# Information Assurance in Critical Infrastructures via Wireless Sensor Networks

Michele Albano Stefano Chessa[1]
Dipartimento di Informatica
Università di Pisa
Largo B. Pontecorvo n. 3
56127-Pisa, Italy
{michele,ste}@di.unipi.it

Roberto Di Pietro[1,2]
UNESCO Chair in Data Privacy
Universitat Rovira i Virgili
Av. Països Catalans n. 26
E-43007 Tarragona, Spain
roberto.dipietro@urv.cat

## Abstract

*Information assurance in critical infrastructure is an issue that has been addressed generally focusing on real-time or quasi real-time monitoring of the critical infrastructure; so that action could be undertaken when anomalies arise, to avoid more severe consequences to the infrastructure. In this paper, we relax the hypothesis of intervening when anomalies are detected: we focus on sensed data survivability. Specifically, we study this problem in a specific critical infrastructure: pipelines. The problem we introduce is how to place sensors in such a way that the sensed data related to the monitoring of the pipeline will survive even in presence of a partial destruction of the infrastructure.*
*The contributions of this paper are twofold. First, we introduce the problem of sensed data survivability in critical infrastructure. In this framework, the goal is to have the sensed data to survive to the infrastructure failure, so that the phenomena that lead to the failure could be better understood and possibly tackled in similar deployment. Second, we provide a model that allows to produce an optimal network topology with respect to the level of information assurance desired, while satisfying deployment constraints, such as available bandwidth and available energy of the sensors. We believe that the work addressed in this paper could foster further research in the field of information assurance in critical infrastructure.*

**Keywords:** Wireless sensor network; critical infrastructure protection; data survivability; modeling; optimization problem.

## 1 Introduction

According to the European Commission, Critical Infrastructures consist of "those physical and information technology facilities, networks, services and assets which, if disrupted or destroyed, would have a serious impact on the health, safety, security or economic well-being of citizens or the effective functioning of governments in the Member States. Critical Infrastructures extend across many sectors of the economy, including banking and finance, transport and distribution, energy, utilities, health, food supply and communications, as well as key government services"[1].

However, the problem of critical infrastructure protection is cross-border sensitive. As an example, United State based water companies are under increasing pressure to improve the management of their ageing assets and to optimize operational and capital expenditure. A recent study by the US Environmental Protection Agency (EPA) estimates that water utilities need $277 billion over the next 20 years (2003-2023) to install, upgrade, and replace infrastructure [2]. Transmission and distribution projects represent the largest component ($184 billion) of this estimate. The threat of contaminant intrusion due to leaking pipes [3] or malicious human action will further increase the projected expenditure. Repairing and securing this infrastructure requires large investments of money and time, and therefore, it is essential to direct efforts to upgrading critical areas. Unfortunately, identifying the highest priority areas is a non-trivial task, because of the scale and age of the pipeline infrastructure and lack of operational data. Failures of large diameter (12" and greater) bulk-water transmission pipelines are of greatest concern as these are supply critical systems. While such failures are infrequent, when they do occur, they have dire consequences, including loss of life, severe interruptions in service, degraded fire fighting ability, damage to adjacent infrastructure and buildings, and multi-million dollar repair bills.

The challenges on protecting the Critical Infrastructures

[1] Also with CNR-ISTI, Via G. Moruzzi n. 1, 56124 - Pisa, Italy.
E-mail: {stefano.chessa, roberto.dipietro}@cnr.it
[2] Also with Dipartimento di Matematica, Università di Roma Tre, L.go S. Leonardo Murialdo n. 1, 00146 - Roma, Italy.
E-mail: dipietro@mat.uniroma3.it

978-0-7695-3324-7/08 $25.00 © 2008 IEEE
DOI 10.1109/IAS.2008.54



IEEE computer society

are numerous and complex, since they are composed by highly interconnected national (and international) software based control systems where a single isolated disturbances can cascade through the system with unexpected consequences. It is then indispensable to have a resilient and robust information infrastructure that could deal with any situation and assure the security of the information, which is of critical importance from a political, economical, financial or social standpoint. It is also important to provide a monitoring system that can issue alerts and warnings even if a problematic situation has yet to occur. Besides, it becomes imperative to create models and simulations that could show how the system should behave in presence of problems, avoiding problems in upgraded systems that could hinder the continuity of the services.

