Learning health systems

Ecarot, Thibaud; Fraikin, Benoit; Lavoie, Luc; McGilchrist, Mark; Ethier, Jean Francois

Published in:
Proceedings - 2021 IEEE 34th International Symposium on Computer-Based Medical Systems, CBMS 2021

DOI:
10.1109/CBMS52027.2021.00038

Publication date:
2021

Document Version
Peer reviewed version

Learning Health Systems: An Anonymous Network Routing Protocol

Thibaud Ecarot*, Benoit Fraikin*, Luc Lavoie*, Mark McGilchrist†, Jean-François Ethier*
*GRIIS, Université de Sherbrooke, Sherbrooke, Québec, Canada
†Division of Population Health Sciences , University of Dundee, Dundee, UK

2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) | 978-1-6654-4121-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CBMS52027.2021.00038

***Abstract*—Learning Healthcare Systems are an emerging approach to healthcare research as translated into practice. For this purpose, a strong interconnection comes to be a necessity when dealing with healthcare services, research and knowledge transfer all at once. Practically, these connections imply that a routing protocol should guarantee anonymity to entities in compliance with both laws and ethical requirements while restricting the quantity of information obtainable had an entity been compromised. In order to bring more protection and meet all the requirements, a new message routing protocol is offered to allow the use of data access paths and to resist traffic analysis security threats. The protocol protects the addresses and roles pertaining to entities from any lurking malevolent minds by implementing proxies into a mix-network. Moreover, flows of synthetic datasets and contents identifiers are handled separately so as to curb any risk of re-identification. A model of this protocol is provided in the form of a multi-objective optimization problem, natively integrating objectives of minimizing both latency and entropy of the information exchanged. The assessment of this model shows that the constrained separation of data flows has a minimal impact on delay times, which not only reveals to be an acceptable compromise but also significantly increases security in data access.**

***Index Terms*—Routing Protocol, Secure Data Access, Protocols, Privacy, E-health**

I. Introduction

Learning Healthcare Systems (LHSs) imply a joint connection between healthcare practices, research and knowledge transfer. To reach their full potential, a flexible, efficient and secure system has to be implemented so that access to needful health data is ensured at the right time. This approach is based on data access paths, typically starting with a patient's healthcare data. Then, research entities will access this data to transfer knowledge to initial care. In order to keep the various stakeholders confident in the system, data access has to be sensibly processed; leaving the data accessible when necessary for a specific purpose and fully visible to all the actors in a planned health activity. Any unplanned change in the initial activity shall require a re-evaluation of all the actors involved. Keeping numbers of entities with access to data-bearing message limited and data sources denominational is part of the requirements from the LHSs. To meet these needs and improve health research and, eventually, patient care, a new protocol suite has been designed, namely Sensitive Data Exchange Protocol Suite (SDEPS) for healthcare [1].

This protocol suite provides access to a great deal of data from several organizations that use heterogeneous systems in different areas (e.g., healthcare, research). On top of that, ethical considerations and social acceptance also have to be part of the automation [2]; working with several data access projects that require the launch of sometimes complex sequences of activities where data access plays a key role (e.g., randomized controlled studies). Thus, the protocol suite must communicate different types of data between the Plan Entities (PEs) according to a data access plan as predefined by the various stakeholders. These communications are carried out between data sources with, for each, a database-anchored connector named Data Connection Entity (DCE). The latter includes the SDEPS communication interface and secure analysis environments. Connectors named Result Connection Entities (RCEs) receive data from other plan entities that show distinct processing features. A data access plan holds three different kinds of data; communications data, synthetic identifiers and content data. For security and efficiency, these types of data should not use the same paths or go through the same entities. In addition, synthetic identifiers are processed by anonymizers applying on information sent to any other plan entities. These SDEPS exchanges relies on a Data Exchange System (DES) which has several protocols from the network layer to the application layer as modeled by Open Systems Interconnection.

This paper focuses on the requirements of a routing protocol between the different entities of a DES. In a DES, there are already end-to-end encryption and certificates between data sources and end users (e.g., a clinical study manager). As a result, this paper deals exclusively with threats from an attacker as a network lurker while purposefully overlooking content encryption and packet authentication. If feasible, a traffic analysis on this type of network would make the identifying of sub-populations of a cohort possible; for instance, this could find application in clinical data from a sparsely populated village. As it comes to confidentiality is concerned, compliance with the European General Data Protection Regulation might be seriously impacted, though. This study focuses on defining the requirements of a new routing protocol which incorporates both the objective of minimizing delays and also the entropy of the messages exchanged.

