Chapter 9
The Ecosystem of Connected RADIO Systems

Angelos Charalambidis, Giannis Mouchakis and Stasinos Konstantopoulos

9.1 Introduction

RADIO home deployments live and act inside the *RADIO ecosystem*. The RADIO home deployments integrate seamlessly in the RADIO ecosystem as nodes to an abstract network regardless of the different communication technologies and of the heterogeneous hardware and software components that comprise them.

The RADIO ecosystem provides the necessary mix of components and interconnections in order to support useful operations such as clinical report inspection, privacy-preserving data analysis, notifications, and home automation. These operations must comply with requirements regarding the protection of the sensitive data produced in each RADIO home.

This design aims to place the local network of each RADIO home in the overall context of the RADIO ecosystem of communicating RADIO homes, caregivers, and care institutions; and to do so in a way that

- Allows only relevant information to be shared and ensures the security of private data and extracted information;
- Can scale to a large number of RADIO home deployments managed by a single-care institution; and
- Can handle heterogeneity in communication technologies and hardware and software components.

A. Charalambidis (✉) · G. Mouchakis · S. Konstantopoulos
Institute of Informatics and Telecommunications, NCSR "Demokritos", Aghia Paraskevi, Greece
e-mail: acharal@iit.demokritos.gr

V. Karkaletsis et al. (eds.), *RADIO–Robots in Assisted Living*,
https://doi.org/10.1007/978-3-319-92330-7_9

9.1.1 Requirements

The design of the RADIO ecosystem architecture takes into account several requirements on the connectivity and availability of RADIO home deployments and the management of the sensitive data transmitted from and to the RADIO home deployments and the clinical sites.

The clinical sites connect and retrieve data from the RADIO home deployment. During connectivity loss, the RADIO home deployment should continue working autonomously and record data that can be later retrieved from the clinical site. However, RADIO home deployment should be able to transmit urgent notifications to the caregiver and clinical institution about the state of the deployment. In such a rare case, the RADIO ecosystem must be able to identify the disconnected home deployments and notify accordingly.

9.1.1.1 Privacy Requirements

The majority of the data produced by the RADIO home deployment are considered private, and certain privacy requirements must be taken into account in every layer of management (i.e., exchanging, retrieving, and processing).

Exchange of private data. Data exchange between RADIO home deployments and clinical institutes should be secured in the sense that no other party except the ones that are communicating can eavesdrop on the data exchanged. This fact includes parties from outside as well as from inside the RADIO ecosystem.

Management of private data. Clinical institute applications can retrieve private data only from the RADIO home deployments that have permission to do so. Moreover, different users of the clinical institute application must have different levels of clearance. The application should be able to validate a user's identity when user credentials are provided.

Processing of private data. Computations over data of all the RADIO home deployments can be defined in order to retrieve potentially useful statistical results for clinical experimentation. Those results should be presented to authorized researchers through the clinical experimentation application. However, the computations over the private data of the RADIO home deployments should be aggregated values and should not reveal individual data points of a known home deployment. In other words, one should be not able to distinguish either an individual data point or the initial RADIO home that has been retrieved from. Moreover, except the result produced, other parties of the RADIO ecosystem should not deduce (or gain access to) other parties of private data. The privacy-preserving processing will be discussed in more detail in Sect. 9.3.

9.2 The Architecture of the RADIO Ecosystem

In this section, we describe the architecture of the RADIO ecosystem and the interactions between its main entities and components.

9.2.1 Entities and Components

A *RADIO home deployment* is the main entity of the RADIO ecosystem. A RADIO home is essentially an appropriate setup space where the robot and the primary subject live and perform their daily activities. The RADIO home is a smart home in the sense that smart sensors and actuators are deployed in that space. A RADIO home can be, for example, a real house building or an appropriately configured room inside a health institution.

A *health institution* is an institution that provides care for the primary subjects and therefore has deployed several RADIO homes. A health institution, for instance, can be a hospital or a rehabilitation center. The medical personnel that monitor and provide care for a primary subject are considered to be the formal caregivers of that RADIO home and should have access to the clinical reports produced by the RADIO home deployment.