A recent technology that can be applied for protecting those Critical Information Infrastructures are Wireless Sensor Networks (WSN) [4].

The main purpose of a Wireless Sensor Network as a whole is to serve as an interface to the real world, providing physical information such as temperature, light, radiation, and others, to a computer system. These types of networks have a simple structure: there are thousands of elements, called sensor nodes or *sensors* that are able to sense the physical features of their surroundings. After such information is processed by these nodes (either on board or in a cooperative fashion), it is sent through a wireless channel to a central system, called "Base Station". One of the key advantages of Wireless Sensor Networks consists on the capabilities of their sensor nodes. All sensor nodes are autonomous from the point of view of energy supply (for example because they employ energy scaveging techniques [5] or because they are powered by batteries), and they can subsist long periods of time (even years) if configured correctly. Moreover, although constrained, they have sufficient computational and storage capabilities (a typical configuration is based on 8-bit processors with up to 1 MB of flash memory) (see for example [6]). As a result, they are totally independent and are able to act autonomously if the context requires them to do so, collaborating with other nodes in pursuing a common goal using their wireless channel. Even more, thanks to their potential to self-configure themselves, it is an easy task to set up a sensor network in a physical context where it is needed without needing any previously existent infrastructure. The services offered by a WSN can be classified into three major categories: monitoring, alerting, and information "On-Demand". Sensor nodes can continuously monitor certain features of their surroundings (e.g. measuring the radiation level). Sensors can also check whether certain physical conditions (e.g. a radiation leak) are taking place, alerting the users of the system if an alarm is triggered. Finally, the network can be queried about the actual levels of a certain feature, providing information "On-Demand".

In this paper we discuss the use of a wireless sensor network for the monitoring of a Critical Infrastructure such as a pipeline, to the purpose of detection of critical situations and of post-failure forensic analysis. In particular we focus on the problem of sensors deployment. More specifically, we identify the constraints given by the shape and topology of the pipeline that affect the size and monitoring capability of the WSN, and we model the problem in terms of linear programming.

The rest of the paper is organized as follows: Section 2 presents the related work, Section 3 introduces the system model, and Section 4 formulates the problem in terms of linear programming and discusses how known algorithms can be used to solve it in polynomial time. The conclusions are drawn in Section 5.

## 2 Related Work

An ever increasing number of works are dealing with *Critical Infrastructure Protection* (CIP). A structured definition of the key concepts has been given in [7], that introduces the problem in terms of "strategies, policies, and preparadness needed to protect, prevent and respond to attacks on these sectors and key assets".

A specific study that addresses security risks in the control systems used to monitor critical infrastructure can be found in [8].

The authors of [9] examined the role of remote sensing and geospatial information technologies in CIPs, specifically in the identification and preservation of Critical Transportation Infrastructure (CTI).

An analysis of the technology (and lack of) needed to protect CIPs is in [10], that points out that there is the need of creating a stronger scientific foundation, the required technology for national-scale cyber defence, and an engineering discipline to provide the required means to face the above challenges.

The survey in [11] covers the general ideas about using a Wireless Sensor Network to ensure the protection of CIPs. The authors argues that, while there are some research issues that a WSN must face in order to protect a CIP, it does provide interesting and essential protection properties due to their intelligent distributed control capabilities alongside with their capability to work under severe conditions.

The particular problem of pipeline protection is firstly addressed in [12], where the authors describe a prototype system that performs automated detection of leaks and bursts of water in water transmission pipelines. The paper focuses on a software infrastructure for distributed high-rate signal processing. Our work, on the contrary, stresses on the *transmission* of the data, the maximization of the throughput for sensed data, and its possibility to survive a disaster

occurring to the infrastructure.

## 3 System model

We model a pipeline as a weighted graph $O$ (called the *Oil Graph*), where the links are the pipes and the nodes are the joints connecting the pipes, namely:

$$\begin{cases} O = & (T, P) \\ T = & \{t_1, ..., t_m\} \\ P = & \{(t_i, t_j) : t_i, t_j \in T\} \end{cases}$$

A pipe $p_k = (t_i, t_j)$ is said to be *incident* to $t_i$ and $t_j$. Hereafter we denote with $\pi$ the cardinality of $P$. In general, the pipes can have a different importance in the *Oil Graph*, for example because a pipe has a larger section and it can transport a larger flux of oil. For this reason each edge $p_k \in P$ is associated with an integer weight $\rho_k \in [0, \infty)$, denoting its importance.