This SDEPS new network layer routing protocol natively supports the here-below high-level requirements which are divided into two groups. The first group of requirements is to guarantee the anonymity of the sources and the entire paths taken by the packets. Group 1 is required to:

- Guarantee anonymity for sources in their addresses on

2372-9198/21/$31.00 ©2021 IEEE
DOI 10.1109/CBMS52027.2021.00038

the network for security and maintenance reasons.
- Avoid multi-source traffic analysis attacks. This implies that a source entity must agree to participate in a study and choose a relay proxy.

Group 2 is about the constraints on information exchange and the ensuing requirements are as follows:

- For sensitive data, ensure that communications messages, synthetic identifier and content data come in through two different paths.
- There are two types of synthetic identifiers, external and internal, and it is necessary that the proxies do communicate with one single anonymizer.

The paper begins with a review of the state of the art of existing routing protocols as used in healthcare, see Section II. Section III describes the protocol design while detailing the network topology. A threat model is defined, then protocol features are presented. Next, a digital model of the protocol is described in Section IV. This model is presented as a linear multi-objective problem modeled as a multicommodity flow problem, strongly NP-hard in general. Finally, an evaluation of the model is performed in Section V. This assessment will make it possible to determine the efficiency of the model using various metrics such as the ratio between the processing capacity and the number of proxies.

II. Related Work

This literature review focuses on solutions that provide routing anonymization for data exchange in healthcare. These solutions entail several techniques to hide metadata such as the use of a VPN, proxy servers or "Tor-like" mix-networks. Practically, the latter solutions, albeit thoroughly studied in theory, show little or no large-scale deployment across healthcare organizations and are mostly dedicated to particular proofs of concept.

Some health organizations worldwide [3] or others like the Baylor University as a member of the LEARN (Lonestar Education And Research Network) consortium [4] all have their own Virtual Private Networks (VPNs) so as to share sensitive datas. Mainly based upon the L2TP protocol, implementing VPNs is not sufficient on its own to meet LHS requirements as there still exist security issues [5].

Some of the here-presented architectures will use proxy servers or components acting as described in such environments as in [6]. This is the case with [7] where the authors use a proxy server to keep a source of data from the healthcare provider hidden to users. The point is, the proxy server can decrypt the data and therein lies a privacy issue. Moreover, this technique is based exclusively on the TLS protocol to ensure its own security, which does not allow security requirements to be met.

In [8] the authors have proposed a protocol for sending data with End-to-End Security while using a mixnetwork during exchanges on a public network. But it is possible to infer the utility of nodes and the location of patients and data sources in hospitals. This can be done because packet metadata are not handled by this protocol [9]. To protect this metadata, some presented Sphinx [10]. It is a specific packet format used in mix-networks with formally proven security requirements. It makes it possible to be very resistant to various attacks by traffic analysis. However, each intermediary can decrypt the payload, which affects the overall privacy of a system. In fact, the literature offers a myriad of mixnet-type protocols such as Tor [11]. These systems have the same inherent problem as in the Tor network. The routing of the packets is not hidden and it is possible not only to carry out passive or active attacks to identify exit nodes close to the final recipient and also to make an inference. Practically, a traffic correlation attack will end up successful, especially when correlating the ingress application to the network and the first node and then between the last node and the target server. Additionally, a traffic size-based correlation will also work as fine when anonymously performed at the ingress and egress traffic boundaries of a network.

Finally, some authors present a review of the state of the art concerning anonymous routing protocols [12]. One might turn to a new promising architecture like Software-Defined Network (SDN) [13] but this technology is deemed immature and still at risk of reconnaissance attacks via traffic analysis attacks [14]. None of these protocols meet the security requirements in terms of traffic analysis, preservation of the identity of the nodes or the confidentiality of the messages in transit.

III. Protocol Design Overview

This section introduces the new routing protocol called ANRP, its design issues and functionalities. The network topology is first presented with the threat model and ensuing risks, then the routing functionalities that meet the requirements are detailed.

A. Data Exchange System Network Topology

A DES must scrupulously follow a particular workflow to meet the requirements of LHSs and to exchange data between DCEs and RCEs. As a reminder, the different types of data exchanged are synthetic identifiers, communication messages and content data. In addition, each entity in the plan is placed behind a proxy to hide its identity. These types of data are processed, in a sequential manner, by the entities of the plan.

