Small Worlds in Security Systems: an Analysis of the PGP Certificate Graph

Srdjan Čapkun, Levente Buttyán and Jean-Pierre Hubaux
Laboratory for Computer Communications and Applications (LCA)
School of Information and Communication Sciences (I&C)
Swiss Federal Institute of Technology Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
srdan.capkun@epfl.ch, levente.buttyan@epfl.ch, jean-pierre.hubaux@epfl.ch

ABSTRACT

We propose a new approach to securing self-organized mobile ad hoc networks. In this approach, security is achieved in a *fully self-organized* manner; by this we mean that the security system does not require any kind of certification authority or centralized server, even for the initialization phase. In our work, we were inspired by PGP [15] because its operation relies solely on the acquaintances between users. We show that the small-world phenomenon naturally emerges in the PGP system as a consequence of the self-organization of users. We show this by studying the PGP certificate graph properties and by quantifying its small-world characteristics. We argue that the certificate graphs of self-organized security systems will exhibit a similar small-world phenomenon, and we provide a way to model self-organized certificate graphs. The results of the PGP certificate graph analysis and graph modelling can be used to build new self-organized security systems and to test the performance of the existing proposals. In this work, we refer to such an example.

Keywords: PGP, small-world graphs, public-key management, self-organization

1. INTRODUCTION

Security in computer networks usually relies on central authorities, certificate directories, or some preinstalled keys and procedures. However, over the last decade, a very important change of paradigm has occurred: The concept of self-organization has appeared in many communication systems. By self-organization, we mean that the system is operated solely by the end-users. The most interesting examples include peer-to-peer systems, such as Gnutella[1], Freenet[2] and P-Grid [1]. In addition to the application layer, self-organization has also emerged on the lower layers of the

---

[1]http://www.gnutella.com/
[2]http://freenetproject.org/

system architecture. An excellent example is a mobile ad hoc network, where self-organization appears at the network layer [11, 4, 5]: By definition, a mobile ad hoc network does not rely on any fixed infrastructure; instead, all networking functions (e.g., routing, mobility management, etc.) are performed by the nodes themselves in a self-organizing manner.

Due to the lack of centralized services in self-organized systems, these systems cannot be secured with the existing security solutions. This is not surprising, as these solutions are devised mainly for the traditional wireline networks. Nevertheless, in devising new security approaches for self-organized systems, we use traditional cryptographic primitives. In our view, the most appealing tool for securing self-organized systems is public-key cryptography because it enables the establishment of secure communication by the exchange of public information and it supports key and user authentication.

The main problem in all public-key security systems is: How can a user u obtain the authentic public-key of another user v in the presence of an active attacker [8, 10]? A common approach to solving this problem is based on public-key certificates. A public-key certificate is a data structure in which a public-key is bound to an identity (and possibly to some other attributes) by the digital signature of the certificate issuer. One of the most common approaches is that certificates are issued by a central authority (called certification authority (CA)) that every user trusts. Every user knows the CA's public-key and thus can verify every certificate that the CA issued. In self-organized systems, the functionality of the CA can be distributed to the users, as proposed in [14, 7]. However, distributing the CA's functionality, means delegating the authority to a subset of users, which makes this approach impractical for fully self-organized systems.

In [10], we propose a different solution to the problem of public key management in self-organized systems. In our approach, each user is her own authority domain and users issue public-key certificates to each other. Like in PGP, key authentication is performed via a chain of certificates. By this we mean that the key authentication is done in the following way: When a user u wants to obtain the public key of another user v, she acquires a chain of public-key certificates such that

1. the first certificate of the chain can be directly verified by u using a public key that u holds and trusts (e.g., her own public key),

2. each remaining certificate can be verified using the public key contained in the previous certificate of the chain, and

3. the last certificate contains the public key of the target user v.

It is assumed that u trusts the issuer of each certificate in the chain to correctly verify that the public key in the certificate indeed belongs to the user name in the certificate.