Apart from the formal caregivers, there exist the roles of the *informal caregiver* that are essentially people with no medical expertise but can be notified in case of an emergency that occurred in the RADIO home.

Last but not least, the *research centers* are also entities of the RADIO ecosystem. Those research centers are interested in conducting statistical analysis and data mining to the clinical data produced by the RADIO home deployments. The certified personnel that can analyze those data will be called health researchers.

RADIO home is the main data provider of the RADIO ecosystem. It processes the raw data retrieved from the deployed sensors and the robot and securely stores the processed data. The main components that reside in a RADIO home with respect to the other RADIO ecosystem components are as follows:

- *Sensing and Recognition System* that is the collection of sensors deployed in the RADIO home and in the robot, and the system of algorithms that can recognize abstract events from sensor measurements. This system produces the events that are stored in the "EventLog" database.
- *Actuation System* that is the collection of actuators in the RADIO home and in the robot that can perform actions in the physical world.
- *Smart Home Controller* that is responsible to integrate the sensor and actuation communication protocols (e.g., WiFi and Z-Wave) and dispatch measurements from sensors to the IoT platform and actuations from the IoT platform to the actuation system.

- *Medical Report Generator* that provides processed clinical data to the authorized personnel of the care institute in order to evaluate the condition and wellness of the subject.
- *Notification Generator* that provides alerts and notifications about urgent events occurring in the home to the registered caregivers and care institute personnel via the appropriate applications.
- *Privacy-Preserving Data Mining Component* that participates in and provides partial computations to a distributed computation initiated by the clinical experimentation application.

RADIO home deployments live and act inside the RADIO ecosystem. The RADIO home deployments integrate seamlessly in the RADIO ecosystem as nodes to an abstract network regardless of the different communication technologies and of the heterogeneous hardware and software components that comprise them.

9.2.2 Topology

Figure 9.1 presents an instance of a RADIO ecosystem network topology. In that topology, a collection of RADIO homes is logically organized as subsidiaries of a health institution. In the physical world, those RADIO homes can either reside in the health institution or be in a remote location. In either case, the health institution is considered to be connected to the RADIO home. The formal caregiver application is considered to be local to the health institution and can access the RADIO homes that are controlled by the health institute. The health researcher is using the clinical experimentation application to perform data analysis and be either located inside a dedicated research institute or inside the health institute.

We can distinguish the following communication channels:

- *Data Channel* that is the main channel of the network. This channel is used by the components of the ecosystem in order to provide or consume data. All the components must be able to connect to the data channel.
- *Notification Channel* that is the channel where the notifications flow. It is mainly used to provide notifications to the informal caregiver's application, and therefore, in principle, the only requirement is the connectivity between the informal caregiver's device and the corresponding RADIO home deployment.
- *Synchronization Channel* is the channel that connects the health institutions and the research centers of the ecosystem in order to synchronize their local registry data. It should be noted that in practice the data and the synchronization channels might share the same physical medium of transportation. However, conceptually there are two separate channels that have different connectivity requirements.

The organization of the RADIO ecosystem assumes that deployed RADIO homes must be associated with exactly one health institution. In other words,

