access to or control of assets with the intent to cause harm
- Threat: A circumstance or event with potential to cause harm
- Exploit: An action who uses vulnerability to cause undesired, unexpected, or unanticipated behavior
- CAN: Control Area Network is a dominant communication network protocol used for intra-vehicle communication

Historical Background

Automotive security related threats initially came to light with the introduction of keyless entry systems. For such vehicles, the primary threat was that of theft, and vehicle manufacturers were keen on implementing security features to prevent a thief from wirelessly unlocking the vehicle and driving off (Biham et al. 2007). Security features included rolling codes as well as immobilizer units that can prevent the vehicle from starting unless the correct key is present. Another area that initially gained recognition was the tire pressure monitoring sensors (TPMS) (Heary 2010). These sensors gave off RF signals that could be used to allow tracking of vehicles, which would violate the driver's privacy. As vehicles added more connectivity features through a telematics unit, USB devices, bluetooth, and Wi-Fi, the attack vectors increased drastically and so did the interest of the research community to uncover potential exploits that could be found in these systems. The work of Checkoway et al. (2011) was instrumental in showing the extent to which attacks could be launched against vehicles through various attack surfaces like OBD tools, CD Roms, and more. This brought cybersecurity to the forefront in the minds of vehicle OEMs who knew that security needed more attention. The attacks by Miller and Valasek (2015) launched remotely on the Jeep in 2015 further demonstrated that the threats were not only academic but real, and lead FCA to ship USB sticks with patches to close the security vulnerabilities uncovered by the attacks. Even before that date, several standardization bodies were aware of the need to address cyber threats on vehicle systems. One of the early ones was EVITA, a European Commission tasked with defining security use cases and deriving security requirements for vehicle systems. EVITA produced 18 use cases and defined a security architecture for on-board automotive networks as well as a threat assessment method.

In addition to EVITA, several other industry initiatives were created to address automotive security. Some of these are provided below:

- SAE J3061: The first cyber-security framework to address all the main security aspects of producing secure automotive systems (Committee SVESS et al. 2016).
- C2C: Car to Car communication consortium aims to create a European industrial standard for communicating cars.
- ETSI C-ITS: A not-for-profit organization aims to achieve standards for Cooperative Intelligent Transport Systems.

- USDOT/CAMP: A joint initiative to produce a security credential management system (SCMS).
- AUTOSAR WP-X: The security working group within AUTOSAR produces software specifications to define security functions in embedded systems.
- UPTANE: An academic and industry co-initiative to standardize Over-the-Air updates in vehicles.
- AUTO-ISAC: An industry and government co-initiative for information sharing across the automotive supply chain relating to cyber threats (Auto ISAC 2017).
- ISO and SAE joint working group to produce a common cyber security standard: ISO-SAE 21434.
- IEEE Standard for Wireless Access in Vehicular Environments (IEEE 1609.2).

Foundations

To study automotive security, one usually starts with threat analysis and risk assessment to identify attack surfaces, attack agents, security objectives, and controls.

Threat Analysis and Risk Assessment processes: In a fashion that parallels the hazard and risk assessment of functional safety (ISO 2011), automotive security requires the enumeration of threats and the qualification of the corresponding risk. There exist several threat assessment methods such as EVITA-THROP (Henniger et al. 2009), OCTAVE (Alberts et al. 2003), TVRA (IMS 2008), HEAVENS (Islam et al. 2014), and NHTSA composite modeling (McCarthy et al. 2014). Once the threats are identified and the corresponding risk is calculated, product developers derive security requirements to address the associated threats. The higher the risk of an attack and its impact, the more stringent the security level would be, which means added hardware countermeasures, and a stricter security profile.

Attack Surfaces: With added connectivity comes the risk of added exposure to security threats due to the ever expanding attack surface. As shown in Fig. 1, today's vehicles offer numerous attack surfaces. In order to address these cyberthreats, automotive security engineers create security defenses both at the vehicle boundary and at the internal vehicle layers. Figure 1 provides an overview of vehicle communication interfaces. Dotted lines represent wireless connections, while solid lines represent wired connections.

