Intersection Attack and Using Dummy Addresses

J. Kabarowski, M. Kutyłowski

15.06.2005

The model

- A network with N participants exchanging messages.
- It should support anonymity of communication (it should hide the link between the source and the destination of a message).

Anonymous communication protocol

- Messages are cryptographically encoded and re-coded on their route, no relationship can be derived from the codes.
- Example - TOR.

Applications

- electronic voting,
- access to databases with medical information,
- business negotiations,
- ...

Dangers — Intersection Attack

- Passive eavesdropper observes which user send encoded messages and which destinations receive the messages.
  Intermediate nodes not traced.
- The adversary cannot knack cryptographic encoding of messages.
- The adversary aims to reveal the destinations that get messages from Alice.

Model assumptions

- In each turn $b = k \cdot N$ users send messages (destinations are chosen uniformly at random, k is a constant less than 1).
- Alice has m *friends*, and each time she sends, the recipient is chosen uniformly at random from the set of her friends.
- Communication in rounds, within a round all messages are delivered.

Consider t communication rounds. Let:

- N_P - a random variable denoting the number of messages received by a non-friend,
- A_F - a random variable denoting the number of messages received by a friend.

Consider t communication rounds. Let:

- N_P - a random variable denoting the number of messages received by a non-friend,
- A_F - a random variable denoting the number of messages received by a friend.

Probability distributions:

$$N_P \sim B(t \cdot b, \frac{1}{N}), \quad A_F \sim B(t \cdot b, \frac{1}{N}) + B(t, \frac{1}{m}),$$

$$E(N_P) = t \cdot \frac{b}{N} = t \cdot k, \ Var(N_P) = t \cdot k \cdot \left(\frac{N-1}{N}\right),$$

$$E(A_F) = t \cdot k + \frac{t}{m} \ Var(A_F) = t \cdot k \left(\frac{N-1}{N}\right) + \frac{t(m-1)}{m^2}.$$

Disclosure attack

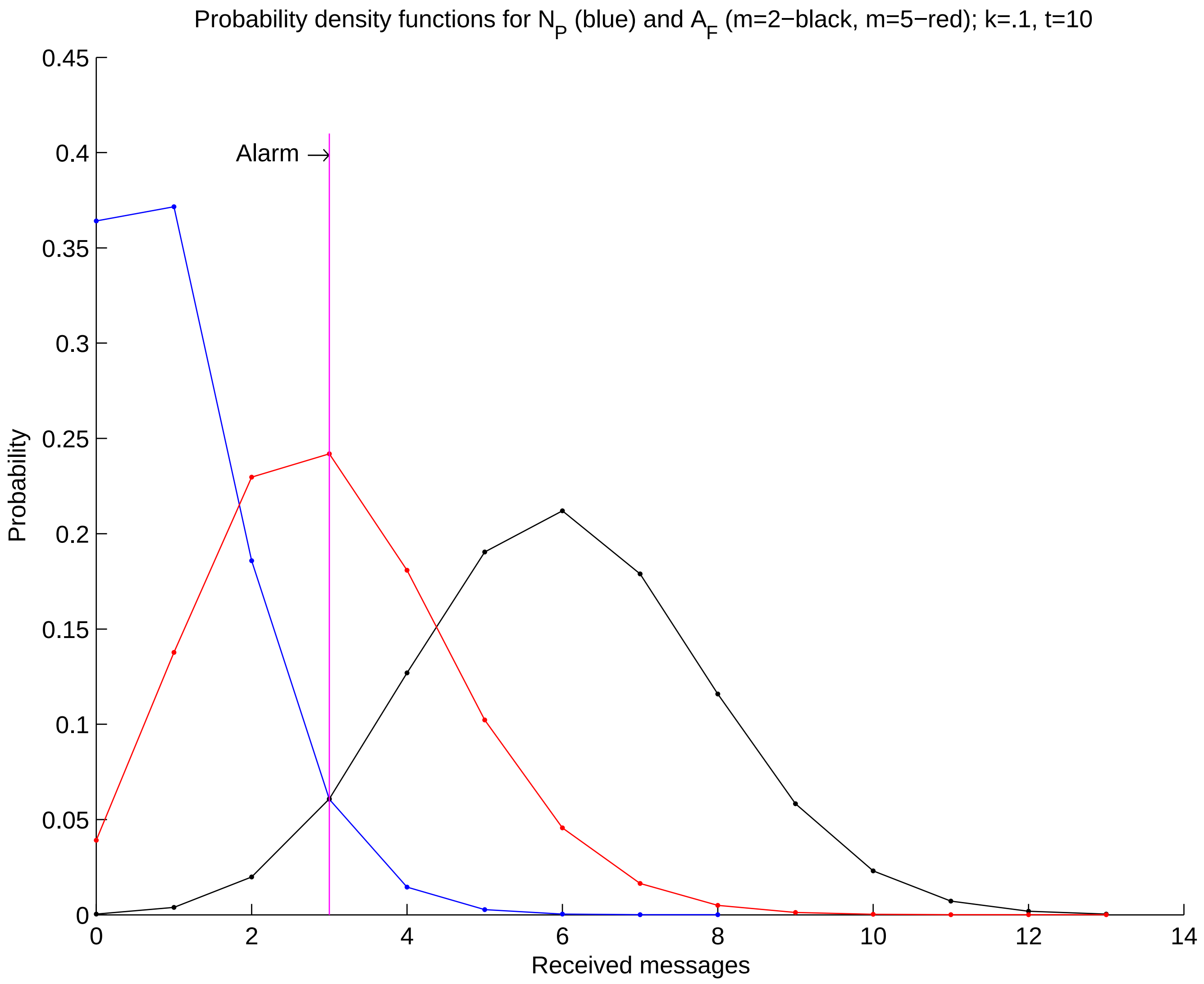
- even if a single round offers a good level of anonymity, the adversary may collect statistical information from mamy rounds, when Alice is active.
- analyse statistical differences between the rounds when Alice active and when not.
- q_α is a α% quantile of random variable X if $P(X \leq q_\alpha) = \alpha$.

