Avoidable colors percolation: Increasing security with heterogeneous paths

Sebastian M. Krause and Vinko Zlatić
Theoretical Physics Division, Rudjer Bošković Institute, Zagreb, Croatia

Michael M. Danziger
Department of Physics, Bar Ilan University, Ramat Gan, Israel

I. INTRODUCTION

Secure communication in networks of servers and communication lines is possible even if a part of servers or lines is faulty. An algorithm was presented already in the 1990s, using sets of paths with disjunct servers [1]. This early study, which gained broad attention in computer science [...], abstracted from the network structure and assumed the existence of the paths a priori. If node and link failure occurs with a given probability, percolation theory on complex networks can be used to determine overall connectivity robustness [2, 3]. In particular, k-core percolation [4] has interesting implications due to the special k-core structure of the AS server network of the Internet [5, 6]. The possibility of secure communication was studied as well for wireless networks using percolation on spatially embedded graphs [7]. However, recent security problems where often due to gaps in the software, and therefore whole sets of servers will likely fault at the same time, if they use the same software version of the OS, transmission software etc.

Here we analyze how the secret sharing method can be used to send messages in a secret way, even if one software version is faulty and it is not known which one. We develop a new kind of percolation theory, where paths avoiding every software version must exist at the same time. For the AS network we find that secure communication is impossible, if there is no heterogeneity of software versions on the highly connected servers.

II. AVOIDABLE COLORS PERCOLATION

Assume a graph G with N vertices and adjacency matrix A_{ij}. Every vertex i has a color $c_i \in \{1, 2, \ldots, C\}$, where C denotes the total number of colors. The colors may stand for software versions on servers, where all servers of the same version are likely to fail at the same time; it may stand for economic entities with correlated failure probability (due to financial dependence, reliance on the same resource); or it may stand for reloading points of transportation (e.g. ports with transferring goods from ship to train, where strikes could hit many ports at the same time). Faced with the possible collective failure of all nodes with one color, two nodes desire to be "connected with avoidable colors": For every color c a connecting path must exist, with all nodes in between not having color c. This is illustrated on the left of figure 1.

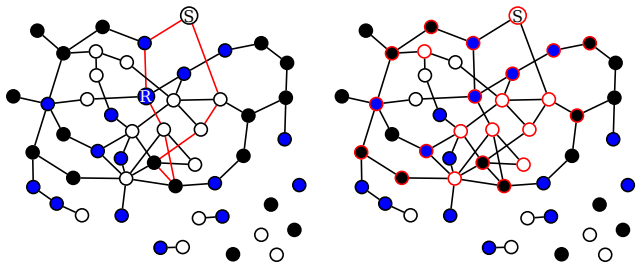

FIG. 1: Left: In this network the sender S and the receiver R can communicate with avoidable colors, as the short path highlighted with red avoids black and white nodes, and the long path avoids blue nodes. Right: All nodes highlighted with red belong to an avoidable colors component, as each pair out of this set is connected with avoidable colors. Notice that some nodes which are needed for connection of other nodes are not included in the component.

In order to characterize noteworthy parts of the network instead of single pairs, we define an "avoidable colors component" as a maximal set of nodes, where every node pair is connected with avoidable colors. Such a component is highlighted with red in the figure on the right. Notice that there are nodes needed for providing connections which themselves not belong to the avoidable colors component. Labeling avoidable colors components can be useful for different tasks: 1) The fraction of nodes being in the largest avoidable colors component tells us whether it would be useful to implement real world routing algorithms (if only a tiny fraction of servers in the Internet is able for avoiding colors, it is not worth of thinking about new protocols). 2) By highlighting the nodes which are connected with many others with avoidable colors, routing algorithms can save time by not searching for impossible connections. 3) The positive effect of adding links, moving colors or increasing color heterogeneity can be quantified and balanced with possible costs or regulation issues.