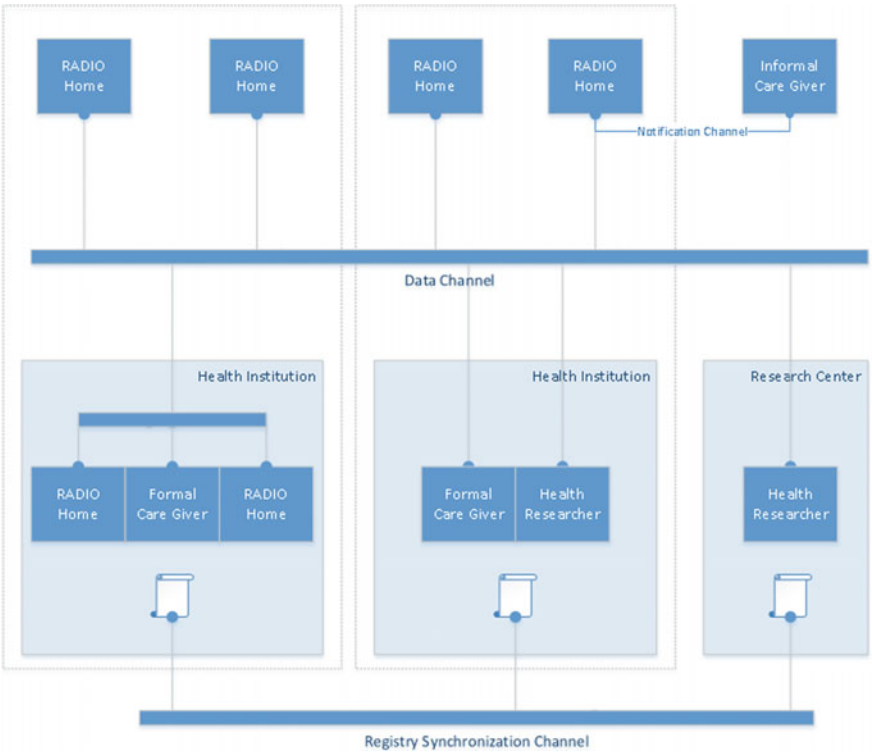

Fig. 9.1 The RADIO ecosystem network topology

RADIO homes cannot exist without an associated health institution that is most probably the institution which is affiliated with the formal caregiver.

The health institution also provides a discovery service to the other components of the RADIO ecosystem. More specifically, the care institution application and the clinical experimentation application can use the discovery service in order to locate the appropriate RADIO homes to contact. In order to provide that kind of service, the health institutes maintain a local registry of the registered RADIO homes. Each health institute synchronized its registry with the other health institutes in the RADIO ecosystem via the synchronization channel. This synchronization yields a global registry that can be used to discover components across the ecosystem.

9.2.3 Security in Communication

The overall communication architecture between the caregiver's environment and the health institution environment, as well as between the health institution environments themselves has to take into account the sensitivity of personal data that must be exchanged between the above-mentioned entities.

The use of Virtual Private Network (VPN) technology, which extends a private network across a public network, such as the Internet, is a necessity in order to seamlessly achieve the communication objectives of the RADIO project in a secure way. The proposed protocol suite to be used for the implementation of the VPNs in the RADIO project is the Internet Protocol Security (IPSec).

IPSec is an open standard protocol suite for secure Internet protocol (IP) communications by authenticating and encrypting each IP packet of a communication session. IPSec includes protocols for establishing mutual authentication between agents at the beginning of the session and negotiation of cryptographic keys to be used during the session. IPSec can be used in protecting data flows between a pair of hosts (host-to-host), between a pair of security gateways (network-to-network), or between a security gateway and a host (network-to-host). It uses cryptographic security services to protect communications over IP networks. IPSec also supports network-level peer authentication, data origin authentication, data integrity, data confidentiality (encryption), and replay protection. It is an end-to-end security scheme operating in the Internet layer of the Internet protocol suite, in contrast with other widespread Internet security systems, which operate in the upper layers at the application layer. This unique feature allows user applications to be automatically secured by IPSec at the IP layer.

The health institution environments may be interconnected by deploying IPSec between their respective security gateways (network-to-network), also known as site-to-site IPSec VPN, by means of either a partial mesh or a full mesh topology. The caregiver's environment may be interconnected with the respective health institution by deploying IPSec between the security gateway of the health institution and the workstation and/or mobile device of the caregiver's environment (network-to-host), also known as remote access IPSec VPN, by means of a hub and spoke topology where the health institutions are considered hubs and the caregiver's environments spokes. In either case, a secure IPSec tunnel, which provides the secure transmission of sensitive personal data in a transparent to the application way, is created between the respective endpoints.

