Cyber Security: Challenges for the Future

Prof. Shambhu Upadhyaya
Department of Computer Science and Engineering
University at Buffalo

Presentation at Science and Technology Forum
October 28, 2009

Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward
- Video Presentations

A Famous Quote on Security

- “If you think technology can solve your security problems, then you don't understand the problems and you don't understand the technology”

Bruce Schneier, computer security specialist

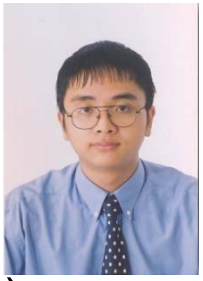
- “Security is only as strong as the weakest link”

Pop Quiz !

- What is the most insecure place on earth?
  - Answer: Internet
- What is the most heavily used network-based application on the Internet?
  - Answer: email
- Who are the most famous hackers of all time?
  - Answer: Jonathan James, Adrian Lamo, Kevin Mitnick
- Who is the father of modern cryptography?
  - Claude Shannon, the information theorist

Acknowledgments

- Graduate students
  - Sunu Mathew (Ph.D.)
  - Duc Ha (Ph.D.)
  - Madhu Chandrasekaran (Ph.D.)
  - Mohit Virendra (Ph.D.)
  - S. Vidyaraman (Ph.D.)
  - Chris Crawford (MS)
- Colleagues
  - Prof. Hung Ngo
  - Dr. Kevin Kwiat
- Funding agencies
  - NSA, NSF, DARPA
- Google Images

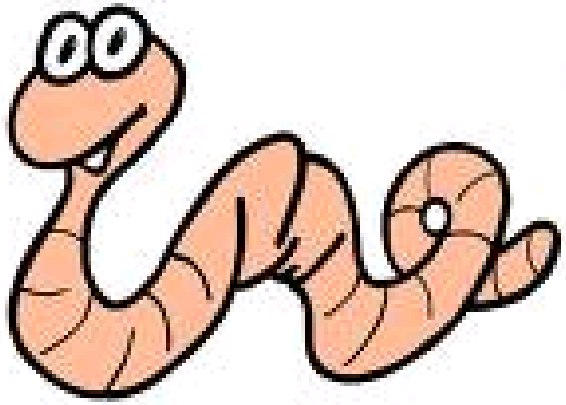



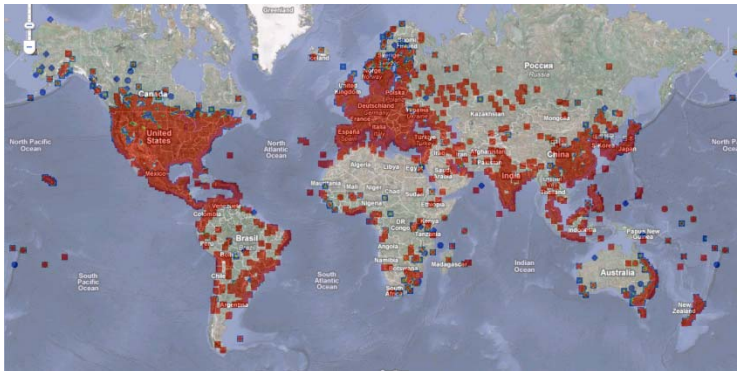



Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward

Computer Security Incident 1

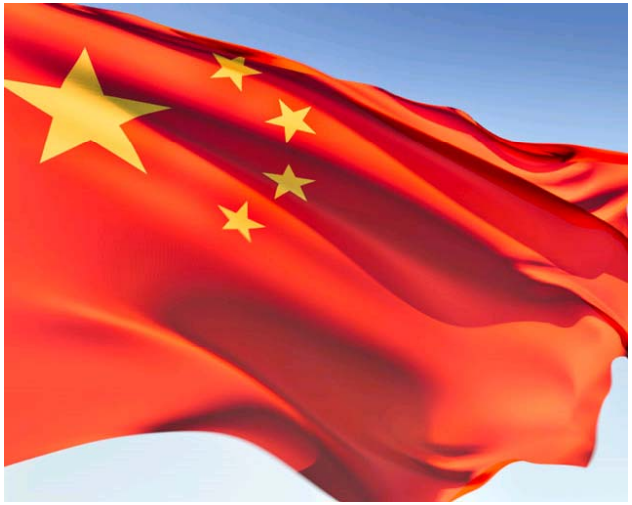




New York Times, January 22, 2009

- A new digital plague (Conficker) has hit the Internet, infecting millions of personal and business computers in what seems to be the first step of a multistage attack. The world's leading computer security experts do not yet know who programmed the infection, or what the next stage will be
- Supposed to have unleashed massive attack on April 1, 2009 – turned out to be a hoax!
- Could mean a digital "Pearl Harbor"

Computer Security Incident 2



Wall Street Journal, April 21, 2009

- Computer Spies Breach Pentagons's Fighter-Jet Project
- Hackers broke into DoD computers and downloaded terabytes of data containing design information about the Joint Strike Fighter, a $300 billion stealth fighter currently under development

Computer Security Incident 3

Wall Street Journal, April 8, 2009

- Electricity Grid in U.S. Penetrated By Spies
- Cyberspies have penetrated the U.S. electrical grid and left behind software programs that could be used to disrupt the system, according to current and former national-security officials

Computer Security Incident 4

- Aldrich Ames (Notorious Insider), a former CIA counterintelligence officer and analyst, sold-out his colleagues to the Russians for more than $4.6 million, was convicted of spying for the Soviet Union and Russia in 1994

- Robert Hanssen (Notorious Insider), Caught selling American secrets to Moscow for $1.4 million in cash and diamonds over a 15-year period, Sentenced for life in prison without the ability for parole in 2002, Photo Courtesy: USA Today
- Have you watched the movie – Breach?
- Try this link: http://www.rottentomatoes.com/dor/objects/868028/breach/videos/breach_020507.html

Security Breach Reports

- CERT Coordination Center
  - Located at CMU's Software Engineering Institute
- US-CERT
  - U.S. Computer Emergency Readiness Team
- NY State Cyber Security and Critical Infrastructure Coordination (CSCIC)

- SANS Institute Storm Center

Types of Attacks

- Threats to national security
- Cognitive hacking
  - Manipulating user's perception
  - "Killing" of Britney Spears (Oct. 2001)
- Worm attacks
  - Sasser Worm (May 2004)
- Virus attacks
  - SoBig.F (Aug. 2003), > $50M damage
  - NIMDA virus in Sept. 2001
- DoS attacks
  - Yahoo, Amazon, eBay, CNN (Feb. 2000)
- SQL injection attacks
  - UN Website defacing (8/12/07)

Web News (Real?)

CNN.com / ENTERTAINMENT

SEARCH GO

MAIN PAGE
WORLD
U.S.
WEATHER
BUSINESS
SPORTS
POLITICS
LAW
SCI-TECH
SPACE
HEALTH
ENTERTAINMENT
TRAVEL
EDUCATION
CAREER
IN-DEPTH

CNN Fe STORE

QUICK NEWS
LOCAL
COMMUNITY
MULTIMEDIA
E-MAIL SERVICES
NEWS ON PDA
ABOUT US

News TV
what's on
show transcripts
Headline News
International
askNews

EDITIONS
CNN.com Asia
CNN.com Europe
set your edition

Languages
Time, Inc.

Singer Britney Spears Killed in Car Accident

October 6, 2001 Posted: 21:56 PM EDT (0156 GMT)

Popular musician Britney Spears was pronounced dead Saturday evening as a result of wounds caused by a car accident. The status of the other passengers in the vehicle is not known at this time.

LOS ANGELES, California (AP) -- A car accident Saturday evening has cost the life of Britney Spears, teen pop music sensation, and has placed fellow passenger and musician Justin Timberlake in critical condition at Los Angeles County Hospital.

Spears and Timberlake were driving through the Los Angeles area late Saturday as both enjoyed a brief moment away from their busy touring and recording schedules when, according to eyewitness reports, their car, a rented Porsche 911, veered suddenly across six lanes of traffic and into a concrete barrier.

Motorists at the scene reported that Spears was ejected from the vehicle at impact and flew into opposing traffic, where she was caught under the wheels of at least one other vehicle before traffic could stop. Timberlake remained safely in the damaged vehicle.

"It was horrible, absolutely horrible," said one witness to the scene, "she was thrown around on the road like a rag doll. We stopped and ran to help but it was obvious there was nothing we could do."

Authorities pronounced Spears dead at the scene, while Timberlake was rushed to emergency surgery to treat internal bleeding.

The cause of the accident is still officially unknown, but witnesses on the scene have told CNN that Timberlake, who was behind the wheel of the vehicle, may have been under the influence of controlled substances, and that Spears may have been engaged in activities that may have distracted him from the road.

"It's too early to say at this point what may have caused the accident," said an officer at the scene, "it appears that the driver lost control of the vehicle suddenly, but we're still determining why that happened."

CNN Shirts and More!
Try Money Magazine Free
Entertainment Weekly
Life Album 2002

A floor lamp that spreads sunshine all over a room...

Scientists adopt NASA technology to create "smart bed" sleep surface...

Micro circuitry technology puts a digital camera, video camera and webcam in your shirt pocket for under $80...

How to make your car invisible to radar and laser...

Power and cyclonic action create one incredible stick vac...

Scientist invents easy solution for hard water problems...

