Network Design Based On Graph Theory

Rado Razafy, Soloniaina Rakotomiraho, Rivo Randriamaroson

***Abstract*— The continuous technology development has been forcing the world to share information or access to some specific shared data. "Being isolated" is no longer a solution "to protect oneself" against any threats. This project aims to define a new design of a network architecture based on the graph theory. The article shows the method used to design a network component.**

***Index Terms*— computer science, graph theory, network architecture, network modelling.**

I. Introduction

Graph theory is part of the field of mathematics discovered in 1736. This theory was born when Euler demonstrated that it was impossible to cross each of the seven bridges of the Russian city of Königsberg (Kaliningrad) exactly once and to return to the starting point (***Figure 1***). The bridges span the arms of the Pregel flowing on both sides of the island of Kneiphof [1].

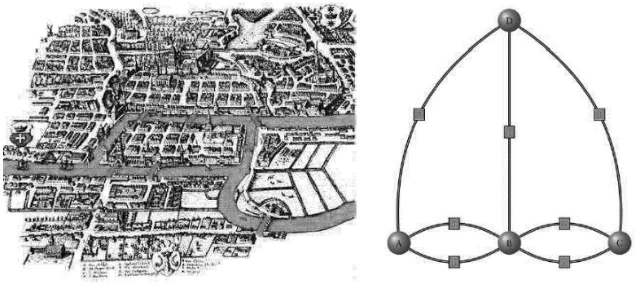

Figure 1: Bridge of Königsberg

II. Definitions and Fundamental Concept

A graph G is visually represented by *vertices* and *edges* which are connecting the vertices. The definition is $G = (V, E)$ with V is the ***set of vertices*** and E is the set of edges, formed by pairs of vertices. E is a *multiset*, in other words, its elements can occur more than once so that every element has a *multiplicity*. Formally, V can be expressed as V= $\{v_1, v_2, \dots, v_n\}$ and E as E=$\{e_1, e_2, \dots, e_m\}$or E=$\{(v_1,v_2), \dots, (v_i, v_k)\}$.

A graph G is *undirected* if G = (V, E) is a pair comprising a finite set of vertices (or nodes) V = {1, 2, …, n} and a set of $E \subset V \times V$ *unordered* pairs called edges (or arcs). G is *directed* (or digraph) if all pairs inE are *ordered*. For example, in the undirected graph (1, 2) = (2, 1) and in the directed graph (1, 2) ≠ (2, 1).

There are some terminologies [2]:

- The graph H = (U, F) is a subgraph of G = (V, E) if U ⊆ V, F ⊆ E and edges in F connect only vertices in U;
- The two vertices u and v are the *end vertices* of the edge (u, v);
- Edges that have the same end vertices are *parallel*;
- An edge of the form (v, v) is a *loop*;
- A graph with no edges is *empty* (E is empty);
- A graph with no vertices is a *null graph* (V and E are empty);
- A graph with only one vertex is *trivial*;
- Edges are *adjacent* if they share a common end vertex;
- Two vertices u and v are *adjacent* if they are connected by an edge, in other words, (u, v) is an edge;
- The graph is a *complete graph* if every pair of vertices are adjacent;
- The *degree* of the vertex v, d(v) is the number of edges with v as an end vertex;
- A *pendant vertex* is a vertex whose degree is 1;
- An edge that has a pendant vertex as an end vertex is a *pendant edge*;
- An *isolated vertex* is a vertex whose degree is 0.

A sequence of edges $e_1, e_2 \dots . e_k$ such that $e_1 = (v_1, v_2), e_2 = (v_2, v_3), \dots, e_k = (v_k, v_{k+1})$ is a *path* from v_1 to v_{k+1}. There are path classifications:

- A path from a vertex to itself is a *cycle*;
- A path is *elementary* if does not contain the same edge twice;
- A path is *simple* if it does not pass through the same vertex twice.

A graph is *connected* if:

- from any node one can get to any other by following a sequence of edges OR
- any two nodes are connected by a path.

A directed graph is *strongly connected* if there is a directed path from any node to any other node.

