

UNIVERSITÉ DE MONTRÉAL

Chit-Hack

Information exchange paths in IRC hacking
chatrooms

David Décary-Héту

9/1/2011

Information is power

Today's world is more than ever focused on information. For most of the 20th century, nations focused on building their military-industrial complex as a way to win the Cold War. With the fall of the iron curtain, it's not the number of tanks or fighter plans than a country can deploy to its borders that matters; it's the number of engineers and scientists it trains every year. Society has come to realize that information is power. In the stock market, for example, those with the most complete information (or even better - inside information) are the best suited to make informed decisions on whether to buy or sell certain stocks. Information enables them to make the right move at the right time – a tool that can be leverage to empower a corporation over its competitors.

Investment is not the only field that depends heavily on reliable and steady sources of information. In the computer security industry, companies live and die by the data they have access to. An attacker who has knowledge on the software packages used in a company can look up known vulnerabilities on the Internet for a specific version and then build an attack accordingly. Alternatively, an attacker with knowledge of internal operations and personnel name of a corporation can pretend to be part of said company to gain inside information or passwords that will allow him/her access to the computer network. As the story of Kevin Mitnick has shown us (Mitnick, 2011), there is very little a defendant can do against a knowledgeable attacker in the IT world.

With the ever increasing importance of information in the computer security field, it is important for security professionals, policy makers and law-enforcement agencies that researchers study the flows of information amongst underground hackers. The aim of this paper will thus be to understand how the computer underground of hackers shares information. Information can either be shared by posting it publicly on forums, websites and blogs or by socializing with others through online communications. This paper will focus on the latter and examine not what information is exchanged but how it is exchanged. In order to do so, we will look at the personal social networks of hackers. The first section of this paper will look at the current literature on hacker relationships. The second section will present an overview of the data and the methodology used in this study. The third section will display the results which we will be followed by a discussion on the structure of hackers' personal networks as well as the mentoring features of IRC. The conclusion will summarize our findings as well as present ideas for future studies.

Exchange paths in hacker networks

Movies have a funny way of presenting hackers: in many cases (see Hackers (1995), The Net (1995) or The Matrix (1999)), it seems that rapidly punching a few keystrokes on a keyboard magically

grants access to whole computer systems which can be manipulated at will. This illusion has been fed by many mediated "hacks" where the hacker merely guessed the password of a Hollywood star or used commercial phone-number spoofing services to gain access to voicemails (Sullivan, 2005). The reality has proven to be quite different.

With the ever increasing complexity of computer networks, hackers now need advanced skills in many fields to pose a real threat. Knowledge of multiple operating systems, networking protocols and security software packages is needed in order to infiltrate well-secured infrastructures. Even hacking less protected systems requires a certain level of skills; the days when no one was running a firewall or an antivirus are now long gone. In order to hone their skills, hackers can access detailed online tutorials on how to hack computers; a simple query on Google with the phrase "how to hack" returned more than 129 million documents. While there is bound to be vast amounts of junk in these results, the dedicated individual will likely find all the needed information to improve their skills or even get started in the hacking field. Hackers can also use resources that are geared towards the legitimate security industry. Websites such as The Academy Pro (<http://www.theacademypro.com>) produces high-quality tutorials on how to configure and use defensive and offensive security software. While these videos are intended for an audience of penetration testers and security professionals, they provide in-depth insight on how software packages work and how to exploit them.

There is no denying the value of these self-learning tools. But self-learning hacking techniques is not at everyone's reach. As most school systems have shown, people learn faster and become more efficient when a teacher (or mentor) is there to provide them with the necessary classes. With the highly technical and constantly changing environment that is the Internet, those that can connect to others with information stand to gain an enormous advantage over those that don't. To understand hackers, it is essential that researchers focus on the mentoring and information sharing in the computer underground.

Rogers (2000) addressed exactly this issue and claimed that: "the area of learning theory may have the best chance at providing an understanding of hacking" (Rogers, 2000: p17). He analyzed theoretical concepts to determine how they applied to the problem and hacking and concluded that Aker's social learning theory, although not a perfect fit, was the best available theory available. Built on top of Sutherland's differential association (1939), Aker's theory can be summarized in four points (Rogers, 2010). According to him, individuals are more likely to commit crimes if they:

1. Differentially associate with other criminals;

2. Receive more reinforcement from their illegal actions than for their legal actions;
3. Are more exposed to deviant thoughts and individuals and normal ones and;
4. Learn that committing crimes is normal and accepted.