Figure 1 shows the exchanges of communications and synthetic identifiers between an initial source entity and an end target entity. The different phases taking place before the phase of content data transmission all are herein called "other phases." This includes the indexation phase of a population and the extraction phase following an evaluation entity during the execution of a study. Each exchange is annotated with a table number referring to the explanations found in the text body. Communication messages are sent exclusively in pairs between two entities. These are mainly controlled and status messages. A message is sent from a source **101** to a proxy, then the message passes through the proxies **10x** to an exit proxy near the destination entity. Last, the communication message

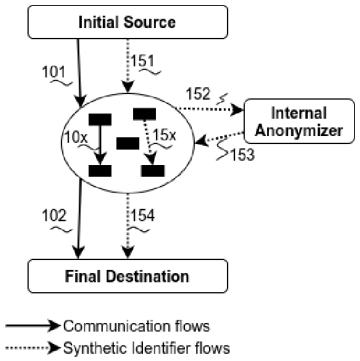

Fig. 1. Other Phases Exchanges

102 ends up delivered. During the other preparatory phases of data transmission, synthetic identifiers are sent to the DCE through a specific entity that is the internal anonymizer that modifies the identifiers. The internal anonymizer will interact with specific entities and the DCE, hence exchanges **151**, **152**, **153** and **154**. There also exist multiple exchanges **15x**, as they may go through multiple proxies.

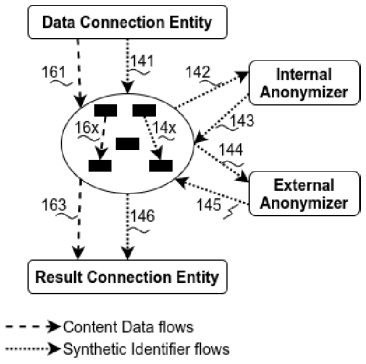

Fig. 2. Transmission Phase Exchanges

When in the transmission phase (see Figure 2), content data are sent from a DCE to an RCE. This can be achieved through exchanges **161 16x** and **163**. Synthetic identifiers are needed for the RCE to be able to interpret the content data. These synthetic identifiers are transited from the DCE to the RCE RCE with Exchanges **141** and **146**. Then they are in turn processed by an internal anonymizer and by an external anonymizer and returned to the proxy pool using messages **142**, **143**, **144** and **145**. The external anonymizer comes to be placed between the RCE and the planned entities In the proxy pool, synthetic identifier exchanges can be multiple and are listed **14x**.

B. Threat Model

Either passive or active attacks may all be carried out in order to observe the traffic on the links. There are no packets modified on the network. This is what makes these attacks extremely difficult to detect. By analyzing the arrival times and exits of packets of a node, as well as the exchange of exchanges, it is possible to deduce information on the architecture or the role of nodes. This paper deals with the worst-case scenario, as the attacker can not only observe the whole of network links buy also perform several types of passive attacks like traffic analysis or active attacks by inserting or mishandling nodes and messages.

C. Routing Protocol Features

The ANRP has several essential features for a routing protocol like selecting nodes or scheduling requests sent to a proxy. The features are as follows:

- **Network structure**: The protocol is made with a partially connected topology because the network nodes connect only to a small subset of the entire network. The connections are asynchronous and there are one-way content data flows. On the other hand, these are two-way exchanges of synthetic identifiers and communications. Proxies have similar roles and responsibilities. Endpoints will never operate as relay nodes. The ANRP nodes have a partial view of the system.
- **Routing Decision**: The protocol uses hop-by-hop routing as the routing decision is made at each node. Moreover, the initiator of the communication only selects the first relay node with which to start, then it will hop to the next node, and so on and so forth until the destination is reached.
- **Demands scheduling**: Incoming demands are assumed to be scheduled on an equitable basis as demands of all types are processed equally.
- **Node selection**: Node selection for a path is non-deterministic and the selection set is made according to the security requirements and restrictions in line with the Data Access Plan. The probability of selecting a proxy node is weighted stably as based on general static parameters.
- **Performance and deployment**: The ANRP is a medium latency protocol and introduces a random delay. It uses the delay-tolerant network methods to increase throughput and reduce latency despite the various constraints.