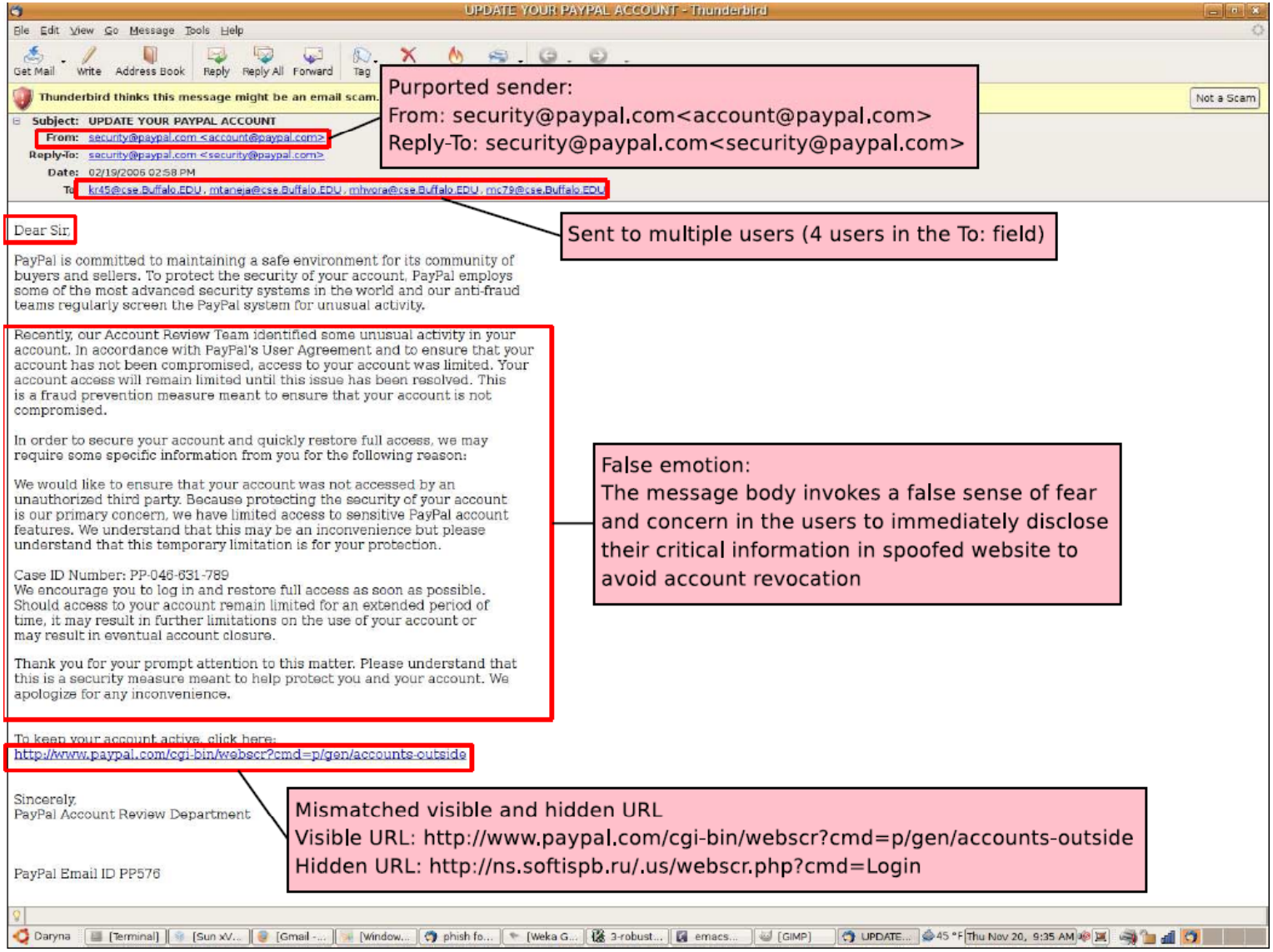
It's time to put all of your photos onto your computer...

If you don't back up your hard drive immediately

Cognitive Hacking

- On Oct. 7, 2001, CNN's top-ranked news story
- Example of a cognitive hacking where you manipulate a user's perception
- These attacks are "hoax" like hoax Virus notifications
- Refer to: IEEE Computer, August 2002 issue:
  - http://www.computer.org/portal/web/csdl/doi/10.1109/MC.2002.1023788
- It began with a spoof of CNN.com
- Through a bug in CNN's software, the article got spread when clicked on "email this article"
- Within 12 hours, more than 150,000 people viewed the spoofed page

Phishing Attacks



Worst Security Mistakes End Users Make

- Failing to install anti-virus, keep its signatures up-to-date, and apply it to all files
- Opening unsolicited e-mail attachments without verifying their source
- Failing to install security patches on favorite applications
- Not making backups
- Using weak passwords

Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward

CSI/FBI Survey

- Annually, the CSI and the FBI release their findings on the survey
- Aims to raise level of security awareness among businesses, educational and medical institutions, and governmental agencies
- Goal to ascertain the type and range of computer crime in the U.S. and to compare annual cybercrime trends with those of previous years

CSI/FBI Survey 2006

- 2006 survey
  - Responses of 616 computer security practitioners in U.S. corporations, government agencies, financial institutions, medical institutions and universities
- The long term trends considered include:
  - Unauthorized use of computer systems
  - The number of incidents from outside as well as inside an organization
  - Types of attacks or misuse detected, and
  - Actions taken in response to computer intrusions

Major Findings

- Virus attacks continue to be the source of the greatest financial losses
- Unauthorized use of computer systems slightly decreased this year
- Use of cyber insurance remains low, but may be on the rise
- The percentage of organizations reporting computer intrusions to law enforcement has reversed its multi-year decline
- Over 80% of the organizations conduct security audits

CSI/FBI Survey 2007

- 2007 survey
  - Average cyber-losses jumping after 5-year decline
  - Average annual loss $168,000 to $350,424 in this year's survey
  - Financial fraud overtook virus attacks as the source of the greatest financial loss
- Additional key findings:
  - 1/5th of respondents said they suffered a targeted attack
  - Insider abuse of network access or e-mail edged out virus incidents as the most prevalent security problem

CSI/FBI Survey 2008

- The most expensive computer security incidents were those involving financial fraud…
- Virus incidents occurred most frequently…
- Almost one in ten organizations reported they'd had a Domain Name System incident…
- Twenty-seven percent of those responding to a question regarding "targeted attacks"…
- The vast majority of respondents said their organizations either had (68%) or were developing (18%) a formal information security policy

Cyber Security Challenges

- Data protection (e.g., data classification, identification and encryption) and application software (e.g., Web application, VoIP) vulnerability, security
- Policy and regulatory compliance (Sarbanes–Oxley, HIPAA)
- Identity theft and leakage of private information (e.g., proprietary information, intellectual property and business secrets)
- Viruses and worms
- Management involvement, risk management, or supportive resources (human resources, capital budgeting and expenditures)
- Access control (e.g., passwords)
- User education, training and awareness
- Wireless infrastructure security
- Internal network security (e.g., insider threat)
- Spyware
- Social engineering (e.g., phishing, pharming) – steal identify

The I3P Report (on Challenges)

- Institute for Information Infrastructure Protection formed a committee to address security R&D challenges
  - Senators Joe Lieberman and Susan Collins
- 4 emerging issues
  - A coordinated and collaborative approach is needed
  - Metrics for security are a broad enabler and must be developed
  - An effective legal and policy framework for security must be created
  - The human dimension of security must be addressed
- http://www.thei3p.org/docs/publications/i3pnationalcybersecurity.pdf

Other Key Documents

- The National Strategy to Secure Cyber Space, Februray 2003 (76 pages)
  - Cyberspace touches practically everything and everyone
  - Leadership from the top
  - http://www.dhs.gov/xlibrary/assets/National_Cyberspace_Strategy.pdf

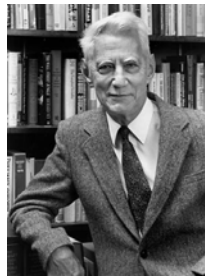
- Cyber Space Policy Review (76 pages)
  - President Obama presented the Cyberspace Policy Review on May 29, 2009
  - Beginning of the way forward towards a reliable, resilient, trustworthy digital infrastructure for the future

Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward

What is Cyber Security?

- Encryption/decryption
  - Symmetric Key and Asymmetric Key Cryptography



- Authentication
  - Kerberos

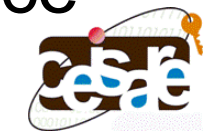
- Program Security
  - Virus, Trojan horse, Malicious code, Covert channels

- Network Security
  - Firewall, Tripwires
  - Electronic mail security, IP security, Web security

- Intrusion Detection
  - Audit trail-based, Concurrent intrusion detection

Security Goals

- Confidentiality
  - Assets are accessible only to authorized parties (privacy)
- Integrity
  - Modification only by authorized parties so that accuracy can be maintained
- Availability
  - Assets accessible to authorized parties always
  - No denial of service
  - Timely response, Fair allocation, Fault tolerance



People Involved – The Bad Guys

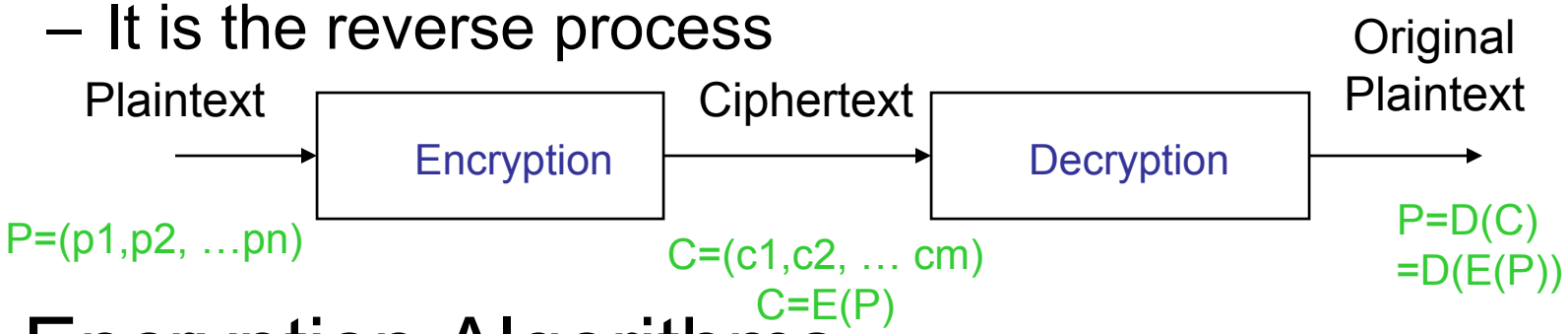
- Ordinary people, teenagers or college students
- Amateurs
  - Most of the crime committed by amateurs
  - They observe a flaw in security and take advantage of
  - There are so many tools publicly available
- Crackers
  - University or high school students
  - Done for no good reason, maybe some kind of self-satisfaction
  - This continues to be an appealing crime, to juveniles
- Career Criminals
  - Do for personal gain, spying

Methods of Defense

- Encryption
  - Coding
  - No encryption is perfect -- weak encryption can actually be worse!
- Software Controls
  - Internal, OS level or Developmental level
- Filters
  - Firewalls

Cryptography Basics

- Encryption
  - A process of encoding a message
- Decryption
  - It is the reverse process



- Encryption Algorithms
  - A key K is generally used
  - Symmetric encryption: $P = D(K, E(K, P))$
  - Asymmetric encryption: $P = D(K_D, E(K_E, P))$