Each edge $p_k$ is monitored by a number of sensors that is proportional to its importance $\rho_k$ and its length $l_k$. Hereafter, for the sake of simplicity, we assume that the number of sensor in a pipe $p_k$ is $w_k = l_k \rho_k$. The sensors hosted by the pipes give rise to another graph $G$, called the *Network Graph* that is a refinement of the *Oil Graph* and that represents the set of sensors and their interconnections:

$$\begin{cases} G = & (V, E) \\ V = & \{v_1, ..., v_n\} \\ E = & \{(v_i, v_j) : v_i, v_j \in V\} \end{cases}$$

The *Network Graph* can be partitioned into a set of subgraphs, one for each edge of the *Oil Graph*. In particular, for each edge $p_k = (t_i, t_j) \in P$ we define the subgraph $G_k \in G$, where $G_k = (V_k, E_k)$, and $V_k = \{v_{k,1}, \cdots, v_{k,w_k}\}$ is the set of sensors hosted by $p_k$ (clearly $|V_k| = w_k$). Let $\beta$ be the maximum bitrate that can be sustained in all communications in which the sensors in $V_k$ are involved.

To ensure that the *Network Graph* is connected the sensors belonging to two different pipes incident into the same joint must be connected, i.e. given $p_k$ and $p_h$ incident in the same joint $t_i$ there must be at least two sensors $u \in V_k$ and $v \in V_h$ such that $(u, v) \in E$. The sensors $u$ and $v$ are called *bridges* between $p_k$ and $p_h$. Given $p_k = (t_i, t_j)$, we define $S(k) = \{p_a \neq p_k : p_a \text{ is incident in } t_i \text{ or in } t_j\}$.

Figure 1 shows an *Oil Graph* with 8 joints and 10 pipes, and it details pipe $p_5$ in terms of the underlying subgraph $V_5$ of the *Network Graph*. In particular, it is shown that $p_5$ is monitored by 7 sensors, and that sensor $v_1$ is a bridge between $p_5$ and $p_4$ while sensor $v_7$ is a bridge between $p_5$ and $p_6$.

The sensors monitor the pipes for maintenance purpose, but also to support forensics analysis in case of infrastructure fault. The *scope* of a fault is pair $(F_P, F_V)$ where $F_P = \{p \in P$ that are affected by the fault $\}$ and $F_V = \{v \in V_k : p_k \in F_P\}$. Thus we assume that a fault in the oil pipes with scope $(F_P, F_V)$ compromises all the pipes in $F_P$ and all the sensors in $F_V$. In particular, all the data stored by the sensors in $F_V$ are lost and the sensors in $F_V$ do not communicate anymore.

Given the task assigned to the sensors in the *Network Graph*, it is mandatory that the data produced by the sensors be stored by the sensors in such a way that it will be still available after a fault.

Each sensor monitors the pipe with a constant rate $\lambda$, hence each sensor produces a new data with a rate $\lambda$ and the pipe $p_k$ produces a new data with a rate $w_k\lambda$. Since the failure of a pipe may compromise all the sensors in the pipe, the sensors must be equipped with a strategy to disseminate the sensed data throughout the network.
We assume that the data is disseminated in the network by using some well known dissemination mechanism. For example the DCS-GHT [13] or other protocols derived from DCS-GHT such as [14, 15] provide efficient mechanisms for the dissemination and retrieval of information in a WSN. These protocols basically use pure data replication to enforce data redundancy, but they can also be combined with erasure codes [16, 17]. In the latter case each data is split into a number of fragments with the property that the data can be reconstructed from an arbitrary subset of the fragments that exceeds a given cardinality threshold.