III. Risk Analysis

An unorganized network can be compared with a random graph. The model provided by Erdös and Rényi can be taken as a reference to illustrate the this type of graph. This model consists of a set of *n* nodes connected by edges that are uniformly randomly placed between node pairs. The most

Rado RAZAFY, PhD Student, Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques (LR-SE-I-MSDE)

Soloniaina RAKOTOMIRAHO, Professor,Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques (LR-SE-I-MSDE)

Rivo RANDRIAMAROSON, PhD , Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques (LR-SE-I-MSDE)

commonly studied is the one called $G_{n,p}$ where between two nodes each edge is independently present with a probability *p* and absent with a probability *1-p*. $G_{n,p}$ is characterized by *z* which is the average degree of *z* of the nodes. The mean number of edges is $\frac{1}{2}\ n\,(n-1)p$ and the connection means corresponds to 2 times this result since each edge has two ends [3]. The average degree of nodes is given by ***Equation 1***.

$$z = \frac{n\,(n-1)p}{n} = (-1)p \cong np \qquad (1)$$

The connected devices can be considered as vertices and the connections as edges. In case of necessity to consider the direction of the flows, we can use the notion of arc. ***Figure 2*** shows the complexity of the schematization of an interaction for a specific social network.

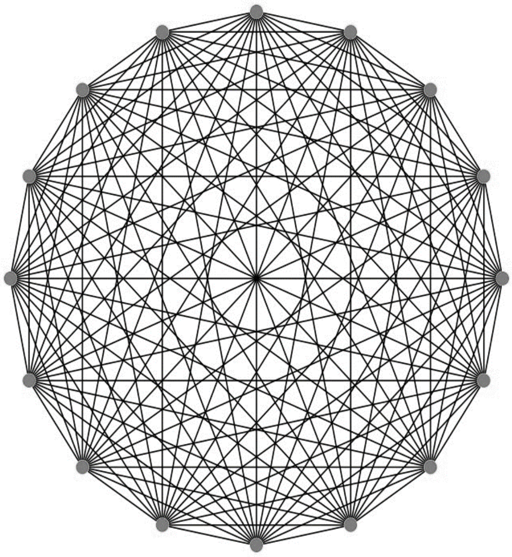

Figure 2: sample graph of social network

Even for a simple information system within a company, a modelling remains complex. In a flat network, all equipments can communicate to each other. This type of network is like a strongly connected graph. A graph G = (X, U) is strongly connected if $\forall\ i, j \in X$, exists a path between *i* and *j*. The speed of a virus propagation, saturation attacks and hacker reconnaissance are difficult to monitor in such a network. Indeed, there are several paths that could be used to target a specific equipment. All equipment can exchange data between them and the implementation of security rules becomes difficult.

if the flow between the nodes is taken aside, a network is the same as a complete graph. A complete graph, ***Figure 3***, is by definition $\forall x_i, x_j \in X, (x_i, x_j) \notin U \Rightarrow (x_i, x_j) \in U$. In this type of graph, the average degree of the node is *z=n*.

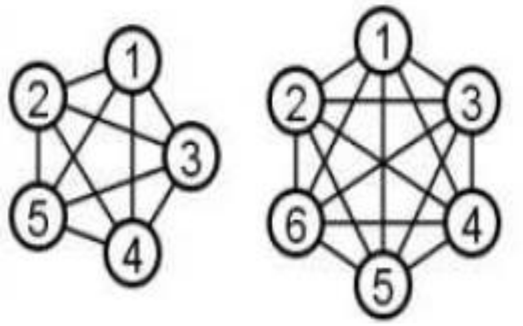

Figure 3: sample of complete graph

A. Access attacks and sniffing

The path in a network is basically ensured by the ARP protocol (Adress Routing Protocol) and the CAM table (Content Addressable Memory). The ARP and the CAM tables do not work on the same layers of the OSI model. Indeed, the ARP protocol is in charge of joiningthe layer 3 (network layer, IP address) and the layer 2 (link layer, MAC address), while the CAM tables do the same thing between layer 2 (link, MAC address) and Layer 1 (physical and physical port number of the switch) [4].

To route a traffic, the switches maintain the CAM table. The paths between the nodes are thus ensured by the switches (if the switches work correctly). Let G = (X, U) be a network representation, $\forall\ i, j \in X$, the path between *i* and *j* is established from the CAM table. For $z \in X$, the path between *i* and *j* will not pass through *z*. The network will be simplified (***Figure 4***) and no longer resembles to a complete graph.