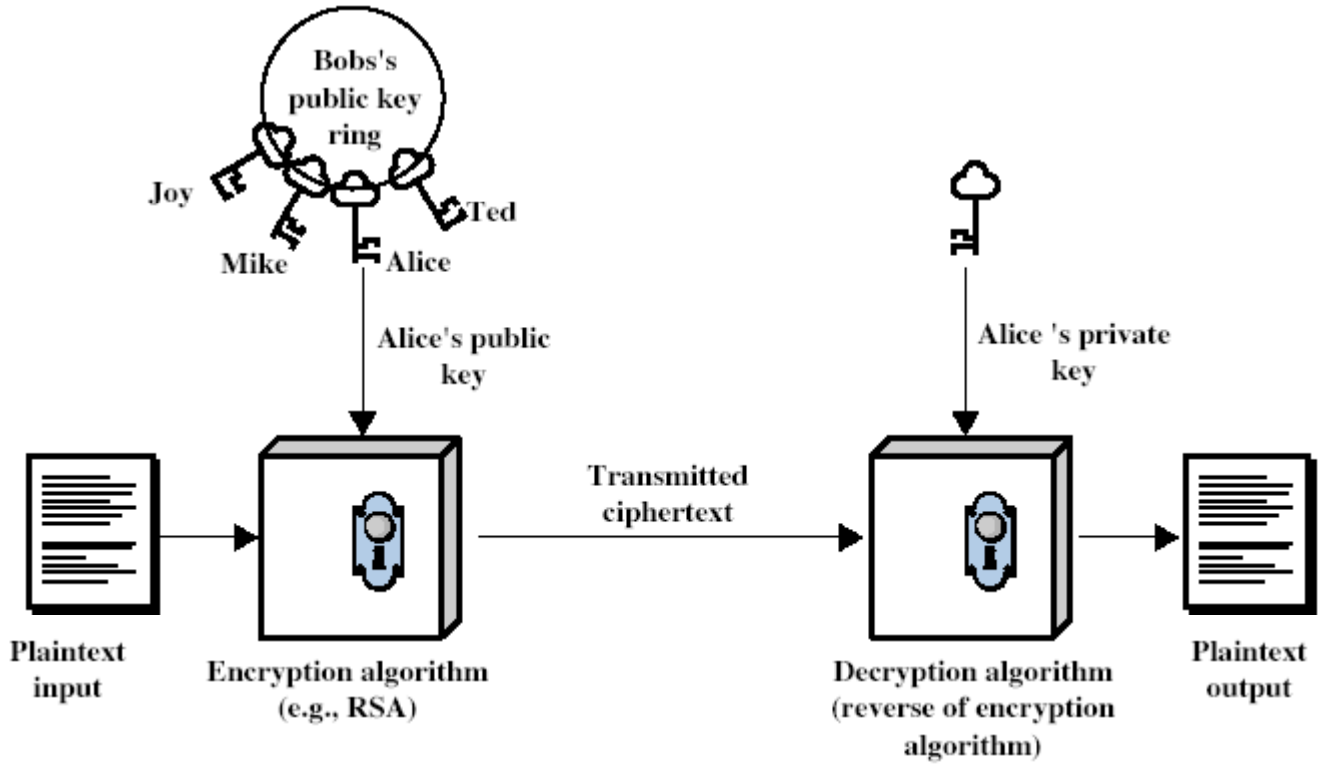
Symmetric Key Encryption

- Data Encryption Standard (DES)
  - Most widely used block cipher in the world, adopted in 1977 by NBS (now NIST)
  - Encrypts 64-bit data using 56-bit key
  - Had widespread use until early 2000
  - Has been subject to considerable controversy over its security
- Advance Encryption Standard (AES)
  - 128-bit data, 128/192/256-bit keys
  - Stronger & faster than Triple-DES
  - Active life of 20-30 years (+ archival use)

Asymmetric Key Encryption (PKI)

- Perhaps the only true revolution in the history of cryptography
- Based on mathematical functions unlike conventional ones
- DES is a significant advance by IBM, but based on substitution and permutation
- PKEs are asymmetric techniques -- use two separate keys
- Enhance confidentiality, key distribution and authentication

PKI Illustration



(Courtesy: William Stallings)

Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward

Computer Science and Engineering

- 25 faculty members, world class researchers
- Ranked 21st in the nation in research funding
- 350 UGs and 220 Grad students

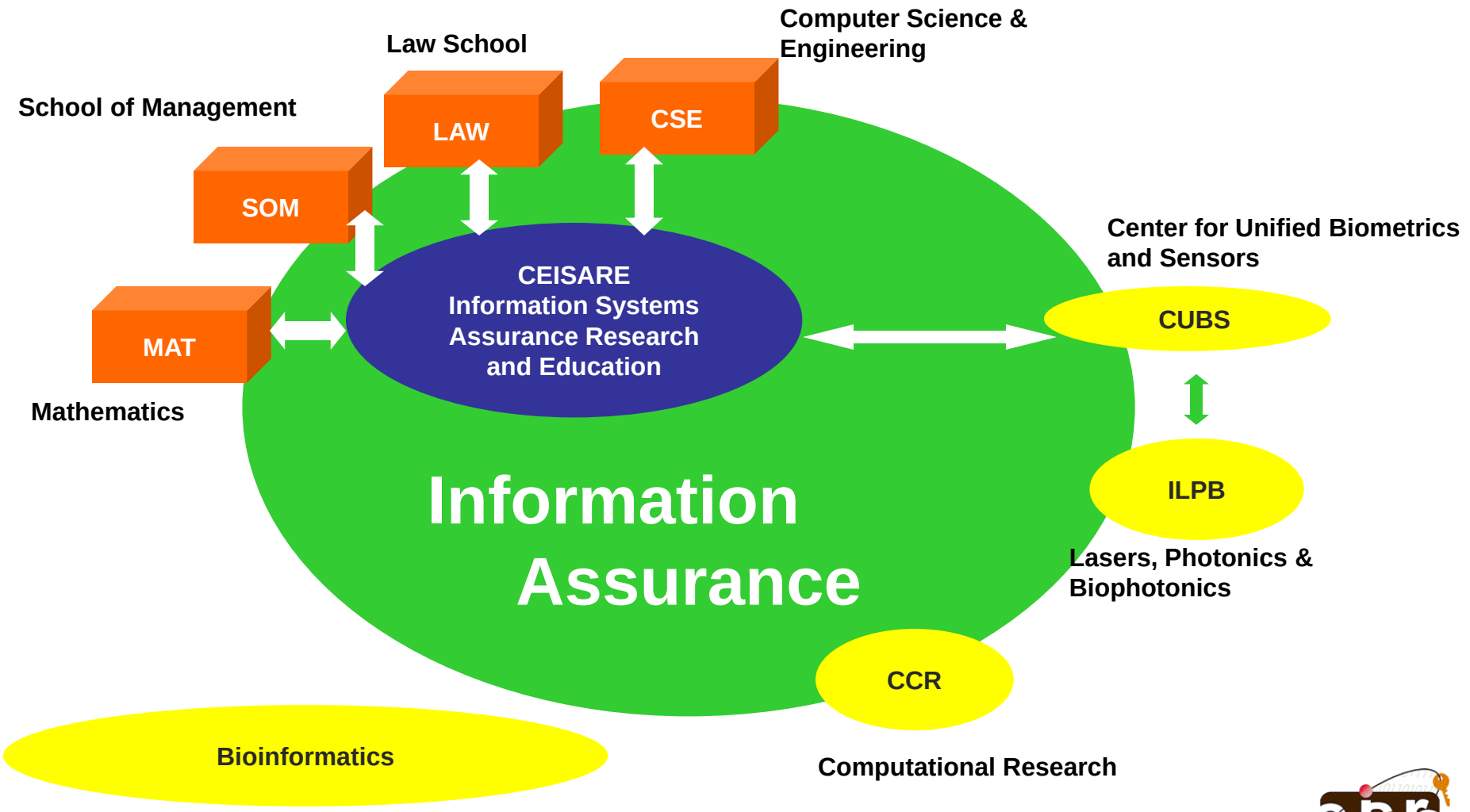
Current Building

Future Building

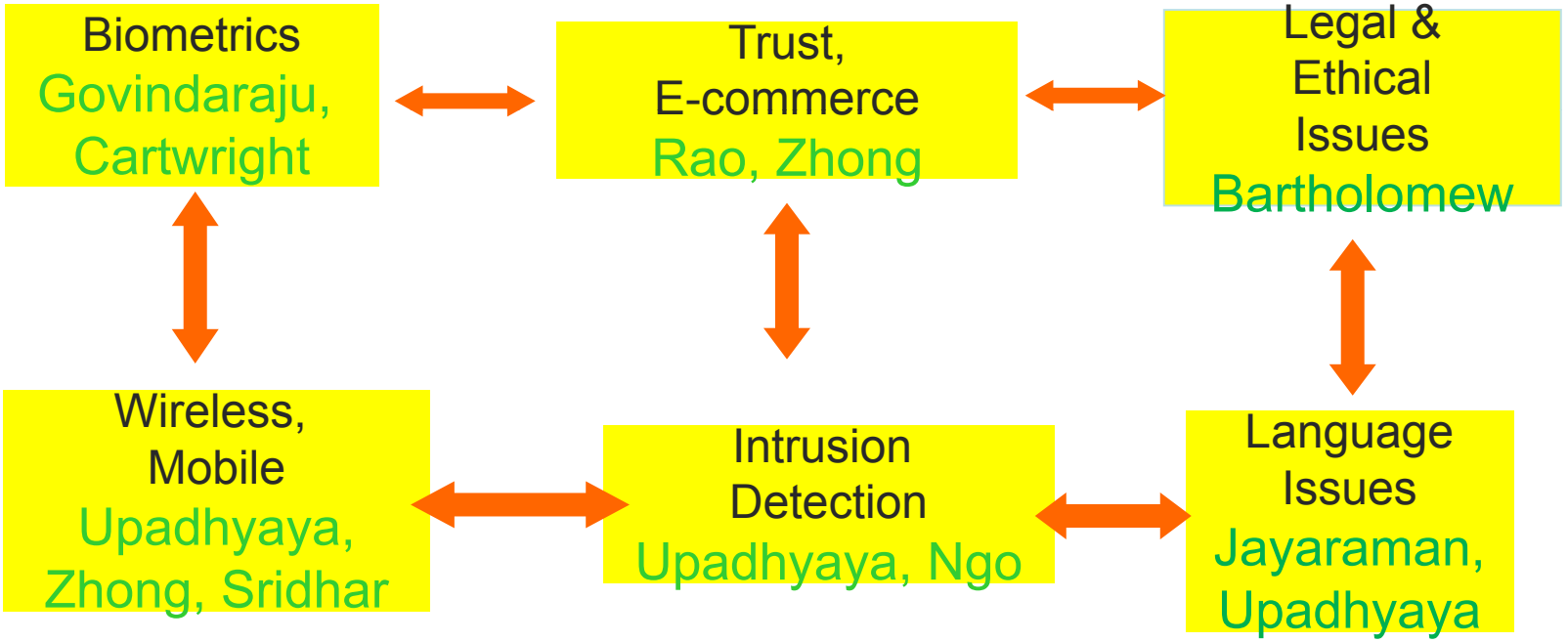
CEISARE

- CEISARE designated as a National Center of Excellence in 2002 by NSA, DHS
  - Through a competitive process
  - We were one of 13 centers designated that year (36 across the country)
  - Today, there are 100+ centers
  -

CEISARE


Computer Science & Engineering
Law School
School of Management
LAW
CSE
SOM
MAT
Mathematics
CEISARE
Information Systems
Assurance Research
and Education
Center for Unified Biometrics and Sensors
CUBS
ILPB
Lasers, Photonics & Biophotonics
Information Assurance
CCR
Bioinformatics
Computational Research

IA Faculty Collaborators



Research & Other Synergistic Activities

- Funding
  - Over 4M from NSF, DARPA, NSA/ARDA, AFRL, DoD (since 2002)
  - Research, education, infrastructure
- Curriculum
  - Cyber security at Ph.D. level
  - Advanced Certificate in IA
  - IASP scholarships (DoD and NSF)
- Workshops
  - SKM 2004, SKM 2006, SKM 2008
  - Local Joint IA Awareness Workshops with FBI, ECC, Local industries
- Outreach Activities
  - High school workshops
  - Minority training