IV. Proposed Anonymous Data Exchange Routing

This section describes the anonymous routing problem for data exchange. First, the routing problem is characterized with these objectives and constraints. Then, definitions of the problem and information entropy is given. Secondly, the problem is formulated as a linear multi-objective problem.

A. Routing Anonymization Problem Definition

The point with routing anonymization in a DES is to minimize system latency and entropy. If a system has high entropy, there will be more information and therefore more likely to correlate. The model must respect capacity and flow constraints (**C1**, **C2**, **C3**) but also data exchange on nodes (**C4**). Constraints **C1** and **C2** require that the edge and node buffer capacities not be exceeded. **C3** is the flow conservation constraint, that means for a node that its total incoming flows

equate its total outgoing flows. **C4** constraint requires that content data and synthetic identifying data cannot go through one same node at once.

In *Flows in Networks* [15], the authors provide a method for integrating time windows into a multi-commodity flow problem. Thus, they construct a flow graph with a whole set of paths within a time interval. The solution to the anonymous routing problem with which this paper deals is based on this type of graph. The anonymous routing problem is defined as follows 1.The resolution of the anonymous routing problem which is treated in this paper is based on this type of graph.

Definition 1 (Routing Anonymization Problem): Given a flow network graph $G = (V, E)$ with different paths over time, where (v, v') is a path. Each edge has capacity $c(v, v')$ and each vertice has buffer $b(v)$. There are $|K|$ messages noted k. Each message is defined by $k_i = (s_i, d_i, w_i)$ where s_i is the source node, d_i the destination node and w_i the weight of the message. The variable $f_i(v, v')$ defines the assignment of message i along edge (v, v'), where $f_i(v, v') \in 0, 1$. Find the paths of all messages through a dynamic topology that satisfies the tradeoff between delay and traffic analysis defense measured by the messages exchanged entropy according to the specified constraints.

The concept of entropy, so essential for the security of our model, is defined as 2. This gives an idea of the total amount of information that has been leaked to any observer.

Definition 2 (Data Exchange Entropy): Given a data exchange system, the entropy of the system represents the amount of information provided by the system. The entropy is calculated using the Shannon function as based on the probability that similar messages follow an identical sub-path in a flux graph.

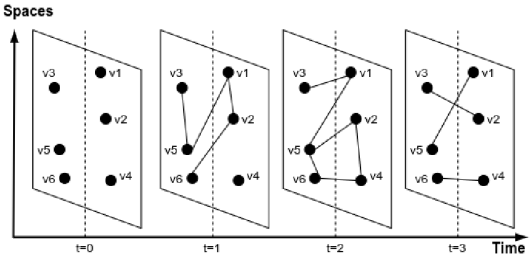

Fig. 3. Dynamic Topology Example

Figure 3 presents an example of a dynamic topology between several nodes within a time window. The topology must evolve following a distribution function chosen by these performances and its capacity to ensure a non-deterministic change.

Figure 4 shows all the path possibilities and an example with communications data (triangles), synthetic identifier data (circles) and content data (squares). The left-side graph indicates all of the possible paths according to the dynamic topology shown in Figure 3. In this example, all the delays for each edge are 1. There is an edge between two nodes of the same proxy, which means the proxy buffer. The graph on the right shows an example of a data transmission phase. Entities send commands along with communications messages, synthetic identifiers and content data. Messages, content data and synthetic identifiers must not go through the same proxies. A proxy can batch-process messages. If a proxy I reached by messages of different types not in line with the requirements and constraints, then the proxy will process the first incoming message while rejecting the next one.

B. Routing Anonymization Formulation

The formulation of the denominalized routing problem is based on a flow network graph with different paths over time as denoted $G = (V, E)$ where V is a set of vertices, the proxies, and E is a set of edges representing the paths. The graph is represented by an incidence matrix I_{ve}. Each node has a maximum buffer size as b_v. Each edge has a maximum capacity c_e and a delay d_e.