At the heart of these propositions lies the notion of exchange. It is by giving tips and demonstrating techniques that the craft of hacking is transferred from an individual to another. As Rogers (2010) notes, hackers do not meet face-to-face. They instead use computer-mediated communications like online chatrooms, online forums, emails and instant messaging. Although many channels are available to them, it appears that hackers are particularly fond of one in particular, IRC. The Internet Chat Relay (IRC) was invented by Jarkko Oikarinen in 1988. It is a synchronous group instant messaging application that allows users (clients) to connect to chatrooms on IRC servers and to exchange messages and files with each other (Reid, 1991).

Figure 1: Example of IRC software

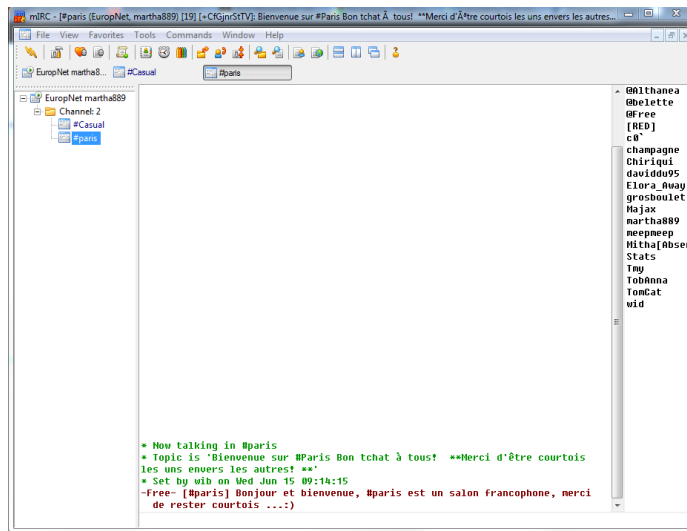


Figure 1 shows an example of an IRC client connected to the *casual* and *Paris* chatrooms on the EuropeNet server. There are hundreds of IRC servers all over the world each sporting their own list of chatrooms. As displayed in Figure 1, all messages are public by default and are posted in the middle section of the client software. It is possible for users to exchange private messages between them. Over the years, users created software programs called *bots* that mimic the human behavior and provide services to users. A bot can be programmed, for example, to ban from the chatroom users who curse too much. It has been known for years that hackers gather and socialize in IRC chatrooms (Bratus, 2007). As a consequence, security researchers, journalists and law-enforcement agencies constantly monitor the

shadier parts of IRC to gather intelligence on hackers. Although only the public messages can be accessed, the data captured in this fashion provides incredible insights in how hackers socialize and exchange information.

This paper aims to determine how information is shared between hackers by looking at the network features of hacking chatrooms. Since IRC is a nexus of interaction for hackers, it is reasonable to suppose that much of the mentoring and information sharing will happen through its chatrooms. By looking at the flow of information inside them, it will be possible to understand how hackers develop. This better understanding of these deviant individuals will enable us to better prevent and stop this flow of information thus limiting the damage that they bring to society.

Data

In order to study how hackers exchange information, we focused on one of the most active socializing environment for hackers, the Internet Chat Relay (IRC). The data for this research was provided to us by a North-American police force that monitored hacking channels.

Table 1: Characteristics of study sample

Length of monitoring	30 months (January 2009 - June 2011)
Number of chatrooms monitored (monthly)	7-15
Number of events	16,978,269
Number of relevant messages	2,232,729
Number of unique relevant messages	1,618
Number of hackers	262
Number of people in social network of hackers	356
Number of ties (reciprocal)	6,168

Table 1 provides the characteristics of our sample data. It covers a 30 months period ranging from January 2009 to June 2011. The number of hacking chatrooms monitored fluctuated over this period with a low of 7 monitored chatrooms in a month to a peak of 15. During our sampling period, more than 16 million events were registered. These events are either messages posted, users that login/logout or users that change their nickname. Since analyzing the millions of messages that were captured was beyond our means, we opted to target specific messages that included at least 1 of 79 keywords related to hacking. This left us with just over two million messages to analyze. When we analyzed those messages, we realized that they included a great deal of duplicates. We came to realize that hackers were using bots to spam the chatrooms with messages advertising their services. Since the bots constantly post the same