However, contrary to PGP, our system does not rely on certificate directories for the distribution of certificates. Instead, certificates are stored and distributed by the users and each user maintains a local certificate repository that contains a limited number of certificates selected by the user according to some algorithm. When user u wants to verify the authenticity of the public key of user v, they merge their local certificate repositories, and u tries to find an appropriate certificate chain from u to v in the merged repository. We have shown that there are algorithms for the construction of local certificate repositories such that any pair of users can find, with high probability, certificate chains to each other in their merged repository, even if the size of the local repositories is small compared to the total number of users in the system. This result shows that it is indeed possible to solve the public-key management problem in a fully self-organized yet scalable way. This approach however provides only probabilistic guarantees and is dependent on the characteristics of the certificate graph on which it operates. A certificate graph is a directed graph $G(V, E)$, where V is a set of vertices, that represent public keys of the users, and E is a set of edges, that represent public key certificates. More precisely, there is a directed edge from vertex K_u to vertex K_v if there is a certificate signed with K_u in which K_v is bound to an identity.

In our work, we assumed that the certificate graphs in self-organized systems exhibit certain properties. Specifically, we assumed that the certificate graphs that might appear in self-organized systems exhibit the small-world phenomenon. A graph exhibits the small-world phenomenon if, roughly speaking, any two vertices in the graph are likely to be connected through a short sequence of intermediate vertices. The small-world phenomenon was first introduced by Stanley Milgram through a series of pioneering experiments that he and his coauthors conducted in the 1960's [9, 12]. The goal of the experiments was to find short chains of acquaintances linking pairs of people in the United States who did not know one another. The striking result was that the average number of intermediate steps in a successful chain was found to lie between five and six. This value has since become popular as the "six degrees of separation" principle.

In this paper, we argue that the small-world phenomenon naturally emerges in self-organized security systems. As an example of such a system, we take PGP, where a certificate graph is created in a fully self-organized manner. We further argue that in self-organized security systems, where
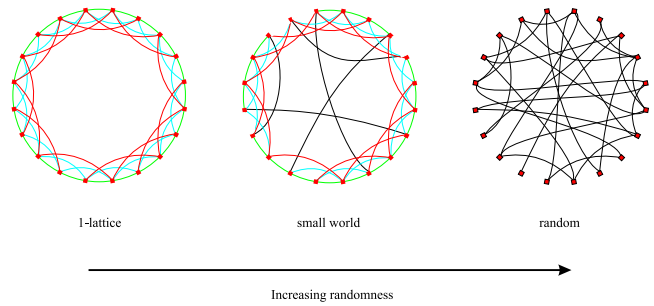

1-lattice   small world   random

Increasing randomness

Figure 1: A transition from the ordered graph (1-lattice) to the random graph.

users issue certificates based on their acquaintances, certificate graphs will exhibit small-world properties as a result of their mutual trust relationships. In order to study different security solutions for self-organized systems, we also propose a certificate graph model. Among other things, our model allows us to construct PGP-like graphs in a random manner for simulation purposes.

A recent work on small-world graph models suggests several reasons for the emergence of this phenomenon, both in social networks, and in the World Wide Web [3, 6, 13, 2]. In their work, these authors propose several models of small-world graphs. Naturally, as certificate graphs reflect the social relations between people, they will exhibit similar characteristics. Our analysis of PGP graphs shows that the existing small-world models only partially model certificate graphs. We therefore propose some modifications to these models, to generate certificate graphs that might appear in self-organized systems.

The four contributions of this work can be summarized as follows: First, we present a detailed study of the characteristics of the PGP certificate graphs. To our knowledge, no such study has been made so far. However, an analysis of some PGP certificate graph properties is available at *http://www.pgpi.org* and *http://pgp.dtype.org*. Second, we show that these graphs exhibit small-world properties and we argue that PGP is a wonderful source of inspiration to predict the way certificate graphs might emerge in fully self-organized systems. Third, we demonstrate that the existing small-world graph models do not model the certificate graphs appropriately. Fourth, we propose modifications of the existing models to better characterize certificate graphs.

