6th USENIX Workshop on Hot Topics in Security (HotSec '11)

August 9, 2011
San Francisco, CA

Welcome

Program Chair: Patrick McDaniel, Pennsylvania State University

Summarized by Rik Farrow (rik@usenix.org)

Patrick McDaniel, the chair of HotSec, explained how he and the PC had decided to revitalize the workshop. Their acceptance rate was 17%, and they included papers that might not otherwise be accepted—for example, for new ideas that are not yet well developed. He said that the format would be three 15-minute presentations followed by 45 minutes of discussion. Session chairs had prepared questions to help get the conversation moving, and he expected the attendees to ask their own questions as well.

New Age System Security

Summarized by Julie Ard (julieard@gmail.com)

Building Secure Robot Applications

Murph Finnicum and Samuel T. King, University of Illinois

Murph Finnicum described how increasing use of robots (Roomba, PR2) requires us to consider their unique security issues. There are many differences between robots and computers, including the fact that robots move around and have inherently probabilistic interactions. The immediate consequences of bad behavior are also much worse, although this line is becoming blurred by cyber-physical systems. For example, improper disclosure of proprietary data or loss of data can result from bad behavior directed at conventional information systems, but a robot's bad behavior could result in your house being burned down or harm to a human being.

Much of the presentation and discussion revolved around fundamental differences between robots. They include probabilistic identification, privacy, and permissions for applications. Because robots will go out into the world and interact, you cannot simply write a program identifying what they can and cannot do. The number of objects, for example, that a robot could pick up is infinite. Orders will be given by one human, and interactions would be with other humans—for example, consider a robot going to get coffee. Facial, voice, and location-based recognition do not guarantee the parameters of an operating environment but provide only probabilistic parameters.

Logging provides a necessary infrastructure for accountability, but this may violate humans' privacy. Robots will have a flawless and complete memory, and one fundamental difference to bear in mind is that unlike computers, humans don't choose where they will interact with robots. A robot could be required to notify humans when it is recording. However, can the infrastructure identify whether surrounding humans are aware that a robot is present and in operation if the robot's presence is not obvious? Perhaps humans could identify their preferences for information sharing, such as "only friends can know my location."

Finally, to carry out a task, a robot would have to take actions on behalf of the user. Permissions would have to involve high-level constructs, such as moving short distances or within a specified area. A discussion of robot behavior and morality followed, based on popular literature and movies, including Asimov's "Three Laws" and the concept of surrogates.

Security Fusion: A New Security Architecture for Resource-Constrained Environments

Suku Nair, Subil Abraham, and Omar Al Ibrahim, Southern Methodist University

Omar Al Ibrahim conveyed how the concept of "security fusion" aims to move complexity from the components to the system level in resource-constrained devices such as sensor and SCADA systems. These devices are characterized by constrained attributes such as gate count, memory, power consumption, bandwidth, physical size, and processing power. The authors propose exploring how these simple structures can lead to emergent security.

Traditional security is not possible on these devices, because the resource constraints may preclude cryptography, energy is limited, there are numerous devices deployed, nodes can be easily compromised, and oftentimes they use a wireless medium.

We were introduced to a "state machine model," which is promising and feasible for this application because resource-constrained devices are less complex than computers. Finite automata concepts will be explored in future work. Inherent in this will be a comparison of the growth of software versus hardware security complexity. The discussion resulted in a suggestion of considering what an adversary could do given a certain number of compromised nodes, in addition to the author's direction of determining how many nodes need to be compromised for an adversary to achieve a particular malicious goal.

DISTROY: Detecting Integrated Circuit Trojans with Compressive Measurements

Youngjune L. Gwon, H.T. Kung, and Dario Vlah, Harvard University

Youngjune L. Gwon began with background information on modern manufacturing methods and third-party involvement making it difficult to determine whether the received silicon is strictly what was ordered. The authors focused on power or current side-channel measurement analysis to detect trojans in integrated circuits (ICs). In particular, they explored driving the IC to a low-power state so that the trojan's power signature would be more pronounced. Their goal is to identify test vectors that will reveal anomalies indicative of trojans.

Compressive sensing is a signal processing technique for recovering data with the number of measurements proportional to the sparsity of data. However, the reduced measurements tradeoff results in an increase in false positives. One method of reducing false positives is by testing multiple chips from the same fabrication process. Additional explorations will include tradeoffs in the number of test measurements needed to reduce false positives to an acceptable level. Scalability of test vectors is a necessary factor for application of this approach. The discussion segued into supply-chain security, which is a human problem.

Privacy and Anonymity

Summarized by Ryan MacArthur (ryan.macarthur@gmail.com)

Privacy-Preserving Applications on Smartphones

Yan Huang, Peter Chapman, and David Evans, University of Virginia

Peter Chapman covered the important topic of smartphone applications that actually consider users' privacy. The phone that you carry around with you contains very personal information, be it contacts, location history, pictures, email, or banking payment records. Chapman covered an application that was built to securely "make friends" with neighboring devices, so-called "mutual contact discovery." It is known that trust is an issue, and, given evil devices, we cannot trust a device with such private data. So the common theme here is to interact with others and secure our data.