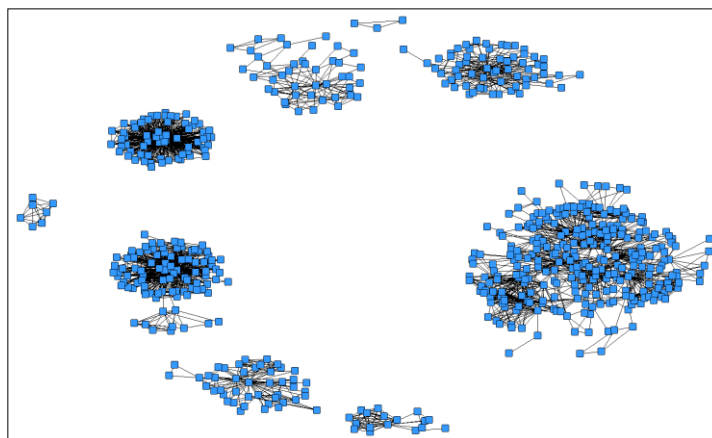
messages over and over again, we calculated, for each user in the chatrooms, an index of the uniqueness of their messages. Anyone who had messages that was less unique than 90% was discarded as a bot. This means that our sample only included people who posted original messages 9 times out of 10. The end result was a sample of 262 individuals who posted 1,618 messages containing at least one keyword between January 2009 and June 2011.

As mentioned before, all the logs come from public chatrooms where many individuals participate in public discussions. It would be almost impossible to determine direct ties between individuals because of the group nature of the conversation. Using a limited sample, we determined that messages containing keywords were generally part of ongoing discussions that included many individuals and that the 10 messages that were posted before and after the flagged messages could be considered as part of this conversation. We thus created the social graph of each user based on the conversations they had and ended with 356 individuals that were involved in discussions with hackers. Since these were ongoing conversations, the ties between the individuals are all reciprocal and not directional since it is almost impossible to determine the initiator of the discussion.

Methodology

Figure 2 displays the structure of the links between the hackers and their social graph. The figure shows that the monitored chatrooms had, in most cases, exclusive members who did not visit the other hacking chatrooms.

Figure 2: Structure of ties in hacking chatrooms



With such a configuration, it would be useless to try and create an artificial network where there is none. As we intend to focus on the individual features of hacker networks rather their structure, an approach focusing on the individual networks appeared as the most sensible.

Studying the personal network of individuals is known in the social network field as ego networks analysis (Wasserman & Faust, 1994). The term *ego* refers to the individual at the center of a network and the *alters* are the people linked to the ego. Although this method has been used in many social science studies (Kalmijn & Vermunt, 2007; Degenne & Lebeaux, 2005; Stefanone & Jang, 2008), it has yet to be adopted in criminology. We were only able to find one reference to an analysis of ego networks in the context of criminal networks, a paper from Malm et al. (2011) on the co-offending of criminal organizations in Canada. In this research, Malm et al. (2011) analyze data relating to co-offending and determine that certain types of organizations tend to co-offend more with other groups while others tend to choose their co-offenders based on their membership to the same group. The authors focus mainly on the centrality, betweenness, composition and density of the groups. Centrality is frequently used to assess the prominence of actors within a network (Wasserman & Faust, 1994: 172). It indicates the number of incoming and outgoing contacts and account for the direction of direct ties around each node (in directed networks). The pattern of ties originating from or sent to a network member is usually a reliable indicator of this person's prestige or status (Wasserman & Faust, 1994: 174) as it helps distinguish people with sought-after expertise. Betweenness measures the extent to which a node mediates between other nodes by its position along the geodesics (the shortest paths between two nodes) within the network, thus providing useful additional information on the structure of a network. The more often a node is located along the geodesics, the higher its betweenness centrality, making it a broker within the network. The position of broker has been associated with the notion of power in networks (Prell et al., 2008; Morselli, 2009; Toral et al., 2009) since these individuals control the flow of information between the different actors. They can decide whether to allow messages to pass through, modify the information, or simply ignore it. Composition is a measure of the heterogeneity of the network (Malm et al., 2011). It measures whether individuals associate with others that share common characteristics (ex: if boys talk more to boys than girls). Finally, density measures the number of actual ties compared to the number of possible ties (Wasserman & Faust, 1994). This metric gives a sense of the implication of individuals in a network; the denser the network, the more involved its participants are.