Outreach/Partnerships

- Govt. Partners
  - AFRL, Rome (Research Associate Professor)
  - FBI Cyber Task Force (Infragard, workshop participation)
  - Local government (advisory board membership)
- Industry
  - HP, M&T Bank (advisory board)
  - Intel Corporation (security research sponsorship)
- Academia
  - Hilbert College (NSF Capacity building grant)
  - Erie Community College (NSF ATE grant)
  - Genesee Community College (NSF Capacity building)
  - Polytechnic University (CSAW contest)
  - Purdue University (Forensics initiative)

Graduate Certificate in IA

- Effort started with funds from DoD, 2003
  - Funding was to create a new integrative course in IA
- Two tracks – technical and managerial
- Requirements
  - 6 credits of core courses in the track
  - 5-6 credits of elective in the dept.
  - 3 credits of required integrative course
- Technical track
  - Core – Intro. to Crypto, Computer security, Wireless networks security (choose two courses)
- Managerial track
  - Core – Network management, E-Commerce security

CEISARE Courses

- Courses with IA Content
  - CSE 565 Computer Security
  - CSE 566 Wireless Networks Security
  - CSE 512 Applied Crypto and Computer Security
  - CSE 671 Security in Wireless Ad Hoc and Sensor Networks
  - LAW 629 Computers, Law, Technology and Society
  - LAW 645 Copyright
  - Law 956 E-Commerce Law
  - MGA 615 Fraud Examination
  - MGS 650 Information Assurance
  - MGS 651 Network Management
  - MGS 659 E-Commerce Security
  - MGT 681 Intellectual Property
  - MHI 512 Ethical, Social & Human Factors in Medical/Health Informatics
  - MTH 529/530 Introduction to the Theory of Numbers I/II
  - MTH 535 Introduction to Cryptography
  - MTH 567 Stream Ciphers
- Other Technical Electives
  - http://www.cse.buffalo.edu/caeiae/advanced_certificate_program.htm

Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward

Research Projects

- Most federally funded
- Some industry funded
- Disciplines ranging from Networks Security to Wireless Networks Security

Real-Time Intrusion Detection with Emphasis on Insider Attacks (2003 - 07)

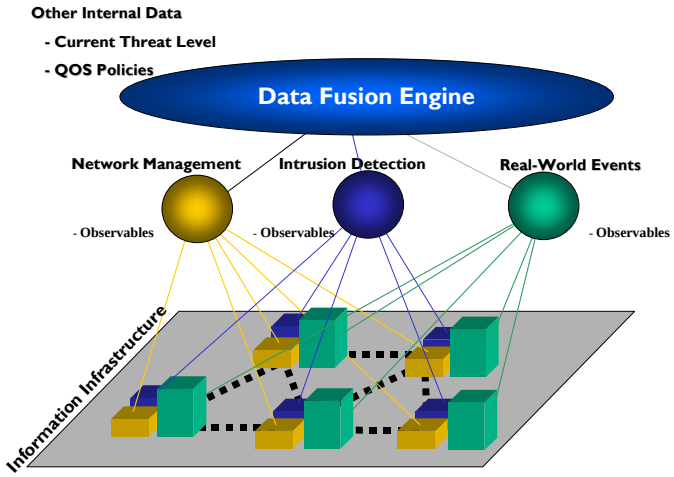
- A novel security system based on the encapsulation of owner's intent
- Can be readily used as a concise reference for monitoring of intrusions
- How – By actively querying the user for his intent
  - Build a small and manageable set of assertions
  - Leads to search space that is more focused
  - System is able to respond faster, make fewer mistakes and scale well
- Moving away from the traditional method of detecting intrusions through low level network and other resource audit, to a much higher level
- Net gain – Capture semantic perspective of what the user wants to accomplish

Impact

- Graduate Students
  - R. Chinchani (Ph.D., May 2005)
  - A. Muthukrishnan (M.S., June 2004)
  - M. Chandrasekaran (M.S., June 2004)
- Publications
- IWIA 2003, ACSAC 2004, Book Chapter on Managing Cyber Threats, Springer 2005
- Funding Agency: DARPA (2003-05), AFRL (2000-06)
- Media
  - Washington Post, NY Times, CBS, NewScientist, Scientific American, etc.

Event Correlation for Cyber Attack Recognition System (ECCARS)

S. Upadhyaya (CSE), Moises Sudit (IE), W. Tagliaferri (Alion Science), NSA/ARDA (9/03 – 12/05)



Goals

- Collect, store and process large amounts of data
  - Cyber sensor observables
  - Real world events
- Fuse the resultant network data into meaningful threat related events
- Perform analysis to find correlations between cyber sensor data and real world events and trends
- Present the information to the analyst in a manner he/she can rapidly make a decision regarding defensive actions
- Test a prototype with basic functionality

Novel Ideas

- Correlation of disparate sources of information, system event information with real-world events, drawn from various information sources using information extraction techniques coupled with a 4-level information fusion framework
- Concept of sensor tasking through the use of mobile sensors that can work in conjunction with the data fusion processes for enhanced threat mitigation
- OOD design with APIs enabling plug-n-play capability for information extraction, fusion and visualization components

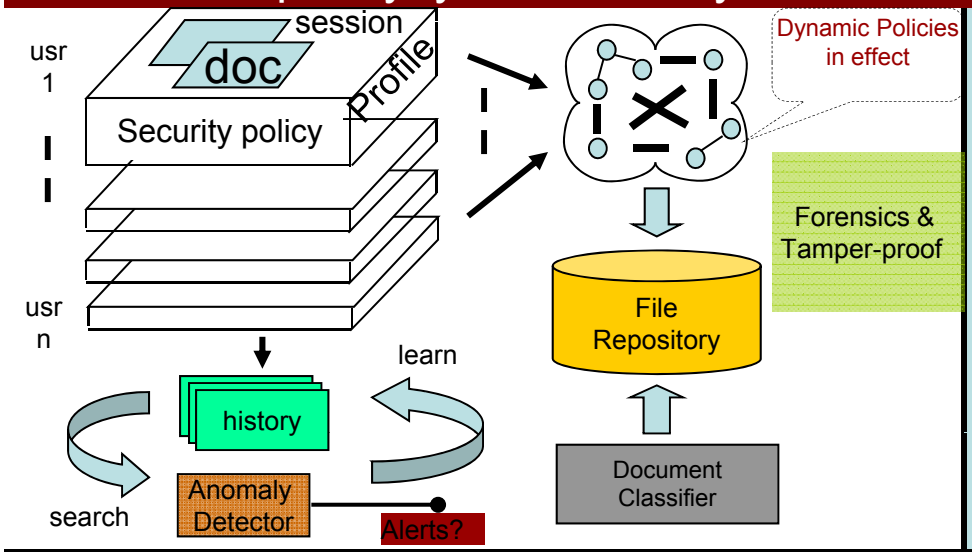
Accomplishments/Milestones

- Prototype ready for testing with data from an experimental testbed
- Papers published:
  - Mathew S., C. Shah and S. Upadhyaya, "An Alert Fusion Framework for Situation Awareness of Multistage Coordinated Attacks", *IEEE International Workshop on Information Assurance*, Washington DC, March 2005.
  - Mathew S., D. Britt, R. Giomundo, S. Upadhyaya, M. Sudit and A. Stotz, "**Real-time Multistage Attack Awareness** Through Enhanced Intrusion Alert Clustering", *to be presented in SIMA 2005, Oct. 2005.*

Impact

- Graduate Students
  - S. Mathew (Ph.D. June 2009), C. Shah (M.S., Jan. 2005)
- Publications
  - IEEE IWIA 2005, IEEE SIMA 2005, ACM VizSec 2006
- Funding Agency
  - NSA/ARDA (2004-06), AFRL (2004-06)
- This work was taken into Phase 2 by CMIF and Alion Science

Multi-phase Approach for Preventing Document Abuse from Malicious Insiders
Shambhu Upadhyaya, Funded by NSA/ARDA, 2003-05



Goals

- **Malicious and masquerading insider threat detection in the Document Control domain**
- Identify importance of documents
- Identify user roles in organizations
- Prevent circumvention and perform trace-back

Novel Ideas

- **Building user profiles at the application level**
- **Usage based document classification**
- **Context & information flow based policy specification for preventing insider abuse**
- **Automated generation of dynamic policies**
- **Papers Published:**
  - **IEEE Information Assurance Workshop, West Point, NY, June 2004**
  - **20th Annual Computer Security Applications Conference, Tucson, AZ, December 2004**

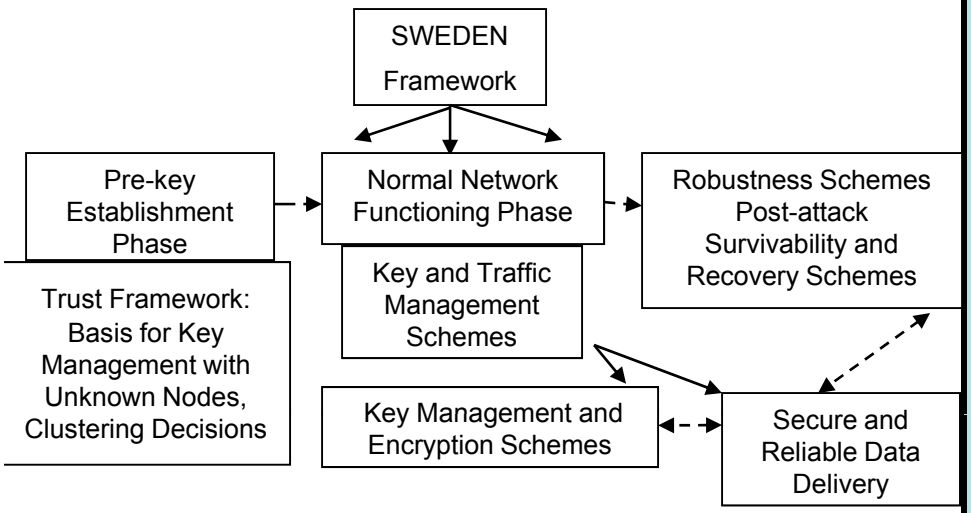
Accomplishments/Milestones

- Prototype for Microsoft Word
  - Monitor and detect masqueraders based on document usage
  - Specify and enforce dynamic policies
- Prototype for dynamic policies generation
- http://www.cse.buffalo.edu/DRM
- Future Plans
  - Detecting the convergence of disparate role structures in collaborating organizations
  - Preventing circumvention of the tools

Impact

- Graduate Students
  - S. Pramanik (Ph.D., Aug. 2007), S. Vidyaraman (Ph.D., Feb. 2008), N. Shah (M.S., June 2004), A. Garg (Ph.D., June 2006)
- Publications
  - IEEE IA Symposium 2004, ACSAC 2004, IA Symposium 2006, IEEE ICC 2006
- Funding Agency
  - NSA/ARDA (2003-05)

SWEDEN: A New Framework for Secure and Trusted Communications in Wireless Data Networks, Shambhu Upadhyaya, Funded by AFRL, NSF/Cisco, 2004-09



Goals

- Design decision making framework for nodes to establish keys with other unknown nodes
- Use this framework for cluster forming decisions in ad-hoc networks
- Improve on existing key management schemes and design secure data delivery schemes for enhanced reliability in data transfer
- Provide schemes for resiliency against attacks and post-failure recovery

