

Research Statement

Debin Gao

School of Information Systems, Singapore Management University

Tel: (65) 6828-0969; Email: dbgao@smu.edu.sg

Updated : 09 Feb 2012

Background

As people rely more on computers, building and maintaining a secure computing environment becomes an important research topic. However, many computer programs remain vulnerable, making intrusions to a computer or a network of computers easy. Vulnerabilities like buffer overflows may permit an attacker to inject attack code and cause the vulnerable machine to run the attacker's program. Automatically detecting the intrusions and analyzing the vulnerabilities and malware are critical in securing a computer system.

My research centers on malware analysis and intrusion detection. Figure 1 shows an overview.

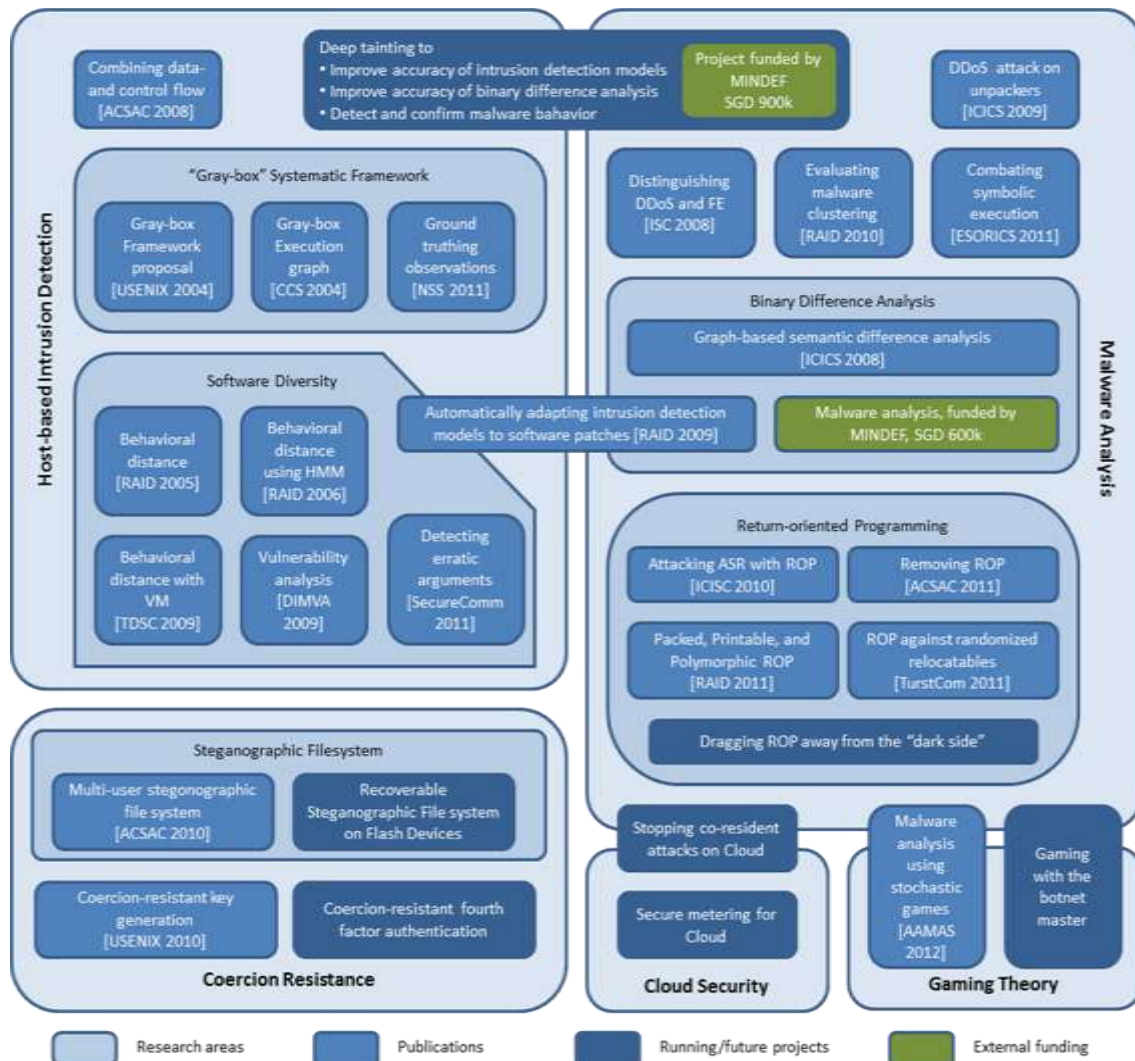


Figure 1: My research interests

I am a leading researcher in malware analysis and host-based intrusion detection with publications in top security conferences. I also work on a couple of new and interesting security research topics, namely coercion resistance, cloud security, and applying gaming theory in malware analysis.

