# Yan Gao


**Contact**

2145 Sheridan Road
Department of EECS
Northwestern University
Evanston, IL 60208, USA

*Voice:* (847)275-8994(C)
*Fax:* (847) 491-4455
*E-mail:* ygao@cs.northwestern.edu
*http://*www.cs.northwestern.edu/∼yga751


**Interests**

- ♢ Network Monitoring and Security
- ♢ Data Mining for Networking and System Issues
- ♢ User Study for System Optimization

**Education**

- ♢ Northwestern University, Evanston, Illinois, USA — 09/2004 - Present
  Ph.D in EECS Department — GPA 4.0/4.0
  Advisor: Prof. Alok Choudhary
- ♢ Xi'an Jiaotong University, Xi'an, Shaanxi, China — 09/2000 - 05/2003
  M.S. in Systems Engineering, ECE Department, May, 2002 — GPA 3.9/4.0
  Advisor: Prof. Guoji Sun
- ♢ Xi'an Jiaotong University (XJTU), Xi'an, Shaanxi, China — 09/1996 - 07/2000
  B.E., Electrical Engineering Department, July, 2000 — GPA 3.7/4.0
  Minor in Business Management

**Honors**

- ♢ IEEE ICDCS student travel grants award, 2006
- ♢ *Morrison* Fellowship, Northwestern University, 2006.6-2006.9
- ♢ IEEE ICNP student travel grants award, 2005
- ♢ *Walter P. Murphy* Fellowship, Department of EECS, Northwestern University, 2004-2005
- ♢ Second Prize of advancement in Science and Technology of Xi'an for the project of "Supply Chain Management System based on B/S", Xi'an, Shaanxi, P.R.China, 2002
- ♢ Outstanding Master Degree Thesis, XJTU, 2003
- ♢ *Hu Baosheng* Fellowship for academic excellence, XJTU, 2001-2002
- ♢ Chinese Academy of Science Research Fellowship for academic excellence, XJTU, 2000-2001
- ♢ University Fellowship for Outstanding Graduate Students, XJTU, 2000-2002.
- ♢ University Fellowship for Outstanding Undergraduate Students, XJTU, 1996-1999
- ♢ *Datang* Scholarship for Outstanding Undergraduate Students, XJTU, 1998-1999
- ♢ *Logic* Scholarship for Outstanding Undergraduate Students, XJTU, 1996-1997
- ♢ Outstanding Freshman Fellowship, XJTU, 1996-1997

**Experience**

- ♢ Electrical Engineering and Computer Science (EECS) Department, Northwestern University — Evanston, IL, USA
  Research assistant. — 09/2004 - Present
  - ◇ The High-Performance Network Anomaly/Intrusion Detection and Mitigation (HPNAIDM) Systems:
    Existing intrusion detection systems (IDS) have three shortcomings: 1) are mostly host-based and not scalable to high-speed networks. Thus they cannot prevent the rapid propagation of the latest viruses/worms which can infect most vulnerable machines in the Internet in only ten minutes; 2) are mostly signature-based and unable to recognize unknown anomalies; and 3) are isolated or centralized systems. To address these limitations, we propose HPNAIDM system with the following features: 1) online traffic recording and analysis on high-speed networks, 2) online statistical anomaly detection, 3) integrated approach for false positive reduction, 4) hardware speedup for real-time detection, and 5) scalable anomaly/intrusion alarm fusion from multiple sources.