The deployment of IPSec VPNs between the RADIO project entities, which need to exchange sensitive data, is a scalable standardized solution that builds secure data channels on top of the Internet which is considered an "untrusted" network. The overall design allows also the creation of multiple connections (using site-to-site VPNs or remote access VPNs) between entities, which provide a level of redundancy in case of communication network failures. The "extension" of the several private networks over the VPN connections facilitates the overall network monitoring of the various devices that are located in disparate networks.

9.2.4 Changes in the Topology

During the lifecycle of the RADIO ecosystem, it is natural to expect that RADIO homes will be deployed or removed from the ecosystem, and health institutes and

research centers will join. We distinguish three different procedures that must be followed when a new site is added to the topology.

Assume first the deployment of a new RADIO home. As mentioned previously, each RADIO home is affiliated with a health institution.

1. After the installation of the physical devices (e.g., Smart home sensors, RADIO robot) in the RADIO home, the RADIO home (specifically the smart home controller) is registered to the IoT platform. This includes the deployment of a signed digital certificate.
2. The health institution registers the newly deployed RADIO home to its local registry. The registration procedure produces VPN credentials for the RADIO home that are transferred in a secure and off-band way and deployed in the RADIO home in order to establish a secure connection with the affiliated health institution.
3. Authorization and authentication of the appropriate medical personnel are defined by the administrator of the health institution.
4. Informal caregivers that will be notified for urgent events are also defined for the specific RADIO home.

Assume now that the health institution decides to join the ecosystem. This requires the following steps:

1. Deploy the site-to-site VPN with other health institutions that require a valid signed digital certificate.
2. Install the registry component and initiate the first synchronization with the rest of the health institutions. This requires to know at least one institution or research center that has already joined the RADIO ecosystem.
3. Start deploying the RADIO homes following the steps described previously.

Lastly, the procedure for a research center is similar to the health institution with the difference that the registry is read only. More specifically, the procedure comprises the following steps:

1. Deploy the site-to-site VPN with other health institutions that require a valid signed digital certificate.
2. Install the registry component and initiate the first synchronization with the rest of the health institutions. This requires to know at least one institution or research center that has already joined the RADIO ecosystem.
3. Define the researchers that are authorized to access the data analysis module of the RADIO ecosystem.

9.3 Remote Medical Assessment

The formal caregivers of the RADIO ecosystem must have access to the recorded data of their assigned subjects for health monitoring reasons. In contrast to the health researchers, the formal caregivers are interested in a more detailed and focused report about the activities of their assigned subjects.

The events that occur in a RADIO home are recorded in a database and can be accessed by the formal caregiver's frontend. Each database corresponds to a specific RADIO home (and, by extension, to a specific subject). The nature of the recorded data nominates the time-series database management system as the predominant choice for using the RADIO ecosystem. More specifically, InfluxDB, an open-source time-series database, is used as the backend of the RADIO ecosystem data management.

The schema of each database consists of multiple time series that contain the following information:

- *Time* of the occurred event.
- *Participant* an alphanumeric identifier used as a field key to identify the subject of the event. This is mainly used for visualization and reporting reasons. Since each database corresponds to a single SubjectID, the field is expected to be constant throughout the time series. However, the existence of this field helps during the reporting of multiple subjects.
- *Measurement* is a field value that contains the measured quantity of the event. The measurement type and value range depend on the type of the event.

The recorded events include

- chair transfer;
- bed transfer;
- four-meter walk;
- pill intake;
- TV watching;
- meal preparation; and
- going out of the room.

The events are visualized as graphs and tables. The visualization tool used in our case is Grafana, an open-source user interface focused on visualizing time series in various ways.

Figure 9.2 depicts an overview of the visualizations implemented for the formal caregiver's data access. The events that contain measurable quantities (chair transfer time, bed transfer time, and time to walk 4 m) are depicted as line graphs over time, while the other events are depicted as tables.