Attack Agents: An important element in automotive security is the agent who is performing the attack. Different agents will have different capabilities, resources, expertise and motivations. For example, a mechanic has far less capabilities and different attack motivations from a nation state hacker. The following is a list of different attack agents who might compromise a vehicle.

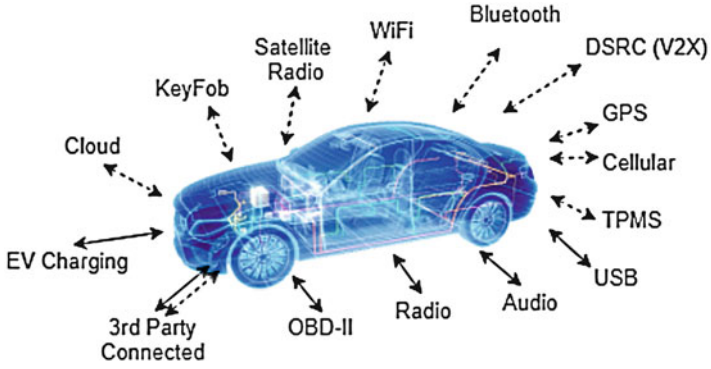
Threat Agent

- Thief
- Vehicle Owner
- Mechanic
- Hacktivist
- Organized Crime
- Researcher
- Terrorist
- State sponsored

Security Objectives: Automotive security is different from IT security due to the dimension of control which results in security threats impacting safety. Still, several security objectives are typical of IT systems with the exception that the assets being protected are automotive specific. Among the primary security objectives for automotive systems are:

- Prevent attacks that have an impact on vehicle safety
- Protect driver privacy and prevent vehicle tracking
- Protect vehicle OEM reputation and brand
- Protect algorithms and IP especially ones related to Autonomous and ADAS systems, from theft through reverse engineering

Throughout the automotive industry, there is a general consensus that automotive security



Automotive Security, Fig. 1 Attack surfaces

requires a security-in-depth approach. This translates to providing security at each layer of a vehicle interface (known as the layered approach). The first layer being the vehicle external interface (telematics unit, OBD, infotainment, etc.). The second layer being the vehicle network gateway that separates different vehicle domains and allows for filtering network data inside the vehicle. The third layer being the vehicle network within each domain. And the fourth layer being the Electronic Control Unit itself, primarily the microcontroller unit, sensors, and the software that goes along with it. Building defenses at each one of these layers is meant to increase the attack difficulty to a point that discourages an attacker from attempting to compromise a vehicle system, as the payoff would not justify the required effort.

Security Primitives and Key Applications

To ensure automotive security, researchers in the industry and academia have come up with a range of security requirements that apply to automotive ECUs:

- Secure Debug: Accessing to JTAG as well as any debug interface shall be locked to prevent reverse engineering and theft of vehicle IP.
- Secure Diagnostics: Vehicles contain extensive diagnostic services meant to allow the OEM to fully diagnose an ECU failure to aid in vehicle maintenance as well as test for emissions. Such diagnostics shall be protected against attackers who can use these services to manipulate the vehicle.
- Secure Programming: Automotive ECUs rely on flash bootloaders to update vehicle software. All application and calibration data shall be cryptographically signed to prevent an attacker from modifying the vehicle software.
- Secure Communication: Automotive systems are highly reliant on CAN as the primary network protocol. Since CAN has no built-in security features, it is important to define security protocols that can prevent an attacker from spoofing control or status messages on the CAN bus.
- Secure Boot: Checking the authenticity of vehicle software after each boot cycle provides added assurance that vehicle software was not tampered with while at rest.

The above security requirements apply to a variety of automotive ECUs. The below list highlights specific security application areas:

- Telematics: Provide various forms of connectivity such as cellular and Wi-Fi. Therefore, they open the door to remote-based attacks. Such systems require strong firewalls to block unwanted traffic into the vehicle.
- Infotainment: Such systems are primarily used to provide audio/video and navigation to the end user. By their nature, they use rich operating systems like Android and Linux. To isolate such applications from critical vehicle systems, sandboxing and trustzones are important

to provide isolation between the trusted and untrusted world.