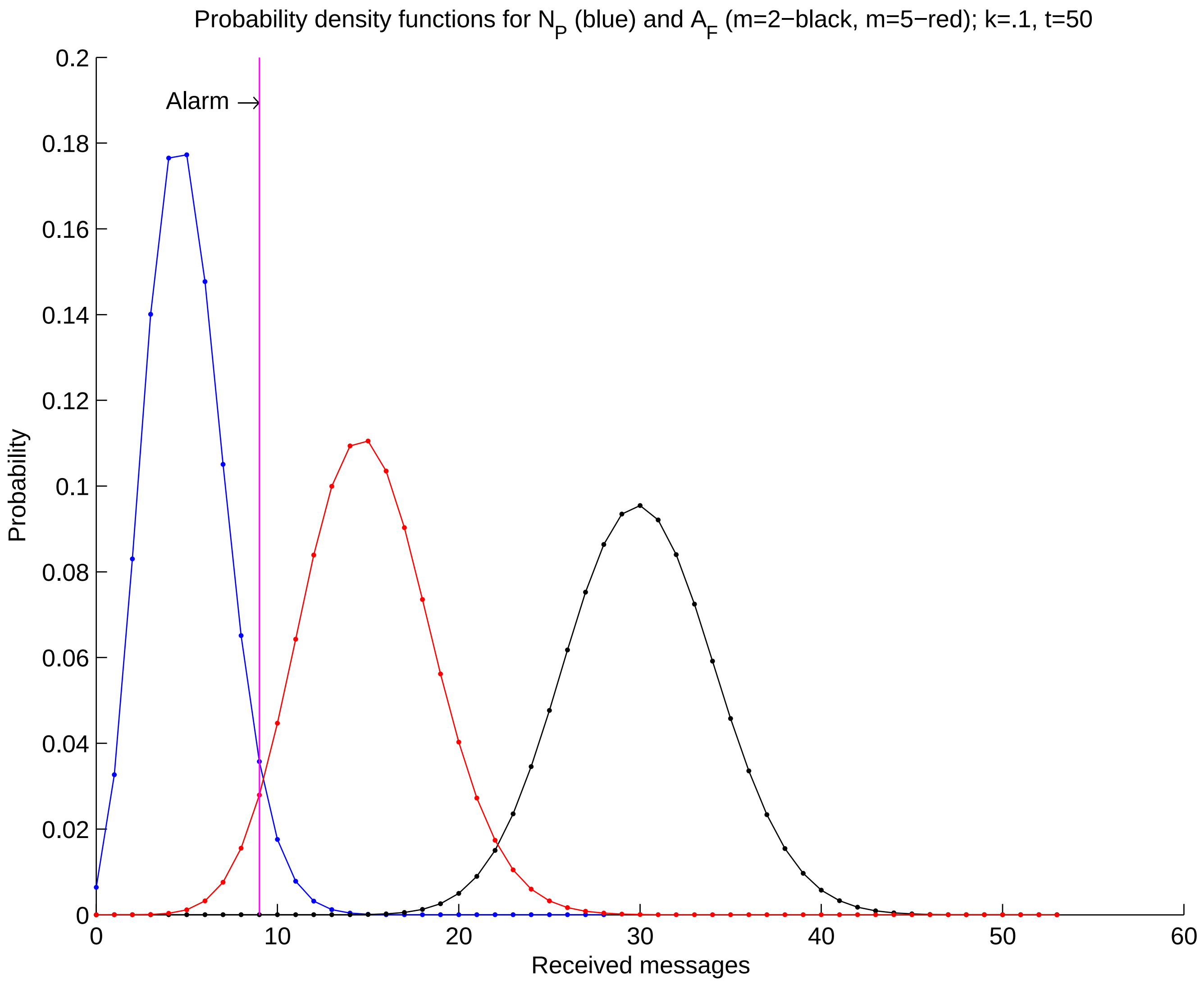
References

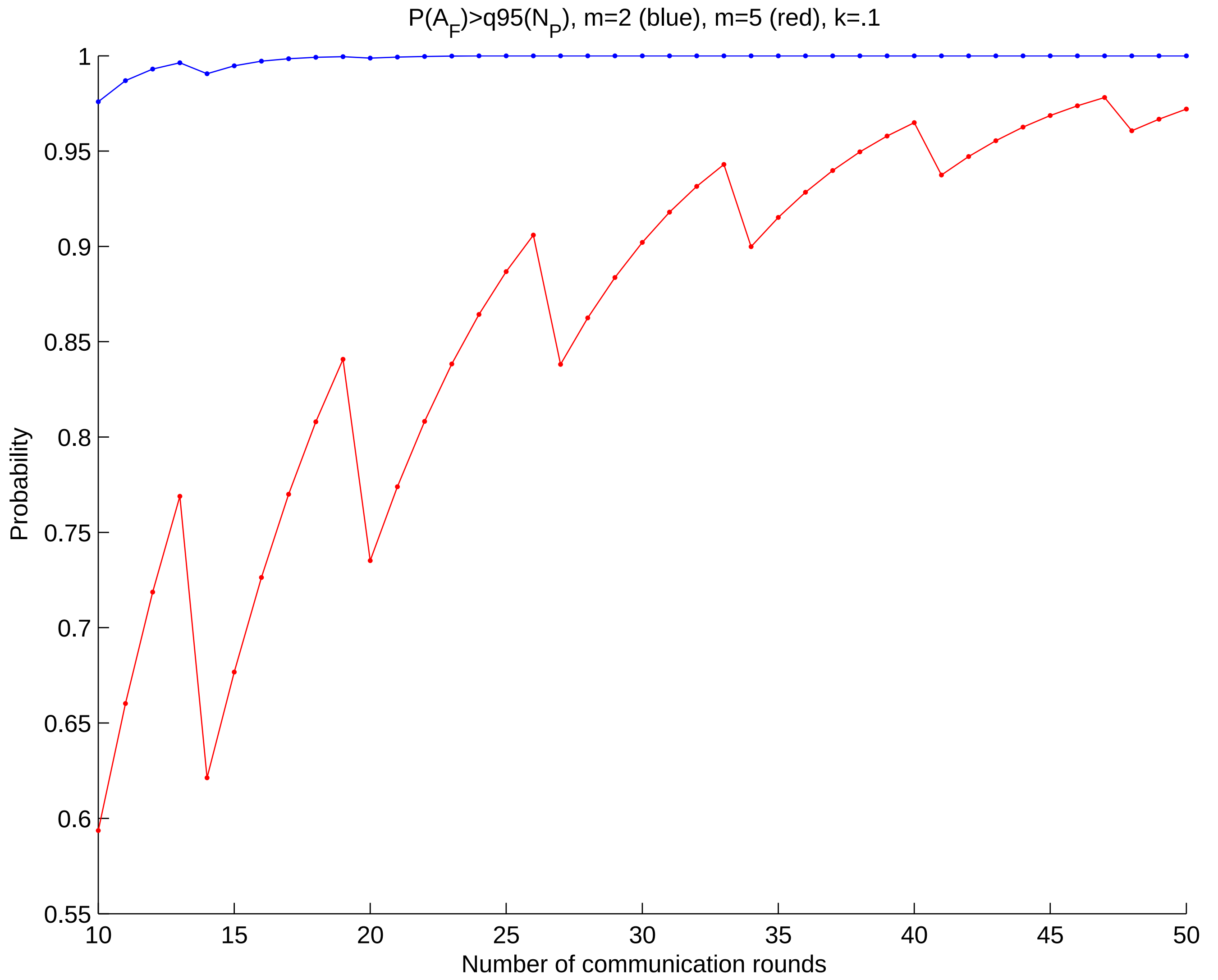
[1] Berthold, O., Pfitzmann, A., Standtke, R., “The disadvantages of free MIX routes and how to overcome them”, Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability,

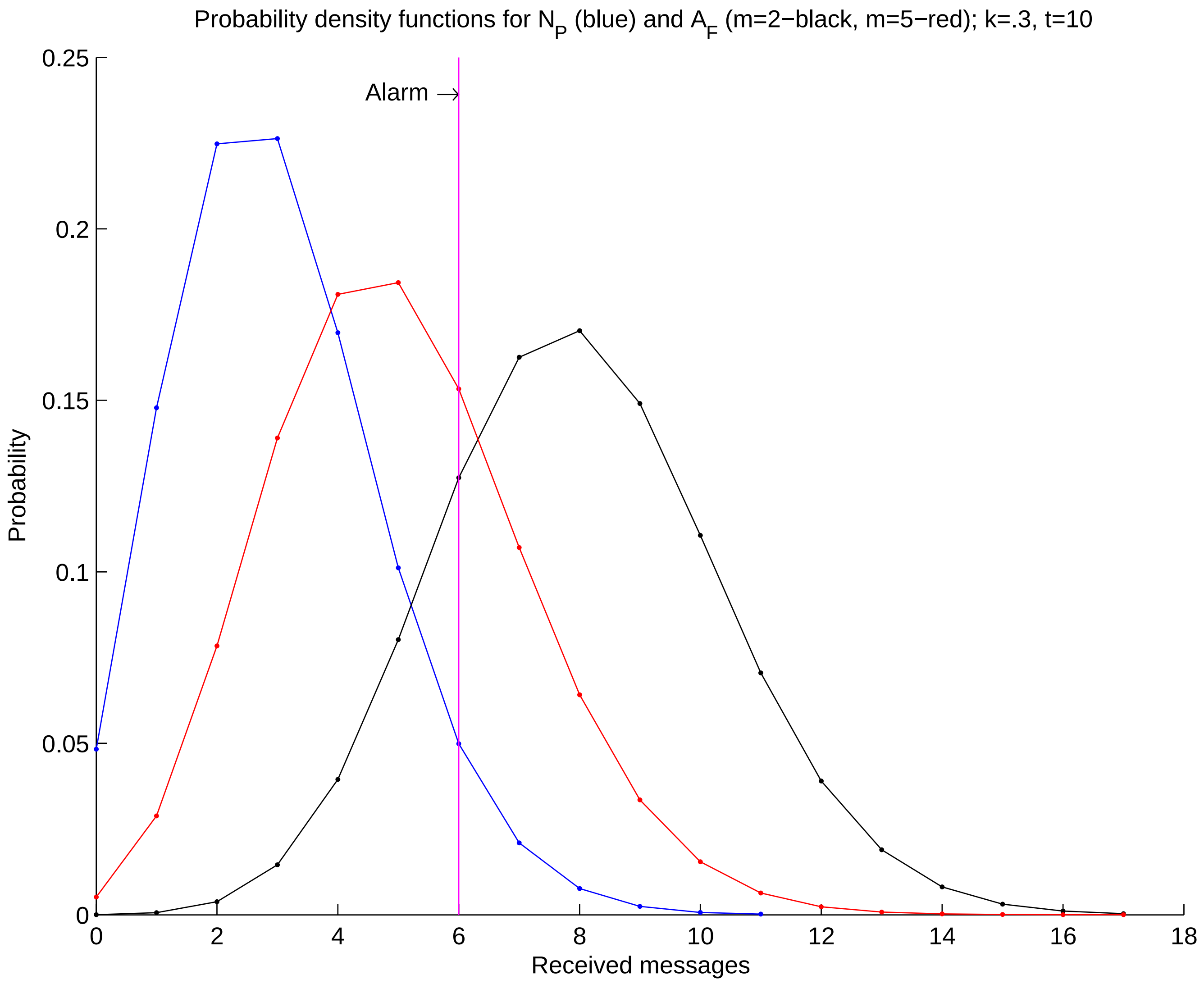
[2] Danezis, G., Serjantov, A.: “Statistical Disclosure or Intersection Attacks on Anonymity Systems”, Information Hiding’2004,

[3] Kesdogan, D., Agrawal, D., Penz, S.: “Limits of Anonymity in Open Environments”, Information Hiding’2002,


Probability density functions for N_P (blue) and A_F (m=2−black, m=5−red); k=.1, t=10
Alarm →
Probability
Received messages
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12
14