The paper is organized as follows. In Section 2, we describe the existing small-world graph models. In Section 3, we present the analysis of the PGP certificate graphs. In Section 4, we propose a model for the creation of the artificial certificate graphs. In Section 5, we conclude and we give some directions for future work.[3]

2. MODELS OF SMALL-WORLD GRAPHS

Milgram's seminal small-world experiment remains one of the most compelling ways to think about the small-world phenomenon. It succeeds in demonstrating that we are not so far apart from each other and that the world is indeed a very small place.

The small-world phenomenon was recently a matter of an extensive study by Watts. In his book [13], Watts studies the small-world phenomenon and he suggests that small-world graphs lay somewhere in between ordered and completely random graphs. He illustrates this through several small-world graph models, out of which the ϕ model, in our opinion, is the most appealing one. In this model, small-world graphs are created such that an ordered structure (a 1-lattice) is rewired randomly. By rewiring the lattice, increasing randomness is introduced into a graph and, at the same time, long range edges are created. This is illustrated in Figure 1. These long range edges are called *shortcuts*, and are defined as follows: a shortcut is an edge such that when it is removed, the shortest path between the vertices previously connected by that edge becomes strictly larger than two. The fraction of the shortcuts in the total number of edges of a graph is called the ϕ coefficient and it determines the level of randomness of the graph.

Here, a *d-lattice* is a labelled, unweighted, simple graph that is similar to a Euclidean cubic lattice of dimension d in that any vertex v is joined to its lattice neighbors, u_i and w_i, as specified by

$$u_i = [(v - i^{d'}) + n](\bmod n); \; w_i = (v + i^{d'})(\bmod n)$$

where $1 \leq i \leq k/2$, $1 \leq d' \leq d$, and it is generally assumed that $k \geq 2d$. Here, k is a vertex degree and d is a lattice dimension. Hence, a 1-lattice with $k = 2$ is a ring, a 2-lattice with $k = 4$ is a two dimensional square grid, and so on (see Figure 1 for an example of a 1-lattice with $k = 4$).

The ϕ model achieves three important properties of small-world graphs: (i) high clustering, which reflects the fact that small groups of people are very well interconnected and that friends of a same friend are very likely mutual friends as well; (ii) small average distance between two vertices; (iii) logarithmic length scaling, that shows the small-world phenomenon, as for all graph sizes, the average length of shortest paths between the vertices is very small compared to the graph size.

These three graph properties are formally expressed through *clustering coefficient*, *characteristic length* and *logarithmic characteristic length scaling*. The clustering coefficient of a vertex v characterizes the extent to which vertices adjacent to vertex v are adjacent to each other. It is defined as the ratio of the number of edges between the vertices adjacent to v to the total number of possible edges between the same group of vertices. The clustering coefficient of a graph G is then defined as the mean of the clustering coefficients of all the vertices in G. The characteristic length of a graph G is defined as the median of the means of the shortest path lengths connecting each vertex to all other vertices in G. Logarithmic length scaling means that the characteristic length of a graph G scales logarithmically with the size of G.
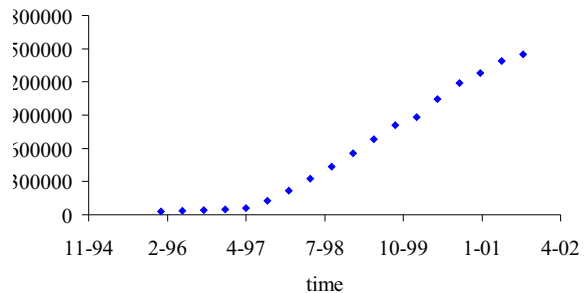

Figure 2: The evolution of the size of the PGP graph.

Therefore, a graph exhibits small-world properties if it has: (i) a high clustering coefficient, (ii) a small characteristic length, and (iii) a logarithmic characteristic length scaling.

One alternative to the model put forward by Watts is that the small-world phenomenon arises not because there are a few long-range connections in the otherwise short-range structure of a social network, but because there are a few vertices in the graph that have an unusually high degree or are linked to a widely distributed set of neighbors. This observation led to a model, proposed by Kasturirangan in [6], in which the graph construction begins with a 1- dimensional lattice, but instead of rewiring edges, a number of extra vertices are added, and are connected to a large number of vertices on the main lattice, chosen at random. This model is still similar to the Watts model in that shortcuts will be created between randomly chosen positions on the lattice and this model will also display the small-world effect.