Novel Ideas

- **Trust between the nodes used as a metric for decision making**
- **Differential encryption (header and payload differently) scheme for ad-hoc networks, and hashing based lightweight techniques for sensor networks**
- **Evaluating security of paths and nodes based on their relative position in the network**
- **Building in survivability in the network architecture proactively for surviving potential attacks**
- **Robustness, Recovery and Survivability Schemes**

Accomplishments

- **Setting up of the NSF and Cisco sponsored Wireless Security Lab**
- **Representative Publications:**
  - **IEEE Conference on Local Computer Networks (LCN), Tampa, FL, Nov 2004**
  - **IEEE ACM IWIA, College Park, MD, Mar 2005**
  - **IEEE Conference on Knowledge Intensive Multi-agent Systems (KIMAS), Boston, MA, Apr 2005**
  - **Secure Knowledge Management (SKM), Sep 2004**
  - **MMM 2007, St. Petersburg, 2007**
- **Future Plans**
  - **Security Schemes for mesh networks**
  - **Performing hands-on experiments at the Wireless Security Lab**

Impact

- Graduate Students
  - M. Virendra (Ph.D., June 2008), M. Jadliwala (Sept. 2009), Ameya Sanzgiri (stated June 2009), Chris Crawford (M.S., June 2009)
- Publications
  - KIMAS 2005, SKM 2006, IEEE ICC 2007, MMM-ACNS 2007, IEEE SRDS 2007, Infocom 2008, WiSec 2009
- Funding Agency
  - Air Force Research Laboratory (2007-09)

Changing HCI Transparency Paradigms to Mitigate the Weak Human Factor in IT Systems, Shambhu Upadhyaya, Funded by AFRL, 2000-08

Goals

- Mitigate the weak Human Factor in IT Systems
- Classify Users as Cooperative, Non Cooperative and Malicious
- Evaluate User Trust Levels
- Provide a technically meaningful process to elicit user cooperation

Novel Ideas

- Logging & Analyzing User Characteristics (Time-Invariant & Role based)
- Changing QoS to elicit user cooperation
- Quantifying Security State based on adherence to best practices
- Papers Published:
  - Sankaranarayanan V., M. Chandrasekaran and S. Upadhyaya, "Towards Modeling Trust Based Decisions: A Game Theoretic Approach", 12th European Symposium on Research in Computer Security (ESORICS 2007), Dresden, Germany, Sept. 2007

Accomplishments

- Game-Theoretic Model for changing QoS
- A HIDS Prototype on Windows
- Dynamic Trust Assignment and Update Mechanisms based on user actions

Impact

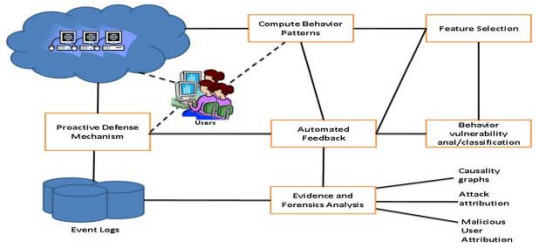
- Graduate Students
  - S. Vidyaraman (Ph.D., June 2008)
- Publications
  - iTrust 2006, Ubisafe 2007, ESORICS 2007
- Funding Agency
  - Air Force Research Laboratory (2004-08)

Fundamental Approaches to Dealing with Malware on the Internet
Shambhu Upadhyaya, State University of New York at Buffalo, Funded by DoD, 2007-09

Objective

A unified behavior based framework for mitigating threats and damaging attacks on the Internet

- Address phishing, zero-day exploits, spyware, email authorship attribution, information leak in documents
- Hardware acceleration to support scalability



The approach

Behavior Capture and Analysis (feature selection, simulated annealing)
Behavior Based Monitoring and Detection (support vector machines)
Attack Attribution and Forensics (causality graphs)
Attack-Agnostic Framework (component based approach, implementing theories in hardware on modern CPUs)
Validation (user studies)

State of the art in the area

Malware on the Internet is rampant
Behavior-based defense used successfully in real-world
Extended to cyber-world by researchers at Columbia U.
Behavior capture and correlation of applications using programming languages
Behavior based monitoring for attack detection using statistical and rule-based algorithms
Behavior based techniques for network forensics using causality graphs
Designing new hardware for content processing and cryptography

Novel ideas

Attack-agnostic framework to address all facets of security – attack protection, detection, response and forensics
A holistic approach
Proof-of-concept prototypes for anti-phishing, handling zero-day exploits, malicious email attribution, anti-spyware, information leak detection
Hardware acceleration techniques to handle “pump and dump” malware
Grounded in theory and preliminary investigation – "Spycon: Emulating user activities to detect evasive Spyware”, IEEE Malware 2007 (Best paper award)

Impact

- Graduate Students
  - M. Chandrasekaran (Ph.D., June 2009), N. Pulera (M.S., June 2008), H. Alkebulan, (M.S., Dec. 2008), N. Campbell (B.S., Dec. 2008)
- Publications
  - Ubisafe 2006, Malware 2007 (Best Paper Award), Albany IA Conference 2007, 2008
- Funding Agency
  - DoD (2007-08)

Accelerating Techniques for Rapid Mitigation of Phishing and Spam Emails

- Phishing scams pose a serious threat to end-users and commercial institutions
- Current software based solutions cannot be implemented on end-user's local computers due to the computation overhead involved with the associated feature selection and data mining algorithms
- We aim at detecting phishing attacks based on the semantic and structural properties present in the content of the phishing emails at the end-user level
- Our solution is hardware based
  - We are implementing some basic theories such as Simulated Annealing, Bayesian Learning, and Associative Rule Mining in the hardware
  - Exploit the inbuilt pipelining, scheduling and other accelerator capabilities and the micro engines of the Intel Tolapai processor

Impact

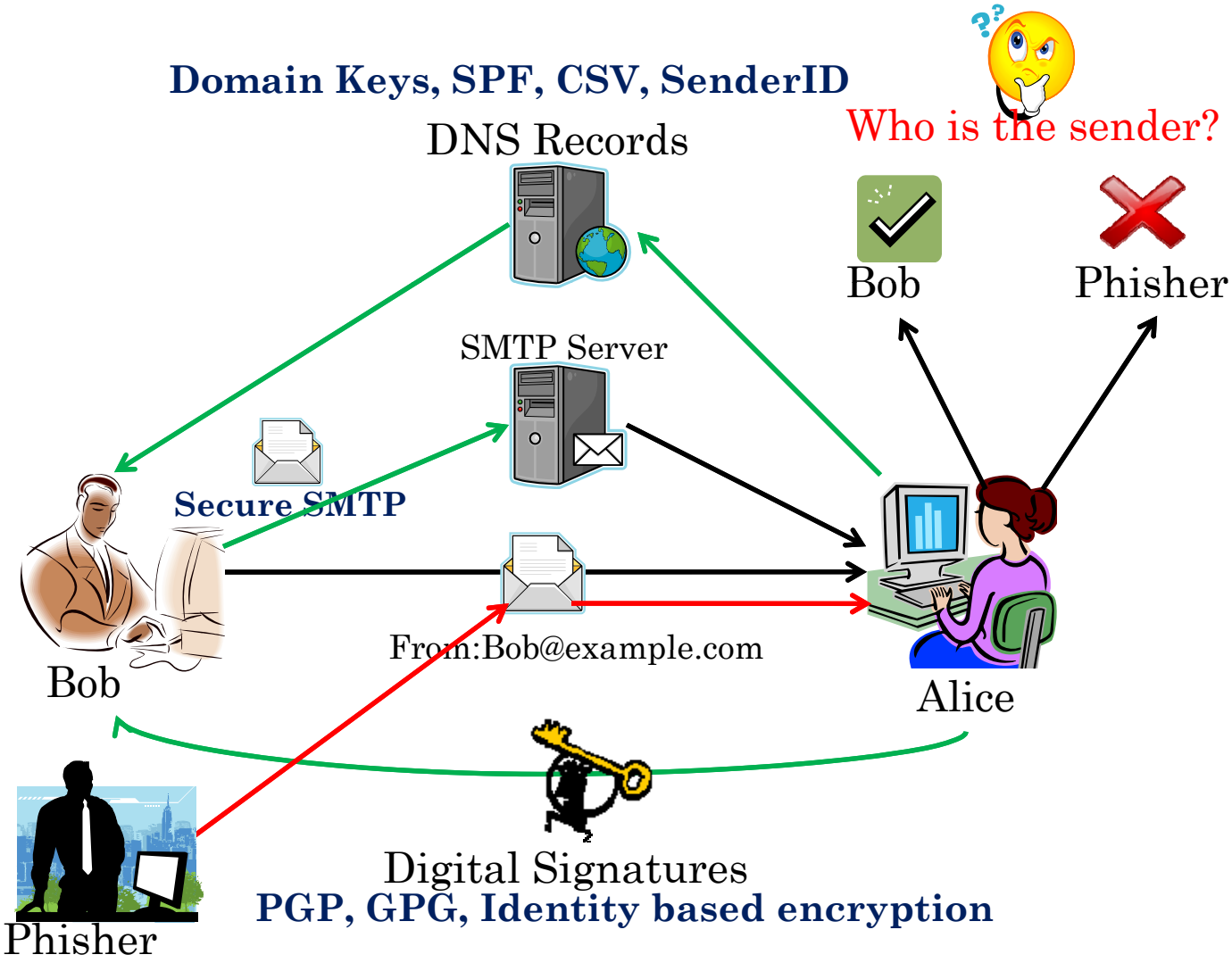
- Graduate Students
  - M. Chandrasekaran (Ph.D., June 2009), Ajay Nagrale (M.S., June 2010), Pranil Gupta (M.S., June 2010)
- Publications
  - Intel Summit, Feb. 2009, ECSC 2009 (in conjunction with IEEE SRDS 2009)
- Funding Agency
  - Intel Corporation (2008-10)