K represents the whole of the messages. The source and target destination for the messages are each found in arrays S^k and D^k respectively. Each message has a weight denoted w_k. The observer will analyze and find out whether any messages are seen as similar. Then the similar messages will be gathered into one message set, denoted K^{sim}. Likewise, the allegedly similar messages which are allocated to one same edge are grouped together in a set denoted P_e^k. Please note this paper does not deal with similarity of messages. However, the similarity function could make use of the cosine similarity which works well for this task. Indeed, the exchanged messages have several vectors representing the different parameters for one packet. The times for incoming and outgoing messages can also be modeled with this method and then integrated into the message vectors.

$$
\begin{aligned}
\underset{y,z}{\text{minimize}} \quad & \left(\textstyle\sum_{k\in K} d_e x_e^k, \quad \sum_{k\in K^{sim}} \mathbb{P}(k) log \frac{1}{\mathbb{P}(k)}\right) \\
\text{subject to} \quad & \sum_{k\in K} w^k x_e^k \leq c_e, \ \forall e \in E \\
& \sum_{k\in K} I_{ve} x_e^k w^k \leq b_v \\
& \forall v \in V, e \in E, I_{ve} = 1 \\
& \sum_{k\in K} x_e^k I_{ve} w^k = a_v^k, \ \forall e \in E \\
& \sum_{k\in K} C_k^C + C_k^D + C_k^I = |K| \\
& \sum_{k\in K} I_{ve} x_e^k (C_k^D + C_k^I) = \sum_{k\in K} I_{ve} x_e^k C_k^D \\
& \forall v \in V, e \in E, I_{ve} = 1 \\
& \sum_{k\in K} I_{ve} x_e^k (C_k^D + C_k^I) = \sum_{k\in K} I_{ve} x_e^k C_k^I \\
& \forall v \in V, e \in E, I_{ve} = 1
\end{aligned}
\tag{1}
$$

A boolean assignment variable x_e^k indicates that the message k will go through Edge e. In order to set the constraints for proxy-processed messages, there are a couple of boolean variables in use C_k^D and C_k^I respectively showing that a message contains content data or synthetic identifier data. Table I summarizes all of the notes used in the model.

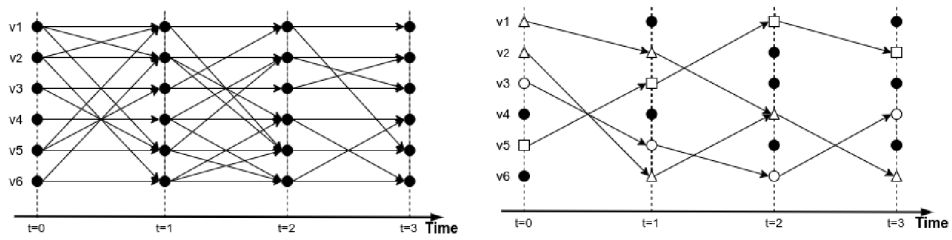

Fig. 4. Graphs of paths over time. Left diagram represents the set of paths according to the dynamic topology. Right diagram is an example of transmission phase with communications (triangles), synthetic identifier (circles) and content data (square) messages.

TABLE I
SUMMARY OF MODEL VARIABLES

	Topologies and requests
$G = (V, E)$	Paths flow network graph
I_{ve}	Incidence matrix
K	Set of all messages
K^{sim}	Set of messages calculated as being similar
S^k	Source node of message k
D^k	Destination node of message k
a_v^k	Flow matrix for each message k and each node v
P_e^k	Set of same edges used by similar message k with $k \in K^{sim}$
	Capacity matrices
c_e	Maximum capacity of edge e
d_e	Delay of edge e
w_k	Size of message k
b_v	Maximum buffer size of vertice v
	Assignment variables
x_e^k	Boolean variable indicating whether the message k through an arc e
C_k^C	Message k contains communication data
C_k^D	Message k contains content data
C_k^I	Message k contains synthetic identifier

The complete formulation of the routing problem is defined by the equation: 1. The multi-objective function consists in minimizing the delays and the entropy of the messages exchanged. The four constraints detailed in the preceding subsection are explained by six constraint equations.

The first constraint represents the maximum capacity on each edge of the graph. The second constraint is the maximum capacity limit on a node buffer. The third constraint is the constraint of conserving flows which will allow the message to reach its destination. Specifically, the flow conservation constraint is equal to the value given by a_v^k. The last three constraints describe the fact that a message necessarily is of only one type. Then, a node can only accept one type of content data message or synthetic identifier over time. Equation 2 is the incidence matrix of the path graph over time. Equation 3 is the variable for assigning messages on a link.

