# Hack Through Your Neighborhood

Rachita Hajela
Cs department, NYU,
NY, USA
rs1997@nyu.edu

Anto Loyola
Cs department, NYU,
NY, USA
al4251@nyu.edu

Nikita Keswaney
Cs department, NYU,
NY, USA
nk1664@nyu.edu

***Abstract*—Based on two data sources - Living Expenditures and internet usage in Different Zip Codes in USA and data provided by Worldwide Intelligence Network Environment (WINE), the project aims at finding out the relationship between the number of attempts at hacking into an internet network with respect to living expenditure of the locality and the internet usage of the average household in the locality. It is useful in identifying any hotspots where hackers go after several people living in the same area.**

***Keywords*—hackers, WINE**

## I. INTRODUCTION

One of the major security threat faced by individuals and organizations today is cyber-crime and internet attacks. According to a majority of internet experts surveyed for a new report by the US-based Pew Research Center, Cyber-attacks on countries and corporations are likely to increase in the next decade. The project aims at finding out the relationship between the number of attempts at hacking into an internet network with respect to living expenditure of the locality and the internet usage of the average household in the locality. The locality is based on the zip codes. Basically it tries to find out if certain areas (with high living expenditure or high internet usage) are more prone to internet attacks than other areas. It is useful in identifying any hotspots where hackers go after several people living in the same area.

## II. RELATED WORK

### *A. Home Network Security*

Computer Security means preventing and detecting unauthorized use of the computer. The document gives home users an overview of the security risks and countermeasures associated with the Internet connectivity. The basic technologies used for internet connectivity are broadband, cable modem access, DSL access. Broadband services are different from traditional dial-up services as in dial up service the computer only connects to internet when it has some request. The IP address is also dynamically assigned. Broadband services are always on services, the computer is always on the network and the IP address will change less frequently, thus making it more of a fixed target for attack.

The three areas where information security is concerned for home users is: Confidentiality, Integrity, Availability (rightful access). The risks include – intentional misuse of computer via Trojan, back door programs. Denial of service, unprotected windows share, mobile code. Email spoofing, hidden file extensions, packet sniffing etc. The actions home users can take to mitigate the risks are many. Some of the recommendations are: if working from home – consult system support personnel, use virus protection software, firewalls, not running programs of unknown origin, disabling hidden filename extension, keeping all applications including operating system patched, disconnecting computer from network when not in use, disabling Java, Javascript. The important thing is to always have a regular backup of critical data.

### *B. Characterizing Transition Behaviors in Internet Attack Sequences*

Sequential patterns exist to characterize attack behavior. So attacks can be analyzed to find out if they have similar attacking strategies. The hacker can perform Host Discovery, Services Scanning, Vulnerability Attempts or a mixture of them. A feature called Attack Transition Action (ATA) is defined to represent the changes on attacked destinations and ports over time. Destination IP and destination port are two key factors to understand attacking behavior. An attacker may change the strategy based on the feedback from the target or even subjective factors, such as his/her personality. The attacker begins with a single vulnerability attempt and then attempts host discovery with random IP and port choices. . A random variable called Attack Transition Action (ATA) is used to show the current attack action shows: 1) no change on IP and port, 2) only change in port, 3) only change in IP, and 4) changes in both IP and port. , sending packets to the same target IP and port gives a sequence of ATA-0, and implies vulnerability attempt. A sequence of packets targeting on different ports of the same host, i.e., a sequence of ATA-1, indicates single host service scanning. Sending a series of packet to different destinations but the same port indicates scanning certain service across hosts, and can be captured by a sequence of ATA-2. A sequence of ATA-3 could make sense if the destination IP are randomly chosen and ports are undefined.