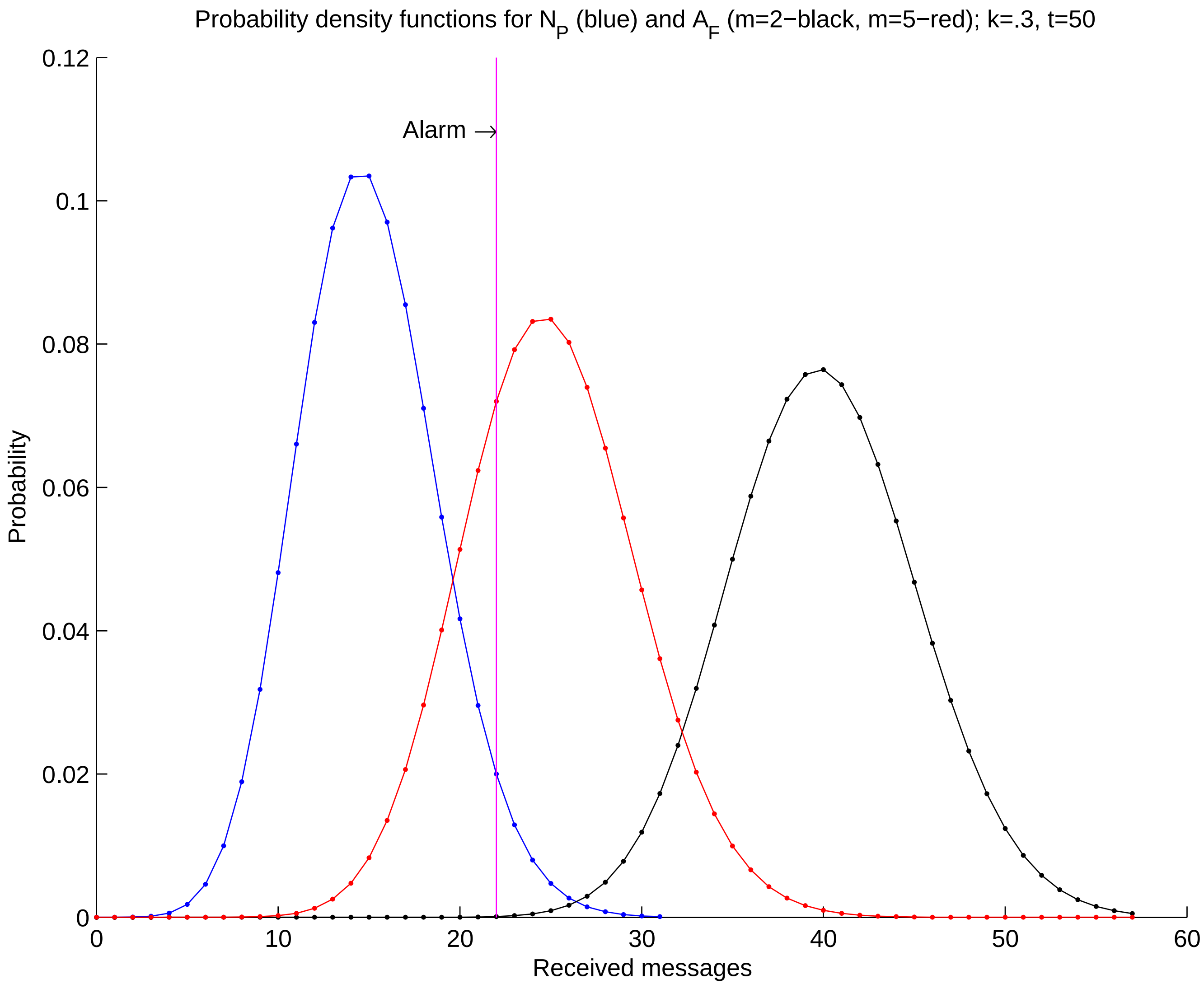

Probability density functions for N_P (blue) and A_F (m=2−black, m=5−red); k=.1, t=50
Alarm →
Probability
Received messages
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
40
50
60

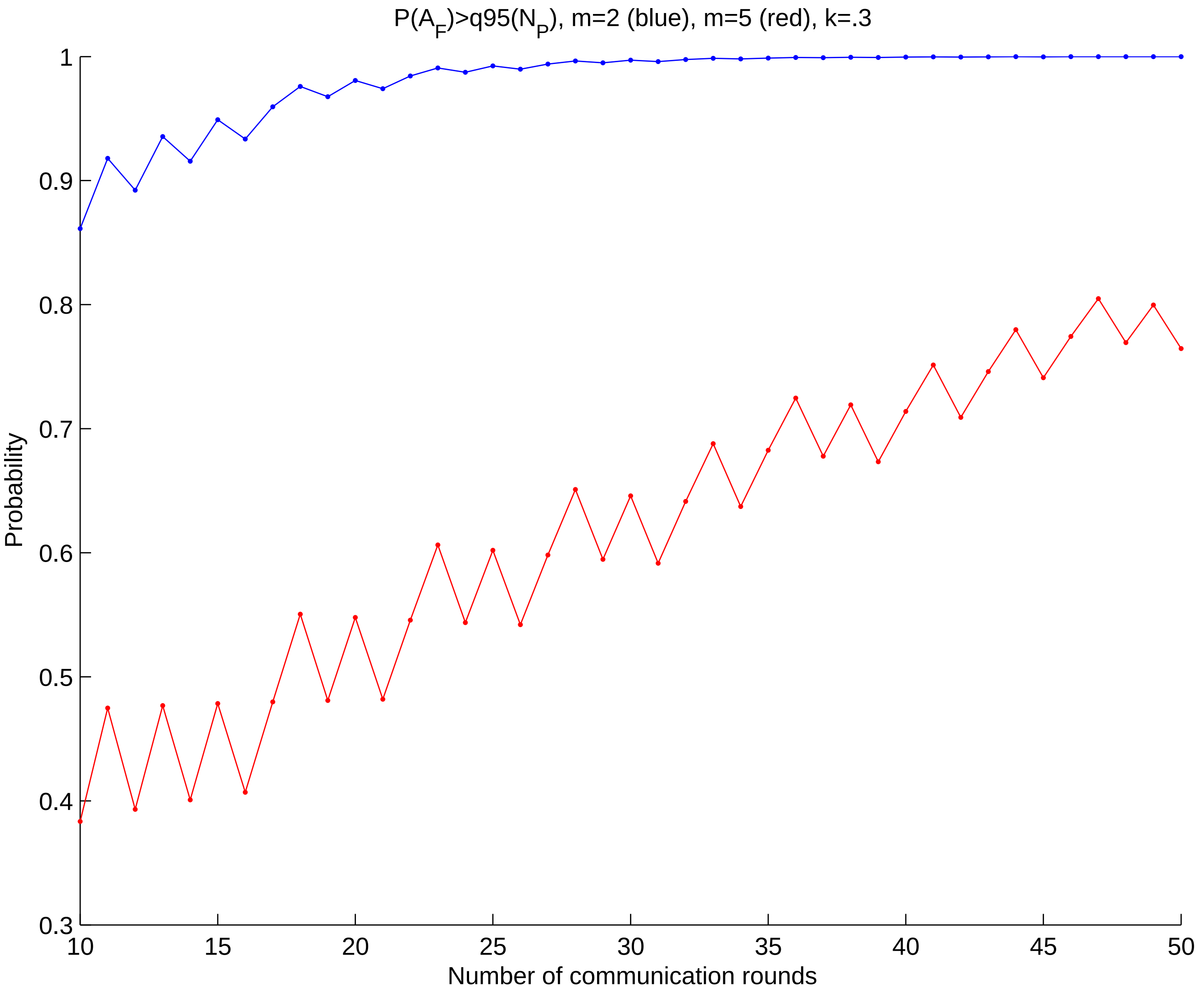

P(A_F)>q95(N_P), m=2 (blue), m=5 (red), k=.1
Probability
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
10
15
20
25
30
35
40
45
50
Number of communication rounds