Note that adopting replicas or erasure codes, results could differ considerably [17]; however, to ease exposition and to fit page limit, in the following we will consider just replicas, leaving an erasure code based approach for further work. In particular, we assume that each data is replicated $r$ times and each replica, as soon as it is produced, is sent to a given destination. Also, all the replicas are stored by a set of $r$ different sensors uniformly distributed in the network. This also means that each sensor produces packets with a rate $r\lambda$. Thus in a time slot of size $1/\lambda$ each sensor produces (and thus sends) $r$ packets. Assuming that each replica is encoded into $\gamma$ bits, the data rate produced by each sensor is $r\gamma\lambda$, and the total data rate produced by the sensors located on a pipe $p_k$ is $r\gamma\lambda w_k$. From the point of view of the storage, all the sensors in a pipe $p_k$ (i.e. all sensors in $V_k$) store a set of replicas $D_k = \cup_{j\in[1,n]} D_{j,k}$, where $D_{j,k}$ represents the set of replicas produced by the sensors in $V_j$ (i.e. the sensors in the pipe $p_j$) and stored by the sensors in $V_k$. We denote with $d_{j,k}$ the cardinality of $D_{j,k}$.

Clearly, the data produced by the sensors in $V_j$ and stored by the sensors in $V_k$ should be routed through the network, hence it implies an overhead for any intermediate sensor that should forward the data. To model this overhead we introduce the variable $x_{j,i\to l}$ that denotes the number of packets produced by $V_j$, that are routed to their destination through two subsequent pipes $p_i$ and $p_j$ (that are incident in

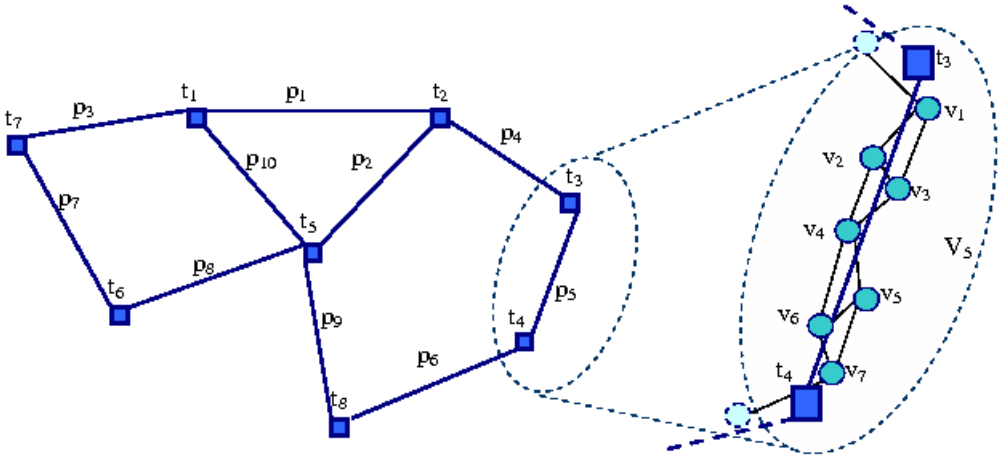



**Figure 1. An *Oil Graph* with $8$ joints and $10$ pipes, and details on pipe $p_5$ in terms of the underlying subgraph $V_5$ of the *Network Graph***

**Table 1.** Parameters for the modeling

| | |
|---|---|
| $O = (T, P)$ | *Oil Graph* |
| $p_i \in P$ | pipe with id $i$ |
| $\pi$ | cardinality of $P$ |
| $\rho_k$ | importance value of pipe $p_k$ |
| $l_k$ | length of pipe $p_k$ |
| $G = (V, E)$ | *Network Graph* |
| $v_i$ | sensor with id $i$ |
| $V_k$ | sensors on pipe $p_k$ |
| $w_k$ | cardinality of $V_k$ |
| $\beta$ | maximum bitrate of $V_k$ |
| $\eta_k(l)$ | *available energy for $V_k$ in round $l$* |
| $S(k)$ | pipes incident on pipe $p_k$ |
| $\lambda$ | sampling rate of the sensors |
| $r$ | number of packet for each produced data |
| $\gamma$ | size of a packet in bits |
| $D_{j,k}$ | packets produced by $V_j$ and stored in $V_k$ |
| $d_{j,k}$ | cardinality of $D_{j,k}$ |
| $x_{i,j\to k}$ | number of packets produced by $V_i$, sent from $V_j$ to $V_k$ |

a common joint).

The notation used in the paper is summarized in table 1.

## 4 Analysis

Given the *Oil Graph* and the *Network Graph*, there is a limit to the maximum sampling rate $\lambda$ that can be sustained by the network, this because increasing the sampling rate increases the number of packets, and, in turn, the bandwidth and the energy that are required. In this section we provide a relationship between network parameters, such as number of sensors per pipe, network topology, and the parameter $\lambda$. To achieve this result we model the system in terms of a Maximum Concurrent Multicommodity Flow (MCMF) problem [18].

