8th CS UNM Student Conference (CSUSC 2012)
April 24, 2012
Albuquerque, NM

The CSGSA is pleased to announce the 8th annual CS UNM Student Conference (CSUSC). The following are lists of talk abstracts.

Keynote-speaker

Dan Wallach, Ph.D. - Rice University

Security architectures for smartphones
Modern smartphones allow a variety of third-party applications to run on them, creating a delicate dance between usability and trust. Without burdening the user with security dialogs, apps must have enough privilege to get their job done, yet with suitable isolation from other possibly hostile apps. Android provides a variety of security features that were engineered to make this possible, but a number of deficiencies have cropped up over the years. This talk considers several problems. Android applications tend to have the ability to make arbitrary Internet connections, making it difficult for remote servers to trust which app might be making the connection. Android applications can similarly make a variety of internal IPCs, leading to "confused deputy" attacks where one app might be tricked into exercising a dangerous privilege on behalf of an untrusted caller. We address these issues with IPC and RPC extensions that can efficiently track the call chain and use this when making security decisions. We will also discuss solutions to the "permission bloat" problem that results from apps, which may not need many permissions themselves, including advertising libraries which require GPS location, Internet access, and more. Our IPC architecture allows us to separate advertisements from their hosting applications, reducing permission bloat and increasing resistance to synthetic click attacks.

Bio: Dr. Dan Wallach is a professor in the Department of Computer Science at Rice University in Houston, Texas and is the acting director of NSF's ACCURATE (A Center for Correct, Usable, Reliable, Auditable and Transparent Elections). His research considers a variety of different computer security topics, ranging from web browsers and servers through electronic voting technologies and smart phones.

Faculty Candidate-Speaker

Stephen Checkoway – University of San Diego
The stereotypical view of computing, and hence computer security, is a landscape filled with laptops, desktops, smartphones and servers; general purpose computers in the proper sense. However, this is but the visible tip of the iceberg. In fact, most computing today is invisibly embedded into systems and environments that few of us would ever think of as computers. Indeed, applications in virtually all walks of modern life, from automobiles to medical devices, power grids to voting machines, have evolved to rely on the same substrate of general purpose microprocessors and (frequently) network connectivity that underlie our personal computers. Yet along with the power of these capabilities come the same potential risks as well. My research has focused on understanding the scope of such problems by exploring vulnerabilities in the embedded environment, how they arise, and the shape of the attack surfaces they expose. In this talk, I will particularly discuss recent work on two large-scale platforms: modern automobiles and electronic voting machines. In each case, I will explain how implicit or explicit assumptions in the design of the systems have opened them to attack. I will demonstrate these problems, concretely and completely, including arbitrary control over election results and remote tracking and control of an unmodified automobile. I will explain the nature of these problems, how they have come to arise, and the challenges in hardening such systems going forward.

Bio: Stephen Checkoway is a Ph.D. candidate in Computer Science and Engineering at UC San Diego and before that he received his B.S. from the University of Washington. He is also a member of the Center for Automotive Embedded Systems Security, a collaboration between UC San Diego and the University of Washington. Checkoway's research spans a range of applied security problems including the security of embedded and cyber-physical systems, electronic voting, and memory safety vulnerabilities.

Student presentations

Antonio M. Espinoza

Title:Work-in-progress: Automated Named Entity Extraction for Tracking Censorship of Current Events

Abstract—Tracking Internet censorship is challenging because what content the censors target can change daily, even hourly, with current events. The process must be automated because of the large amount of data that needs to be processed. Our focus in this paper is on automated probing of keyword-based Internet censorship, where natural language processing techniques are used to generate keywords to probe for censorship with. In this paper we present a named entity extraction framework that can extract the names of people, places, and organizations from text such as a news story. Previous efforts to automate the study of keyword-based Internet censorship have been based on semantic analysis of existing bodies of text, such as Wikipedia, and so could not extract meaningful keywords from the news to probe with. We have used a maximum entropy approach for named entity extraction, because of its flexibility. Our preliminary results suggest that this approach gives good results with only a rudimentary understanding of the target language. This means that the approach is very flexible, and while our current implementation is for Chinese we anticipate that extending the framework to other languages such as Arabic, Farsi, and Spanish will be straightforward because of the maximum entropy approach. In this paper we present some testing results as well as some preliminary results from probing China's GET request censorship and search engine filtering using this framework.