Since our dataset is slightly different from that of Malm et al. (2011), we adapted their methodology slightly. In the first part of the results, we looked at the ego networks through four social network metrics: centrality, betweenness, density and longest distance. This last metric measures the number of steps any node needs to reach any other node in the ego network. We will present the minimum and maximum value of these metrics as well as the quartile distribution to show the distribution of the results in our sample. In the second part of the results, we ran a partial correlation that controlled for the number of actors in each ego network. This correlation will present the variables that are associated with the number of unique keywords and the number of conversations of hackers. These include the number of ties, the density, the normalized betweenness and the 2 step reach. This last metric measures the number of alters that egos can reach within 2 ties.

Results

Table 2 presents a summary of the social network metrics for the ego networks of hackers who visited hacking chatrooms. Each metric will be analyzed separately in this section.

Table 2: Percentiles of social network metrics

	Centrality	nBetweenness	Density	Longest Distance
Min	1	0.00	0	0
25th	4	0.00	55	1
50th	7	0.00	100	1
75th	12	9.86	100	2
Max	65	100.00	100	6

N = 262

Centrality

Centrality represents the number of alters that are connected to an ego. In our sample, hackers have contacts with a limited number of alters in general. The minimum number of ties is 1 and the maximum is 65. 25% of egos are tied to 4 alters or less and 75% of egos are connected to 12 alters or less. Considering how our data was analyzed, these values are much lower than what we would have expected. Since each conversation we flagged included a potential of 21 individuals (10 messages before and after plus the ego), a median of 7 connections is fairly low. Our distribution does show that some hackers have large social networks and that they have access to more information.

Betweenness

Betweenness reflects the importance of undirected connections by calculating the number of times an ego sits on the shortest path between two alters. In this case, we used the normalized betweenness value to eliminate the differences in the number of nodes in each ego network. Our results show that our hackers are very poor brokers for the most part. With half of our sample showing a betweenness of 0.00, the number of ties between the alters that transit through the egos are very low. Some egos are more essentials than others. The last 25% of our sample have a betweenness score that varies between 9.86 and 100.00.

Density

Density is the number of actual ties divided by the number of possible ties. Giving the limited size of most ego networks, it is not surprising to see that more than half of our sample displays density scores of 100. Once again, a select few hackers have different network configurations with lower densities than the others. Such high density in general means that the ego networks are closed groups where everyone knows each other. It could be affected by the undirected nature of our network which increases the number of ties between the nodes.

Longest Distance

This last social network metric measures the number of ties that separates any node from any other node in the ego network. With half the sample showing a longest distance of 1, most ego networks are once again homogenous entities with low centrality and high density. It is interesting to see that the maximum number of steps needed to reach any node on the largest ego network is 6, a special number in social network analysis ever since Watts book *Six Degrees* (Watts, 2004). Our data suggests that it is true that anyone on the planet (or in this case, chatrooms) can be reached within 6 ties.

Table 3: Partial correlation of diversity of interests and involvement in hacking chatrooms

	Nb of ties	Density	nBetweenness	2 Step Reach	Nb of keywords	Nb of conversations
Nb of ties	1.000					
Density	**0.257	1.000				
nBetweenness	**-.0574	**-.0661	1.000			
2 Step Reach	**-.0361	0.030	**0.162	1.000		
Nb of keywords	**-.0352	**-.0274	**0.369	**0.102	1.000	
Nb of conversations	**-.0140	**-.0113	**0.169	0.054	**0.819	1.000

N = 618

Our partial correlation presents the correlation between the number of unique keyword (a measure of the diversity of interests) as well as the number of conversations (a measure of the involvement in hacking chatrooms). Our results show that the two variables are correlated with almost all of the social metric variables excluding the 2 step reach for the number of conversations.

