Università di Pisa
Scuola di Ingegneria

Performance Evaluation of Computer Systems and Networks

Epidemic Broadcast

Supervisors
Prof. Giovanni Stea
Ing. Antonio Virdis

Students
Marco Pinna
Rambod Rahmani
Yuri Mazzuoli

January 14, 2021

Contents

Chapter 1

Introduction

In this work a study on the broadcast of an epidemic message is carried out.
The specifications are the following:

> ## Epidemic broadcast
>
> Consider a 2D floorplan with N users randomly dropped in it. A random user within the floorplan produces a *message*, which should ideally reach all the users as soon as possible. Communications are *slotted*, meaning that on each slot a user may or may not relay the message, and a message occupies an entire slot. A *broadcast radius* R is defined, so that every receiver who is within a radius R from the transmitter will receive the message, and no other user will hear it. A user that receives more than one message in the same slot will not be able to decode any of them (*collision*). Users relay the message they receive *once*,according to the following policy (*p-persistent relaying*): after the user successfully receives a message, it keeps extracting a value from a Bernoullian RV with success probability p on every slot, until it achieves success. Then it relays the message and stops. A sender does not know (or cares about) whether or not its message has been received by its neighbors.
>
> Measure at least the broadcast time for a message in the entire floorplan, the percentage of covered users, the number of collisions.
>
> In all cases, it is up to the team to calibrate the scenarios so that meaningful results are obtained.

The work is organized as follows:
firstly, an initial overview and a presentation of the problem are given. Here, meaningful parameters to be tweaked and useful scenarios are identified and some considerations about them are made.
Secondly, a graph-based modelling technique, commonly used in literature, is proposed and some simplified scenarios are analysed.

Then, in chapter 3 the real model is considered and some assumptions about it are made, such as its performance when one or more parameters are near the ends or such as its asymptotic behaviour with different configurations of the parameters.
In chapter [TODO] the construction of the simulator is described and the results of its validation are presented.
Chapter [TODO] concerns the full simulation and performance evaluation of the system.

Chapter 2

Overview

To perform the analysis of message of a message broadcast among users in a floorplan, the following hypotheses have been made:

- the floorplan always has a rectangular shape and it is empty, i.e. there are no obstacles such as walls or pillars in it.
- each user takes zero area and does not move inside the floorplan.
- the transmission of a message is instantaneous and it happens at the beginning of every time slot; the whole apparatus can therefore be considered a *Discrete Time System*.

Depending on the performance metrics to be analysed, different choices can be made about the parameters to be adjusted. According to the specifications, the main three metrics for this study are:

- the broadcast time `T` for a message in the entire floorplan
- the percentage of covered users `U`
- the number of collisions `C`

The following parameters have been identified: the transmission range of the devices, the *per-slot* transmission probability, the floorplan size and shape and finally the density of users per square metre.

More in detail:

- the radius of transmission `R`: it represents the maximum distance between two devices such that the message sent from one is detected by the other. It is the same for every device on the floorplan. Realistic values for R have been taken from Bluetooth Low Energy standard and range from a minimum of 5 metres to a maximum of 50 metres.

- the *per-slot* transmission probability `p`: it is the success probability for the Bernoulli random variable associated to the transmission

- the side `L` of the rectangle that models the floorplan. The set of values for L was chosen to range from 10 metres to 100 metres.

- the *aspect ratio* of the rectangle `a`, defined as the ratio between the longer side L and the shorter side W. Its values were chosen to range from 1 (square floorplan) to 8. Because of the radial symmetry of the transmission phenomenon, there is no actual need to consider values for `a` lower than 1; it is enough to only consider "wide" rectangles and discard "tall" rectangles, or vice versa, since, for the purpose of the study, a $50m \times 100m$ rectangle would behave exactly the same as a $100m \times 50m$ one.

- the population density `d`, defined as the average number of users in a square metre of the floorplan.