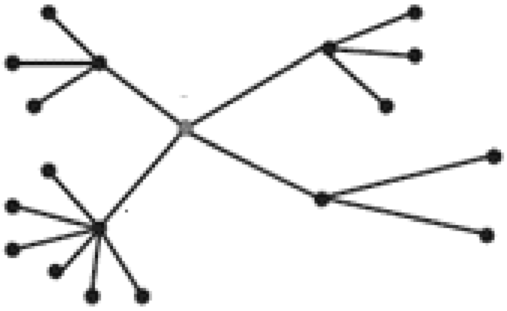

Figure 4: Representation of network with switch

It is possible to exploit a memory overflow on a certain devices by saturating the MAC address table, in order to capture the packets. Indeed, when the memory is full, switches behave as a hub and send packets to all hosts connected to the network [4]. In this case, the network will be modelled as a complete graph. All the equipment will receive the information circulating within network. By changing the configuration of the network card in promiscuous mode, an attacker can use a sniffing software to analyse and subtract confidential information on the network.

Apart from the use of sniffing techniques, the risk of scanning by a vulnerability scanner is also high in such a network. Let a strongly connected graph G = (X, U) represent a flat lattice. $\forall\, i, j \in X$, there exists a path between i and j. A computer with a vulnerability scanner can therefore scan all network devices. The probe program sends over a data and analyses the received responses. The program scans all ports to determine their behaviour. An attacker can perpetrate attacks by having a complete network mapping and components.

B. *Attack by Denial Of Service*

The denial of service can occur in three ways:

- Attacking a target from several sources (flooding),
- Attack from one source towards a set of target
- Rebound attack

The load of any target in the whole network can be calculated from the average degree of the nodes. The average degree of nodes is given in Equation 1.

Let a graph G = (X, U) be a representation of a flat network, let *i* be a target in the network, $\forall\, j \in X$ there exists a path between *i* and *j*. The average degree of the nodes is *z* = *n* with *n* the number of nodes. For a flooding attack, the attacker exploits all available devices on the network in a way that once exploited they send commands to the given target. Suppose each device sends *r* requests per second to the target, the load supported by the target will be *nr*.

For example, the TCP-SYN flooding is a variant of flooding that relies on a TCP fault. The attack is at level 4 (transport) of the OSI layer. The principle is to send a large number of connection requests to the server (SYN) from several machines. The server will return and respond with the SYN-ACK packet and wait in return for an ACK response that will never happen. At the tail saturation to store the connections awaiting the end of opening, the machine will not accept any new requests. For example, in a network with 100 client computers and one server as a target, if each workstation sends simultaneously *r* connection requests per second, the server must satisfy *r * 100* connection requests per second.

In another case, an attack may come from a single source to the network set. Given a strongly connected graph G = (X, U) represents a flat network, let $j \in X$ be the source of the attack, $\forall\, i \in X$, there exists a path between *i* and *j*.

The last possible attack alternative is to make rebound attacks which is the combination of the two previous attacks. The principle is to make the fraudulent requests pass through another network equipment.

C. *Attack of repudiation*

As a reminder, repudiation is an attack against responsibility. In other words, repudiation consists in attempting to give false information or to deny that an event or transaction actually occurred. The IP spoofing also known as mystification, for example, is not an attack as such but a method to dissipate the identity. The method is to perform an action by falsifying the IP address. The source communicates with the target machine as a trusted machine. Bounce attack is a method of hiding identity and causing the attack to pass through another machine. The goal is to hide the traces that go up to the source machine. In the case of a network in the form of a random graph, it is impossible to perform an identity check. In addition, as all equipment can communicate to each other, the hackers can use other equipment as a pivot to hide their real identity.

IV. Building network

A. *VLAN concept*

VLAN is a Virtual Local Area Network. To reduce complexity, "community" algorithms such as "waltrap" and "fastgreed" can be used to form distinct groups. The first is based on a probability proportional to the weights of the edges and the degrees of the neighbouring vertices. The second is based on a hierarchical decomposition of the edges of the graph to highlight the communities [5]. ***Figure 5*** is an example of the representation of communities in a graph.

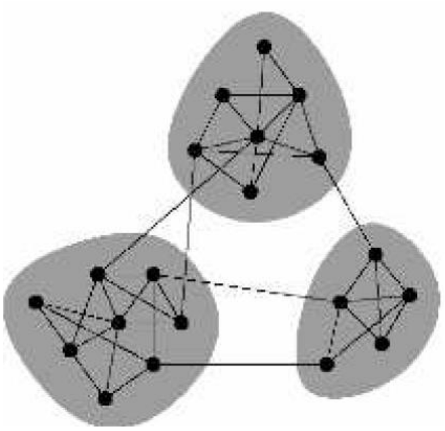

Figure 5: Graph community

The degree of correlation or the average degree of the neighbours of a vertex *i* is highlighted in ***Equation 2***.

$$k_{nn,i} = \frac{1}{k_i}\sum_j a_{ij}k_j \qquad (2)$$

Here is an example of a corresponding graph Figure 6.