TECHNIQUES FOR RAPID MITIGATION OF PHISHING AND SPAM E-MAILS

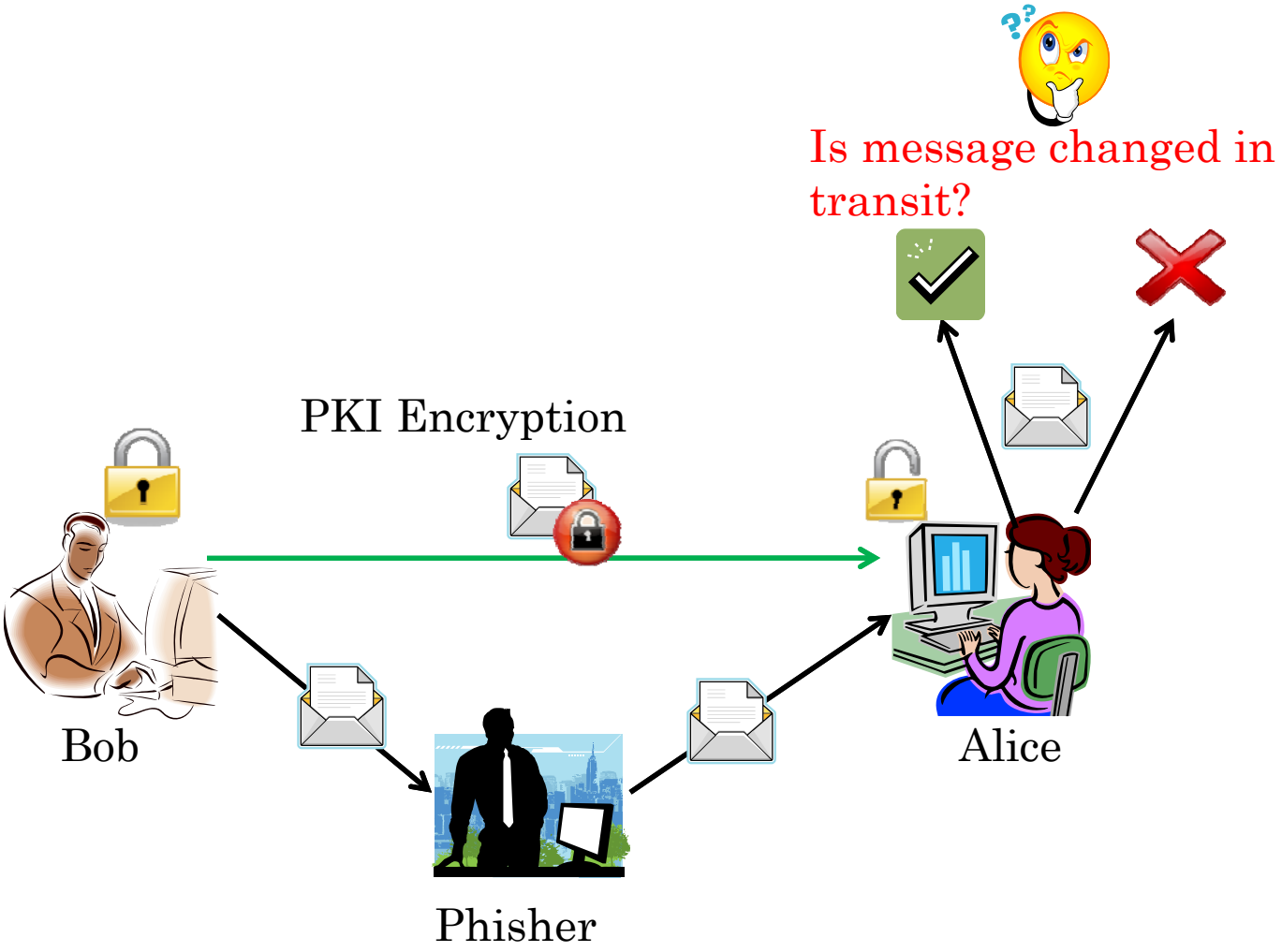
INTRODUCTION

- E-mail is the most popular tool for exchanging personal and business information
- Simple Mail Transfer Protocol (SMTP, RFC 5321; RFC 821) defines standard for sending messages
- Drawbacks with SMTP protocol
  - Authentication
  - Integrity
  - Non-repudiation

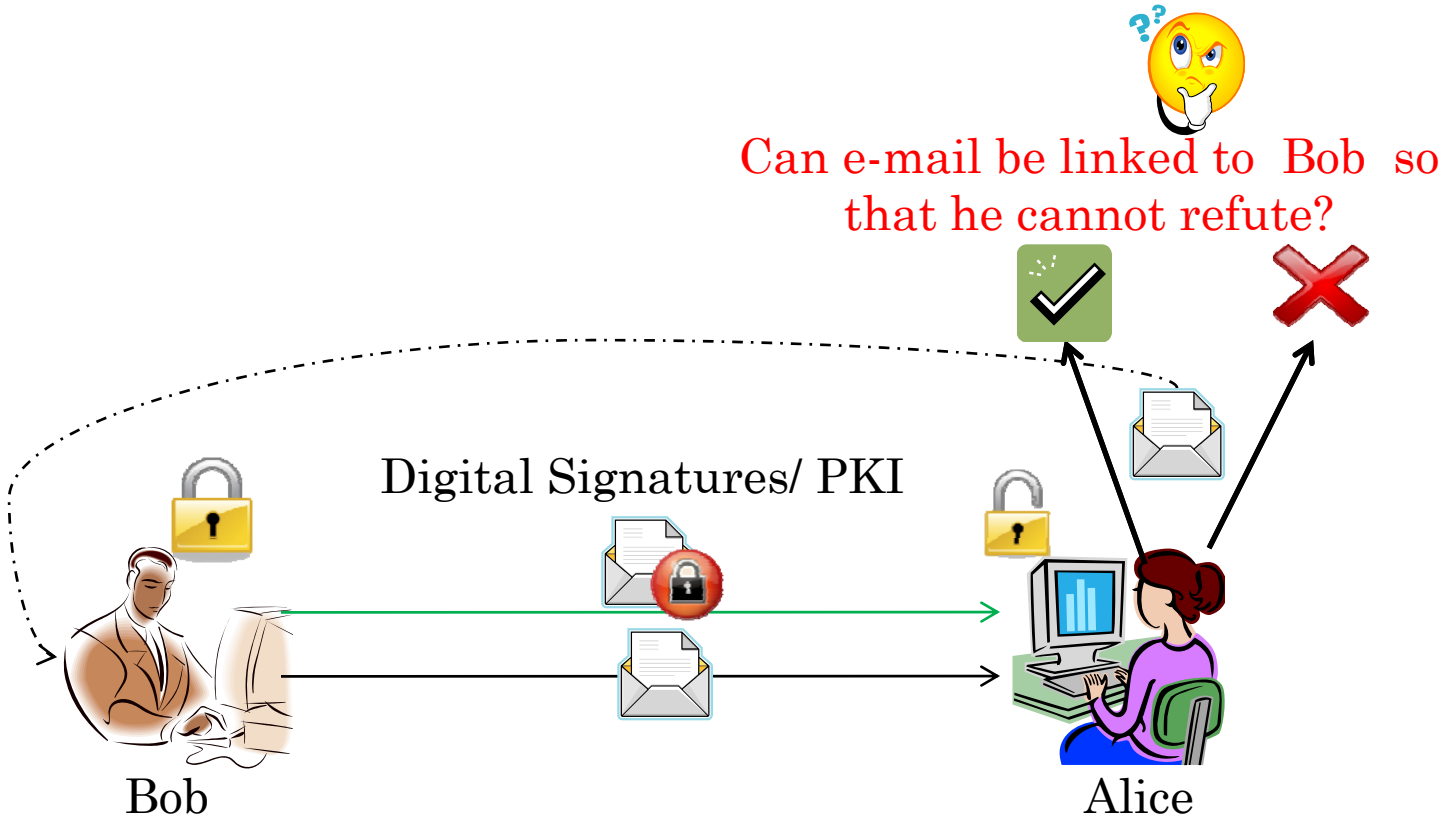
AUTHENTICATION



INTEGRITY


Is message changed in transit?
PKI Encryption
Bob
Alice
Phisher

NON-REPUDIATION

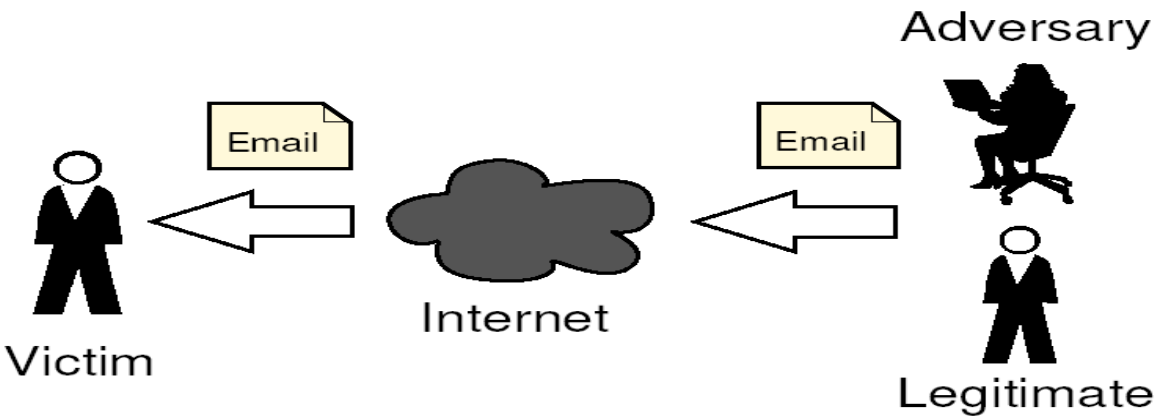


Drawbacks of Secure Email Tools

- Adoption is not straightforward
  - Requires revamp of existing e-mail infrastructure
- Human factors Issues
  - Users are bad in encrypting, decrypting and signing messages
- Suffers from key exchange problem
  - Need for third-party CA or trusted platform
- Encryption/Decryption of messages is not supported by most Web-based clients
  - Yahoo! and Gmail use DKIM, however still most other services don't

Email Phishing Attacks

- Lack of authentication, integrity, repudiation = Phishing attacks
- Victim is sent an email with links referring to corresponding spoofed website
- Efficient coverage and ease of execution
- Preys upon user's inability to make correct decision