The number of keywords is moderately and negatively correlated with the number of ties and the density (-0.352 and -0.274 respectively). This could be the result of the structural holes paradigm (Burt, 2010). It is thought that close-knit networks have access to more redundant data since all the actors in the network know each other and talk about the same subjects. It could be the case that these smaller and denser networks are experts in a limited range of subjects and thus only mentioned a limited number of our keywords. Betweenness and 2 step reach are moderately and poorly but positively correlated to the number of keywords (0.369 and 0.102 respectively). As mentioned before, those with greater betweenness usually display more structural holes in their networks and can access more diverse sources of informations. They can then expand their knowledge on a variety of subjects. This increased number of alters is reflected in the 2 step reach metric which is correlated positively to the number of keywords. It shows once again that having access to a greater number of nodes (directly or indirectly) increases the available knowledge base.

The number of conversations is negatively and poorly correlated to the number of ties and the density. Individuals with a greater number of ties have fewer conversations. As we have mentioned before, the conversations in hacking chatrooms seem to be concentrated amongst tight and small networks. It is thus normal to see that larger ego networks show lower levels of conversations. It is surprising though that the density is negatively correlated with the number of conversations. This could be explained by the fact that we are only monitoring the public messages on IRC chatrooms and not the complete communications channels between the individuals. Some of the conversations amongst them could be conducted over private messages or instant messaging. It could also be that certain smaller units would rather move their discussions to more private settings rather than discuss publicly. The number of conversation is positively correlated to the betweenness. With a greater level of communication, the chances that a node will sit in the middle of a dyad is higher and it is thus normal to see these two figures positively correlated though we might have expected a slightly higher correlation between these two metrics.

Discussion

The aim of this paper is to determine exactly how information is shared in hacking chatrooms. Our data suggests that visitors of hacking chatrooms tend to limit the number of rooms they hang in. There is thus a clear interference in the knowledge transfer from chatroom to chatroom; information in one room rarely crosses from one chatroom to the other. Since each room has its own topic, it is not surprising to see these iron walls set in between chatrooms. What might of interest to A might not be to B. Each chatroom is thus its own private club with select members.

These individuals do share some characteristics. They interact with other in what appears to be small and dense network where direct connectivity is more important than indirect connections. The number of contacts in the ego networks is limited in most cases and the number of brokers is fairly poor. Alters and egos evolve in very dense ego networks where everyone knows everyone. Each node of these ego networks can rapidly and easily reach any other member of the network with a few steps. In such a setting, it is expected that the bandwidth of information will be higher meaning that any information will be quickly accessible to any member of the group. This is done at the expense of the diversity of information. Nodes will have less access to new and innovative information and will be limited in the topics they can discuss with others. This would suggest that hackers who visited these chatrooms were interested in one topic only – perhaps the topic of discussion of the chatroom – and used other sources of information for their other needs. When it comes to learning hacking skills, hackers could be using web forums and online tutorials and then using the direct and interactive features of IRC to get a better understanding of specific subjects which might be harder to learn on their own. This ego network configuration also shows that researchers or law-enforcement officials who would like to infiltrate hacking groups will need time and technical skills before they can be accepted in these chatrooms. Passively monitoring is easy enough but getting accepted in tight networks is much harder. This ego network configuration is different than what we have seen elsewhere in the hacking world (Décary-Hétu et al., 2011) where the centrality and density of networks was much lower.

Our partial correlation confirmed a number of social networking paradigms about structural holes particularly. It proves that IRC is an environment in which there can be intense exchanges of information but that this information is shared mostly in small groups. The number of subjects discussed in each of these groups is limited but its quality and quantity is usually high. This means that IRC chatrooms should be excellent learning classrooms for any hacker who would like to hone his skills on a variety of subjects. In this rich environment where people can be contacted in real time, learning how tools and software

packages work will be much easier. This advantage is offset by the dedication and time that outside hackers will have to invest to join these groups. It has been shown that online hacking communities are not always open to newcomers and that they must prove themselves before they are accepted in the group. This is particularly true in smaller communities. It is our belief that those who succeed in tapping into the resources available on IRC will greatly enhance their hacking skills.