Benjamin Edwards

Title:Internet Topology over Time

Abstract—There are few studies that look closely at how the topology of the Internet evolves over time; most focus on snapshots taken at a particular point in time. In this paper, we investigate the evolution of the topology of the Autonomous Systems graph of the Internet, examining how eight commonly-used topological measures change from January 2002 to January 2010. We find that the distributions of most of the measures remain unchanged, except for average path length and clustering coefficient. The average path length has slowly and steadily increased since 2005 and the average clustering coefficient has steadily declined. We hypothesize that these changes are due to changes in peering policies as the Internet evolves. We also investigate a surprising feature, namely that the maximum degree has changed little, an aspect that cannot be captured without modeling link deletion. Our results suggest that evaluating models of the Internet graph by comparing steady-state generated topologies to snapshots of the real data is reasonable for many measures. However, accurately matching time-vaDanny Z. Chenriant properties is more difficult, as we demonstrate by evaluating ten well-known models against the 2010 data.

Benjamin M. Gordon

Title:Progress in Spoken Programming

Abstract—The dominant paradigm for programming a computer today is text entry via keyboard and mouse, but there are many common situations where this is not ideal. For example, tablets are challenging the idea that computers should include a keyboard and mouse. The virtual keyboards available on tablets are functional in terms of entering small amounts of text, but they leave much to be desired for use as a keyboard replacement. Before tablets can can become truly viable as a standalone computing platform, we need a programming environment that supports non-keyboard programming. An introduction to this research was presented at the UNM CS Student Conference in 2011 [8]. In this paper, we describe progress and lessons learned so far.

Chayan Chakrabarti

Title:Enriching Chatter Bots With Semantic Conversation Control

Abstract—Businesses deploy chatter bots to engage in text-based conversations with customers that are intended resolve their issues. However, these chatter bots are only effective in exchanges consisting of question-answer pairs, where the context may switch with every pair. I am designing a semantic architecture that enables chatter bots to hold short conversations, where context is maintained throughout the exchange. I leverage specific ideas from conversation theory, speech acts theory, and knowledge representation. My architecture models a conversation as a stochastic process that flows through a set of states. The main contribution of this work is that it analyses and models the semantics of conversations as entities, instead of lower level grammatical and linguistics forms. I evaluate the performance of the architecture in accordance with Grice's cooperative maxims, which form the central idea in the theory of pragmatics.

Dewan Ibtesham Shafi

Title:On the Viability of Compression for Reducing the Overheads of Checkpoint/Restart-based Fault Tolerance

Abstract—The increasing size and complexity of high performance computing (HPC) systems have lead to major concerns over fault frequencies and the mechanisms necessary to tolerate these faults. Previous studies have shown that state-of-the-field checkpoint/restart mechanisms will not scale sufficiently for future generation systems. Therefore,

optimizations that reduce checkpoint overheads are necessary to keep checkpoint/restart mechanisms effective. In this work, we demonstrate that checkpoint data compression is a feasible mechanism for reducing checkpoint commit latency and storage overheads. Leveraging a simple model for checkpoint compression viability, we show: (1) checkpoint data compression is feasible for many types of scientific applications expected to run on extreme scale systems; (2) checkpoint compression viability scales with checkpoint size; (3) user-level versus system-level checkpoints bears little impact on checkpoint compression viability; and (4) checkpoint compression viability scales with application process count. Lastly, we describe the impact checkpoint compression might have on projected extreme scale systems.

Jeffrey Knockel

Title:Three Researchers, Five Conjectures: An Empirical Analysis of TOM-Skype Censorship and Surveillance

Abstract—We present an empirical analysis of TOM-Skype censor- ship and surveillance. TOM-Skype is an Internet tele-phony and chat program that is a joint venture between TOM Online (a mobile Internet company in China) and Skype Limited. TOM-Skype contains both voice-over- IP functionality and a chat client. The censorship and surveillance that we studied for this paper is specific to the chat client and is based on keywords that a user might type into a chat session. We were able to decrypt keyword lists used for censor- ship and surveillance. We also tracked the lists for a period of time and witnessed changes. Censored keywords range from obscene references, such as 二女一杯 (two girls one cup, the motivation for our title), to specific passages from 2011 China Jasmine Revolution protest ınstructions, such as 成都 春熙路麦当劳门前 (McDonald's in front of Chunxi Road in Chengdu). Surveillance keywords are mostly related to demolitions in Beijing, such as 灵境胡同拆迁 (Ling Jing Alley demolition). Based on this data, we present five conjectures that we believe to be formal enough to be hypotheses that the Internet censorship research community could potentially answer with more data and appropriate computational and analytic techniques.