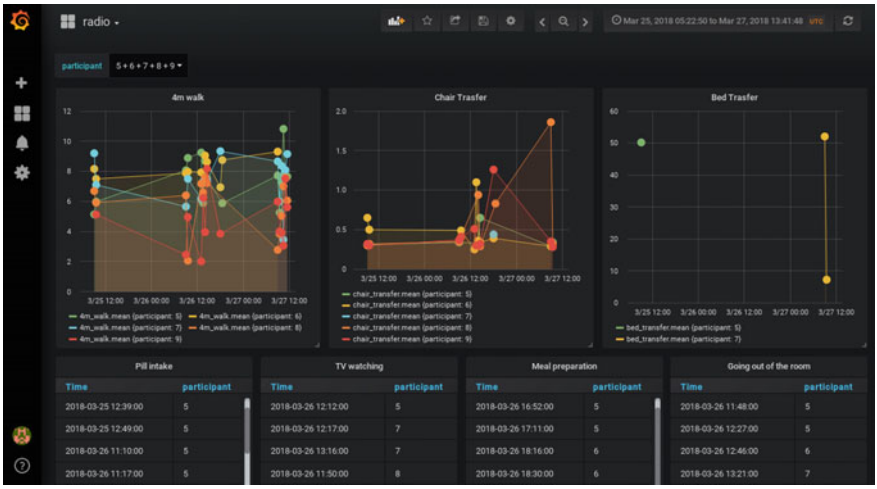

Fig. 9.2 An overview of visualizations provided to the medical assessment personnel

9.4 Privacy-Preserving Statistical Analysis

The insights gained by the large-scale analysis of health-related data can have an enormous impact in public health and medical research, but access to such personal and sensitive data poses serious privacy implications for the data provider and a heavy data security and administrative burden on the data consumer. The discussion on what exactly it means to not disclose private data [1] and the discussion on policies for balancing between scientific advancement and privacy [2] are very relevant but should be complemented by the equally relevant discussion of whether there is tension at all between data privacy and data-driven research. In other words, it is not straightforward if private data can be insulated from medical research workflows without compromising either.

As anonymization has been repeatedly proven to be inadequate [3], attention has turned to research in cryptography and distributed computation. These fields can provide methods for computing aggregates and statistics without revealing the specific data values involved in the computation, offering a much stronger guarantee of privacy than anonymization. However, from the perspective of the data mining practitioners and the medical researchers, there is still a residue of functionality missing between their workflows over anonymized data and what is technically possible to achieve without accessing specific data points.

The RADIO ecosystem architecture foresees the statistical analysis of the data produced by the RADIO home deployments in a way that no requirements are violated. In particular, a health researcher should be able to use the clinical experimentation application in order to pose statistical queries against the collection of the data that reside in the RADIO ecosystem. However, the data produced in each RADIO home deployment are considered private, and as such the

privacy-preserving data mining (PPDM) component must respect the privacy requirements for managing and processing such data.

The proposed architectural approach is reduced into two main ideas:

- The set of the valid statistical queries that are allowed by the system must be appropriately restricted in order to avoid exposing individual data points but only aggregated data. A wide range of existing statistical method depends only on aggregation of data, and therefore can also be formalized in that restricted query set.
- Instead of fetching raw data in order to compute the aggregation, the computations will be performed local to the data and only the result will be transmitted. In order words, the RADIO home deployment will contain the processing units in order to perform partial computations on its private data. Moreover, multiple RADIO homes must collaborate in such a way that can produce the final result of the computation and at the same time will not expose any of their private data points.

The scope of our discussion here is restricted to the data and processing required to empirically validate an already formulated hypothesis over a larger dataset than what can reasonably be made available to research. Naturally, part of the researchers' workflow involves browsing data in order to formulate a hypothesis. This initial hypothesis formulation remains in the scope of smaller experimental data specifically collected and licensed to be shared.