Another model of small-world graphs has been proposed by Albert et al. [3]. In their work, the authors analyzed the World Wide Web graph and concluded that the Web is dominated by a small number of very highly connected sites. They found that the degree distribution of the WWW graph is a power law: the probability that a node has a degree k is proportional to $1/k^p$ for some positive $p > 1$, where p is called the power factor. Therefore, they suggested the following graph model: a normal random graph with an average degree k is rewired such that its degree distribution becomes a power law. In each iteration of this graph construction algorithm, a pair of vertices is chosen randomly and an edge is added between them, only if a newly created edge brings the overall degree distribution closer to the required power law. By repeating this process long enough, a graph is generated with the correct power law. But, this graph is in other respects a random graph with a very low clustering coefficient.

The analysis of Adamic [2] confirms that WWW graphs are indeed small-world graphs, but shows that these graphs do have a larger clustering coefficient than random graphs. She concludes that from this point of view, Albert's model is unrealistic for the modelling of WWW graphs.
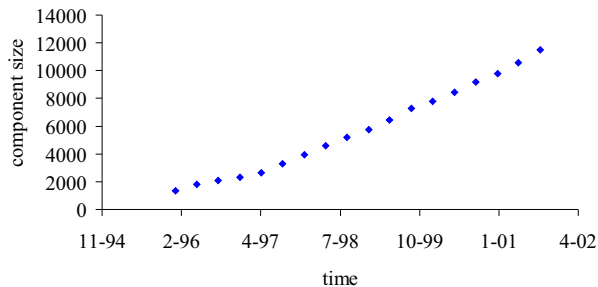
Figure 3: The evolution of the size of the largest strongly connected component of the PGP graph.

Figure 4: The directed characteristic length of the largest strongly connected component of the PGP graph.

3. ANALYSIS OF THE PGP CERTIFICATE GRAPH

In the introduction, we explained our motivation for exploring the characteristics of the certificate graphs for the provision of security services in self-organized systems. In this section, we present the analysis of the PGP certificate graphs that is, to our knowledge, the only known example of a self-organized certificate graph creation.

We extracted the database containing the PGP certificate graph from the public PGP certificate directories. One of these directories can be found at *http://pgp.dtype.org*. This database contains the information about public-keys and public-key certificates issued and revoked from the launch of the PGP project (early 1990's) until today. This information enables us to follow the evolution of the PGP certificate graph and to observe its characteristics as they change over this time period. The graph analysis that we present makes sense only if connected graphs are analyzed. Therefore, in our analysis, we focus on the characteristics of the largest strongly connected component of the PGP certificate graphs. The evolution of the sizes of both the PGP certificate graph and its largest strongly connected component is shown in Figures 2 and 3. Observing the characteristics of the second largest strongly connected component of the PGP certificate graph did not make sense as it contained few vertices (32 vertices, for the 2001 PGP graph).

To show that the PGP certificate graphs exhibit small-world properties, we observed the following graph characteristics of its largest strongly connected component:

- *directed clustering coefficient*
- *directed ϕ-coefficient*
- *directed characteristic length*
- graph diameter
- length scaling
- vertex degree distribution
- median and mean in-degree and out-degree
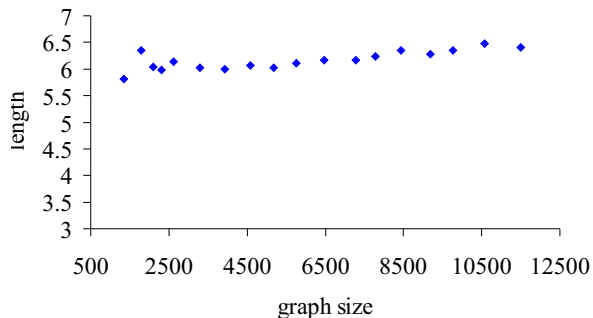- largest strongly connected component size

where directed ϕ-coefficient and directed characteristic length are generalizations of the Watts [13] definitions of the same graph characteristics to directed graphs. The *directed ϕ coefficient* is defined as a fraction of *directed shortcuts* in the total number of edges. The *directed shortcut* is defined as a directed edge that, when it is removed, the shortest *directed* path (in the same direction as the edge that is removed) between two nodes previously connected by that edge becomes strictly larger than two. The *directed characteristic length* of a graph G is defined as the median of the means of the directed shortest path lengths connecting each vertex to all other vertices in G.