- ◇ Pollution Resilience for Internet Caches:
  In this project, we investigate and develop efficient methods to detect a class of pollution attacks that aim to degrade a proxys caching capabilities, either by ruining the cache file locality, or by inducing false file locality.
- ◇ Detecting Stealthy Spreaders Using Online Outdegree Histograms:
  We consider the problem of detecting the presence of a sufficiently large number of hosts that connect to more than a certain number of unique destinations within a given time window, at high-speed networks. Previous techniques have focused on detecting the sources with an extremely large outdegree. However, such techniques will fail to detect spreaders such as bot scans in which each scanning host will scan only a moderate, fixed number of destinations. In contrast, our scheme maintains a small, fixed size memory usage, and is still able to detect stealthy spreader scenarios by approximating outdegree histograms from continuous traffic.
- ◇ Image Spam Hunter:
  The newest image-based spam uses simple image processing technologies to vary the content of individual messages. Thus, they pose great challenges to conventional spam filters. In this project, we propose a system using a probabilistic boosting tree to determine whether an incoming image is a spam or not based on global image features. The system identifies spam without the need for OCR and is robust in the face of the kinds of variation found in current spam images.
- ◇ Learning Relationship Between Operating System Level Measurements and End User Satisfaction:
  In interactive application domain, there exists a variation for user expectations and satisfaction relative to the real operating system performance. In this project, we aim to study the relationship between operating system level features and user satisfaction. By leveraging this variation, we propose an efficient system prototype that can customize dynamic voltage and frequency for different end users in terms of reducing the CPU power consumption.
- ◇ User Perception Measurements for Automatic Configuration of BitTorrent Tools:
  BitTorrent(BT) tools are widely used in the world, but the configuration is complicated and generally the default configuration can not reach the satisfaction of all end users. In this project, we measure the relationship between user satisfaction and the automatic configuration choices of BT tools, and further propose a prototype to predict the onset of user dissatisfaction without user interaction, and design a customized configuration adjustment scheme by using data mining results to improve the BT performance.

♢ Microsoft Research Intern — Redmond, USA
Cybersecurity and Systems Management Research Group — 06/2007 - 08/2007

- ◇ Project: Downloading Internet Webpage Anonymously and Efficiently
  The design of Internet does not evolve with the privacy, thus the protocols that provide the fundamental functions of the Internet are inherently non-anonymous. Due to the intense competition and large profit, some website administrators use IP cloaking and IP blocking schemes to prevent the well-intentioned active measurements, e.g. web crawlers, from accurate information recently. In this project, we propose to deploy the active measurement tools through anonymous proxy servers as a result of defeating these threats.

♢ Schlumberger Limited — Beijing, China
Network and Infrastructure Solution Engineer — 09/2003 - 06/2004

- ◇ Involved in the project of the Credit Card System for the Bank of China. My main work is related to Data Management.

♢ Institute of Systems Engineering, XJTU — Xi'an, Shaanxi, China
Research assistant — 07/2000 - 05/2003
Advisor: Prof. Guoji Sun

- ◇ Integrated Network Information Security Defense System, funded by the hi-tech research and development program of China(863 plan). As one of the main designer, I build the Host Intrusion Detection System by using statistical and information fusion algorithms.

- ◇ Supply Chain Management System based on B/S. My main work is the system analysis and design, the database design and the development of the key software modules
- ◇ Virtual Campus on web. My main work is the system modeling and simulation by using 3Dmax and OpenGL.

PUBLICATIONS

♢ Journal Articles and Conference Papers

- ◇ **Yan Gao**, Ming Yang, Xiaonan Zhao, Bryan Pardo, Ying Wu, Thrasos Pappas, Alok Choudhary, "Image Spam Hunter", in Proc. of *the 33th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.
- ◇ **Yan Gao**, Yao Zhao, Robert Schweller, Shobha Venkataramany, Yan Chen, Dawn Songy and Ming-Yang Kao, "Detecting Stealthy Spreaders Using Online Outdegree Histograms", in Proc. of *the 15th IEEE International Workshop on Quality of Service (IWQoS)*, 2007.
- ◇ **Yan Gao**, Leiwen Deng, Aleksandar Kuzmanovic and Yan Chen, "Pollution Resilience for Internet Proxy Caches", in Proc. of *the 14th IEEE International Conference on Network Protocols (ICNP)*, 2006.
- ◇ **Yan Gao**, Zhichun Li and Yan Chen, "A DoS Resilient Flow-level Intrusion Detection Approach for High-speed network", in Proc. of *The International Conference on Distributed Computing Systems (ICDCS)*, 2006.
- ◇ Robert Schweller, Zhichun Li, Yan Chen, **Yan Gao**, Ashish Gupta, Elliot Pearson, Ying Zhang, Peter A. Dinda, Ming-Yang Kao, and Gokhan Memik, "Reversible Sketches: Enabling Monitoring and Analysis over High-speed Data Streams", to appear in *ACM/IEEE Transaction on Networking*.
- ◇ Robert Schweller, Zhichun Li, Yan Chen, **Yan Gao**, Ashish Gupta, Ying Zhang, Peter A. Dinda, Ming-Yang Kao, and Gokhan Memik, "Reverse Hashing for High-speed Network Monitoring: Algorithms, Evaluation, and Applications", in Proc. of *IEEE INFOCOM*, 2006.
- ◇ Pin Ren, **Yan Gao**, Zhichun Li, Yan Chen and Benjamin Watson, "IDGraphs: Intrusion Detection and Analysis Using Stream Compositing", in *IEEE Computer Graphics & Applications*, special issue on Visualization for Cyber Security, 2006.
- ◇ Pin Ren, **Yan Gao**, Zhichun Li, Yan Chen and Benjamin Watson, "IDGraphs: Intrusion Detection and Analysis Using Histographs", in Proc. of *IEEE Workshop on Visualization for Computer Security (VizSEC)*, in conjunction with *Visualization'2005* and *InfoVis'2005* conferences, Octobor, 2005.
- ◇ Li Feng, Xiaohong Guan, Sangang Guo, **Yan Gao** and Peini Liu, "Predicting the Intrusion Intentions by Observing System Call Sequences", *Computers and Security, Elsevier Science*, Volume 23, Issue 3, pp.241-252, May, 2004.
- ◇ **Yan Gao**, Xiaohong Guan and Guoji Sun, "Host Intrusion Detection System based on Real-Time Keystroke", *Chinese Journal of Computers*, Volume 27, No.3, pp.396-401, March, 2004. (in Chinese)
- ◇ Li Feng, Xiaohong Guan, Sangang Guo, **Yan Gao** and Peini Liu, "Plan Recognition Based Method for Prodicting Intrusion Intention of System Call Sequences", *Chinese Journal of Computers*, Volume 27, No.8, pp.1083-1091, August, 2004. (in Chinese)
- ◇ **Yan Gao**, Gang Hua and Guoji Sun, "A New Hybrid Method or Fast Vector Quantization", *Microelectronics& Computer*, Vol.20, No.2, pp.56-59, Feb., 2003. (in Chinese)