The system architecture can be perceived as a stack of three layers, and each layer depends on the functionality provided from the layer at the lower stage. The upper layer, called the Medical Researcher's interface, accepts from the medical researcher the method with the initial parameters to be executed by the system. The purpose of this interface is to provide a familiar environment to the researcher, and therefore in our current implementation this layer is developed in the R language. The initial parameters are transformed appropriately in order to be passed to the next layer, which is the compilation layer. At that stage, the high-level parameters and commands of the statistical method are transformed into low-level instruction for accessing the private databases of the agents. An instruction represents an aggregation over a selection of data. Currently, the aggregation operation is summation. However, the aggregations that are both feasible by the system and safe for preserving privacy depend on the secure protocol used. These instructions will be eventually evaluated by the lowest layer of the architecture, the privacy protocol layer. Figure 9.2 depicts the system architecture and the information exchanged between the layers.

9.4.1 The Compilation Layer

This layer is responsible for the communication between the two other layers. Specifically, it translates the arguments of the secure statistic to a suitable format,

and thus it defines the appropriate data that are going to be used for the statistic computation. Moreover, it converts the simple statistic equations to a set of summations, a compatible format to achieve the secure summation protocol. Therefore, a set of instructions is composed where each instruction represents a summation equation of the statistic with the appropriate parameter's set for its computation. During the execution, the compilation layer gives the privacy protocol layer a single instruction at a time and it receives its result. After the execution of the whole set, it computes the statistic and the analysis parameters. The statistic result is sent back to the Medical Researcher's interface layer.

9.4.2 The Aggregation Protocol

This layer executes the privacy protocol between the AAL agents. To deal with the concurrent computation of each instruction, we model our agents as actors. Each actor makes the appropriate computations with respect to the given instruction and its private data. These computations can easily be done since every AAL agent controls its corresponding health records. After the computation of the value, which represents the initial secret, the privacy protocol is executed. The protocol may involve all the actors to work collaboratively in order to compute the aggregation of their secrets without revealing the actual secrets to each other or the agent requesting the aggregation. The aggregated result is collected a designated actor. The selection of such actor is irrelevant and can be done randomly.

9.4.3 Discussion

The proposed system architecture assumes that

- the statistical analysis that is to be carried out can be implemented using the set of aggregation instructions provided by the aggregation protocol. In other words, the algorithm should not depend on individual data points.
- a summation protocol exists that guarantees privacy.

The first assumption holds, since the most commonly used class of data mining algorithms can be expressed as an iteration of summation expressions [4]. If needed, categorical operators can be implemented based on summation [5]. Regarding the second assumption, we will now proceed to discuss the summation protocols that can be used in our architecture.

Most of the related studies guarantee their privacy by utilizing encryption or differential privacy techniques. These approaches do not fit in our problem, because we deal with medical history data that are distributed across AAL agents. In homomorphic techniques, a master agent shares a public key with the rest of the

agents, in order to encrypt their data, and keeps a private key for the final decryption. Such a mechanism is privately weak in the case of collaborative computations, because if the medical researcher (master agent) and one AAL agent collude, they can learn another AAL agent's private value. This makes the technical protocol weak, as it places a heavy burden on non-technical policies and protocols to guarantee the integrity of the medical researcher. Since our main aim is to alleviate the need for non-technical policies and protocols and to make it easier for medical researchers to run statistics over data point, they are not meant to access directly, and homomorphism encryption does not cover our requirements.

In addition, differential privacy is also not applicable, from both the perspective of the medical researcher and from that of the AAL agent. From the perspective of the medical researcher, differential privacy computes approximations. From the perspective of the AAL agent, the secret value can be approximated by its repeated querying, since a different perturbation of the real secret needs to be computed for each query. The AAL agent cannot produce a single perturbed value and use this for all queries, since it needs to be re-computed to follow the distribution parameters requested by the medical researcher. This might be less of a problem in time-series data (such as power grid data or traffic data) but can result in substantial information leaking in static historical data, such as health records.

9.4.4 The RASSP Protocol

The RADIO data mining system is unaware of the underlying privacy-preserving protocol that it is using. In this section, we will present the RASSP (RADIO Secure Summation Protocol) that satisfies the requirements needed by the system to ensure privacy preservation.