The rationales behind the choices of the parameters were the following:
- the transmission radius clearly has a great impact on all the performance metrics; the greater the radius, the faster the message moves across the floorplan and the greater the number of users that can be reached. On the other hand, a greater radius is likely to cause more collisions than a smaller one.
- the higher the transmission probability, the faster a message "moves away" from a device. At the same time, a high transmission probability implies a high collision probability in a local area where two or more nodes are transmitting.
- a bigger floorplan area, all else being equal, will need a longer time to be covered entirely from the broadcast.
- a very long and very narrow floorplan will probably cause less collisions than a square one with the same area, as the average number of users in range of another user decreases.
- population density `d` was chosen over total population `N` as it was deemed to be more suitable for a 2^kr factorial analysis. Given `d`, `L` and `a`, the total population `N` can be computed as $N = d \cdot A_{tot} = d \cdot L \cdot W = d \cdot \frac{L^2}{a}$.

The effective duration of the timeslot obviously has an effect on the total broadcast time, but it is only a scaling factor on the total number of slots. `T` can therefore be measured in terms of slots and converted to units of time accordingly.

Chapter 3

System modelling

Wireless communication networks have been extensively study in scientific literature and one of the most used mathematical tools to model them and their behaviour is *graph theory*. In this work the same approach was followed.
As the broadcast goes on and the message is transmitted among the nodes, the system goes through different states where each node could be transmitting, listening for an incoming message or could have already sent the message and has stopped. These different behaviours were modelled by means of three different states:

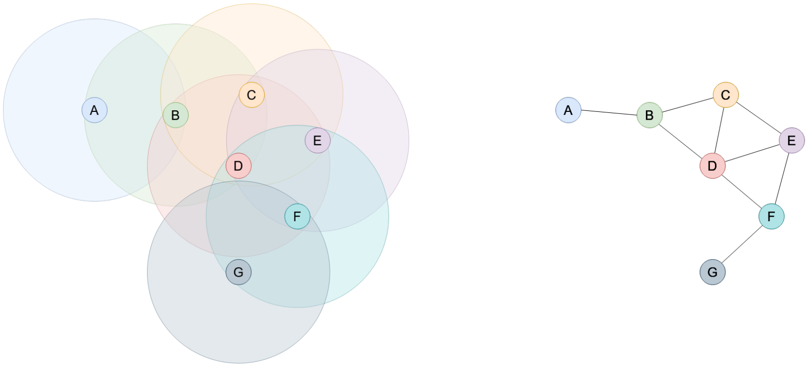
- `listening`: the node has not received the broadcast yet and therefore it is still listening for incoming messages from other nodes.

- `transmitting`: the node has received the message and during each slot it is trying to transmit it to adjacent nodes. A node in `transmitting` state will sometimes be referred as *active* node.

- `sleeping`: the node has already received the message and transmitted it in turn. Once a node is in a sleeping state, it has no effect on the system any more.

3.1 Graph model for wireless systems

The N users dropped in a floorplan make up the set of vertices V of a graph G, whose set of edges E is composed by all the connections between two nodes in reach of each other. Let us take a simplified scenario to exemplify this approach.



(a) Ranges representation (b) Equivalent graph representation

Figure 3.1

In Figure 3.1 (a) devices A, C and D are within device B transmission radius. In the equivalent graph, there will be edges that connect B to A, B to C and B to D. The same goes for all the other vertices.
In general, the existence of an edge from vertex i to vertex j means that nodes i and j are within reach of each other. Two vertices connected by an edge are said to be *adjacent*. The set of a vertex v together with all its adjacent nodes forms a subgraph called the *neighbourhood of v*.
During the broadcast, a node can only receive from and transmit to its neighbourhood.

Once a node has transmitted the message and has gone into `sleeping` state, it disappears, along with all the edges connected to it. Therefore, the set of vertices V changes with time. This is what in literature is called a *dynamic graph* or, more specifically, a *node-dynamic graph*.