The first constraint that the system must respect is the source/sink balance. All the packets produced by sensors in $V_j$ must be stored somewhere into the network. Thus, for $\forall V_j$, with $j \in [1, \pi]$, the following equality must hold:

$$\sum_{k\in[1,\pi]} d_{j,k} = \lambda r w_j$$

The flow conservation constraint dictates that all the flows that enter a pipe or that are produced in the pipe must either be stored into the pipe or routed away. That is, with respect to the packets produced by the sensors in $V_j$ and an arbitrary pipe $p_k$, we have:

$$\sum_{p_a\in S(k)} x_{j,k\to a} - \sum_{p_b\in S(k)} x_{j,b\to k} = \delta(j,k)\lambda r w_j - d_{j,k}$$

where $\delta(j,k)$ is the Kronecker delta, defined as:

$$\delta(j,k) = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$$

The third constraint that must be coped with is the maximum bandwidth constraint. The total bandwidth consumed

by the data routed through $V_j$ must be lower than $\beta$. That is, $\forall V_j$, we have:

$$\sum_{i\in[1,\pi]} \sum_{p_k\in S(j)} (x_{i,j\to k} + x_{i,k\to j})\gamma \leq \beta$$

The last constraint that the system must respect is related to energy availability. For each data sampled by the sensors (recall that the sensors sample the status of the pipes with a rate $\lambda$), we define a round, and for a given $V_j$ the available energy for a given round $l$ is $\eta_j(l)$. Thus the total energy consumed by each sensor in $V_j$ must be lower than $\eta_j(l)$. That is, $\forall V_j$, for each round $l$, we have:

$$\sum_{i\in[1,\pi]} \sum_{p_k\in S(j)} (x_{i,j\to k} + x_{i,k\to j})\gamma \leq \eta_j(l)$$

Summarizing, the problem can be modeled as the following linear programming problem:

$$\begin{cases} \max & \lambda \\ \forall V_j & \sum_{k\in[1,\pi]} d_{j,k} = \lambda r w_j \\ \forall V_j \forall p_k & \sum_{p_a\in S(k)} (x_{j,k\to a} - x_{j,a\to k}) = \\ & \quad \delta(j,k)\lambda r w_j - d_{j,k} \\ \forall V_j & \sum_{i\in[1,\pi]} \sum_{p_k\in S(j)} (x_{i,j\to k} + x_{i,k\to j})\gamma \leq \beta \\ \forall V_j \forall l & \sum_{i\in[1,\pi]} \sum_{p_k\in S(j)} (x_{i,j\to k} + x_{i,k\to j})\gamma \leq \eta_j(l) \end{cases}$$

This problem is known in the literature as the Maximum Concurrent Multicommodity Flow (MCMF)[18] and it can be solved by exploiting the algorithms also presented in [19] and [20]. These papers provide $O(\log^2 \pi)$ running time algorithms to find out a minimal capacity approximately balanced cut; this result is further leveraged to implement a "Divide and Conquer" strategy. In this strategy, two smaller subproblems are (recursively) solved on the two sides of the (approximately) optimal cut, then the two solutions are recombined: the almost balanced sides ensure at most a logarithmic number of recursion steps, while the small capacity limits the amount of work required to paste the two partial solutions.

## 5 Conclusions

The contributions of this paper are twofold. First, we have introduced the problem of sensed data survivability in critical infrastructure. In particular, we have relaxed the assumption that the critical infrastructure cannot fail, while focusing on having the sensed data to survive, so that the study of the collected data could prevent similar accidents from happening. Second, we have studied a specific problem in the introduced framework: data survivability in pipelines. In particular, we have modelled the distribution of sensors in a pipeline so that an optimal placement of sensor could be achieved —where the optimum to achieve is to maximize the amount of collected information—, while satisfying deployment constraints (such as available bandwidth).

We believe that this work paves the way for further research. For instance, to enrich the model with further constraints, to adapt the sensed data survival problem in other contexts other than pipelines (such as power grid distribution), and to develop a tool to provide an implementation of the described model.