Host-based intrusion detection and malware analysis

I see these two research areas closely related to one another. Host-based intrusion detectors focus on mechanisms a defender could take to detect an intrusion to make it more difficult for malware to exploit a system, while malware analysis tries to understand what malicious programs do to better defend against them. Both areas are rich in research potential and are on my long-term research strategy.

Host-based intrusion detection: My research in host-based intrusion detection focuses on anomaly detection techniques as opposed to signature-based ones. Anomaly detection techniques have the advantage of being able to detect zero-day attacks, and are what future intrusion detectors would use. (Currently the most widely used intrusion detectors and anti-virus tools are signature-based.)

We propose a gray-box systematic framework for host-based anomaly detection techniques [USENIX 2004]. This gray-box framework not only captures most existing host-based intrusion detectors, but has become the framework under which new techniques are proposed. Execution graph [CCS 2004] is one of them and has a nice feature of conforming to the control flow graph of the program (static) while being built from dynamic training. We also research on ground truthing observations to obtain better training samples for the detector [NSS 2011].

Many host-based anomaly detection techniques, including the execution graph we introduced, are susceptible to mimicry attacks in which the injected code masquerades as the original software (including returning the correct responses) while conducting its attack. To make such mimicry attacks more difficult, we introduce a notion, behavioral distance, for evaluating the extent to which processes — potentially running different programs and executing on different platforms — behave similarly in response to a common input. This is a novel idea of using software diversity to help improving accuracy of intrusion detectors. We propose the first system of behavioral distance [RAID 2005], improve its accuracy by using a customized Hidden-Markov Model [RAID 2006], and realize it with virtual machines running on one physical computer [TDSC 2009]. We also challenge the assumption in using software diversity for improving security and perform a systematic analysis on existing software vulnerabilities [DIMVA 2009], and propose enhancements to behavioral distance to detect erratic arguments [SecureComm 2011].

An interesting area in host-based intrusion detection, which the research community had not looked into, is what to do when a new version of the program being protected is released. In the current software industry, it is common to have new versions every few weeks, if not days. Automatically updating the intrusion detection model becomes important for practical reasons. We introduce BinHunt, a novel technique for finding semantic differences in binary programs [ICICS 2008], and use it to automatically adapt intrusion detection models to program patches [RAID 2009]. BinHunt bases its analysis on the control flow of the programs using a new graph isomorphism technique, symbolic execution, and theorem proving.

Malware analysis: Interestingly, BinHunt is useful in malware analysis, too, and has since had important impact on my research in the last three years in that it draws the connection between my research in intrusion detection and malware analysis. In malware analysis, an important research topic is malware clustering where we try to group malware samples, potentially polymorphic and metamorphic, into clusters or families for better understanding. BinHunt can find semantic differences between binary executables and is resistant to many obfuscation techniques; therefore it is a good technique to help malware clustering. This idea is highly supported by Ministry of Defense, Singapore, who provides funding of SGD 600,000 for a two-year research project on malware analysis using BinHunt. The project has come to an end, and our system is now in use by the Ministry of Defense and other government agencies.

I am also active in other research directions on malware analysis. For example, we systematically analyze how people evaluate malware clustering techniques, find limitations in them, and propose better evaluation mechanisms [RAID 2010]. We analyze state-of-the-art malware unpackers and find vulnerabilities in them, propose a working denial of service attack, and caution the community in developing and using such unpackers [ICICS 2009].

Return-oriented Programming: Return-oriented Programming (ROP) is one of the latest attacking strategies used by malware writers. Being a very new attacking technique, ROP has attracted a lot of attention in the research community.

We show that return-oriented programming can be used to attack address space randomization systems, making it fragile in the sense that if any one of the important system objects is not randomized properly, the entire system will lose its security property [ICISC 2010]. We further analyze the capability of ROP, and find that it could be made packed, printable, and polymorphic [RAID 2011], and be used to attack randomized relocatables [TrustCom 2011].

Having realized the power of ROP, we propose an automatic system to remove it from any malicious program so that the large body of existing software analysis tools can be used to analyze ROP-based malware [ACSAC 2011].