The results of our analysis, presented on Figures 4, 5 and 6, show that PGP certificate graphs have (i) a small directed characteristic length, (ii) a high clustering coefficient, and (iii) slower than logarithmic length scaling. These results clearly show that the PGP certificate graphs do exhibit small-world characteristics.

A high ϕ-coefficient (Figure 6) can be explained by the fact that the trust relationships in the PGP community are often created between very distant (e.g. geographically distant) users. Figure 7 shows the PGP certificate graph diameter.

To compare PGP graphs with the graphs generated by the existing small-world models, we investigated the in-degree, out-degree and degree distributions of the PGP certificate graphs. The results of this analysis are shown on Figure 9. Clearly, the vertex degree distribution of the PGP graph is neither bimodal (like in Kasturirangani's small-world model), nor does it obey the power law (like in Albert's small-world model), nor do all vertices have approximately the same degrees (like in Watts' ϕ-model), but it resembles to the Zipf's distribution. Zipf's law, named after the Harvard linguistic professor George Kingsley Zipf (1902-1950), is the observation that frequency of occurrence of some event (P), as a function of the rank (k) when the rank is determined by the above frequency of occurrence, is a power-law function $P_k \sim 1/k^p$ with the exponent p close to unity.
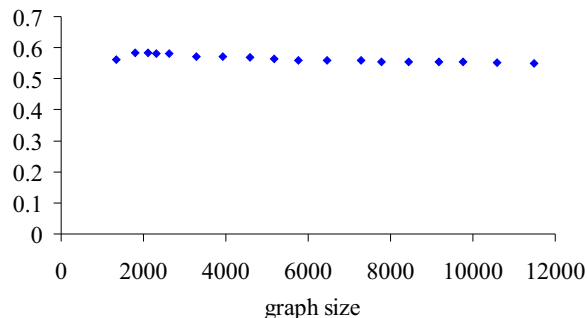
Figure 5: The directed clustering coefficient of the largest strongly connected component of the PGP graph.
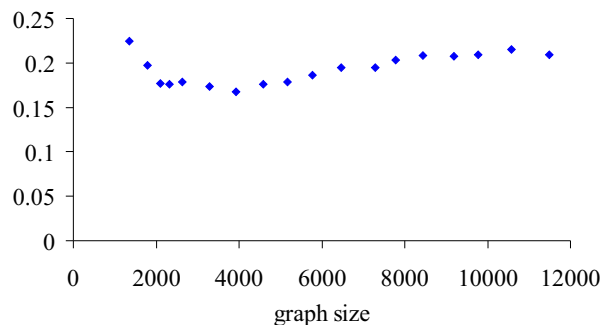

Figure 6: The directed ϕ-coefficient of the largest strongly connected component of the PGP graph.

4. CONSTRUCTING A MODEL OF THE CERTIFICATE GRAPH

In the previous section, we studied the characteristics of the PGP certificate graphs, as these graphs represent the closest existing example to the certificate graph creation in fully self-organized systems. In this section we propose a model for the generation of this kind of graphs, which will make it possible to generate certificate graphs of an arbitrary size.

Here, we argue that the certificate graphs in self-organized systems will exhibit small-world properties, as the creation of the certificate graphs depends on the existing trust relationships between people. In self-organized systems, if users are properly stimulated, their mutual acquaintances will eventually result in public-key certificates. We do not claim that the certificate graph will precisely map users' acquaintances, as the certificates do not express friendship, but the user confidence in a given user-key binding. It is however, reasonable to expect that most of the acquaintances will result in public-key certificates. It has been already shown by Milgram's experiment and argued in [13] that people's acquaintances do form a small-world graph. All this gives us reasonable assurance that the certificate graphs in self-organized systems will closely resemble graphs of user acquaintances. The presented PGP certificate graph analysis gives us additional assurance in our claim that certificate graphs in self-organized systems are indeed small-world graphs. Here, we do not consider the certificate graph creation where users are given other incentives, besides their acquaintances, to issue public key certificates, as these incentives could unpredictably change the certificate graphs.