Modelling the system with graphs also allows for the easy computation of a lower bound for the broadcast time `T`, which can be useful for the validation of the simulator. Given a graph $G(V, E)$ that represents the users in the floorplan, let v^* be the starter of the broadcast, i.e. the first node with the message.
In a best case scenario, the system evolves with no collisions at all and the message moves along the paths of the graph, reaching all nodes.
Let $d(u, v)$ be the *distance* between two vertices u and v, i.e. the length of a shortest directed path from u to v consisting of arcs, provided at least one such path exists.
Then, the lower bound for the broadcast time is given by the greatest distance between v^* and any other vertex. This quantity, in graph theory, is called *eccentricity* of the vertex v^*. More formally, the eccentricity of v^* is defined as follows:

$$\epsilon(v^*) = \max_{v \in V} d(v^*, v) \tag{3.1}$$

3.2 Simplified models

3.2.1 Single queue configuration

Let us consider a configuration where devices arranged in a line, as shown in Fig. 3.2. Each device only has two neighbours, except for the outer ones that only have one. Let assume A to be the broadcast starter. A is in `transmitting` state while all the other nodes are initially in `listening` state.

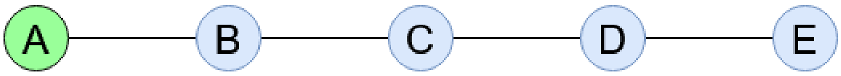


Figure 3.2

It is clear that, with this configuration, each listening node has a maximum of one active node in its neighbourhood and thus cannot possibly receive the message from two different sources at the same time. This guarantees the absence of collisions.

In such a scenario, 100% asymptotic coverage is ensured: during each slot, the active node extracts a Bernoulli RV with success probability p. The probability of the active node not transmitting for k consecutive slots is a geometric distribution:

$$P(X = k) = (1\text{-}p)^k \tag{3.2}$$

The successful transmission of the message from an active node to its neighbour is thus guaranteed since $\lim_{k\to\infty}(1\text{-}p)^k = 0$

As for the total broadcast time `T`, on average it is equal to the mean value of the Bernoulli RV, p, times the number of hops needed to reach the last node.

$$E[T] = p \cdot (N - 1) \tag{3.3}$$

3.2.2 Star configuration with one active node

Another useful simple configuration worth analysing is a star-shaped configuration. In this setup, there is a central node A connected to N-1 nodes, all of which are non-adjacent to each other. Let us suppose A to be the broadcast starter.

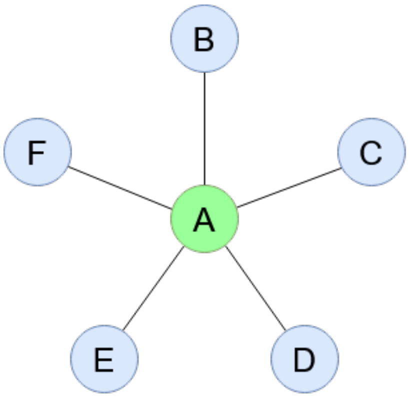


Figure 3.3

The absence of collisions is ensured in this scenario as well, for the same reason as the previous example.
At each slot, A keeps extracting a Bernoulli RV. When the extraction is successful, A broadcasts the message to all its neighbour and total coverage is reached. Hence, 100% asymptotic coverage is ensured in this case too, as the probability of A not transmitting for k consecutive slots is Eq. 3.2 and goes to 0 as k goes to infinity.
In this case, the average total broadcast time E[`T`] is simply equal to the mean value of the Bernoulli RV p.

If the broadcast starter was one of the "rays" of the star, instead of the center, there would not be much difference: absence of collisions and total coverage would be ensured as well.
As for `T`, its expected value would just be $2p$, since the are now **two** hops involved in the broadcast: one from the starter to the center of the star and the other from the center node to all the N-2 remaining ones.