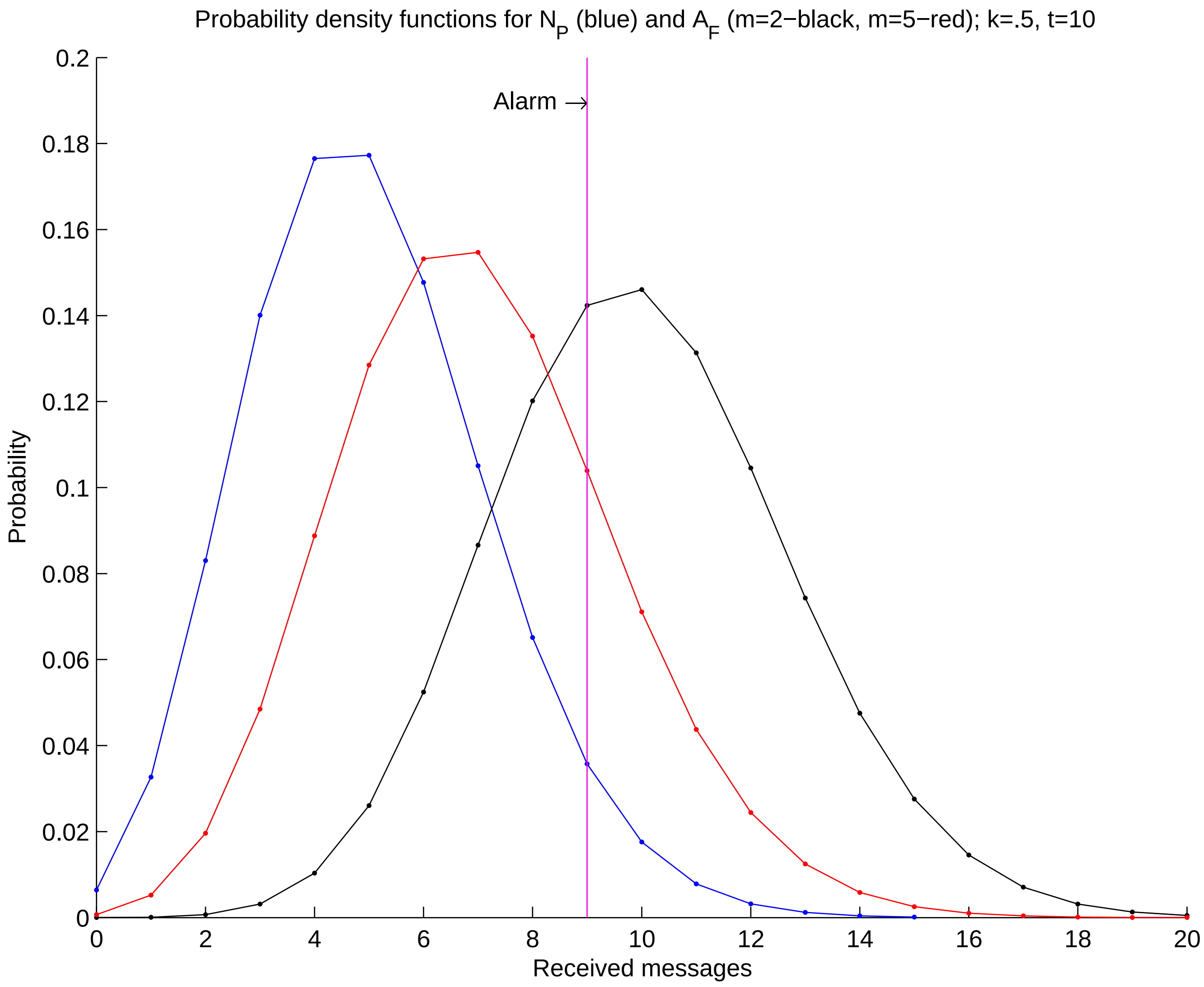

Probability density functions for N_P (blue) and A_F (m=2−black, m=5−red); k=.3, t=10
Alarm →
Probability
Received messages
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12
14
16
18

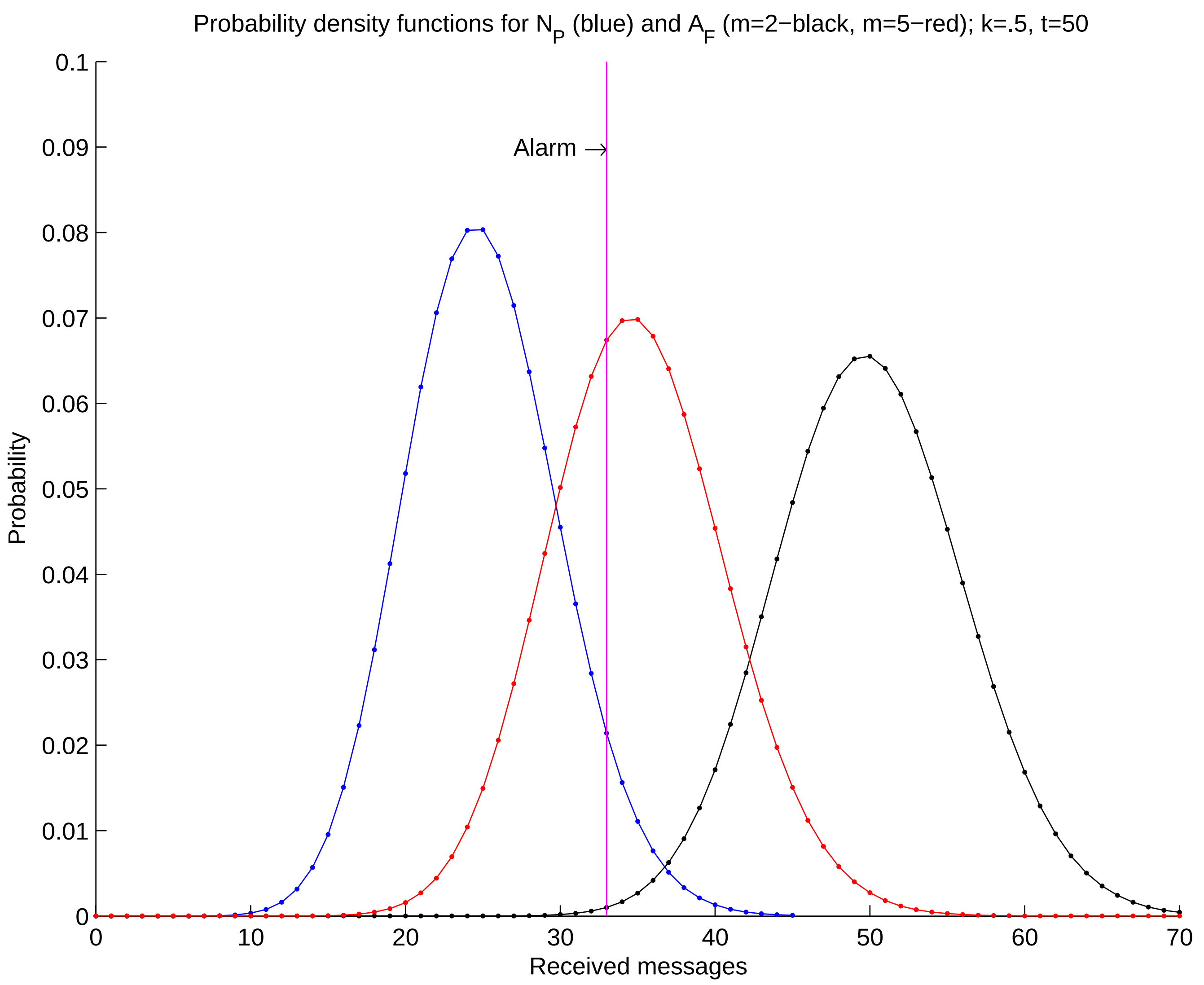
Intersection Attack and Using Dummy Addresses

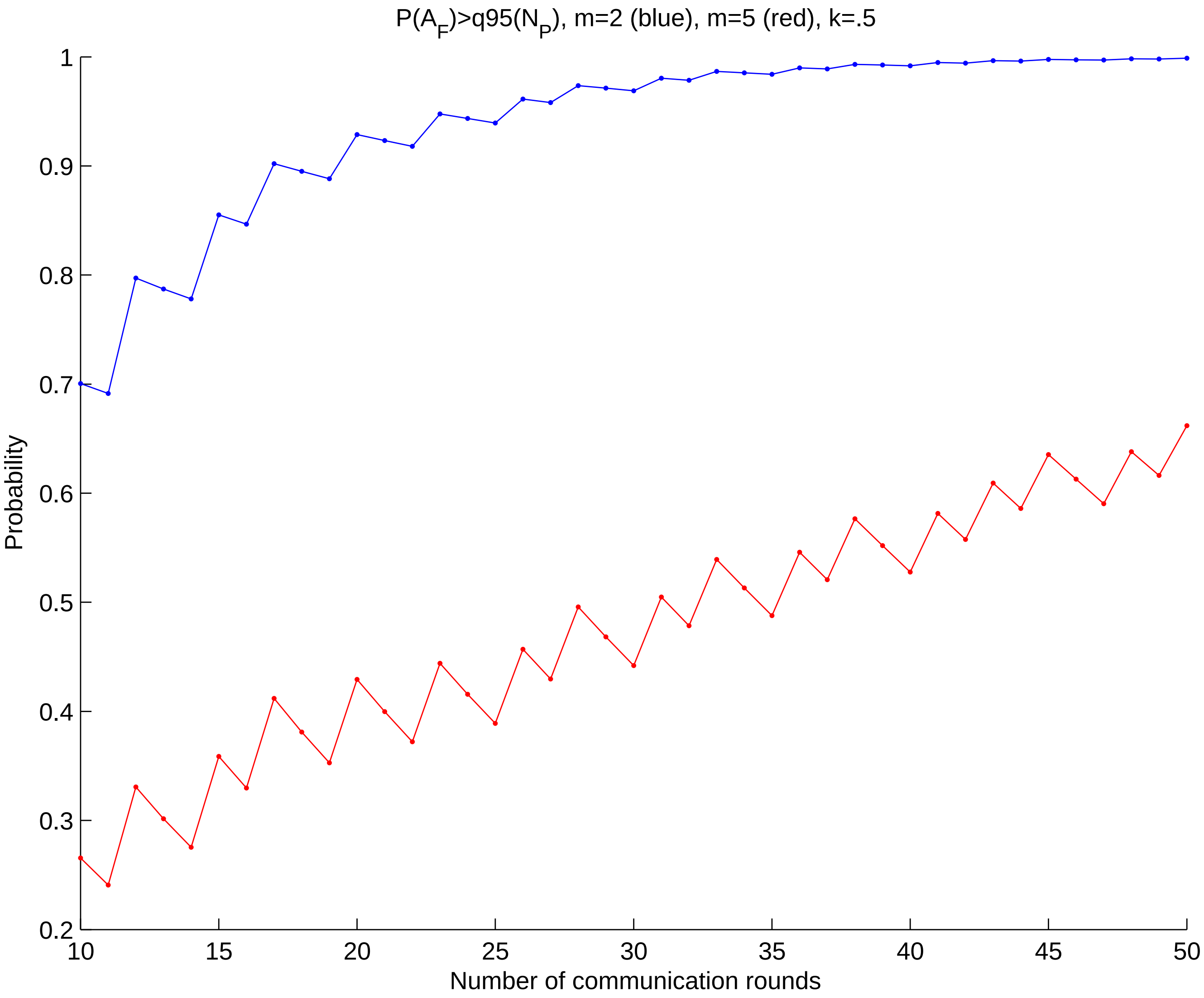
J. Kabarowski, M. Kutyłowski




P(A_F)>q95(N_P), m=2 (blue), m=5 (red), k=.3
Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
10
15
20
25
30
35
40
45
50
Number of communication rounds


Probability density functions for N_P (blue) and A_F (m=2−black, m=5−red); k=.5, t=10
Alarm →
Probability
Received messages
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
2
4
6
8
10
12
14
16
18
20


Probability density functions for N_P (blue) and A_F (m=2−black, m=5−red); k=.5, t=50
Alarm →
Probability
Received messages
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50
60
70


P(A_F)>q95(N_P), m=2 (blue), m=5 (red), k=.5
Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
10
15
20
25
30
35
40
45
50
Number of communication rounds

Defense strategy — dummy addresses

- Alice chooses at random a group of d dummy destinations,
- she devotes 50% of her communication bandwidth for dummy traffic.