As illustrated in figure 2, there is a way to find a candidate set of nodes L_{color} for the largest avoidable colors component: 1) For every color c, we delete all nodes with color c and find the largest component in the remaining graph, lets call this set of node as $L_{\bar{c}}$. 2) A node belongs to L_{color}, if it has for every color c at least one link to $L_{\bar{c}}$

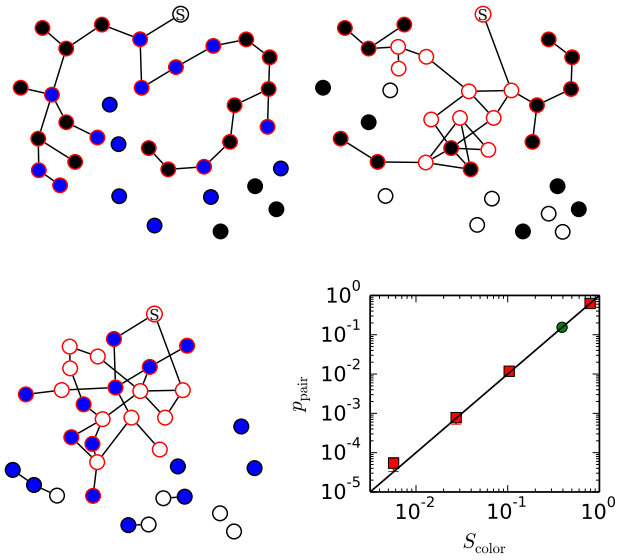

FIG. 2: Illustration of the construction of the set of nodes L_{color} which is in many cases of large networks the largest avoidable colors component. The largest components without white ($L_{\bar{1}}$), without blue ($L_{\bar{2}}$) and without black nodes ($L_{\bar{3}}$) are highlighted in red, and the test node S is connected to all of them and therefore belongs to L_{color}. Lower right: Estimation of the fraction of successful pairs for quenched graphs with different values $S_{\text{color}} = N_{\text{color}}/N$. Red squares show Poisson graphs with increasing degree and $C = 3$ colors, the green circle shows the autonomous systems network with $C = 2$ colors. The black line indicates the case where only node pairs in L_{color} are connected with avoidable colors. As numerical results are close, L_{color} indeed dominates the secure communication abilities of many graphs.

(or belongs to it). It is easy to see that every node pair in L_{color} can communicate with avoidable colors. If L_{color} includes for every color c at least one node out of $L_{\bar{c}}$, it is maximal and therefore it is an avoidable colors component. There is no easy way to test whether L_{color} is the largest avoidable colors component (as shown in figure 3, avoidable colors components can exist due to different mechanisms and they can largely overlap). However, we will see that L_{color} might scale with system size and in this case can be considered as a giant avoidable colors component. If the number of nodes in L_{color} is N_{color}, then at least $N_{\text{color}}(N_{\text{color}} - 1)/2$ out of all $N(N-1)/2$ possible node pairs in the network are connected with avoidable colors. This is a macroscopic fraction if L_{color} scales linear with system size. We can use this fact to test whether L_{color} dominates the secure communication abilities of a network by plotting the fraction of pairs connected with avoidable colors in the whole network p_{pair} against $S_{\text{color}} = N_{\text{color}}/N$. In figure 2 on the lower right we see that secure connectivity is indeed dominated by L_{color}. With red squares, results for Poisson graphs with $N = 10^5$ nodes and average degrees $\bar{k} = 1.6;\ 1.7; 1.9; 4.0$ are shown, where $C = 3$ colors where distributed over the nodes with the same probability of 1/3 for every color. Results fit well even for small S_{color}. This is important for the critical behavior, as will be discussed below. Notice that even for the smallest value shown, $N_{\text{color}} = 570$ has reasonable size. The blue circle shows the network of autonomous systems with two colors distributed over the nodes. Our network snapshot of the year 2006 contains $N = 22963$ nodes. p_{pair} was approximated with samples of up to 5×10^5 pairs, error-bars are smaller than the symbols in most of the cases. As we have seen that L_{color} may describe random networks and the real world example of autonomous systems, we will base a theory for calculating the size of the giant avoidable colors component on that idea.