### *C. Building a Secure Home Network*

A home computer has tons of data collected about the owners. It keeps track of where they have been, what they have been browsing, who they are, personal information, bank account numbers, passwords, email, identity, etc. Leaving a computer open and unsecured allows intruders to view every detail. Some attacks can cause damage to the system and its data making that information inaccessible to even the owner. Ways to secure home network: Always bind TCP/IP only to the external adapters. Most importantly, all network services (like Microsoft Logon, File and Print Sharing, and Client for Microsoft Networks) should NEVER be bound to TCP/IP to prevent unauthorized access. Protect yourself from renegade viruses by updating virus definitions once a week and scanning your PC for viruses. Use a Personal Firewall. Firewalls allow you to monitor the requests coming into your PC and provide alerts at unauthorized accesses. Use encryption where necessary. Utilize safe-surfing and know what you are installing. Maintaining confidentiality and integrity of your data should always be of critical Importance to never get complacent about your home security. Use Special Care with Sensitive Data

### *D. Use of the Internet in higher-income households*

These papers bring to notice the dependence that the average person (in America) has on the Internet. A shocking 99% of all Americans use Internet at home and this number is growing every day. 95% of people with an income of $75000 or more use their cellphone to access the Internet.71% of the higher income Americans pays their bills online and this shows that we have a lot of our personal information being tossed back and forth across the Internet in what we can hope are safe transactions. This leads to the necessity to know if the amount of money we earn is somehow influencing how many people are trying to hack into our network. Which is the idea behind our project.

### *E. Cyber-bullying*

This paper takes into account that teens are major consumers of the Internet in the present lifestyle This means that there is another way for hackers to get access to sensitive information and use this in many twisted ways This shows that there are no limits to what kind of messages teens receive from random strangers and due to puberty and curiosity they make stupid decisions and share images, videos, private files of the family, etc with these hackers It shows that woman are more prone to cyber-bullying and this leads to a lot of problems that they cannot get over through their life This also shows that older girls receive threats that leave them cornered without knowing whom to approach for help Intense Internet users are bullied more than an average user In this age of Cyber-Bullying is would be helpful to know the factors that urge network hacks

### *F. From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases*

Internet access is an important resource. However it has been observed that groups with higher levels of access to the Internet were the same groups (whites, men, and residents of urban areas) that had greater access to education, income and other resources. The phrase ‘access to Internet’ has been redefined, not meaning ‘who can find a network connection at home, work or public center’ but instead, ‘what are people doing, and what are they able to do, when they go on-line.’ As access diffuses to parts of public who were initially excluded, dimensions related to quality of use become important. At the point, when Internet penetration will reach levels enjoyed by telephone, access to Internet could no longer contribute significantly to social inequality. Some policy analysts believe that the ‘digital divide’ will be overcome at such stage. While others anticipate that high rates of Internet Penetration will not eliminate inequality so much as increase the new kinds of inequality - inequality among Internet users in the extent to which they are able to reap benefits from their use of the technology. The critical dimensions of inequality can be technical (hardware and software), autonomy of use, skill, availability of social support, variation in use. As the digital revolution is taking place, it is a large agenda but not an impossible to anticipate the consequences of technological change.