Secret sharing schemes divide a secret into many shares which can be distributed to n mutually suspicious agents. The initial secret can be revealed if any k of these n agents combine their shares. We will call such schemes as *(k, n)*-threshold schemes. If such a scheme also possesses the homomorphism property, then multiple secrets can be combined by direct computation only on the shares. Such schemes are usually called composite secret sharing schemes [6].

More specifically, assume n mutually suspicious agents and each agent holds a secret s_i. The desired computation is combined into a super-secret s under an operation $\oplus$. Using a secret sharing scheme, each s_i can be split into k shares d_{i1}, …, d_{ik} such that given a known function F_I it is the case that $s_i = F_I(d_{i1}, \ldots, d_{ik})$.

The composition of the shares d_1, d_1' yields a *super-share* $d_1 \otimes d_1'$. In other words, the $(\oplus, \otimes)$-homomorphism property implies that the compositions of the shares under the operator $\otimes$ are shares of the composition under the operator $\oplus$.

Overall, the advantage of having a composite secret sharing scheme is that secret cannot be obtained, only if k or more agents collude and combine their sub-shares. In addition, this protocol is suitable to our approach from the AAL agent's point of view, because it does not use a trusted third party or depends on cryptographic

assumptions, while at the same time it is *k*-secure. This approach represents a secure summation protocol that can easily be applied to collaborative agent systems. Based on this mathematical foundation, we will now proceed to present the RASSP protocol, a ($\oplus$, $\otimes$)-homomorphic composite secret sharing scheme.

Figure 9.3 gives an example of the above description for a system of three AAL agents. In this example, *House1* has the private value v_1 and produces three numbers: r_{11}, r_{12}, and r_{13}. Then, it shares r_{12} and r_{13} with *House2* and *House3*, keeping r_{11} hidden. *House1* receives two numbers (r_{21}, r_{31}) from the other AAL agents. It then shares the computed Y_1, so that F_I can be computed by summing all Y_i. $F_I(Y_1, Y_2, Y_3)$ computes the sum of all three AAL agents' secret values (Fig. 9.4).

The described secure summation protocol is suitable for computing medical statistics and preserves privacy at the same time. The only constraint is that the resulted outcome is a sum of the private values, and thus the statistic equations should be converted into a summation form. The summation form results in accurate values and not approximations, while simultaneously it can easily be

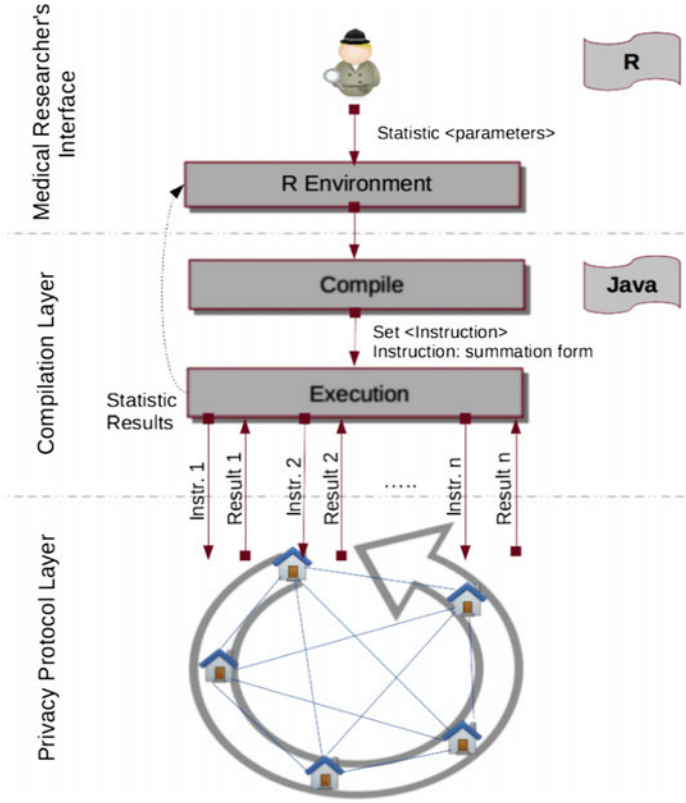

Fig. 9.3 The system architecture and the information exchanged between the layers