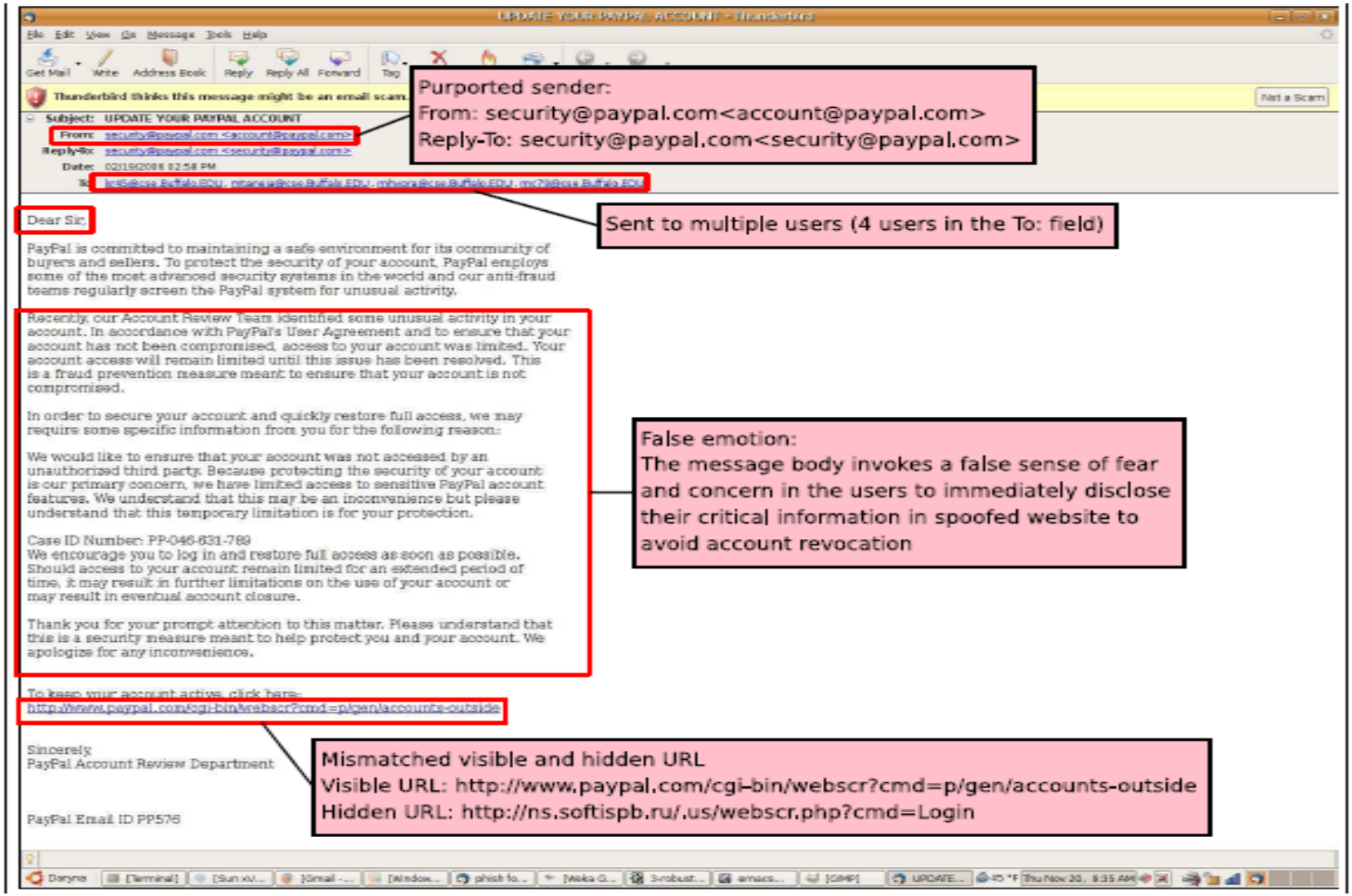
Existing Filtering Techniques

- Browser add-ons and third-party toolbars
  - Not scalable, black/white listing info. may not reach clients in time
  - Zhang et al. study shows poor performance
- Using spam filters
  - 1-gram words, IP-based URL, different hidden and visible links
- Content based filters
  - Textual and structural features
- Phishing Website Analysis
  - Domain validation, URL de-obfuscation, link analysis and image analysis
- Suffer from false positive and false negative rates

Arms Race

User takes the bait

Anatomy of Phishing E-mail



Phishing – A Social Engineering Attack

- Uses threat, concern, reward to attract users
- Threat
  - Provide password within 24 hours to avoid account revocation
- Concern
  - Your account has been compromised. Please provide new username and password
  - Password insecure, provide strong one
- Reward
  - Provide your details to get $10,000,000
  - Nigerian Scammers

Features Used for Detection

- Three classes of features used
- Textual features (individual words)
  - 1gram, 2gram, and 3gram words
  - Normalized using stopword elimination and stemming
- Linguistic features
  - Grammatical statistics to indicate ratio of nouns, pronouns, verbs, and 10 more similar features
- Structural features
  - Salutation, IP based URLs, difference in URLs

Supervised ML Algorithms

- Features from e-mails are fed into supervised machine learning algorithms for classification
- Has two phases:
  - Training phase – learns the classification model
  - Testing phase – Uses the learned model to classify incoming e-mails into phishing and ham
- Three popular supervised machine learning algorithms are used
  - Naïve Bayesian Classifier
  - Decision Trees
  - Support Vector Machines (SVM)

NAÏVE BAYESIAN CLASSIFIER

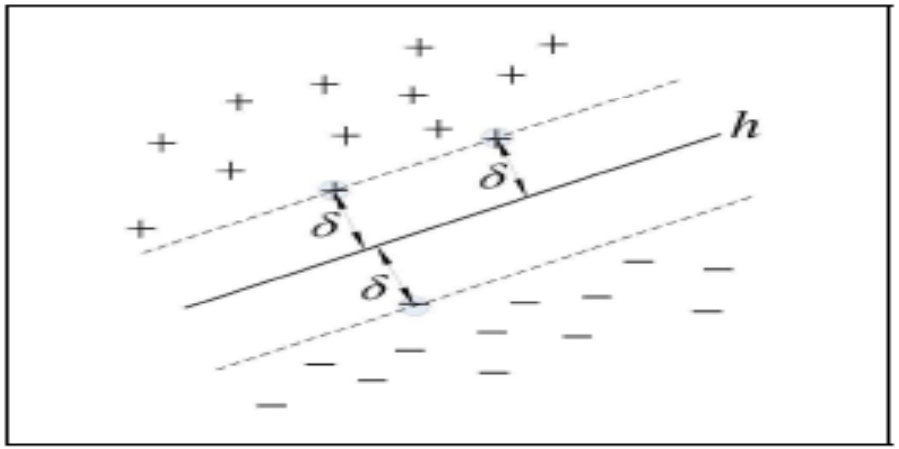
- Based on Bayes theorem and conditional independence assumption
- Given a set of input features $X = \{x_1, x_2, \ldots, x_n\}$ and class labels $C = \{c_1, c_2, \ldots, c_n\}$, naïve Bayes classifier assigns X class label such that
  - $Pr(c_i \mid X) > Pr(c_j \mid X)$, for all $i \neq j$
- Bayes rule relates $Pr(c_i|X)$ to $Pr(X|c_i)$
- Application of conditional independence, equation becomes

Prior probability obtained from training

$$Pr(c|X) = \arg\max_{c \in C} Pr(c) \prod_{i=1}^{n} Pr(x_i|c)$$

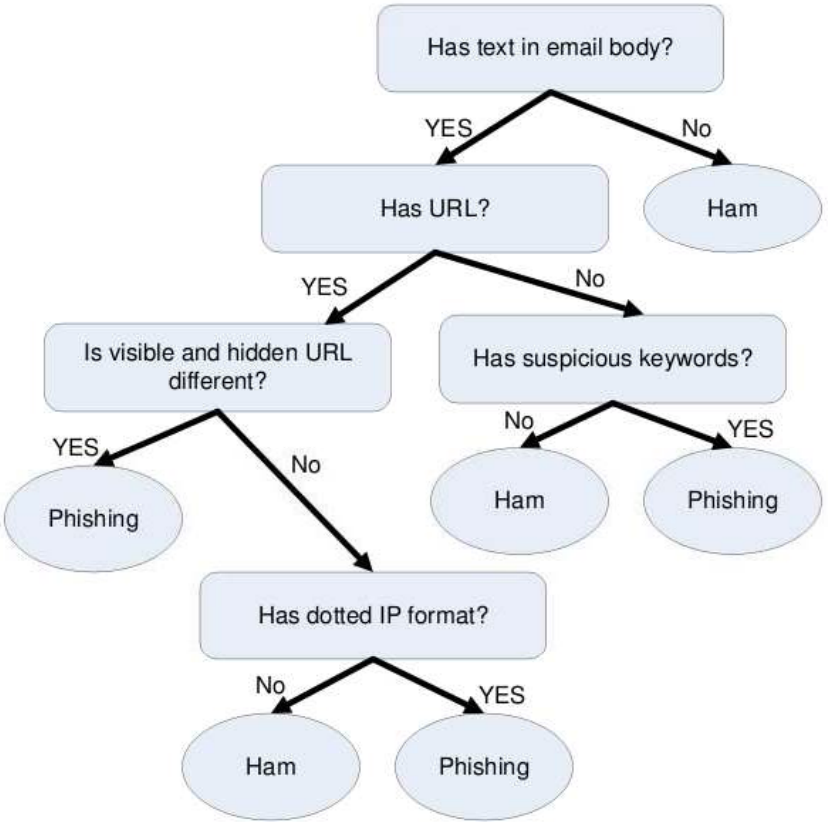
SUPPORT VECTOR MACHINES (SVM)

- Used in binary classification problem,
  - C = {+, }, + → ham, → phishing
- Input features are plotted in high dimensional vector space
- Learns to build a separating hyperplane h to separate two classes



DECISION TREES

- Builds decision trees with class label along with leaves
- Non-leaf nodes represent constraint on some input feature
- Is viewed as series of if-else tests
- ID3 algorithm is used to construct decision trees



Experimental Setup

- Dataset
  - Phishing corpus – 4550 phishing e-mails obtained over period of three years (2004 – 07)
  - Ham dataset – 6950 e-mails from SpamAssassin dataset
- Setup
  - 10-fold cross folding is used (90% training, 10% testing)
- Performance is measured using detection rate, false positive rate, precision, recall, f_1 statistic

Results

Feature Set Used	Detection Algorithm	Detection Rate	False Positive Rate	Precision	Recall	F-measure
F_{3gram_all}	Naïve Bayes	68.9	0.10	99.7	69.0	80.1
	SVM	97.2	0.59	99.0	97.2	98.1
	Decision Tree	97.3	1.17	98.1	97.3	97.7
F_{2gram_all}	Naïve Bayes	87.6	1.6	87.6	92.1	93.7
	SVM	98.6	0.4	99.2	98.6	98.9
	Decision Tree	97.9	1.0	98.4	97.9	98.1
F_{1gram_all}	Naïve Bayes	90.8	2.4	95.8	90.8	93.2
	SVM	99.1	0.3	99.6	99.1	99.4
	Decision Tree	98.4	0.9	98.5	98.4	98.4
F_{3gram_wrds}	Naïve Bayes	67.9	0.10	99.8	67.9	80.9
	SVM	89.1	0.6	98.8	89.1	93.7
	Decision Tree	86.5	1.2	97.8	86.5	91.8
F_{2gram_wrds}	Naïve Bayes	77.4	0.3	99.4	77.4	87.1
	SVM	98.9	1.0	98.4	98.9	98.7
	Decision Tree	94.5	0.8	98.6	94.5	96.5
F_{1gram_wrds}	Naïve Bayes	89.7	2.4	95.9	89.7	92.7
	SVM	99.0	0.4	99.3	99.3	99.3
	Decision Tree	97.8	1.4	97.6	97.8	97.7
F_{disc}	Naïve Bayes	91	6.6	89.3	91	90.1
	SVM	76.7	2.7	94.6	76.7	84.7
	Decision Tree	95.3	2.5	95.9	95.3	95.6

>99% detection rate & <0.3 false positive rate using linguistic features

Limitations

- Building and deploying decision trees, SVM and naïve Bayesian classifiers is computation-intensive
- When spam load increases, solutions may prove to be sluggish
- Should be deployed at MTA for better speed and efficiency
- Has number of subcomponents that can be done in parallel
  - Different branches of decision trees can be traversed independently
  - Feature extraction, probability estimation and classification in NB and SVM can be done in parallel