Joshua Hecker

Title:Formica ex Machina: Ant Swarm Foraging From Physical to Virtual and Back Again

Abstract— Ants use individual memory and pheromone communication to achieve effective collective foraging. We implement these strategies as distributed search algorithms in robotic swarms. Swarms of simple robots are robust, scalable and capable of exploring for resources in unmapped environments. We test the ability of individual robots and teams of three robots to collect tags distributed at random and in clustered distributions. Teams of three robots that forage based on individual memory without communication collect RFID tags from all three distributions approximately twice as fast as a single robot using the same strategy. Adding pheromone-like communication in the teams of three robots improves foraging success. Our simulation system mimics the foraging behaviors of the robots and replicates our results, with slight improvements in the three robot teams. Simulated swarms of 30 and 100 robots collect tags 8 and 22 times faster than teams of three robots. This work demonstrates the feasibility of programming large robotic swarms for collective tasks such as retrieval of dispersed resources, mapping and environmental monitoring. It also lays a foundation for evolving collective search algorithms in silico and then implementing those algorithms in machina in robust and scalable robotic swarms.

Kimberly Kanigel Winner

Title:Ovarian cancer relapse: micro-carcinomas growing from microscopic residual disease vary in form with peritoneal niche

Abstract— In ovarian cancer, the majority of patients are not diagnosed until surgery is needed to remove large tumor masses. Therefore the earliest stage at which chemotherapy and other therapies can be given to such patients is after debulking surgery, at which time surgeons describe the state of the cancer as "microscopic residual disease". Morphology of subsequent tumors is highly dependent on local physical and chemical characteristics of tissues to which they attach in the peritoneal cavity. We use an integrated experimental and modeling approach to study the microscopic tumor growth. A mouse xenograft model of an intraperitoneal injection of SKOV3ip ovarian cancer cells recapitulates microscopic residual disease. Morphometric and rate-of-process data from the mouse model is used to parameterize mesoscopic (cell-scale) cellular Potts models of micro-tumor morphologies on two surfaces: mesothelium overlying muscle, and dual mesothelia overlying a thin, fatty membrane (mesentery). We incorporate fundamental tumor development processes of cell growth, division, chemotaxis, extracellular matrix degradation, tissue invasion, nutrient consumption, and angiogenesis driven by tumor VEGF secretion. Model results suggest that tumor morphology and location depend upon a."tightness" of cellular junctions in deeper tissue and b.chemotactic chemical gradients from different tissues affecting spheroid travel. We have also created a model of angiogenesis in early tumors. These models show great potential of modeling the response by ovarian micro-carcinomas to drugs and immunotherapy in peritoneal micro-environments.

Mahnush Movahedi

Title:Breaking the O(nm) Bit Barrier: Secure Multiparty Computation with a Static Adversary

Abstract— We describe scalable algorithms for secure multiparty computation (SMPC). We assume a synchronous message passing communication model, but unlike most related work, we do not assume the existence of a broadcast channel. Our main result holds for the case where there are n players, of which a $1/3-\epsilon$ fraction are controlled by an adversary, for any positive constant. We describe a SMPC algorithm for this model that requires each player to send $\tilde{O}(\frac{n+m}{n}+\sqrt{n})$ messages and perform $\tilde{O}(\frac{n+m}{n}+\sqrt{n})$ computations to compute any function f, where m is the size of a circuit to compute f. We also consider a model where all players are selfish but rational. In this model, we describe a Nash equilibrium protocol that solve SMPC and requires each player to send $\tilde{O}(\frac{n+m}{n})$ messages and perform $\tilde{O}(\frac{n+m}{n})$ computations. These results significantly improve over past results for SMPC which require each player to send a number of bits and perform a number of computations that is $\Theta(nm)$.

Michael Janes

Title:Life Wont's Wait! (on the Slowdown of Asynchronous Automata Networks)

Abstract— Nakamura [1974], and later independently, Toffoli [1978] and Nehaniv [2002] proved that any synchronous cellular automaton can be simulated by an asynchronous cellular automaton. Their constructions effectively discard many of the updates in order to create a limited artificial synchrony between the cells. We consider the over- head cost of this procedure, assuming that cells update in random order. In particular, we prove explicit upper bounds on this overhead that depend only on the maximum degree of the network.