- V2X: Vehicle-to-Everything is another emerging set of communication protocols. Such systems have special requirements for processing a large number of ECC calculations to authenticate other entities. Due to the high potential impact to safety, the hardware in such systems is expected to support FIPS 140-2 level 3 or higher HSMs.
- Vehicle Gateway: Network domains separate vehicle functions into separate CAN bus systems. In this scenario, methodologies such as network isolation and intrusion detection systems are used to detect and prevent outside and inside intruders (Kaja et al. 2017; Zhang et al. 2018).
- In-Vehicle Network: In-Vehicle networks enable ECU-to-ECU communication via robust and reliable networks. In order to secure such networks, different mechanisms are used such as AUTOSAR SecOC, for authentication, and Freshess.
- Sensors: The number of sensors in a vehicle has ballooned to allow understanding of its surrounding. Authenticating sensor data and source is important to prevent sensor spoofing or sensor counterfeiting.
- AUTOSAR based ECUs: As one of the most widely prevalent automotive software platform, AUTOSAR now presents a rich area of study in automotive embedded security. Therefore, applying best practices within AUTOSAR for increasing the security resilience of such systems is mandatory (Nasser et al. 2017)

Cross-References

- ▸ Dedicated Short-Range Communication (DSRC)
- ▸ Industrial Cyber-physical Systems
- ▸ Internet of Things: Architecture, Key Applications, and Security Impacts
- ▸ Network Security
- ▸ Vehicular Networks
- ▸ Wireless Data Security

References

Alberts C, Dorofee A, Stevens J, Woody C (2003) Introduction to the octave approach. Technical report. Carnegie-Mellon University

Auto ISAC (2017) Automotive information sharing and analysis center. https://www.automotiveisac.com/. Accessed 6 May 2018

Biham E, Dunkelman O, Indesteege S, Keller N and Preneel B (2007) How to steal cars – A practical attack on keeLoq. In: Crypto 2007, Santa Barbara, California

Checkoway S, McCoy D, Kantor B, Anderson D, Shacham H, Savage S, Koscher K, Czeskis A, Roesner F, Kohno T et al (2011) Comprehensive experimental analyses of automotive attack surfaces. In: USENIX security symposium, San Francisco

Committee SVESS et al (2016) SAE j3061-cybersecurity guidebook for cyber-physical automotive systems. SAE-Society of Automotive Engineers

Heary J (2010) Defcon: Hacking tire pressure monitors remotely. https://www.networkworld.com/article/2231495/cisco-subnet/defcon—hacking-tire-pressure-monitors-remotely.html

Henniger O, Ruddle A, Seudié H, Weyl B, Wolf M, Wollinger T (2009) Securing vehicular onboard IT systems: the EVITA project. In: VDI/VW automotive security conference, Ingolstadt

IMS (2008) Final draft etsi es 282 007 v2. 0.0 (2008-03)

Islam M, Sandberg C, Bokesand A, Olovsson T, Broberg H, Kleberger P, Lautenbach A, Hansson A, Söderberg-Rivkin A, Kadhirvelan S (2014) Deliverable d2-security models. HEAVENS Project, Version 1

ISO (2011) 26262: road vehicles-functional safety. International Standard ISO/FDIS 26262

Kaja N, Shaout A, Ma D (2017) A two stage intrusion detection intelligent system. In: The International Arab Conference on Information Technology, Tunisia

McCarthy C, Harnett K, Carter A (2014) Characterization of potential security threats in modern automobiles: a composite modeling approach. Technical report

Miller C, Valasek C (2015) Remote exploitation of an unaltered passenger vehicle. Black Hat USA 2015

Nasser AM, Ma D, Muralidharan P (2017) An approach for building security resilience in AUTOSAR based safety critical systems. J Cyber Secur Mobil 6(3):271–304

Zhang L, Shi L, Kaja N, Ma D (2018) A two-stage deep learning approach for CAN intrusion detection. NDIA ground vehicle systems engineering and technology symposium, Michigan