## III. DESIGN

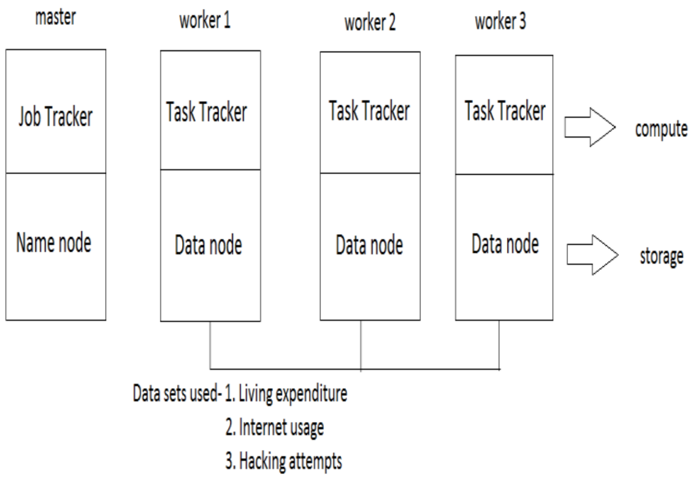

Fig 1. Software architecture

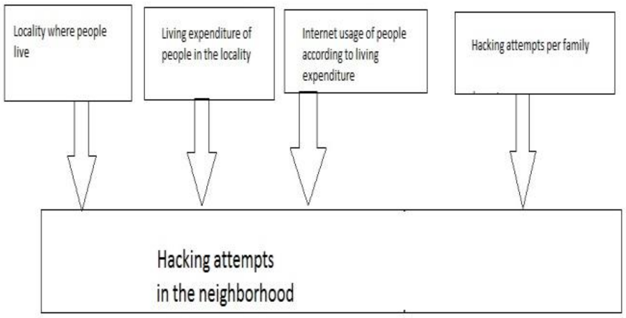
Locality where people live
Living expenditure of people in the locality
Internet usage of people according to living expenditure
Hacking attempts per family
Hacking attempts in the neighborhood

Fig 2. Data Flow Diagram

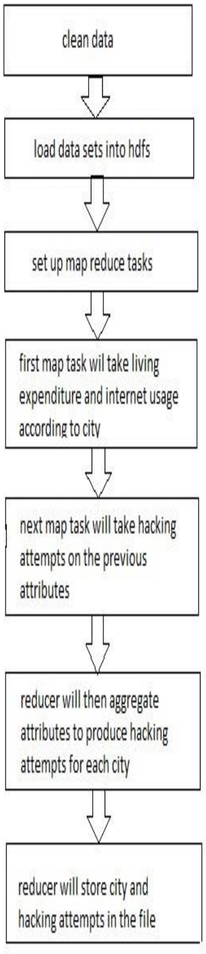
clean data
load data sets into hdfs
set up map reduce tasks
first map task wil take living expenditure and internet usage according to city
next map task will take hacking attempts on the previous attributes
reducer will then aggregate attributes to produce hacking attempts for each city
reducer will store city and hacking attempts in the file

Fig 3. Map reduce diagram

The first map reduce task will work on two data sets: living expenditure and internet usage. The next map reduce tasks will work on hacking attempts which will be applied on the result obtained from the first map reduce task. The final output containing hacking attempts per neighborhood will then be stored in the output file.

## IV. RESULTS

(Future… In this section, you can describe: Your experimental setup/issues with data/performance/etc. Describe your experiments, describe what you learned. Did you prove or disprove your hypothesis? Were some results unexpected? Why? )

## V. FUTURE WORK

(Future… Given time, how would you expand your analytic? Could it be applied to other areas? Etc…)

## VI. CONCLUSION

(Future… One or two paragraphs about the value/accuracy/goodness of your analytic.)

## ACKNOWLEDGMENT

(This section is optional. It can be used to thank the people/companies/organizations who have made data available to you, for example. You can list any HPC people who were particularly helpful, if you used the NYU HPC.)

## REFERENCES

[1] Paul DiMaggio and Eszter Hargittai, From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases, Center for Arts and Cultural Policy Studies.

[2] Jim Jansen, Use of the internet in higher income households, PewResearch Internet Project,

http://www.pewinternet.org/2010/11/24/use-of-the-internet-in-higher-income-households/

[3] Haitao Du and Shanchieh Jay Yang, Characterizing Transition Behaviors in Internet Attack Sequences, IEEE 2007

[4] Amanda Lenhart. Cyberbullying, PewResearch Internet Project, http://www.pewinternet.org/2007/06/27/cyberbullying/

[5] Kim Thomas, Building a secure home network, SANS Institute InfoSec ReadingRoom, http://www.sans.org/readingroom/whitepapers/hsoffice/building-secure-home-network-611

[6] Home network security,CERN, http://www.cert.org/historical/tech_tips/home_networks.cfm?