$$I_{ve} = \begin{cases} +1 & if\ e = (v, v')\ for\ some\ v' \in V \\ -1 & if\ e = (v', v)\ for\ some\ v' \in V \\ 0 & otherwise \end{cases} \quad (2)$$

$$x_e^k = \begin{cases} 1 & if\ k\ is\ assigned\ to\ edge\ e \\ 0 & otherwise \end{cases} \quad (3)$$

Equation 4 is the matrix which represents the incoming and outgoing flows on a node. It is used for the flow conservation constraint.

$$a_v^k = \begin{cases} w_k & if\ v = S^k \\ -w_k & if\ v = D^k \\ 0 & otherwise \end{cases} \quad (4)$$

Equation 5 represents the probability that similar messages will follow an identical path over time. Indeed, the scenario in which several non-identical messages pass through different paths would give out clues on what the messages contain and what the nodes are used for $|P_e^k|$ is the total number of similar messages passing through a subset of identical paths. $|E'|$ is a subset of identical paths in E. An identical path is a path that bridges two similar nodes in a given time interval.

$$\mathbb{P}(k) = \frac{|P_e^k|}{|E'|^2} \quad \forall k \in K^{sim}, e \in E' \subseteq E \quad (5)$$

For instance, two paths are considered similar as long as the edge connects the same entities at different time intervals.

V. NUMERICAL EXPERIMENTS

The digital experiment, as based on the model in this paper, is designed to assess the impact of the flow separation constraint between synthetic and content identifier data upon both delay and entropy. The implementation of this assessment centers on five scenarios. The substrate used is found in Figure 4. The scenarios are based on the amount of messages increasing from 3 messages to 40 messages. These messages need to be routed from source to target destination and they have a data type of payload. The messages all have the same size while the edge capacities between nodes are similar. In order to set a relevant comparison, the problem will be solved by applying and removing the flow separation constraints. This constrained multi-objective optimization problem is solved numerically using an evolutionary algorithm. More specifically, the NSGA-ii algorithm is used where a set of a non-dominated population is selected to obtain the Pareto-optimal set. The chosen solution is that which has the shortest distance from the initial points of the objective functions.

Figure 5 shows the analysis of delay and entropy according to the different scenarios. In the scenario with no flow constraint, the delay will depend on both the number of messages sent and distance from the target destination. The delay function is linear. The delay-impacting parameters are the number of message types and the distance gone through to be delivered. In fact, when the flow separation constraints are

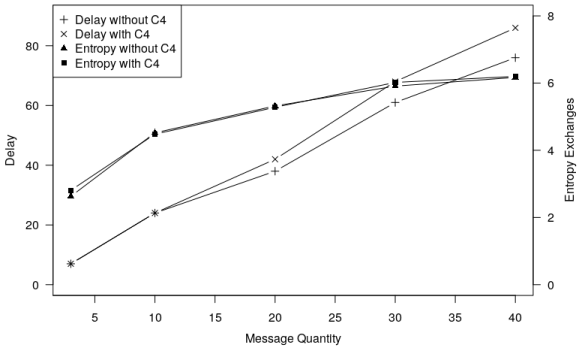

Fig. 5. Analysis of delay and entropy according to 5 scenarios

enabled, the messages that cannot go through a node take another – often longer – path. This is the reason why the average delay takes longer under the active constraint. The objective of minimizing the exchange entropy is a logarithmic function. This objective will force the grouping of messages containing the same type of data. It appears that entropy is higher in most scenarios when the constraint is active as the messages take a link that once exists between the nodes. In the twenty-message scenario, the entropy is lower with the constraint because several messages are gathered together on a specific between-node link that has been set before, meaning more than once. As a conclusion, this assessment reveals that the trade-off between security, with the addition of a constraint, and latency time is acceptable. As a matter of fact, this experiment tries entropy as a way of quantifying the security performance of the systems. This entropy underlies that identical messages might follow along the same paths. This allows to gauge how much information might be disclosed to any attacker. It appears to be a good metric to our scenarios. It may be worthy to put this metric to the test when under different contexts or in the light of additional studies.

VI. Conclusion

Data exchange between LHSs implies a deal of security requirements, including minimizing the risks of patient re-identification and entity targeting while information is being routed. The existing message routing protocols as seen earlier do not meet these requirements for data exchange. A new protocol named ANRP natively supports mitigation means against attacks by traffic analysis. In addition, an entropy metric of the messages exchanged on a specific path allows for the integration of this security as an objective as well as the minimization of the latency. The results of the experiment show that it is possible to add mitigation means while facing a reasonable increase in the delay. In the next work more specific analyses shall be carried out to define how efficient the dynamic topology between the proxies is and how it affects the message delay. Moreover, an assessment of the Pareto front for this problem would sound interesting to study. Data protection is a fundamental point in Learning Health Systems, which is why great care must be taken to secure each logical layer with the help of mitigation means.