The problem:

What is the best value for d, given k and m?

Let A_D be a random variable denoting the number of mesages received by Alice's dummy destination. Then

$$
\begin{aligned}
N_P &\sim B(t \cdot b, \frac{1}{N}), \\
A_F &\sim B(t \cdot b, \frac{1}{N}) + B(t, \frac{1}{2m}), \\
A_D &\sim B(t \cdot b, \frac{1}{N}) + B(t, \frac{1}{2d}).
\end{aligned}
$$

Role of dummy destinations

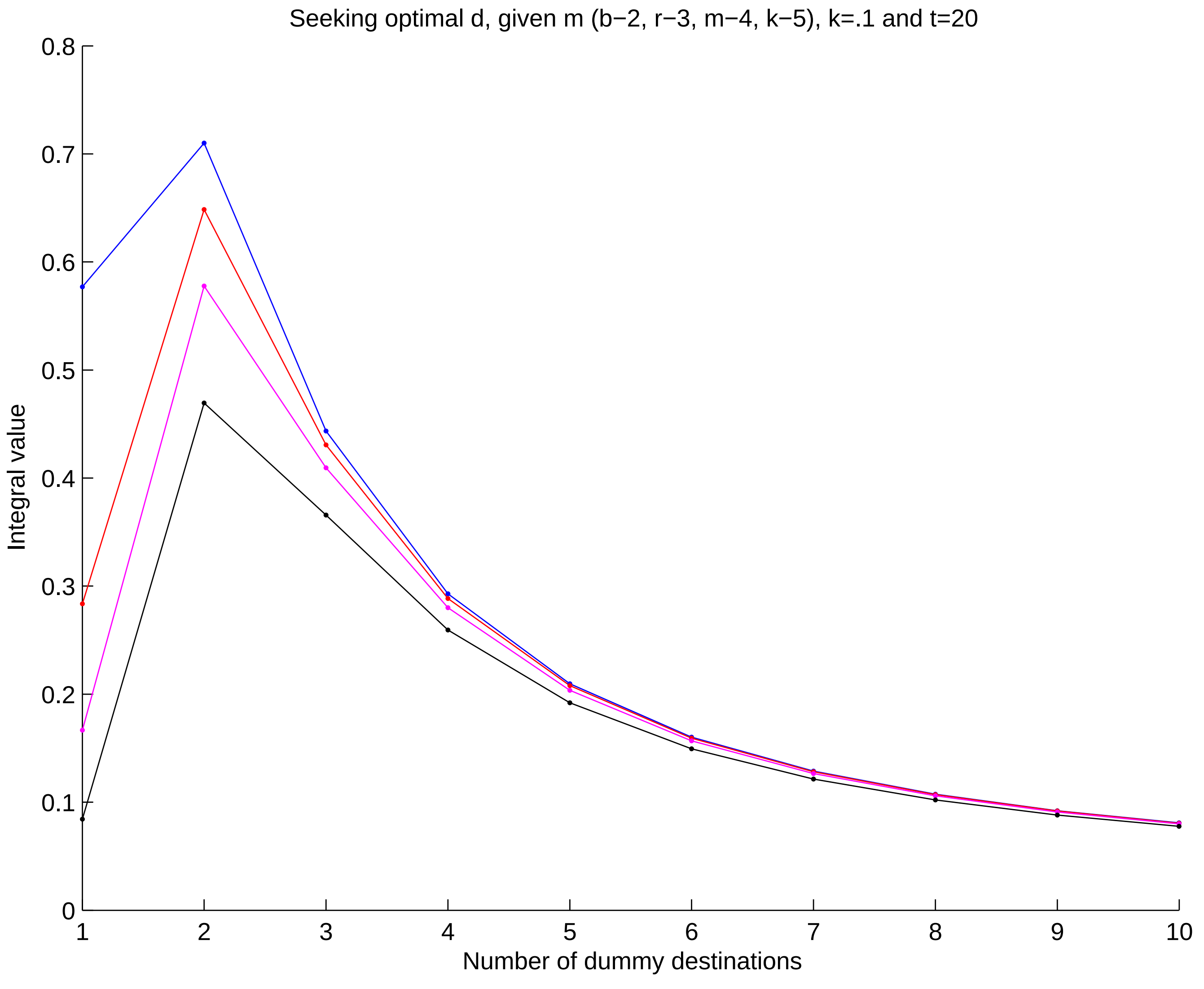
- Dummy should get a sufficient number of messages in order to become a suspect (therefore, $A_D \geq q_{95}(N_P)$),
- Dummy should receive similar amount of messages comparing to a real Alice's destination (therefore, $q_{05}(A_F) \leq A_D \leq q_{95}(A_F)$).

Role of dummy destinations

- Dummy should get a sufficient number of messages in order to become a suspect (therefore, $A_D \geq q_{95}(N_P)$),
- Dummy should receive similar amount of messages comparing to a real Alice's destination (therefore, $q_{05}(A_F) \leq A_D \leq q_{95}(A_F)$).

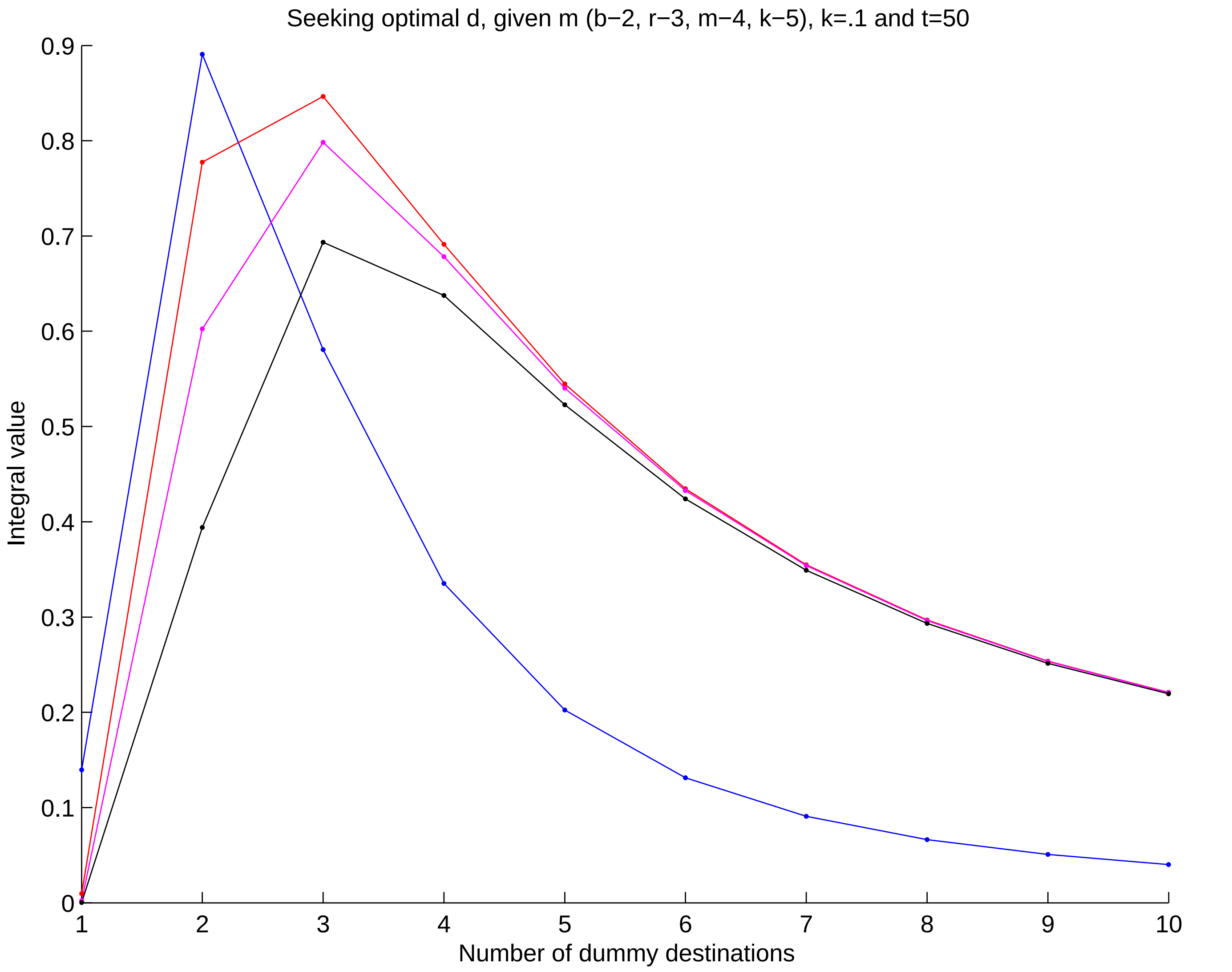
So we propose the following formula for the best choice for d:

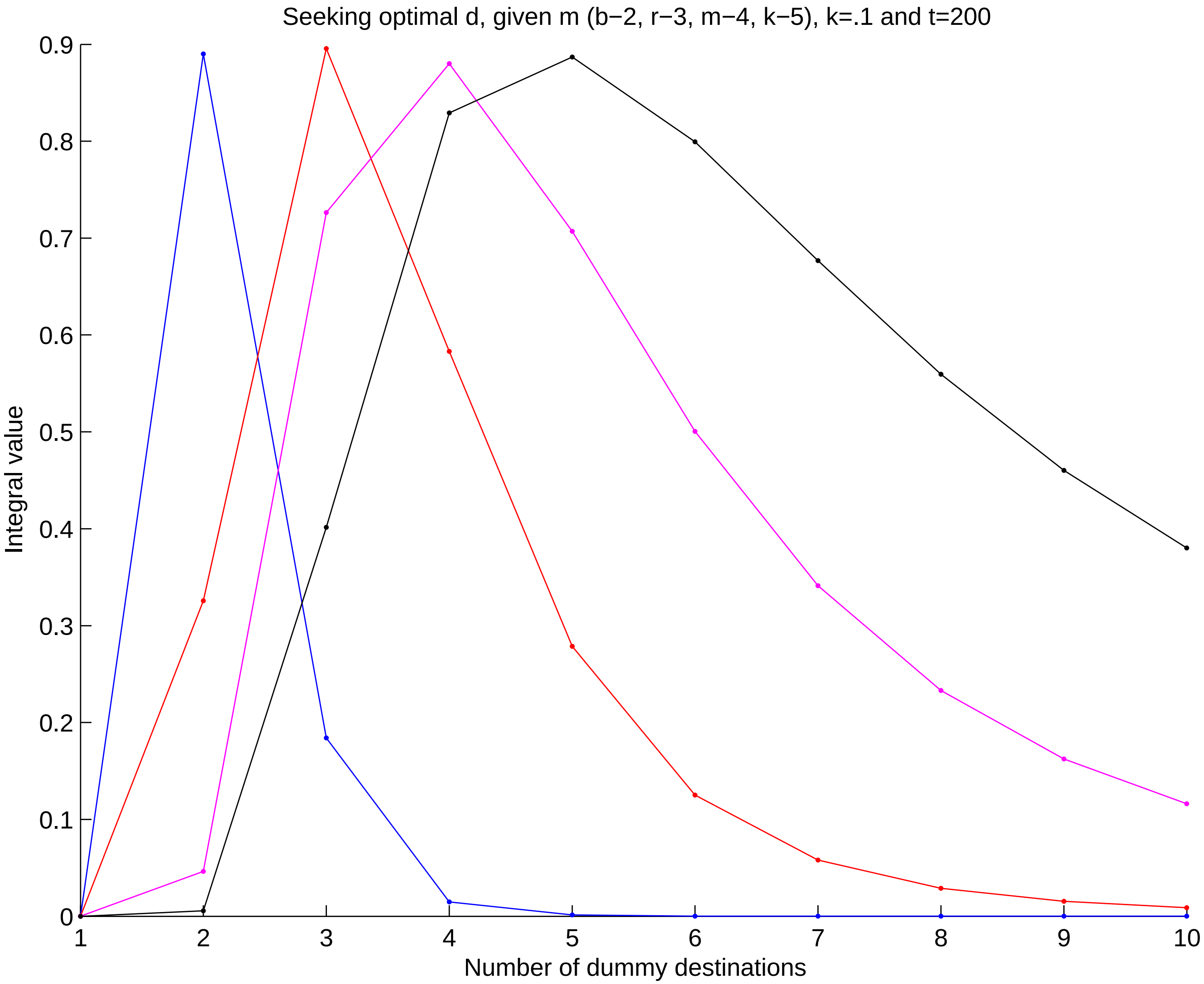
$$d_{best} = \max_{d} \int_{max(q_{05}(A_F),q_{95}(N_P))}^{q_{95}(A_F)} g_{A_D}\, dA_D,$$


Seeking optimal d, given m (b−2, r−3, m−4, k−5), k=.1 and t=20
Integral value
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
Number of dummy destinations

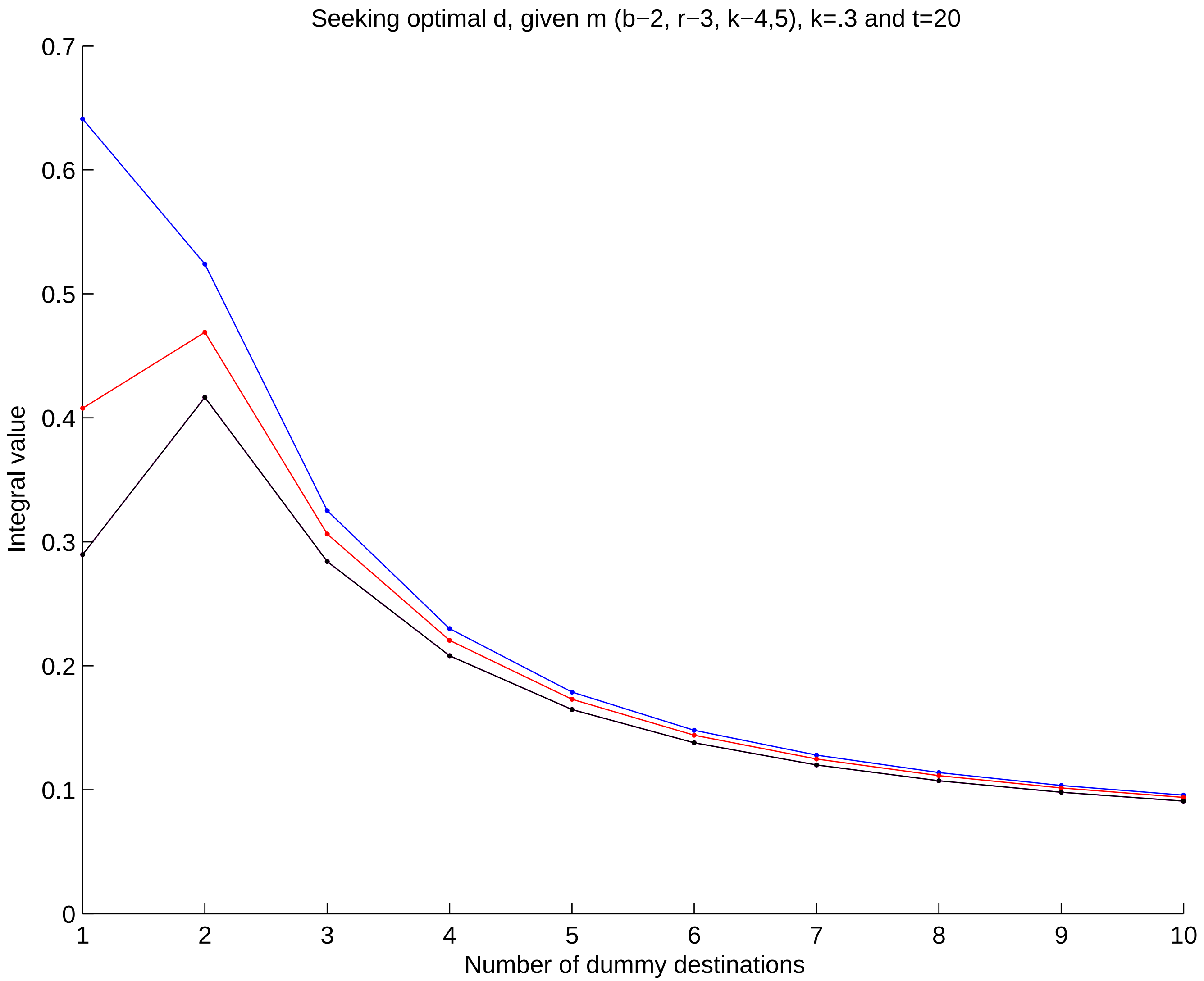
Intersection Attack and Using Dummy Addresses