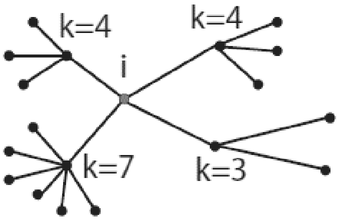

Figure 6: Sample graph

$$k_i = 4$$
$$k_{nn,i} = \frac{1}{4}(3+4+4+7) = 4{,}5$$

The subgraphs in the communities can also be obtained by the "clustering method" or by the detection of subgroups by k-cores (Seidman, 1983). The first method is to search for entities which are sharing the same relationships or having the same properties. The second principle is based on the degree of the nodes. A *k-core* is a subgraph where each sum is adjacent to a minimum of *k* other vertices in the subgraph. To apply these principles in a network architecture, sub-graphs or communities can perform the same by separating the networks. In practice, the virtual network technique or VLAN is used. For example, each department in a given company can be classified as a specific VLAN.

For $G = (X, U)\ et\ X_s\ C\ X. G_s = (X_s, V)$ is a subgraph of G, where V the is the restriction of the characteristic function from U to X_S . $V = \{(x, y)/(x, y) \in U \cap X_s \times X_S\}. \forall x_i \in X_s, \Gamma_s(x_{i)}) = \Gamma_s(x_{i)}) \wedge X_s$

.

B. Traffic management

Grouping in a sub-graph already increases the complexity of a network. It offers the possibility to define security rules but is still insufficient. In order to minimize the risk of attack, it must be completed by the block principle. The basic method makes it possible to decompose a graph into sub-components and block. ***Figure 7*** illustrates an example of a three-block graph. All the paths to go from block 1 to block 3 or vice versa must pass through block 2. The security rules and permissions will be defined in block 2 to ensure communications. Block 2 acts as a firewall.

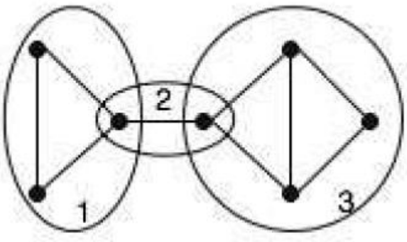

Figure 7: graph with three blocks

The routing algorithm provided by Kleinberg defines the transmission mode of the message from and towards each node [6].

```
Imput : source s and target t.
Initialisation : x ← s. (x is the current bearer of the message)
if x ≠ t do
    x ← y where y is the local or long-distance contact of x
which minimizes |y – t| (uniform random choice in case of
equality).
end if
```

In this case, communication between two nodes of two different blocks must pass through the central block. Flow filtering and monitoring will be performed in this block. Firewalls and roads provide filtering and data routing between blocks.

A stateless firewall intervenes on the network and transport layers of the OSI model. The filtering rules are based on the source and / or destination IP address and the protocol used (port). They are kept in an "Access Control Lists". The limitation of this type of firewall is that for a port opened by the administrator, is also open for the hackers to exploit. It cannot stop attacks of IP Spoofing and SYN Flooding.

A statefull firewall is an extension of the stateless firewall. In addition to the previous features, it also provides packet controls. The attributes stored in memory are the IP addresses, port numbers and sequence numbers of the packets that have passed through the firewall. It can then allow or deny new connections without reading the access control lists. The advantage is that it can stop DoS attacks by removing unanswered fake packets. This type of protocol has the operating modes of the standard protocols but there is a limitation for protocols that are using specific port numbers.

C. Flow management

After the VLAN and the firewall, the final step is to define the rules of exchanging data between the different subnets. The principle is to create an exchange matrix between subgroups. The matrix will be translated by the rules at the firewall level or the development of VLAN access lists. Two things are important in establishing the rules: the authorized protocol and the direction of the communication.

The flow can be represented with an *adjacency matrix*. The adjacency matrix of the graph G = (V, E) is an $n \times n$ matrix $D = (d_{ij})$ where n is the number of vertices in $G, V = \{v_1, \dots, v_n\}$ and

$d_{ij} = number\ of\ edges\ between\ v_i\ and\ v_j$

With $d_{ij} = 0$ if (v_i, v_j) is not an edge in G.