Acknowledgments

We are grateful to all the contributors for revising the model and the analysis. Last but not least, we would like to acknowledge the support from the Unité de Soutien SRAP du Québec and Health Data Research Network Canada with the Canadian Data Platform.

References

[1] T. Ecarot, B. Fraikin, F. Ouellet, L. Lavoie, M. McGilchrist, and J.-F. Ethier, "Sensitive data exchange protocol suite for healthcare," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, 2020.

[2] A. Cumyn, A. Barton, R. Dault, A.-M. Cloutier, R. Jalbert, and J.-F. Ethier, "Informed consent within a learning health system: A scoping review," *Learning Health Systems*, Dec 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lrh2.10206

[3] S. Guinez-Molinos, J. M. Andrade, A. Medina Negrete, S. Espinoza Vidal, and E. Rios, "Interoperable platform to report polymerase chain reaction sars-cov-2 tests from laboratories to the chilean government: Development and implementation study," *JMIR Med Inform*, vol. 9, no. 1, p. e25149, Jan 2021. [Online]. Available: http://medinform.jmir.org/2021/1/e25149/

[4] J. Zurawski and J. Schopf, "Baylor university campus-wide deep dive," 2021.

[5] A. K. Singh, S. G. Samaddar, and A. K. Misra, "Enhancing vpn security through security policy management," in *2012 1st International Conference on Recent Advances in Information Technology (RAIT)*, 2012, pp. 137–142.

[6] S. Desai, T. Vyas, and V. Jambekar, *Security and Privacy Issues in Fog Computing for Healthcare 4.0*. Cham: Springer International Publishing, 2021, pp. 291–314. [Online]. Available: https://doi.org/10.1007/978-3-030-46197-3_12

[7] S. Sabitha and M. S. Rajasree, "Anonymous-cpabe: Privacy preserved content disclosure for data sharing in cloud," in *Architecture of Computing Systems – ARCS 2015*, L. M. P. Pinho, W. Karl, A. Cohen, and U. Brinkschulte, Eds. Cham: Springer International Publishing, 2015, pp. 146–157.

[8] E. Marin, M. A. Mustafa, D. Singelée, and B. Preneel, "A privacy-preserving remote healthcare system offering end-to-end security," in *Ad-hoc, Mobile, and Wireless Networks*, N. Mitton, V. Loscri, and A. Mouradian, Eds. Cham: Springer International Publishing, 2016, pp. 237–250.

[9] A. Al-Hababi and S. C. Tokgoz, "Man-in-the-middle attacks to detect and identify services in encrypted network flows using machine learning," in *2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet)*, 2020, pp. 1–5.

[10] G. Danezis and I. Goldberg, "Sphinx: A compact and provably secure mix format," in *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, ser. SP '09. USA: IEEE Computer Society, 2009, p. 269–282. [Online]. Available: https://doi.org/10.1109/SP.2009.15

[11] S. Baek, S.-H. Seo, and S. Kim, "Preserving patient's anonymity for mobile healthcare system in iot environment," *International Journal of Distributed Sensor Networks*, vol. 12, no. 7, p. 2171642, 2016.

[12] F. Shirazi, M. Simeonovski, M. R. Asghar, M. Backes, and C. Diaz, "A survey on routing in anonymous communication protocols," *ACM Comput. Surv.*, vol. 51, no. 3, Jun. 2018. [Online]. Available: https://doi.org/10.1145/3182658

[13] D. P. Isravel, S. Silas, and E. B. Rajsingh, "Sdn-based traffic management for personalized ambient assisted living healthcare system," in *Intelligence in Big Data Technologies—Beyond the Hype*, J. D. Peter, S. L. Fernandes, and A. H. Alavi, Eds. Singapore: Springer Singapore, 2021, pp. 379–388.

[14] A. Alshamrani, "Reconnaissance attack in sdn based environments," in *2020 27th International Conference on Telecommunications (ICT)*, 2020, pp. 1–5.

[15] D. R. Ford and D. R. Fulkerson, *Flows in Networks*. USA: Princeton University Press, 2010.