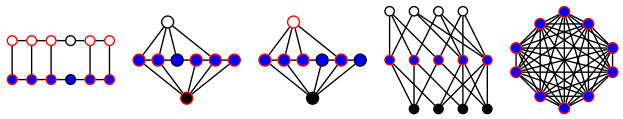

FIG. 3: Avoidable colors components, as highlighted with red, can be due to different scenarios. On the left, we see a scalable case similar to random graphs. In the second graph, the high degree black node serves as an alternative paths provider for the blue nodes. In the third graph an alternative avoidable colors component is highlighted for that graph, showing that components might overlap. The second graph from the right does not need any connection among the blue nodes, however, there is a massive overhead of nodes and connections. On the right, we see that a clique is an avoidable colors component.

To illustrate the possibly rich phenomenology of avoidable colors components, in figure 3 some different mechanisms are shown which establish such components. On the left, we see a case which is similar to random graphs: All nodes which are neighbors to largest components without the color white and without the color blue can connect securely. In the second graph, the black node serves as an alternative paths provider for the blue nodes. It needs to have high degree for that. In the third graph an alternative avoidable colors component is highlighted. This shows that they might overlap and it is not straight forward to find the largest one. The second graph from the right does not need any connection among the blue nodes and the connecting white and black nodes have lower degree, however, there is a massive overhead of nodes and connections. On the right, we see a clique. In this case, no node of a different color is needed, but instead a high number of links.

III. RESULTS

We can find analytical results for random graph ensembles with randomly distributed colors in the limit of infinite graphs. The detailed derivation and explanation can be found in the supplements. We use the general-

ized configuration model graph ensemble with N nodes, where each degree sequences $\{k_i\}$ occurs with probability $\prod_i p_{k_i}$ with the degree distribution p_k. Additionally we want to assign to every node i a color $c_i \in 1, 2, \ldots, C$. For given degree sequence k_i, the color sequence $\{c_i\}$ has probability $\prod_i \tilde{r}_{c_i,k_i}$ with the degree-dependent color distribution $\tilde{r}_{c,k}$ ($\sum_c \tilde{r}_{c,k} = 1$ for every degree k separately).

We calculate S_{color} in the limit of $N \to \infty$ as the probability of a single node to belong to L_{color}. This problem can be decomposed into two parts. First, all possible cases of neighborhoods are summed over with the according probabilities. We need to specify the numbers κ_c for all colors c. κ_c counts the number of neighbors having color c and at the same time being in the normal giant component. Second, with the vector $\vec{\kappa} = (\kappa_1, \ldots, \kappa_C)$, the conditional probability $P_{\vec{\kappa}}$ that these links suffice to connect to L_{color} can be calculated. This can be realized as follows:

$$S_{\text{color}} = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \sum_{\kappa_1,\ldots,\kappa_C=0}^{k'} M_{k',\vec{\kappa}} P_{\vec{\kappa}}, \qquad (1)$$

$$P_{\vec{\kappa}} = \prod_{c=1}^{C} [1 - (U_{\bar{c}})^{\sum_{c' \neq c} \kappa_{c'}}]. \qquad (2)$$

The binomial factor $B_{k,k'}$ (equation S...) accounts for the probability that out of k links k' links connect to the normal giant component. The multinomial factor $M_{k',\vec{\kappa}}$ (equation S...) gives the multinomial probability of having the color distribution $\vec{\kappa}$ among the neighbors belonging to the normal giant component. In the second equation, $U_{\bar{c}}$ (equation S...) gives the conditional probability that a link fails to connect to $L_{\bar{c}}$, while this link fulfills the precondition of connecting to a node in the normal giant component having a color $c' \neq c$. Using $U_{\bar{c}}$, $P_{\vec{\kappa}}$ is the probability that for every color c at least one link succeeds to connect to $L_{\bar{c}}$