As already discussed at the end of the previous section, the existing small-world graph models generate certificate graphs with unrealistic degree distributions. Even without observing specifically the PGP certificate graphs, it seems very counterintuitive that all the people have an approximately equal number of acquaintances, like it is proposed in the Watts model. It is more realistic to assume that, like in the PGP degree distribution, a majority of people have a small number of acquaintances, fewer people have more acquaintances, and just a few people are very well socially connected and have a large number of acquaintances.
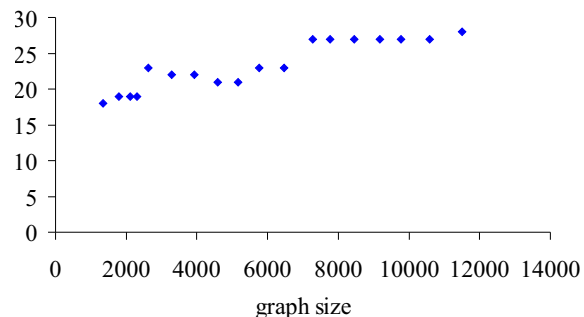

Figure 7: The diameter of the largest strongly connected component of the PGP graph.

Albert's model, on the other hand, takes into account the vertex degree distribution, but the graphs generated by this model are random graphs with a small clustering coefficient.

For the reasons stated above, we find that the existing small-world graph models do not correctly model certificate graphs. Therefore, we propose a new model for the generation of small-world certificate graphs, which we call *certificate graph model*. This model generates certificate graphs that have: (i) a small characteristic length that scales logarithmically or slower than logarithmically with the size of the graph, (ii) a high clustering coefficient and (iii) predefined in-degree, out-degree and degree distributions, similar to those of the PGP certificate graphs.

Our certificate graph model follows the same logic as Watts ϕ-model, as it singles out shortcuts as the main cause of the small-world effect. However, in our model, we use a different substrate. Instead of rewiring a regular 1-lattice, where each vertex has the same degree, we rewire an irregular lattice-like graph in the same fashion as in the ϕ-model. By irregular we mean that vertices on that graph do not necessarily have the same degrees.
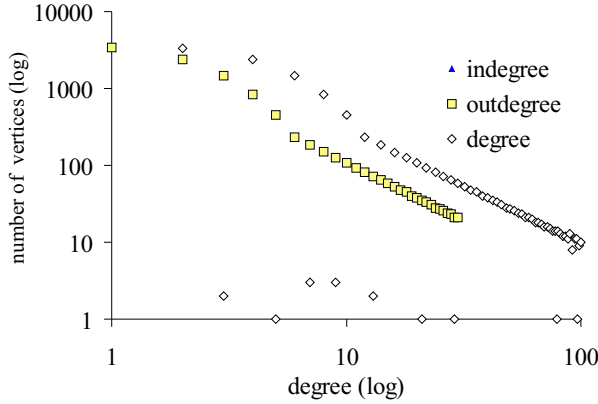
Figure 8: The in-degree, out-degree and degree distributions of the small-world certificate graph generated by the certificate graph model, with the model parameters that match the characteristics of the PGP September 2001 certificate graph.

We construct our substrate graph in the following way:

- Given a set V of vertices and given in-degree kin_v and out-degree $kout_v$ for each $v \in V$.