Neal Holtschulte

Title:Optimal Population Size in Island Model Genetic Algorithms

Abstract—American Proverb: Two's company, three's a crowd. Genetic Algorithm Proverb: One's a hill climber and a thousand's random search. Population size is one of the key parameters affecting the success of genetic algorithms (GAs). Assuming a limited number of fitness evaluations (the most time-intensive factor in virtually all optimization problems), there exists an optimal population size for a genetic algorithm for a given application. Intuitively, a GA with population size one is a hill climber and a GA with maximal population size performs random search. Somewhere in between lies the sweet spot. The Island Model GA divides a single population into semi-isolated subpopulations connected by migration. On the extreme of high migration, the subpopulations function as a single large population. On the extreme of no migration, the subpopulations mights as well be independent runs of smaller population size GAs. Somewhere in between lies the sweet spot. In this paper we propose to explore the dynamics of optimal population size as a function of migration in island model GAs.

Nick Malone

Title:Implementation of an Embodied General Reinforcement Learner on a Serial Link Manipulator

Abstract— BECCA (a Brain-Emulating Cognition and Control Architecture software package) was developed in order to perform general reinforcement learning, that is, to enable unmodeled embodied systems operating in unstructured en- vironments to perform unfamiliar tasks. It accomplishes this through automatic paired feature creation and reinforcement learning algorithms. This paper describes an implementation of BECCA on a seven Degree of Freedom (DoF) Barrett Whole Arm Manipulator (WAM) undergoing a series of experiments designed to test the reinforcement learner's ability to adapt to the WAM hardware. In the experiments, the following is demonstrated, 1) learning to transition the WAM between states, 2) learning to perform at near optimal levels on one, two and three dimensional navigation tasks, 3) applying learning in simulation to hardware performance, 4) learning under inconsistent, human-generated reward, and 5) combining the reinforcement learner with Probabilistic Roadmap Methods (PRM) to improve scalability. The goal of the paper is to demonstrate both the scalability of the BECCA reinforcement learning approach using different formulations of the state space and to show the approach in this paper operating on complex physical hardware.

Soumya Banerjee

Title:Quantication of Uncertainty in Parameters Characterizing Within-Host West Nile Virus Infection

Abstract—West Nile virus (WNV) is a neurotropic avivirus that has emerged globally as a signicant cause of viral encephalitis. Currently, little is known about the within-host viral kinetics of WNV during infection. We used a series of mathematical models of increasing complexity to examine WNV dynamics in mice and birds. To the best of our knowledge, this is the rst eort to model within-host dynamics of WNV. We use a computationally intensive method to

quantify the uncertainty in parameter estimates given uncertainty in input parameters. We set up a framework to explore really large search spaces after imposing constraints from biology. Our method of quantifying uncertainty estimates of model parameters in terms of uncertainty in input parameters could be more generally applicable to modeling of other diseases where precise estimates of input parameters are hard to obtain.

Sunny Fugate

Title:These go to eleven: Cranking up the knobs on IDS scaling performance

Abstract—Signature-based intrusion detection system (IDS) approaches represent the brunt of modern threat detection methods. This is primarily due to their specificity and low false-positive rates and in spite of scalability issues. The inherent scaling issues have meant that measurements of these systems generally ignore the scaling of systems beyond conventional parameter spaces. In particular, while signature-based systems are conventionally measured for total packet processing throughput and false alarm rates, performance is highly dependent on ruleset size. While IDS packet processing performance may appear to be well understood, IDS scaling performance has not been adequately characterized beyond available rulesets. In this paper I present my measurement methods, describe a straightforward method for generating large random rulesets, and present an analysis of the scaling performance of the Snort IDS system.

Yaojia Zhu

Title:Oriented and Degree-generated Block Models: Generating and Inferring Communities with Inhomogeneous Degree Distributions

Abstract—The stochastic block model is a powerful tool for inferring community structure from the topology of a network. To deal with the fact that, unlike most real-world networks, the stochastic block model predicts a Poisson degree distribution within each community, Karrer and Newman recently generalized it to the degree-corrected block model, which can accommodate inhomogeneous degree distributions within communities. However, because it takes the node degrees as parameters rather than generating them, the degree-corrected block model cannot use the node degrees to help it classify the nodes, and its natural generalization to directed graphs cannot even use the orientation of the edges. In this paper, we present variants of the block model with the best of both worlds: they can take advantage of node degrees and edge orientations in the classification process, while tolerating heavy-tailed degree distributions. We show that for certain kinds of networks, including synthetic networks and networks of word adjacencies in English text, these new block models achieve a higher accuracy than the standard or degree corrected block model.