In figure 4 on the top, the dependence of S_{color} on the average degree is shown for Poisson graphs with different numbers of colors C ($\tilde{r}_{c,k} = 1/C$ is chosen homogeneous and independent of k). Comparing to the standard giant component size S (dashed black line in the figure), the percolation sets in at increasing $\bar{k}$ with smaller numbers of colors, and the component size grows slower to the saturation value of one. The symbols show numerical results with $N = 1000$ and 100 network realizations, the lines show results of equation 1, both correspond well.

The suppression of the number of securely connected nodes can be understood as a combination of two effects. The first effect is purely topological and can be understood with $S_{\text{color},\infty}$ of eq. S**??** (shown with dotted red line). It means that only nodes can belong to S_{color}, which are connected to the normal giant component over at least two links. We can confirm that S_{color} comes close to $S_{\text{color},\infty}$ for high numbers of colors C with the results for $C = 10$. $S_{\text{color},\infty}$ is remarkably reduced compared to S for small k, but has the same critical parameter. For the Poisson graph we show in the supplements that

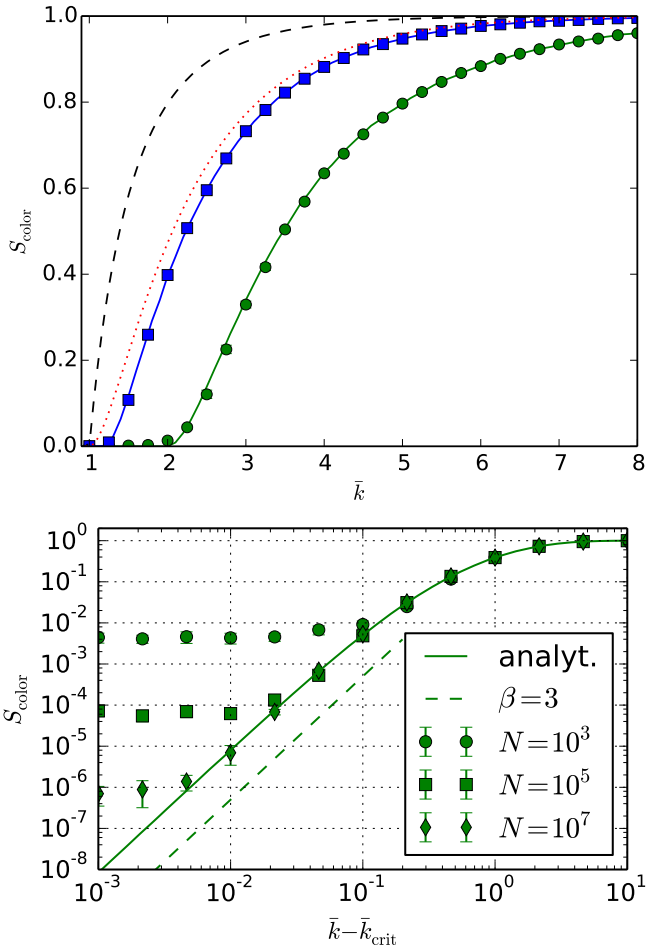

FIG. 4: Top: Dependence of S_{color} on the average degree for Poisson graphs with different numbers of colors. Symbols show numerical results for networks of size $N = 1000$ (blue squares for $C = 10$ and green circles for $C = 2$ colors), the straight lines show the according analytical results. For comparison, the giant component size S is shown (black dashed). S_{color} is reduced due to two mechanisms: First, every node has to be connected to the giant component via two links. The according fraction of nodes $S_{\text{color},\infty}$ is shown with a red dotted line. Second, increasing color frequencies further decrease S_{color}. Bottom: Finite size scaling for $C = 3$ colors emphasizes the dependence of the critical exponent β on the color distribution, here $\beta = C = 3$.