J. Kabarowski, M. Kutyłowski

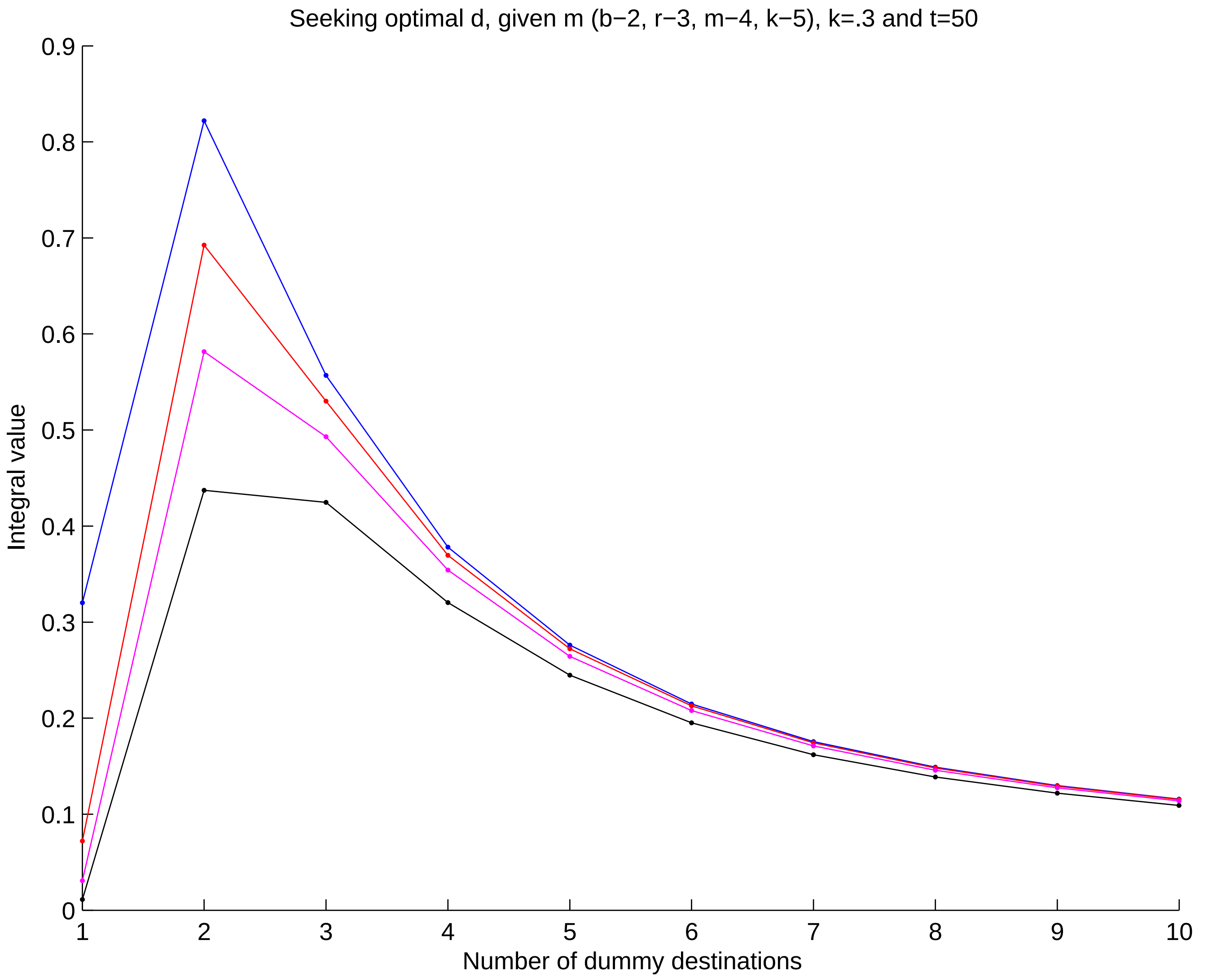


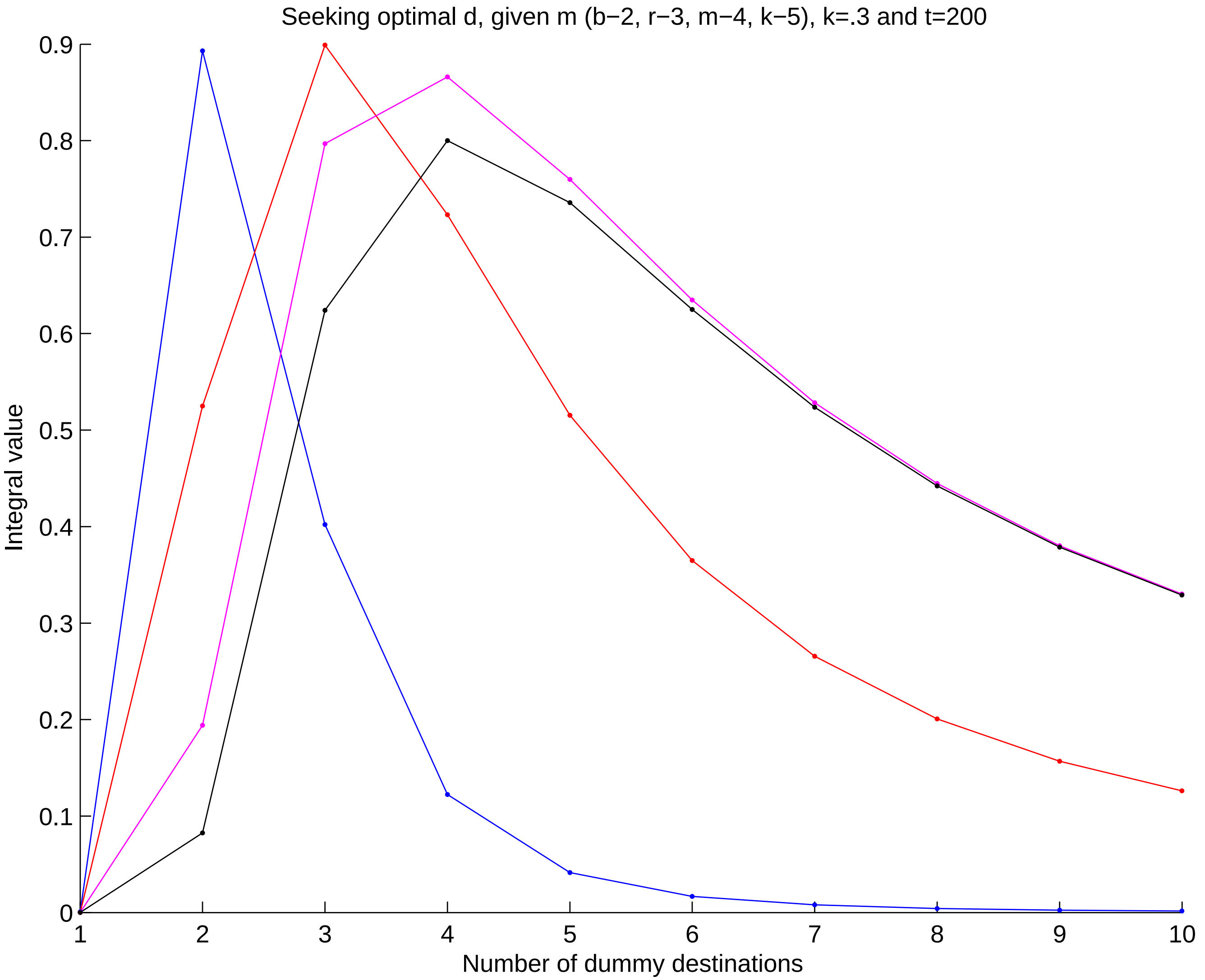


Seeking optimal d, given m (b−2, r−3, m−4, k−5), k=.1 and t=200
Integral value
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
3
4
5
6
7
8
9
10
Number of dummy destinations



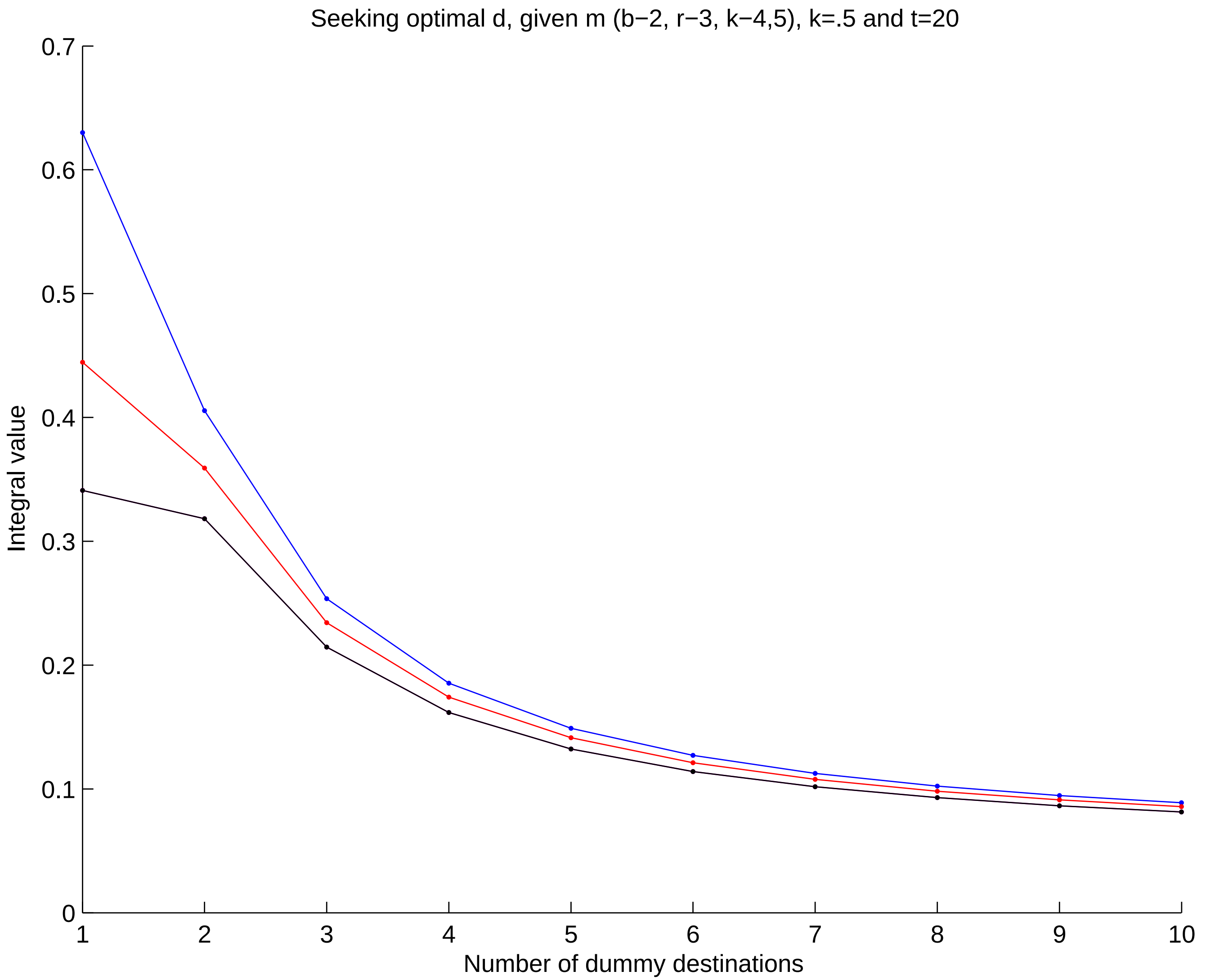
Seeking optimal d, given m (b−2, r−3, k−4,5), k=.3 and t=20
Integral value
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
2
3
4
5
6
7
8
9
10
Number of dummy destinations