3.2.3 Star configuration with all but one active nodes

This configuration can be seen as the complement of the previous one: every "ray" of the star is active and trying to transmit the message to the center node.

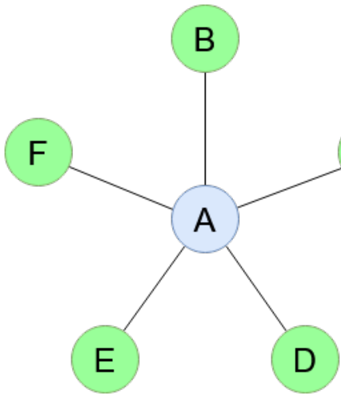


Figure 3.4

Now there is the possibility of collisions and a non-zero probability that there will never be total coverage.

To simplify the analysis of this system and obtain some more insight, it is useful to model it by means of a discrete-time Markov chain.

Discrete-time Markov chain model for N nodes transmitting to a target

Since the state of the system evolves only once per slot, it can be modelled with a discrete-time Markov chain (DTMC).

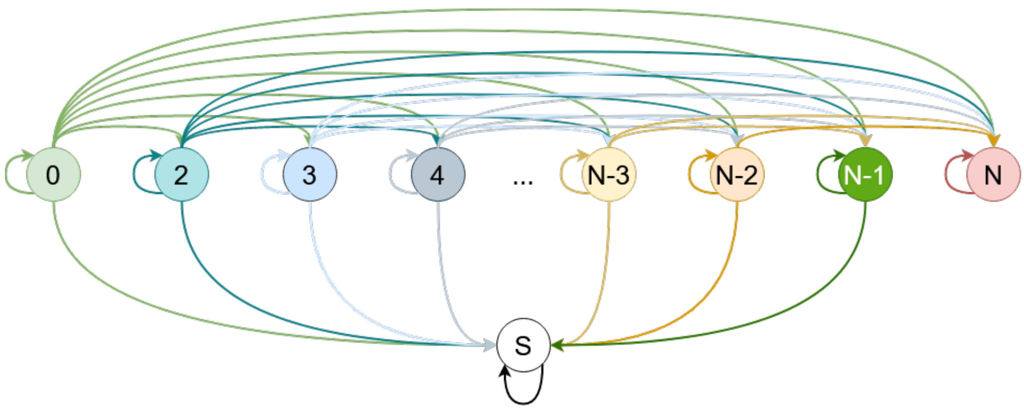


Figure 3.5: Discrete-time Markov chain for a scenario with N active devices

Figure 3.5 shows the DTMC for a generic configuration with N transmitters (transition probabilities are not shown for the sake of clarity).
State 0 and states 2 to N represent the number of `sleeping` devices, namely devices that have already sent the message and have stopped. State S represents the successful

transmission state, which the system transitions to when only one device has transmitted the message during the previous slot.

The initial state X_0 is 0. Any transition from a state i to a state j, $j \neq S$, means that more than one device has transmitted the message and consequently the target device has detected a collision. Both state S and state N are *absorbing states*, i.e. states that, once entered, cannot be left (as can be seen in Fig. 3.5, where the only outgoing arrow from each of the two goes back to the state itself).

If the system transitions to state N, it stays in it indefinitely since all the devices would be `sleeping` and they cannot become active again. This implies that there will never be total coverage since the target device will forever stay in a `listening` state.
On the other hand, state S, although being an absorbing state as well, should actually be considered an "exit" instead of a "sink": if the system transitions to it, the target device has successfully received the message and the DTMC does not model the system any more.
Another interesting observation concerns state N-1: once the system reaches this state, it would be guaranteed that the target device will sooner or later receive the message, for the same reason set forth in 3.2.1.

Transition probabilities

Let us now address the challenging part: computing the transition probability.
During the first slot, the probability that j devices out of N transmit the message is the following:

$$P_1(j, N) = \binom{N}{j} p^j (1-p)^{N-j} \tag{3.4}$$

Derivation of 3.4 can be found in Appendix A.