## Acknowledgements

We would like to thank Professor Antonio Frangioni for his help on the analysis of the system model.

This work was partly supported by the Spanish Ministry of Science and Education through projects TSI2007-65406-C03-01 E-AEGIS and CONSOLIDER CSD2007- 00004 ARES, and by the Government of Catalonia under grant 2005 SGR 00446.

## References

[1] Commission of the European Communities: Communication from the Commission to the Council and the European parliament: critical infrastructure protection in the fight against Terrorism In: COM (2004) 702 final, Brussels; 20 October 2004

[2] Environmental Protection Agency (EPA): Drinking Water Infrastructure Needs Survey and Assessment In: Third Report to Congress, EPA 816-R-05-001 (2005)

[3] M. Friedman, L. Radder, S. Harrison, D. Howie, M. Britton, G. Boyd, H. Wang, R. Gullick, M. LeChevallier, D. Wood, J. Funk: Verification and Control of Pressure Transients and Intrusion in Distributed Systems In: ISBN: 1843398966 AwaaRF Report Series (2005)

[4] P. Baronti, P. Pillai, V. Chook, S. Chessa, A. Gotta, and Y. F. Hu: Wireless Sensor Networks: a Survey on the State of the Art and the 802.15.4 and ZigBee Standards In: Computer Communications, 30 (7), 26 May 2007, pp. 1655-1695.

[5] E.M. Yeatman: Energy scavenging for wireless sensor nodes In: Advances in Sensors and Interface, 2007 (IWASI 2007), pag 1-4 (2007)

[6] Crossbow Technology: Mica wireless measurement system datasheet In: www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA.pdf.

[7] Ted G. Lewis: Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation In: John Wiley & Sons, Inc. (2006)

[8] United States General Accounting Office: Challenges and Efforts to Secure Control Systems In: Report to Congressional Requesters (March 2004)

[9] D. Fletcher Spatial Information Technologies in Critical Infrastructure Protection In: National Consortium on Remote Sensing in Transportation

[10] O. S. Saydjari: Cyber Defence: Art to Science In: Communications of the ACM, Vol 47, No 3 (March 2004)

[11] R. Roman, C. Alcaraz, J. Lopez: The role of Wireless Sensor Networks in the area of Critical Information Infrastructure Protection In: Information Security Technical Report 12, pag. 24-31 (2007)

[12] I. Stoianov, L. Nachman, S. Madden, T. Tokmouline: PIPENET: A Wireless Sensor Network for Pipeline Monitoring In: Information Processing in Sensor Networks, 2007 (IPSN 2007), pag 264-273 (2007)

[13] Ratnasamy S., Karp B., Shenker S., Estrin D., Govindan R., Yin L., Yu F.: Data-centric storage in sensornets with GHT, a geographic hash table. Mob. Netw. Appl. (MONET) **8**(4) (2003) 427–442

[14] Albano, M., Chessa, S., Nidito, F., Pelagatti, S.: Q-NiGHT: Adding QoS to Data Centric Storage in Non-Uniform Sensor Networks. Mobile Data Management (MDM) 2007, Mannheim, Germany, 2007.

[15] Newsome, J., Song, D.: GEM: Graph EMbedding for Routing and Data-Centric Storage in Sensor Networks Without Geographic Information. In: Proc. of the First International Conference on Embedded Networked Sensor Systems, Los Angeles, (2003) 76–88

[16] A. G. Dimakis, V. Prabhakaran, K. Ramchandran: Ubiquitous Access to Distributed Data in Large-Scale Sensor Networks through Decentralized Erasure Codes In: Information Processing in Sensor Networks (IPSN 2005), Los Angeles, California, USA, April 25-27 2005.

[17] M. Albano, S. Chessa: Replication vs Erasure Coding in Data Centric Storage for Sensor Networks In: University of Pisa, Technical Report TR-08-09, pag. 26, May 21, 2008

[18] A. Frangioni Dual-Ascent Methods and Multicommodity Flow Problems In: PhD Thesis, 1997

[19] F. T. Leighton: Complexity Issues in VLSI: Optimal Layouts for the Shuffle-Exchange Graph and Other Networks In: MIT Press, 1983

[20] S. N. Bhatt, F. T. Leighton: A Framework for Solving VLSI Layout Problems In: J. Comp. and Syst. Sci. 28(2), p. 300-343, 1984