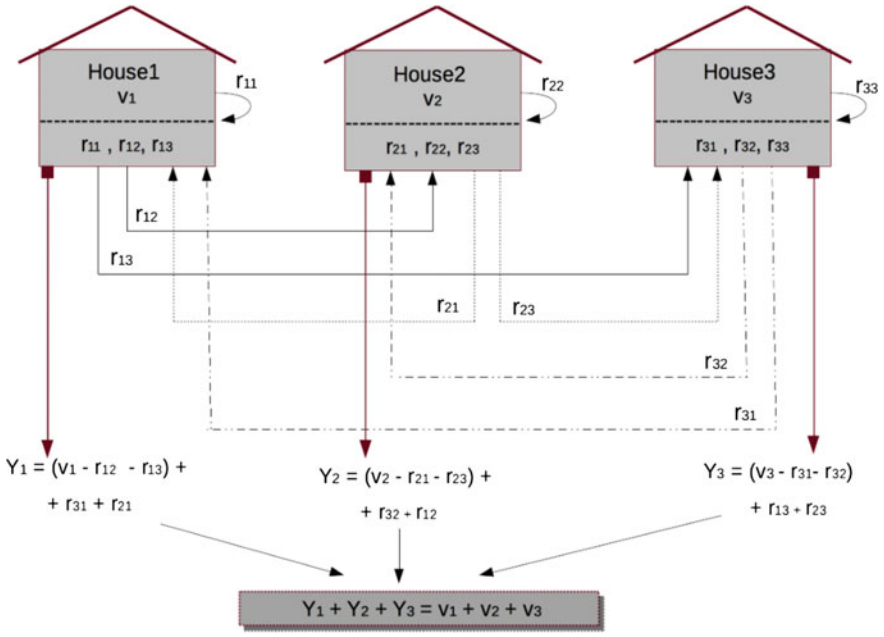

Fig. 9.4 An example of the RASSP protocol

parallelized [7]. Besides, medical researchers typically use descriptive statistics which utilize numerical descriptors such as mean and standard deviation. These descriptors can easily be converted into a summation form and thus computed by our system.

9.5 Conclusions

RADIO proved the concept of connecting RADIO homes and medical institutions into the RADIO ecosystem, adding value to the health data collected by RADIO homes by making it available not only for medical monitoring but also for medical research. Specifically, RADIO developed network security and access control guidelines for direct access to health data by the competent medical personnel, as well as the RASSP protocol for the privacy-preserving mining of the data collected in each home's database. These expose appropriate programmatic interfaces, so that, and depending on one's access and use case, individual data and time series can be visualized to monitor particular end users and statistical data aggregations can be visualized or used by R programs to carry out medical research.

The RADIO ecosystem is a central part of the overall solution offered in particular in light of the future project exploitation and commercialization, as it touches

upon uptake by medical institutions and public bodies. The most innovative outcome is the ability to serve sensitive health data to medical research without compromising privacy.

References

1. Clifton, C., Kantarcioglu, M., & Vaidya, J. (2002). Defining privacy for data mining. *National Science Foundation Workshop on Next Generation Data Mining, 1*(26).
2. Horvitz, E., & Mulligan, D. (2015). Data, privacy, and the greater good. *Science, 349*(6245), 253–255.
3. Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57,* 1701.
4. Kearns, M. (1998). Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM), 45*(6), 983–1006.
5. Kissner, L., & Song, D. (2005). Privacy-preserving set operations. In *Annual International Cryptology Conference*. Berlin: Springer.
6. Benaloh, J. C. (1986). Secret sharing homomorphisms: Keeping shares of a secret secret. In *Conference on the Theory and Application of Cryptographic Techniques*. Berlin: Springer.
7. Chu, C.-T., Kim, S. K., Lin, Y.-A., Yu, Y. Y., Bradski, G., Ng, A. Y., et al. (2007). Map-reduce for machine learning on multicore. *Advances in Neural Information Processing Systems*.