As for the probability of j devices transmitting at the same time during slot k, be it $P_k(j)$, we can model the system as if it was in the first slot, with the total number of active devices now being equal to N - t, where t is the total number of devices that have transmitted up to the (k-1)-th slot.
The problem with this formulation is... Is what?

A better way to compute the probability of having j devices transmitting at slot k, is to use the *stochastic matrix* of the Markov chain.

If the probability of moving from state i to j in one time slot is $Pr(j|i) = P_{i,j}$, the stochastic matrix P is given by using $P_{i,j}$ as the i-th row and j-th column element, e.g.

$$P = \begin{bmatrix} P_{0,0} & P_{0,S} & P_{0,2} & \dots & P_{0,j} & \dots & P_{0,N} \\ P_{S,0} & P_{S,S} & P_{S,2} & \dots & P_{S,j} & \dots & P_{S,N} \\ P_{2,0} & P_{2,S} & P_{2,2} & \dots & P_{2,j} & \dots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,0} & P_{i,S} & P_{i,2} & \dots & P_{i,j} & \dots & P_{i,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,S} & P_{N,2} & \dots & P_{N,j} & \dots & P_{N,N} \end{bmatrix}$$

Since S and N are absorbing states, $P_{S,j} = 0$ for j $\neq$ S and $P_{N,j} = 0$ for j $\neq$ N.
Moreover, all the possible transitions can only generate a non-decreasing sequence of states, hence $P_{i,j} = 0 \ \forall i > j$.
All the other elements of the matrix can be computed using formula 3.4:

$$P_{i,j} = \binom{N-i}{j} p^j (1-p)^{N-i-j} \tag{3.5}$$

Therefore the stochastic matrix becomes:

$$P = \begin{bmatrix} P_{0,0} & P_{0,S} & P_{0,2} & \dots & P_{0,j} & \dots & P_{0,N} \\ 0 & 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & P_{2,S} & P_{2,2} & \dots & P_{2,j} & \dots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & P_{i,S} & 0 & \dots & P_{i,j} & \dots & P_{i,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$

Let x_0 be the *initial state vector*, i.e. an $N \times 1$ vector that describes the probability distribution of starting at each of the N possible states.
To compute the probability of transitioning to state j in **k** steps, it is now sufficient to multiply the initial state vector x_0 by the stochastic matrix raised to the k-th power, e.g.

$$P_k(j) = x_0 \cdot P^k \tag{3.6}$$

In our case, the system always starts in state 0, so we have

$$x_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

which yields

$$P_k(j) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot P^k = (P^k)_{0,j} \tag{3.7}$$

Calculating the k-th power of a matrix can be an intensive task from a computational point of view. To improve the complexity of the computation, rows and columns of P can be rearranged, moving the S row and the S column as penultimate, thus obtaining

$$P = \begin{bmatrix} P_{0,0} & P_{0,2} & P_{0,3} & \dots & P_{0,N-1} & P_{0,S} & P_{0,N} \\ & P_{2,2} & P_{2,3} & \dots & P_{2,N-1} & P_{2,S} & P_{2,N} \\ & & P_{3,3} & \dots & P_{3,N-1} & P_{3,S} & P_{3,N} \\ & & & \ddots & \vdots & \vdots & \vdots \\ & & & & P_{N-1,N-1} & P_{N-1,S} & P_{N-1,N} \\ & & & & & 1 & 0 \\ 0 & & & & & & 1 \end{bmatrix}$$

P is now an upper triangular matrix and this allows for faster computation of its powers in 3.7 .