Coercion-resistant systems

This is a relatively new research area that I started working on since about two years ago. We are the first to propose a coercion-resistant system of generating cryptographic keys from biometrics [USENIX 2010]. A very nice property of such a system is that when users are being coerced, they naturally lose the capability of generating the correct cryptographic key. Therefore, an informed attacker (who understands how the system works) would not coerce users in order to get access, which in turn protects the users. We presented this work in the top security conference USENIX Security in 2010 and received very positive feedback.

We also provide similar properties in multi-user file systems. Dr-Steg is a new file system we propose in which users can plausibly deny the existence of data on a shared storage [ACSAC 2010]. When a user of Dr-Steg is under coercion, he/she can claim that certain important data does not exist, and an attacker would not be able to observe or show evidence otherwise even if he has multiple snapshots of the file system.

Future research

I plan on continuing working on host-based intrusion detection, malware analysis, and coercion-resistance, as well as starting working on cloud security and applying gaming theories in malware analysis.

MINDEF has approved the funding our project on “deep taint” to advance the research of host-based intrusion detection and malware analysis. This would become another connection of the two areas. “Deep taint” employs the same idea of taint analysis where the propagation of insecure data is traced in the system, but differs from it in its granularity. Deep taint can assign different taint tags to different parts of program input and trace the propagation of each of them. When applied to the training of a host-based intrusion detection system, it allows one to learn more accurate rules among system call arguments and return values and therefore improves the accuracy of the model learned. When applied to malware analysis of comparing two malicious programs, it provides specific and accurate hints on the mapping of basic blocks from them.

Return-oriented programming is one of the hottest research areas and I plan on continuing my research in it. Since the introduction of return-oriented programming, it has been considered as an attacking technique. We just started the research of dragging return-oriented programming away from the “dark side”, i.e., we research applying ROP in non-attack scenarios. For example, we research using ROP for software obfuscation, as well as using ROP in compilers to produce binaries that are more difficult to be exploited.

Coercion resistance is a very special and useful property we had proposed, and we plan on extending/modifying selected security protocols to make them coercion-resistant. One project we are working on right now is to add coercion resistance to the protocol of fourth factor of authentication. A user study has been conducted to see how people react to such a scheme. We also plan on extending this property to authenticating mechanisms on mobile devices.

I have identified two new areas, both of which are related to my research on malware analysis. One direction is on cloud security, in which we have just finished a project on stopping co-residency attacks on cloud. Co-residency attack is a sophisticated attack where an attacker locates the physical machine on which the victim virtual machine is running, and performs side-channel attacks against it. We propose stopping such attacks by duplicating virtual machines on multiple replicas, and applying the median among all replicas so that attackers could not locate the victim. Another security concern in cloud is secure billing, where a service provider and a client might have disputes upon time spent on executing certain jobs. We plan on proposing a secure metering system so that a proof can be revealed in case of billing disputes.

The other new direction we just started is in applying gaming theories on malware analysis. Malware is intelligent in that it may conceal its behavior when an analyzer is detected. The interaction between a malware and the corresponding analyzer is just like a game (e.g., chess), and we propose an interactive malware analyzer using gaming theories [AAMAS 2012]. We are also working on applying this idea in the interaction between a bot master and an analyzer.

Selected publications

[USENIX 2004] Debin Gao, Michael K. Reiter and Dawn Song, "On Gray-Box Program Tracking for Anomaly Detection", in *Proceedings of the 13th USENIX Security Symposium (USENIX Security 2004)*, San Diego, CA, USA, August 2004

[CCS 2004] Debin Gao, Michael K. Reiter and Dawn Song, "Gray-Box Extraction of Execution Graphs for Anomaly Detection", in *Proceedings of the 11th ACM Conference on Computer and Communications Security (CCS 2004)*, Washington, DC, USA, October 2004

[RAID 2005] Debin Gao, Michael K. Reiter and Dawn Song, "Behavioral Distance for Intrusion Detection", in *Proceedings of the 8th International Symposium on Recent Advances in Intrusion Detection (RAID 2005)*, Seattle, WA, USA, September 2005

[RAID 2006] Debin Gao, Michael K. Reiter and Dawn Song, "Behavioral Distance Measurement Using Hidden Markov Models", in *Proceedings of the 9th International Symposium on Recent Advances in Intrusion Detection (RAID 2006)*, Hamburg, Germany, September 2006

[ISC 2008] Hyundo Park, Peng Li, Debin Gao, Heejo Lee and Robert H. Deng, "Distinguishing between FE and DDoS using Randomness Check", in *Proceedings of the 11th Information Security Conference (ISC 2008)*, Taipei, September 2008

[ICICS 2008] Debin Gao, Michael K. Reiter and Dawn Song, "BinHunt: Automatically Finding Semantic Differences in Binary Programs", in *Proceedings of the 10th International Conference on Information and Communications Security (ICICS 2008)*, Birmingham, UK, October 2008