♢ DISSERTATION AND THESIS

- ◇ **Yan Gao**, "Online Scalable Intrusion Detection Systems for High-speed Networks", *Master Degree Thesis*, Northwestern University, Press 2007.
- ◇ **Yan Gao**, "Design and Study of Host-based Anomaly Detection System", *Master Degree Thesis*, XJTU, Press 2003. (in Chinese)
- ◇ **Yan Gao**, "Development and Realization of Virtual Host Management System based on Web", Bachelor Degree Thesis, XJTU, Press 2000. (in Chinese)

## Courses

| | | |
|---|---|---|
| Advanced data mining | Machine learning | Data structure & management |
| Operating systems | Introduction to networking | Introduction to computer security |
| Internet security | Advanced networking | Design and analysis algorithms |
| Resource visualization | Distributed computing systems | Algorithmic techniques for bioinformatics |
| Internet measurement | Computer architecture | |

## Computer Skills

- ♢ Programming with C/C++, Perl, Matlab under various environments.
- ♢ Skilled in Latex, Microsoft Office and Microsoft Visio etc.
- ♢ Proficient on database designing, programming and web based programming.
- ♢ Proficient with TCP/IP communication protocol.

## Teaching

♢ Electrical Engineering & Computer Science Dept., Northwestern University
Teaching assistant.

- ◇ CS395/495 "Basic Information Security: Technology, Business and Law". Fall 2005
- ◇ CS340 "Introduction to Networking". Winter 2006
- ◇ CS395/495 "Internet Measurement and its Reverse Engineering". Spring 2006
- ◇ CS395/495 "Advanced Networking". Spring 2006
- ◇ EECS203 "Introduction to Computer Engineering". Spring 2007
- ◇ EECS221 "Fundamentals of Circuits". Winter 2008

♢ Institute of Systems Engineering, XJTU
Teaching assistant. 03/2001 - 07/2002

- ◇ Supervising the final thesis of undergraduate student,
- ◇ Assisting teaching of the course "Modeling and simulation of complex systems"

## Activities

- ♢ Reviewer, IEEE ICDCS 2008
- ♢ Reviewer, IEEE ICNP 2007
- ♢ Reviewer, ACM MobiCom 2007
- ♢ Reviewer, IEEE INFOCOM 2007
- ♢ Reviewer, IEEE GlobleCom 2006
- ♢ Reviewer, IEEE IWQoS 2006
- ♢ Reviewer Committee, ACSAC 2006
- ♢ Reviewer, IEEE IM 2006
- ♢ Reviewer, IEEE INFOCOM 2005
- ♢ Reviewer, ACM SIGCOMM poster, 2005