Seeking optimal d, given m (b−2, r−3, m−4, k−5), k=.3 and t=50
Integral value
Number of dummy destinations
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
3
4
5
6
7
8
9
10


Seeking optimal d, given m (b−2, r−3, m−4, k−5), k=.3 and t=200
Integral value
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
Number of dummy destinations

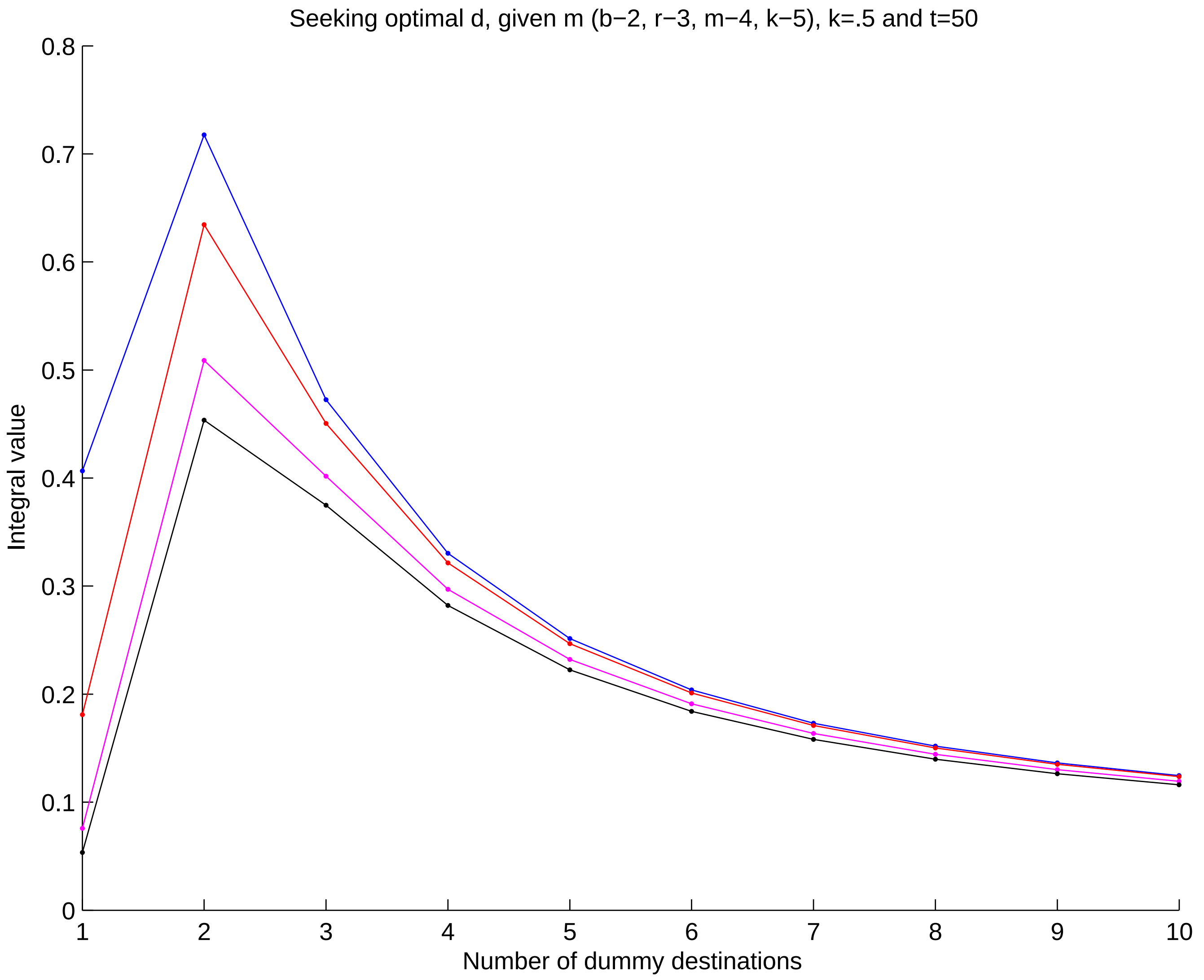
Intersection Attack and Using Dummy Addresses

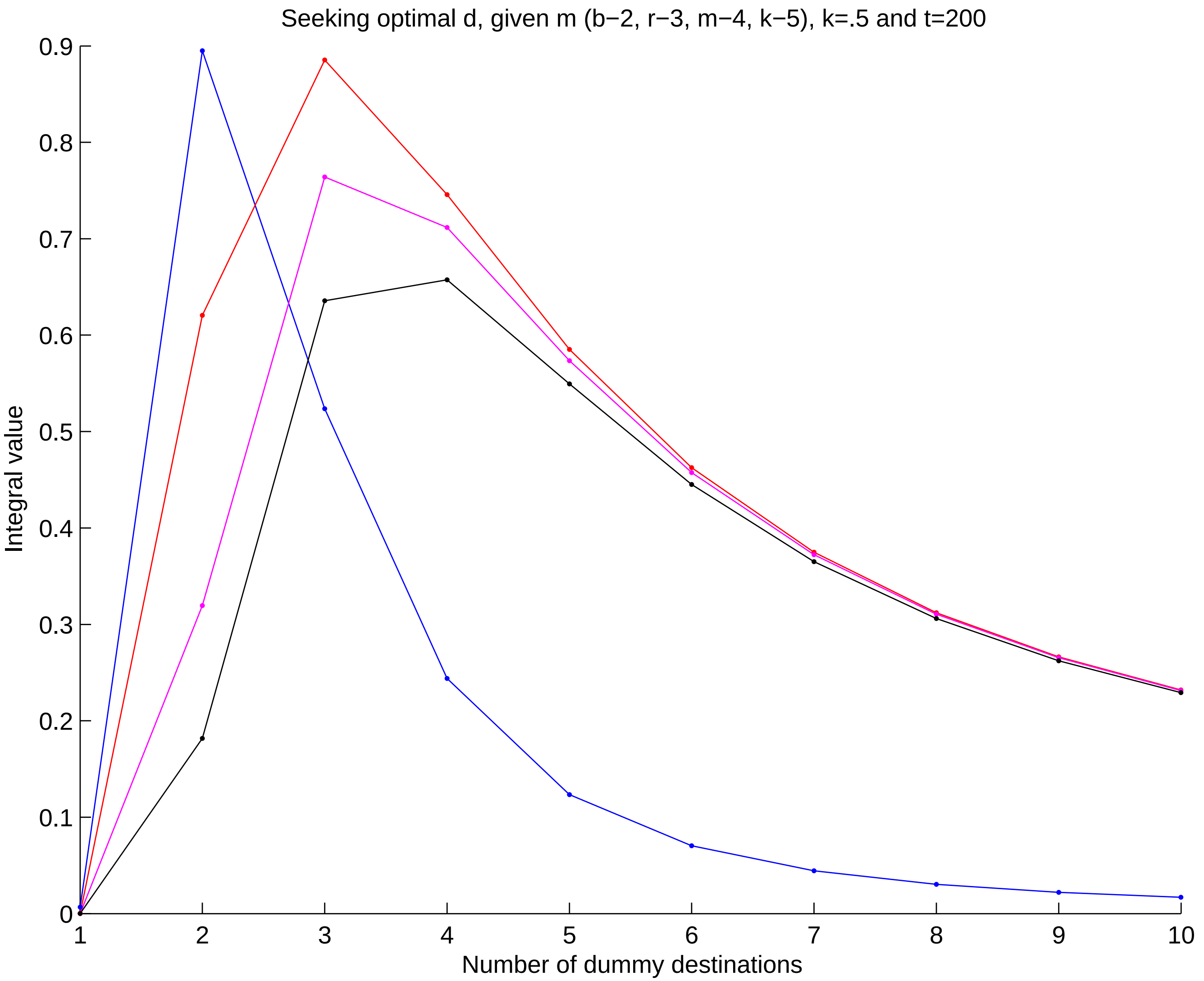
J. Kabarowski, M. Kutyłowski



Intersection Attack and Using Dummy Addresses

J. Kabarowski, M. Kutyłowski




Seeking optimal d, given m (b−2, r−3, m−4, k−5), k=.5 and t=200
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Integral value
1
2
3
4
5
6
7
8
9
10
Number of dummy destinations

Conclusions

- greater t makes Intersection Attack easier (as expected),
- higher communication (parameter k) slows down the attack (as expected),
- **if t is small, then d should be smaller than m**
  (a common belief is that $d = m$ is the optimal choice)
- when t grows, then the optimal choice of d converges to m.