Chapter 4

Real model theoretical analysis

Chapter 5

Simulator and validation

Chapter 6

Simulation

Chapter 7

Appendices

Appendix A

Given a scenario with N transmitter devices and a target device T in reach of all the transmitters, let us define the probabilities $P_1(j, N)$ as the probability of j devices out of N transmitting at the same time during slot 1 and $P_i(j)$ as the probability of j devices transmitting at the same time during slot i.
By specification, the successful reception of the message by device T happens if and only if **one** of the transmitters sends the message during the slot. Furthermore, the successful transmission of a device is "*a Bernoullian RV with success probability* p *on every slot, until it achieves success*"; therefore we can model $P_1(j, N)$ as follows:

$$\left.\begin{aligned}
&P_1(0,N) = (1-p)^N \\
&P_1(1,N) = Np(1-p)^{N-1} \\
&P_1(2,N) = \binom{N}{2} p^2 (1-p)^{N-2} \\
&P_1(3,N) = \binom{N}{3} p^3 (1-p)^{N-3} \\
&\ldots \\
&P_1(N-1,N) = \binom{N}{N-1} p^{N-1} (1-p) \\
&P_1(N,N) = \binom{N}{N} p^N
\end{aligned}\right\} P_1(j,N) = \binom{N}{j} p^j (1-p)^{N-j}$$

As for $P_i(j)$ we can model the system as if it was in the first slot, with the total number of active devices now being equal to $N - t$, where t is the number of devices that have transmitted in the (i-1)-th slot.

Therefore, for $i = 2$ we have:

$$
\begin{aligned}
P_2(0) &= P_1(0,N)P_1(0,N) + P_1(2,N)P_1(0,N-2) + ... + P_1(N-1,N)P_1(0,1) = \\
&= P_1(0,N)P_1(0,N) + \sum_{k=2}^{N-1} P_1(k,N)P_1(0,N-k) = \\
&= \sum_{k=0}^{N-1} P_1(k,N-k)P_1(0,N) - P_1(1,N)P_1(0,N-1) \\
P_2(1) &= P_1(0,N)P_1(1,N) + P_1(2,N)P_1(1,N-2) + ... + P_1(N-1,N)P_1(1,1) = \\
&= P_1(0,N)P_1(1,N) + \sum_{k=2}^{N-1} P_1(k,N)P_1(1,N-k) = \\
&= \sum_{k=0}^{N-1} P_1(k,N)P_1(1,N-k) - P_1(1,N)P_1(1,N-1) \\
P_2(2) &= P_1(0,N)P_1(2,N) + P_1(2,N)P_1(2,N-2) + ... + P_1(N-2,N)P_1(2,2) = \\
&= P_1(0,N)P_1(2,N) + \sum_{k=2}^{N-2} P_1(k,N)P_1(2,N-k) = \\
&= \sum_{k=0}^{N-2} P_1(k,N)P_1(2,N-k) - P_1(1,N)P_1(2,N-1) \\
P_2(3) &= ... = \sum_{k=0}^{N-3} P_1(k,N)P_1(3,N-k) - P_1(1,N)P_1(3,N-1) \\
&... \\
P_2(N-2) &= \sum_{k=0}^{2} P_1(k,N)P_1(N-2,N-k) - P_1(1,N)P_1(N-2,N-1) \\
P_2(N-1) &= \sum_{k=0}^{1} P_1(k,N)P_1(N-1,N-k) - P_1(1,N)P_1(N-1,N-1) \\
P_2(N) &= P_1(0,N)P_1(N,N)
\end{aligned}
$$

which has the general form:

$$
P_2(j) = \begin{cases} \sum_{k=0}^{N-1} P_1(k,N)P_1(0,N-k) - P_1(1,N)P_1(0,N-1) & \text{j = 0} \\ \sum_{k=0}^{N-j} P_1(k,N)P_1(j,N-k) - P_1(1,N)P_1(j,N-1) & \text{0 < j < N} \\ P_1(0,N)P_1(N,N) & \text{j = N} \end{cases} \tag{7.1}
$$

where the term with the minus sign is due to the fact that, if only one device transmitted during slot i, the target device T will have correctly received the message and therefore, starting from slot $i+1$ onwards, it will not be listening for incoming messages any more but it will be itself transmitting instead.