Conclusion

In the computer security world, information is everything. Knowledge on vulnerabilities, exploitation techniques and software configuration can save or destroy any target. Hackers are very familiar with this axiom and have been claiming, unsurprisingly, for greater access to information. This paper has shown that hackers meet on IRC in specialized chatrooms to discuss and exchange on various hacking techniques and tools. They form close-knit groups with high level of densities and low levels of density. Their discussion tends to focus on a limited number of subjects and each group discussion could be seen as an advanced class in a particular category of hacking.

Studying hackers on IRC, although easier today, is still very complicated. Access to more underground chatrooms is protected by passwords or by bots that ban unknown users. Accessing these channels would provide even more detailed information on the most specialized hackers who believe that their discussion is so sensitive that it needs to be protected of the public eye. The number of public chatrooms in our sample was also limited to a single IRC server and a handful of chatrooms. Using more resources, it would be possible to monitor a much greater number of chatrooms and come up with a more representative portrait of the computer underground on IRC.

Studies on IRC individuals are by no means new. Computer professionals have been using them for years but with their own technical angle. This paper brings a theoretical framework that can be harnessed to better understand the social interactions of hackers online. This hybrid study which features the most up-to-date technical tools with tested criminological theories offers a new take on the problem of hackers, and one that we hope will be picked up and replicated amongst more datasets in the upcoming years.

References

Bratus, S. (2007). *Hacker Curriculum: How Hackers Learn Networking*. IEEE Distributed Systems Online 8(10).

Burt, R. S. (2010). *Structural Holes In Virtual Worlds*.

- Décary-Héту, D. & C. Morselli & S. Leman-Langlois (2011). *Welcome to the Scene: A Study of Social Organization and Recognition among Warez Hackers*. *Journal Of Research On Crime And Delinquency*.
- Degenne, A. & M. Lebeaux. (2004). *The Dynamics Of Personal Networks At The Time Of Entry Into Adult Life*. *Social Networks* 27: 337-358.
- Kalmijn, M. & J. Vermunt. (2007). *Homogeneity Of Social Networks By Age And Marital Status: A Multilevel Analysis Of Ego-centered Networks*. *Social Networks* 29: 25-43.
- Malm, A. & G. Bichler & R. Nash. (2011). *Co-offending Between Criminal Enterprise Groups*. *Global Crime* 12(2): 112-128.
- Mitnick, K. (2011). *Ghost In The Wires: My Adventures As The World's Most Wanted Hacker*. New York, NY: Little, Brown & Company.
- Morselli, Carlo. (2009). *Law-Enforcement Disruption Of A Drug-Importation Network*. *Studies Of Organized Crime* 8:1-17.
- Prell, Christina and Klaus Hubacek and Mark Reed. (2008). *Who's In The Network? Systemic Practice And Action Research*. 21: 443-458.
- Reid, E. (1991). *Electropolis: Communication And Community On Internet Chat Relay*. Masters at University of Melbourne, Australia.
- Rogers, M. (2000). *Psychological Theories Of Crime And Hacking*.
- Rogers, M. (2010). *The Psyche Of Cybercriminals: A Psycho-Social Perspective*. IN Ghosh, S. & E. Turrini. *Cybercrimes: A Multidisciplinary Analysis*. New York, NY: Springer.
- Stefanone, M. A. & C. Jang. (2008). *Writing For Friends And Family: The Interpersonal Nature Of Blogs*. *Journal Of Computer-Mediated Communication* 13: 123-140.
- Sullivan, B. (2005). *Cell Phone Voicemail Easily Hacked*. Retrieved on August 12th, 2011: http://www.msnbc.msn.com/id/7046776/ns/technology_and_science-wireless/t/cell-phone-voicemail-easily-hacked/.
- Sutherland, E. H. (1939). *Principles Of Criminology*. Philadelphia, PA: Lippincott.

Toral, Sergio and M.R. Martinez-Torres and Federico Barrero. (2010). *Analysis Of Virtual Communities Supporting OSS Projects Using Social Network Analysis*. Information & Software Technology. 52(3).

Wasserman, S. & K. Faust. (1994). *Social Network Analysis: Methods And Applications*. Cambridge, UK: Cambridge University Press.

Watts, D. (2004). *Six Degrees: The Science Of A Connected Age*. New York, NY: W.W. Norton & Company.