$S_{\text{color},\infty} \propto (\bar{k} - 1)^2$ which grows slowly for small parameter $\bar{k} - 1$.

The second effect is connected to finite color frequencies $\tilde{r}_{c,k}$ which further reduces the percolating fraction of nodes. This also changes the critical value $\bar{k}_{\text{crit}}$ and the critical exponent β. The critical behavior is discussed in detail in the supplements, for the general case of heterogeneous color distributions. Approximations in equation 1 allow us to understand the critical behavior. Applied to the homogeneous color distributions discussed

here, the results reduce to

$$S_{\text{color}} \propto (\bar{k} - \bar{k}_{\text{crit}})^{\beta} \tag{3}$$

$$\beta = C, \quad \bar{k}_{\text{crit}} = C/(C-1). \tag{4}$$

This behavior can be confirmed with numerical results. On the bottom of the figure, a finite size scaling for $C = 3$ colors establishes both the critical value of 3/2 and the critical parameter. For large numbers of colors, we observe the interesting phenomenon of high critical exponent β. With this, the system shows an effectively shifted transition between vanishing and finite S_{color}, as the growth of the giant avoidable colors component $\frac{\mathrm{d}}{\mathrm{d}\bar{k}} S_{\text{color}} \propto \beta(\bar{k} - \bar{k}_{\text{crit}})^{\beta-1}$ is close to zero for small arguments. To our knowledge, this is a new kind of behavior, and a more detailed analysis of other quantities at the phase transition seems promising.

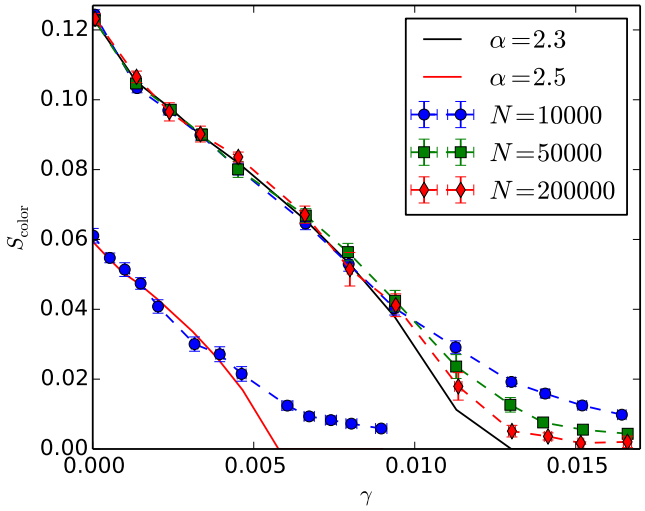

FIG. 5: On graphs with broad degree distribution, S_{color} drops fast, if a fraction γ of nodes with the highest degree is restricted to one color, while the nodes with smaller degree can have one of two colors. This is shown for networks with $\alpha = 2.3$ and $\alpha = 2.5$ having $N = 10000$ with symbols. Analytical results need the modified version of equation 1 as described in section **??**. Results are shown with straight lines. Without diversity on the hubs, nodes cannot communicate in the desired way.

Lets now discuss graphs with broad degree distributions with $p_k = nk^{-\alpha}$. n is a normalization constant. For broad degree distributions, color distributions can show an additional type of heterogeneity, as a dependence of frequencies on the degree of a node can strongly influence the behavior. We used two colors, where the first color has a frequency of $\tilde{r}_{1,k} = 1$ for all degrees $k \geq k_{\text{step}}$ larger than a certain k_{step}. These nodes have a probability of $\gamma = \sum_{k=k_{\text{step}}}^{\infty} p_k$. Accordingly $\tilde{r}_{2,k} = 0$ for $k \geq k_{\text{step}}$, and probabilities for smaller degrees are chosen as $\tilde{r}_{c,k} = 1/2$. Figure 5 shows results for an ensemble with $\alpha = 2.3$ and $\alpha = 2.5$. The analytical results for $\alpha = 2.3$ show that already for a portion of $\gamma = 1.4\%$ of the largest nodes occupied by the first color exclusively, S_{color} vanishes.