[ACSAC 2008] Peng Li, Hyundo Park, Debin Gao and Jianming Fu, "Bridging the Gap between Data-flow and Control-flow Analysis for Anomaly Detection", in *Proceedings of the 24th Annual Computer Security Applications Conference (ACSAC 2008)*, Anaheim, California, USA, December 2008

[TDSC 2009] Debin Gao, Michael K. Reiter and Dawn Song, "Beyond Output Voting: Detecting Compromised Replicas using HMM-based Behavioral Distance", in *IEEE Transactions on Dependable and Secure Computing (TDSC)*, April 2009

[DIMVA 2009] Jin Han, Debin Gao and Robert H. Deng, "On the Effectiveness of Software Diversity: A Systematic Study on Real-World Vulnerabilities", in *Proceedings of the 6th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 2009)*, Milan, Italy, July 2009

[RAID 2009] Peng Li, Debin Gao and Michael K. Reiter, "Automatically Adapting a Trained Anomaly Detector to Software Patches", in *Proceedings of the 12th International Symposium on Recent Advances in Intrusion Detection (RAID 2009)*, Saint-Malo, Brittany, France, September 2009

[ICICS 2009] Limin Liu, Jiang Ming, Zhi Wang, Debin Gao and Chunfu Jia, "Denial-of-Service Attacks on Host-Based Generic Unpackers", in *Proceedings of the 11th International Conference on Information and Communications Security (ICICS 2009)*, Beijing, China, December 2009

[USENIX 2010] Payas Gupta and Debin Gao, "Fighting Coercion Attacks in Key Generation using Skin Conductance", In *Proceedings of the 19th USENIX Security Symposium (USENIX Security 2010)*, Washington, DC, USA, August 2010

[RAID 2010] Peng Li, Limin Liu, Debin Gao and Michael K. Reiter, "On Challenges in Evaluating Malware Clustering", In *Proceedings of the 13th International Symposium on Recent Advances in Intrusion Detection (RAID 2010)*, Ottawa, Ontario, Canada, September 2010

[ICISC 2010] Zhi Wang, Renquan Cheng and Debin Gao, "Revisiting Address Space Randomization", In *Proceedings of the 13th Annual International Conference on Information Security and Cryptology (ICISC 2010)*, Seoul, Korea, December 2010

[ACSAC 2010] Jin Han, Meng Pan, Debin Gao and HweeHwa Pang, "A Multi-User Steganographic File System on Untrusted Shared Storage", In *Proceedings of the 26th Annual Computer Security Applications Conference (ACSAC 2010)*, Austin, Texas, USA, December 2010

[RAID 2011] Kangjie Lu, Dabi Zou, Weiping Wen and Debin Gao, "Packed, Printable, and Polymorphic Return-Oriented Programming", In *Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection (RAID 2011)*, Menlo Park, California, USA, September 2011

[ESORICS 2011] Zhi Wang, Jiang Ming, Chunfu Jia and Debin Gao, "Linear Obfuscation to Combat Symbolic Execution", In *Proceedings of the 16th European Symposium on Research in Computer Security (ESORICS 2011)*, Leuven, Belgium, September 2011

[SecureComm 2011] Jin Han, Qiang Yan, Robert H. Deng and Debin Gao, "On Detection of Erratic Arguments", In *Proceedings of the 7th International ICST Conference on Security and Privacy in Communication Networks (SecureComm 2011)*, London, United Kingdom, September 2011

[NSS 2011] Jiang Ming, Haibin Zhang and Debin Gao, "Towards Ground Truthing Observations in Gray-Box Anomaly Detection", In *Proceedings of the 5th International Conference on Network and System Security (NSS 2011)*, Milan, Italy, September 2011

[TrustCom 2011] Limin Liu, Jin Han, Debin Gao, Jiwu Jing, and Daren Zha, "Launching Return-Oriented Programming Attacks against Randomized Relocatable Executables", In *Proceedings of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011)*, Changsha, China, November 2011

[ACSAC 2011] Kangjie Lu, Dabi Zou and Debin Gao, "deRop: Removing Return-Oriented Programming from Malware", In *Proceedings of the 27th Annual Computer Security Applications Conference (ACSAC 2011)*, Orlando, Florida, USA, December 2011

[AAMAS 2012] Simon Williamson, Pradeep Varakantham, Debin Gao and Chen Hui Ong, "Active Malware Analysis using Stochastic Games", In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Valencia, Spain, June 2012, to appear