Currently this is achieved through a trusted third party such as a social media site, bank, or video game producer. The trusted third party has become the "untrusted" third party, with cases of major corporations losing massive amounts of data (e.g., Sony, Citi, Sega). To remove the third party, Chapman discusses the usefulness of the "garbled circuit protocol" proposed by Yao in the '80s. You can think of it as collective voting, implemented securely in Java. Implementation problems arise using certain immutable Java classes, and novel optimizations were developed to achieve impressive speedups. The beta version of their application is able to anonymously find common contacts with a peer, with a performance of 128 contacts in 150 seconds. Future directions are leveraging the carrier for peer discovery, software-based attestation, and lower-level (OS) support to handle secure communications.

Public vs. Publicized: Content Use Trends and Privacy Expectations

Jessica Staddon and Andrew Swerdlow, Google

For this talk, Jessica described the studies conducted on users concerning privacy expectations during their normal interactions with Internet-based services they use on a daily basis. They apparently took pains to use a diverse pool of global candidates, creating a diverse human study on current interest in privacy. There seems to be a common misconception as to where users' data actually goes and how it can be used. Staddon proposed three major categories to improve privacy expectations: transparency—in-context awareness of where data is going; control—data-use settings that users understand and can find; utility—users being given the data they need to make informed choices.

Herbert West—Deanonymizer

Mihir Nanavati, Nathan Taylor, William Aiello, and Andrew Warfield, University of British Columbia

Mihir delivered a comical talk describing efforts toward identifying authors of critical paper reviews. They collected reviews from program committees and utilized machine learning through a naive Bayes classifier utilizing NLTK in Python. They trained on unigrams, bigrams, and trigrams scored on an authorial basis using TF-IDF. The results were very interesting, as they were able to mimic the voice of PC members. It seems that simple machine classifiers are capable of identifying supposed anonymous reviews. Someone suggested that humans are good classifiers; you know whose paper it is 90% of the time. Ted Faber (USC/ISI) followed up by asking whether humans really are such good classifiers. Mihir replied that to reduce the set of possible candidates, you should use both computer and human techniques. Sandy Clarke disagreed with fingerprinting, citing Rachel Greenstadt's work at Drexel, where they found that if people disguised their own styles detection becomes impossible.

Discussions

Summarized by Rik Farrow (rik@usenix.org)

Peter Chapman was questioned about their privacy-preserving Android app. Bill Aiello (UBC) wanted to know if they had examined other work, such as fair play, as opposed to the garbled circuit implementation that they had used. Bill also wanted to know whether the authors could imagine a library of best implementations to help other developers solve these issues. Peter responded that fair play is the most famous of the garbled circuits, and he suggested checking out the Telex paper (Wustrow) that was presented on Friday. Patrick Traynor wondered whether there was a semantic difference in the results of searches performed this way. Peter said that there was not and added that their approach to performing the calculations were orders of magnitude faster. Franzi Roesner (University of Washington) considered a denial-of-service attack where the attacker would make lots of trivial changes to her address book. Peter responded that they could limit the number of times the protocol could be run with a particular partner.

Patrick Traynor asked Jessica Staddon whether we aren't already warning users of the potential for publicizing their posts or responses. Jessica replied that privacy policies are, for the most part, impenetrable. Mike Ryan (USC/ISI) pointed out that Google+ has summaries of parts of the EULA in the plainest language, such as "Google will not resell your pictures." Perry Metzger (University of Pa) said that using simply and clearly worded privacy policies is totally legal and it is just custom to word them in impenetrable legalese. John Springer of USC mentioned that Laurie Kramer at CMU has done a lot of work on copyright, a related area. Patrick McDaniel wondered about some of the results in the paper: that over half of the participants didn't realize that their posts would become public. Jessica said that a surprising number of people are not as aware as they should be. Vern Paxson commented that people don't value their private discussions and also pointed out that there is no visibility for the cost of giving away your privacy. Jessica summarized by saying we could be doing far more than we are doing now, particularly in social networks.

The paper on de-anonymizing referees' reviews generated discussion among PC members about the culture of reviewing. Ted Faber asked if the researchers had compared results from humans to results from their classifier. Mihir Nanavati said that they hadn't, as they only read those reviews where their classifier had failed in an attempt to discover why, and that algorithmic classifiers work differently from people: people tend to pick out features that algorithms ignore. Patrick Traynor thought that perhaps he should get his graduate students to write his reviews, causing Patrick McDaniel to quip, "They don't already?" Traynor responded that he isn't tenured yet.