- every vertex v is joined to its neighbors u_k and w_k as specified by:
    forall $v \in V$
        $k = 0$
        for $i = 1$ to $|V|$
            $u = [(v - i) + |V|] \ (mod \ |V|)$
            if $indeg(u) \leq kin_u$, then $u_k = u$
                and $k = k + 1$
            if $k \geq \lfloor kout_v/2 \rfloor$ then stop
        $k = 0$
        for $i = 1$ to $|V|$
            $w = (v + i) \ (mod \ |V|)$;
            if $indeg(w) \leq kin_w$, then $w_k = w$
                and $k = k + 1$
            if $k \geq \lfloor kout_v/2 \rfloor$ then stop

The obtained substrate graph is created in the same fashion as the 1-lattice, but with a specified degree distribution. We call a graph generated by this construction an *irregular lattice*.

Here, we define our certificate graph model:

1. Construct an irregular lattice graph according to the predefined degree distribution.

2. Specify a desired ϕ.

3. "Randomly rewire" the irregular lattice with the constraint that $\phi \cdot |E|$ of its edges are forced to be shortcuts. Specifically:
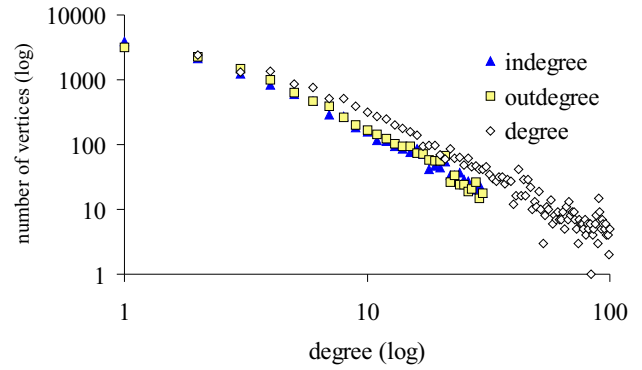

Figure 9: In-degree, out-degree and degree distributions of the largest strongly connected component of the 2001 PGP graph.

(a) Pick a vertex u at random

(b) Pick a neighbor v such that an edge between u and v is not a shortcut.

(c) Delete edge (u, v)

(d) Create shortcut (u, w), such that w is chosen randomly

(e) Choose a non-shortcut edge e connected to w

(f) Disconnect e from w and connect it to v

We validated our model by generating artificial certificate graphs with input parameters that are similar to those of the PGP certificate graphs. As a predefined degree distribution, we used the approximation of the vertex degree distributions of the PGP certificate graphs. The other graph construction parameters (number of vertices and edges, ϕ coefficient) were taken from the characteristics of the PGP certificate graphs.

We generated 18 artificial certificate graphs with the construction parameters obtained from 18 different PGP graphs (1996-2001, three graphs from each year). We repeated this experiment 20 times for each set of characteristics. The results of our simulations are shown in the Figures 8, 10 and 10.

In Figure 8 we display the degree distributions of the certificate graph generated with the same parameters as in the PGP September 2001 certificate graph and with the approximation of the vertex degree distribution as previously described. This figure shows that our model successfully generates graphs with given vertex degree distributions. A comparison of clustering coefficients of the PGP graph and the graphs generated by our model is given in Figure 10. We can see that the graphs generated by our model have a slightly larger clustering coefficient then the PGP certificate graphs. Figure 11 shows that the graphs generated with our model exhibit the same length scaling as the PGP certificate
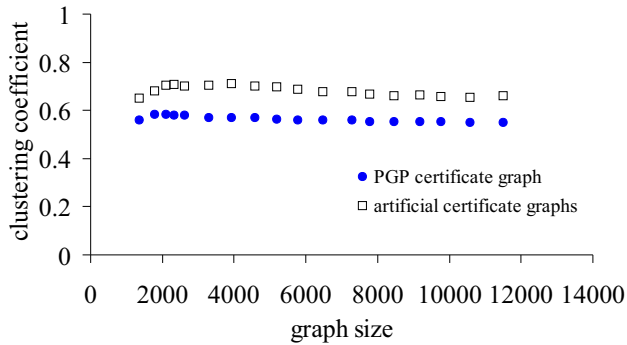
Figure 10: Directed clustering coefficients of PGP graphs and artificial graphs generated by the certificate graph model when the artificial graphs are generated with the same model parameters as PGP graphs.
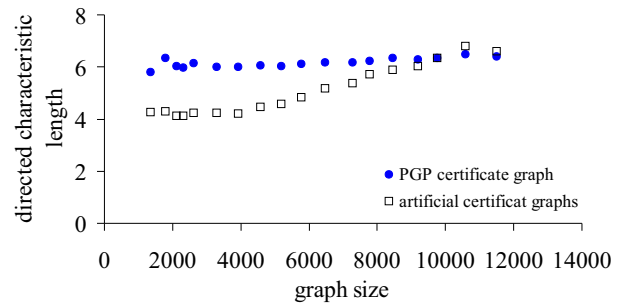