To configure a firewall, the direction of the flow is very important. The network has to be represented like a directed graph. The *all-vertex incidence matrix* of a non-empty and loopless directed graph G is $A = (a_{ij})$,

$$a_{ij} = \begin{cases} 1\ if\ v_i\ is\ the\ initial\ vertex\ of\ e_j \\ -1\ if\ v_i\ is\ the\ terinal\ vertex\ of\ e_j \\ 0\ otherwise \end{cases}$$

For example,

$$A = \begin{array}{cc} \begin{matrix} e_1 & e_2 & e_3 \end{matrix} & \\ \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} & \begin{matrix} v_1 \\ v_2 \\ v_3 \end{matrix} \end{array}$$

The table can be used to represent VLANs relationships. The ***Table 1*** show an example.

Table 1: example of flow diagram

	VLAN1	VLAN2	VLAN3
VLAN1		ftp, http	rdp
VLAN2	file sharing		database
VLAN3		ftp	

V. Monitoring Network

The implementation of the basic and well-known means of protection is not usually sufficient to ensure safety full network security. There is a need for monitoring and auditing. Monitoring is basically examining the behaviour of various equipment and applications, in real-time. The audit consists of evaluating what measures have been applied. The actions to be taken are as follows:

- Analyse the access to the specified resources and access to various devices;
- Detect any unsuccessful attempts to disregard a security policy;
- Detect any successful attempts to override a security policy;
- Detect any privilege exploitation (escalation)
- Identifying the offenders.

In order to understand the exploitation of security information, it is necessary to identify the best sources and the associations of the components. The firewall can be taken as a source. The complexity can be obtained with the number of partitions according to the number of devices which is obtained using the number of Bell B_n (***Equation 3***) [7].

$$B_n = \frac{1}{e}\sum_{k=1}^{\infty}\frac{k^n}{k!} \qquad (3)$$

For example if n = 4 point of control, $B_n = 15$ with:

- 1 partition with 1 class (abcd),
- 7 partitions with 2 classes (ab,cd), (ac,bd), (ad,bc), (a,bcd), (b,acd), (c,bad), (d,abc),
- 6 partitions with 3 classes (a,b,cd), (a,c,bd), (a,d,bc), (b,cad), (b,d,ac), (c,d,ab),
- 1 partition with 4 classes (a,b,c,d).

For n = 30, $B_n = 8{,}47 * 10^{23}$.

The sample algorithm based on graph theory are:

- Social network analysis;
- Algorithm of “Subdue”;
- Links analysis;
- Design of graph;
- Minimum spanning tree;
- Articulation detection;
- Clustering method.

VI. Conclusion

The graph theory is a tool to design a network. It will serve as a reference but the steps depend on each other. Securing a network will have an impact on the business investment and performance. A design must begin with a risk assessment and identification. The goal is to put in place a right and effective solution. It is necessary to refer to the best practice and the existing norm.

References

[1] V. Levorato, "Contributions à la Modélisation des Réseaux Complexes: Prétopologie et Applications", 2010
[2] K. Ruohonen, "Graphe Theory", 2013
[3] P.Erdös et A.Rényi., "On random graphs", Publicationes Mathematicae, 1959
[4] https://www.aldeid.com/, 2016
[5] A. Laure et T. Zufferey, "Intégration de méthodes de data mining dans le renseignement criminel", 2009
[6] E.Lebhar, "Algorithmes de routage et modèles aléatoires pour les graphes petits mondes", 2006
[7] S. Tuffery, "Cours de data mining", 2014

Rado RAZAFY, PhD Student
Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques
(LR-SE-I-MSDE)
Ecole Doctorale en Sciences et Techniques de l’Ingénierie et de l’Innovation (ED-STII)
Ecole Supérieure Polytechnique - Université d’Antananarivo
BP 1500 - Antananarivo 101 – Madagascar

Soloniaina RAKOTOMIRAHO, Professor
Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques
(LR-SE-I-MSDE)
Ecole Doctorale en Sciences et Techniques de l’Ingénierie et de l’Innovation (ED-STII)
Ecole Supérieure Polytechnique - Université d’Antananarivo
BP 1500 - Antananarivo 101 – Madagascar

Rivo RANDRIAMAROSON, PhD
Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques
(LR-SE-I-MSDE)
Ecole Doctorale en Sciences et Techniques de l’Ingénierie et de l’Innovation (ED-STII)
Ecole Supérieure Polytechnique - Université d’Antananarivo
BP 1500 - Antananarivo 101 – Madagascar