Information Protection

Summarized by Ryan MacArthur (ryan.macarthur@gmail.com)

Towards Practical Avoidance of Information Leakage in Enterprise Networks

Jason Croft and Matthew Caesar, University of Illinois at Urbana-Champaign

Jason Croft points out that we need to differentiate between sensitive and nonsensitive data, network-wide. Problematically, protecting and configuring data against theft is challenging, as has been indicated by recent attacks on large amounts of sensitive data. Better protection is needed. Previous work (Tightlip) tends to focus on data protection at the machine level. This limits the functionality of applications that demand that data be shared, and also incurs high overhead, as each machine needs to be configured properly.

Croft presented a technique using shadow processes to compare between the original process and one where sensitive data has been scrubbed. The two processes are synchronized at system calls, where both receive the same results. When compared, if the two streams of data match, then it is safe to share. One problem they encountered is false positives relating to data that is nonsensitive. The current implementation achieves a 2x slowdown, where they hook read/write APIs to compare data. An audience member was concerned with encrypted data, but since this implementation marks data as sensitive before encryption, it is not a concern. A majority of questions hinged on the fact that managing such a system is an administrative nightmare.

Towards Client-side HTML Security Policies

Joel Weinberger, University of California, Berkeley; Adam Barth, Google; Dawn Song, University of California, Berkeley

The landscape of local HTML security offerings was detailed in this talk by Joel Weinberger. A history of attacks was given, most notably the Samy worm that wreaked havoc on MySpace. The argument was made that we need to segregate elements of content on Web pages into trusted and untrusted as a first step. The next step would be to implement a policy to deal with both types of data. Web application frameworks like RoR and Django were mentioned as proving weak amounts of policy relating to trust levels of data. It was also made clear that sanitization is hard, and we have been failing to do it properly for a while. It is for these reasons that explicit policies on how to manage both kinds of data need to be implemented. Weinberger introduced three off-the-shelf solutions—BEEP, BLUEPRINT, and Content Security Policy (CSP)— and listed the pros and cons of each. BEEP focuses on preventing XSS by whitelisting scripts. BLUEPRINT uses its own trusted JavaScript parser and the blueprint, a security policy. CSP is actually included in Firefox 4. The authors ported two applications, Bugzilla and HotCRP, to determine the impact on developers and on performance. The porting effort was substantial, because CSP does not support dynamic script generation. The performance hit was between 35% and 55%. In conclusion, Weinberger suggested that a combination of whitelisting, as found in BEEP for inline scripts, and CSP might work.

An audience member asked, "Does performance really matter?" Weinberger's response was that performance is a big deal. Rewriting the application is part of the performance issue.

TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion

Liang Cai and Hao Chen, University of California, Davis

A novel proof-of-concept keylogger was presented by Hao Chen. The technique utilizes hardware devices normally thought of as safe, such as accelerometers and gyroscopes. Using a custom keyboard, Chen described how they are able to track which finger is tapping which section of the screen and were able to recreate entered text. The concept is in its early stages and was not implemented on a stock touchscreen keyboard. The hardware utilized by Chen et al. is readily accessible through JavaScript, which is a non-privileged interpreter. An audience member suggested discovering the handedness of the target, then optimizing for the detected hand. It was also clarified that the test trials had the targets sitting still with phones in hand, so any movement of the person holding the phone, such as walking or riding, was avoided.

Emerging Areas in Security

Summarized by Ryan MacArthur (ryan.macarthur@gmail.com)

On Dynamic Malware Payloads Aimed at Programmable Logic Controllers

Stephen McLaughlin, Pennsylvania State University

Stephen McLaughlin tackled the tough problem of generating a process dependency graph for logic variables, with the goal of exploiting interlocking variables in PLCs. The interlocking variables may represent safety controls, never exceeding, for example, a particular speed in a controlled device. He reviewed common systems that utilize these controllers. The Stuxnet sample was explained, as it contained a precompiled PLC payload. This indicated that the Stuxnet authors had a priori knowledge of the system they were attacking.

McLaughlin postulates that writing malware to overcome the obscurity of process control systems is an engineering problem.

He has created a system that takes binary code and translates it into an intermediate language code, and then translates that even further into Boolean expressions, all with the intent of inferring device types and interlocking variables. The goal would be to create an intelligent exploit that would determine how to manipulate key controls to wreak havoc.

Effective Digital Forensics Research Is Investigator-Centric

Robert J. Walls, Brian Neil Levine, and Marc Liberatore, University of Massachusetts Amherst; Clay Shields, Georgetown University

Walls argued that digital forensics lacks a solid scientific foundation. Without such a foundation, it becomes difficult to successfully prosecute alleged offenders. Digital forensics is inherently investigator-centric, and as such the research should be driven by the investigator, not the prosecutor. The problem we all seem to face is that forensics and the law are inseparable, yet the law is always struggling to keep up.

Investigations are about people and their actions, and intent is left out of the security domain. Walls provided simple rules, which hold close to Occam's razor, to follow for creating new policies around digital forensics. Someone made the comment that forensics show that a suspect did something with a computer, but computers do things without the owner taking action, so it is hard to prove ownership over many low-level computing functions. One open question was around the underlying issues in forensics, such as the burden of proof: how do we support the law?