Figure 11: Directed characteristic lengths of PGP graphs and artificial graphs generated by the certificate graph model when the artificial graphs are generated with the same model parameters as PGP graphs.

graph and a very small characteristic length. These results demonstrate that with the proposed certificate graph model we can successfully generate artificial certificate graphs of arbitrary size and of appropriate graph characteristics.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have observed the emergence of the small-world phenomenon in fully self-organized systems, and its significance to system security. As self-organized security systems do not exist today, it is difficult to predict what the characteristics of the certificate graphs that would be created in these systems will be. Fortunately, PGP provides a very useful analogy and reveals the way in which certificate graphs will likely emerge in fully self-organized systems.

This paper provides a detailed analysis of the PGP certificate graph. It shows that PGP certificate graphs exhibit small-world characteristics. It provides a study of the appropriateness of existing small-world graph models for the representation of these kind of graphs. It concludes that some modifications to the existing models are necessary and proposes the corresponding modifications. The proposed model can be used to construct certificate graphs of any size.

In terms of future work, we intend to study in more detail the mechanisms by which trust emerges in fully self-organized systems. In this way, we can check to what extent the PGP certificate graph is really representative. We believe that this analysis can be an excellent basis to devise and compare solutions to the challenging problem of providing security in fully self-organized systems.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] K. Aberer, M. Punceva, M. Hauswirth, and R. Schmidt. Gridella, a P2P system based on the P-Grid approach, improves on Gnutella's search performance while reducing bandwidth requirements. *IEEE Internet Computing*, 2002.

[2] L. Adamic. The small world web. In *Proceedings of the European Conf. on Digital Libraries*, 1999.

[3] R. Albert, H. Jeong, and A.-L. Barabasi. Diameter of the World Wide Web. *Nature*, 401:130–131, 1999.

[4] L. Blažević, L. Buttyán, S. Čapkun, S. Giordano, J.-P. Hubaux, and J.-Y. Le Boudec. Self-organization in mobile ad hoc networks: The approach of Terminodes. *IEEE Communications Magazine*, June 2001.

[5] J. Jubin and J. Turnow. The DARPA packet radio project. *Proceedings of the IEEE*, 1987.

[6] R. Kasturirangan. Multiple scales in small-world networks. In *MIT AI Lab Memo 1663*, 1999.

[7] J. Kong, P. Zerfos, H.Luo, S. Lu, and L. Zhang. Providing robust and ubiquitous security support for mobile ad-hoc networks. In *Proceedings of the Int. conference on Network Protocols*, 2001.

[8] A. Menezes, P. van Oorschot, and S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1997.

[9] S. Milgram. The small world problem. In *Psychology Today*, volume 61, 1967.

[10] J.-P. Hubaux, L. Buttyán, and S. Čapkun. The quest for security in mobile ad hoc networks. In *Proceedings of MOBIHOC*, 2001.

[11] J.-P. Hubaux, T. Gross, J.-Y. Le Boudec, and M. Vetterli. Toward self-organized mobile ad hoc networks: The Terminodes Project. *IEEE Communications Magazine*, January 2001.

[12] J. Travers and S. Milgram. An experimental study of the small world problem. In *Sociometry*, volume 32, 1969.

[13] D. Watts. *Small Worlds*. Princeton University Press, 1999.

[14] L. Zhou and Z. Haas. Securing ad hoc networks. *IEEE Network*, 13(6):24–30, November/December 1999.

[15] P. Zimmermann. *The Official PGP User's Guide*. MIT Press, 1995.