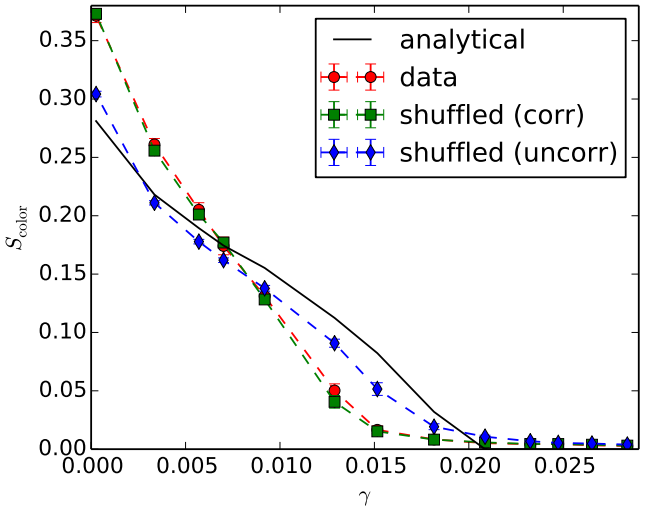

FIG. 6: Red circles show results for the network of autonomous systems, where colors are distributed in the same way as in figure 5. Averages were taken over 10 color distributions. Analytical results are shown with the black line and reproduce the results qualitatively. Deviations are due to degree-degree-correlations which are reserved in shuffled networks shown with green squares, while results with ignoring correlations are shown with blue diamonds.

The red circles in figure 6 show results for the autonomous systems network, where colors where distributed with degree-dependence over the nodes as described at the end of the last section. Averages where taken over 10 realizations of the color distributions. As expected from our results for scale free degree distributions, S_{color} drops to 0 even for small fraction γ which is exclusively of one color. That means that if there is no heterogeneity in the highly connected servers, it is not possible to avoid e.g. software versions. This is also interesting in the following sense: It is known that secret services try to store all decrypted data running through servers to decrypt it later. As this is connected to technical afford, the services will more likely monitor the large servers. Therefore it would be beneficial, if once using encryption, to sent parts of the message using small servers. Unfortunately, this seems to be impossible, the services only have to monitor a low percentage of servers to hinder alternative paths.

In order to assess the predictive power of our analytical method, we used a model ensemble with using the degree frequencies of the autonomous systems network as degree distribution p_k. Results for the according ensemble are shown with the black line. The qualitative behavior is represented well. To understand deviations, we compared to data from shuffled networks starting with the original data. Shuffling with ignoring degree-degree-correlations while only keeping the degree sequence gives results close

to the analytical results (blue diamonds). Shuffling with also keeping degree-degree-correlations gives results close to the original network (green squares). Therefore, deviations between our theory and the data arise mainly due to degree-degree-correlations.

IV. SUMMARY AND OUTLOOK

[1] D. Dolev, C. Dwork, O. Waarts, and M. Yung, J. ACM **40**, 17 (1993).

[2] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, 2010).

[3] M. Newman, *Networks: an introduction* (OUP Oxford, 2010).

[4] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. Lett. **96**, 040601 (2006).

[5] S. Tauro, C. Palmer, G. Siganos, and M. Faloutsos, in *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, Vol. 3 (2001) pp. 1667–1671 vol.3.

[6] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, Proceedings of the National Academy of Sciences **104**, 11150 (2007).

[7] P. Pinto, J. Barros, and M. Win, Information Forensics and Security, IEEE Transactions on **7**, 125 (2012).