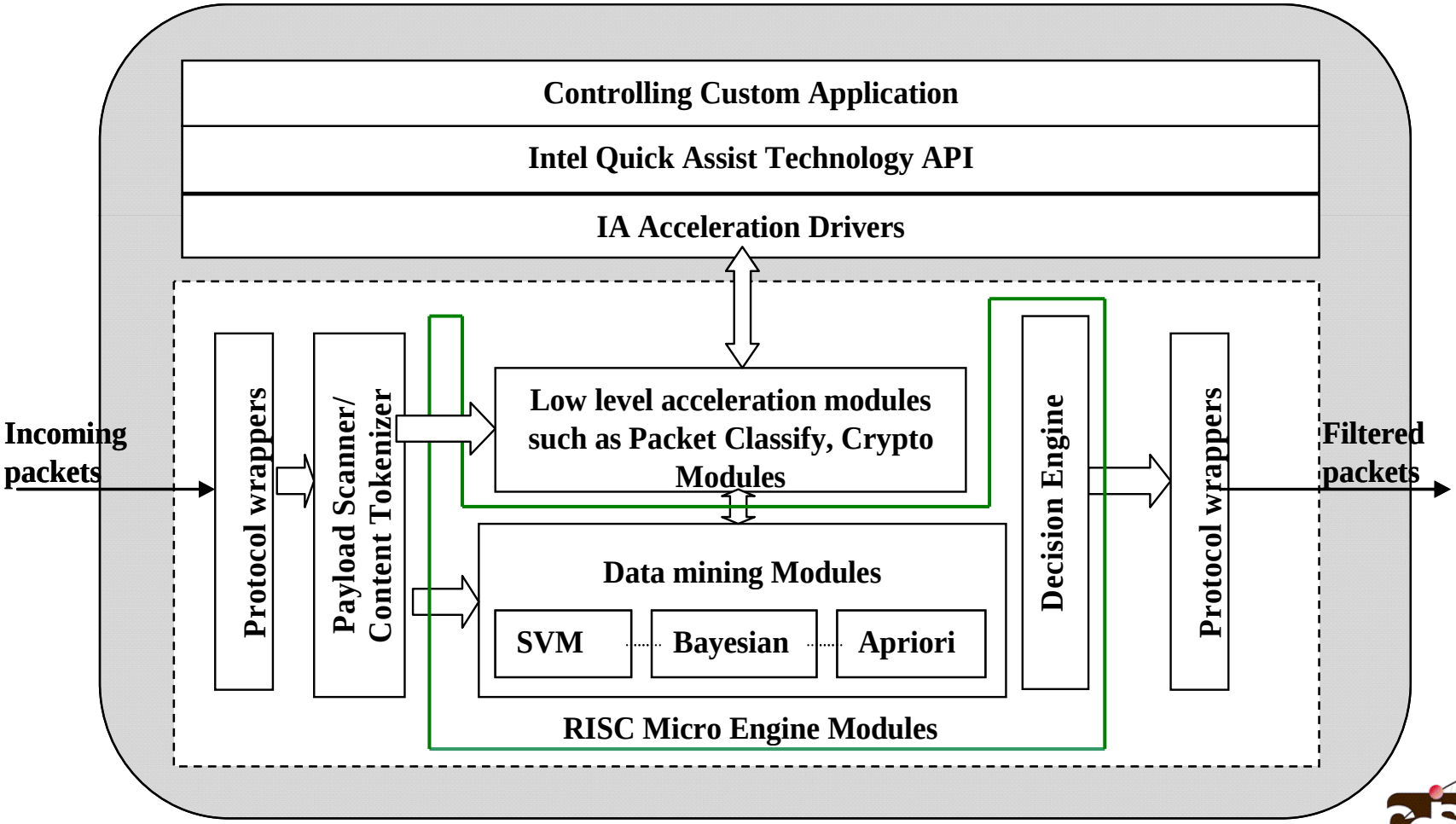
Intel Tolapai Salient Features

- Gigabit Ethernet (GbE) Controller
  - Highly integrated, high-performance Ethernet LAN Device
  - Implements hardware acceleration capabilities
  - Offloads checksum capabilities from host processor
  - Filters packets based on checksum errors
  - Minimizes I/O accesses and interrupts
  - Supports various address filtering

Tolapai Salient Features Contd.

- Acceleration of cryptographic operations
  - Symmetric Operations:
    - Cipher: AES, DES, 3DES, NULL, ARC4
    - Hash/Authentication: SHA, MD5, AES
    - SSL/TLS Key Generation
  - Public Key Operations
    - Diffie-Hellman Key Generation
    - RSA, DSA
    - Primality Tests, Large Number Operations (Modular Inversion & Modular Exponentiation)
  - Random Number Generation

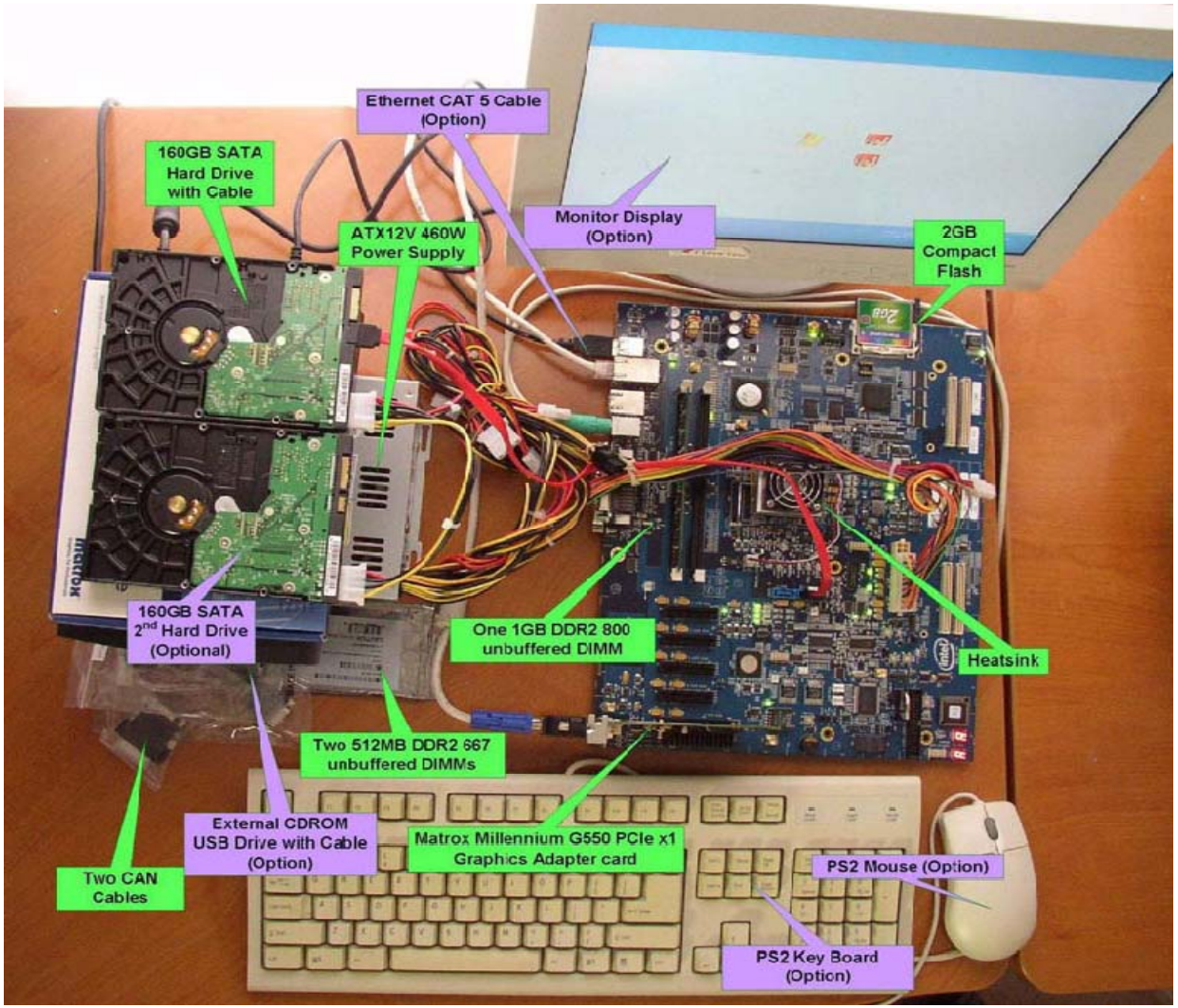
Integrating with Tolapai



Setting Tolapai Board

- Installed CentOS 5.2 with Gcc 4.1 and Glibc 2.5
- Rebuilt Kernel (applied new RPM packages & patches)
- Prepared fresh Kernel 2.6.18EP805XX
- Installed all the Embedded & Cryptographic drivers (sys. calls – setkey, encrypt_r, setkey_r)
- Enabled Crypto and other Debug tools
- Added Tolapai machine in LAN & started *sendmail* and *procmail* to support Email flow
- Analyzed Cryptographic API & new system calls
- Hacked GbE code to understand packet processing

Board Assembly (Courtesy: Intel)



NB Algorithm

- Whether spam filtering or phishing detection, NB is used as a text classifier
- Training phase and classification phase
- Training phase
  - Parsing email
  - Tokenizing
  - Hash maps – separate tables for spam and ham emails and a third table for mapping probabilities to tokens
- Classification phase
  - Essentially hash table lookup
- Acceleration in tokenizing and hash computation
  - Hashing is moved to hardware using the APIs

Results

No. of Words	hardware (S)	software (S)	%gain w.r.t software
10000	0.0210	0.0282	25.52%
20000	0.0414	0.0563	26.45%
30000	0.0630	0.0844	25.38%
40000	0.0830	0.1126	26.33%
50000	0.1041	0.1411	26.19%
60000	0.1247	0.1702	26.74%
70000	0.1469	0.1969	25.38%
80000	0.1678	0.2253	25.52%
90000	0.1893	0.2533	25.28%
100000	0.2095	0.2817	25.61%
200000	0.4201	0.5627	25.34%
300000	0.6297	0.8483	25.77%
400000	0.8355	1.1262	25.82%
500000	1.0236	1.4082	27.31%
600000	1.2355	1.6880	26.81%
700000	1.4367	1.9716	27.13%

Summary

- Traditional SMTP protocol has security design flaws
- Phishers and spammers exploit them to pump-and-dump spurious e-mails
- Spam detection approaches fail miserably
- Linguistic and structural features encapsulate phisher's intent and improve detection rate
- Software implementation of classifiers is computationally expensive
- Work is preliminary (only hash function implemented)
- Currently working on full implementation

Outline

- Acknowledgments
- Cyber Security, Current Status
- Challenges for the Future
- A Cyber Security Primer
- What are we doing at UB?
- Selected Research Projects
- Path Forward

Conclusions

- “This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.”

Sir Winston Churchill

- We need to look forward

Looking Forward

- CEISARE has been designated as CAE-R (2009-14)
- Aligning with Cyber Security Act of 2009 (S.773)
  - address our nation's vulnerabilities to cyber crime, global cyber espionage, and cyber attacks
- We need to train cyber sleuths
  - At UB, we graduated 6 scholars through DoD program
  - In 2008, we received a $868,000 grant from NSF Federal Cyber Service Program
  - We have 5 scholars, funds to educate 4 more in 2010-11
- "Cybersecurity is one of our most urgent priorities" says DHS chief Janet Napolitano, Oct. 2, 2009
  - DHS to hire up to 1,000 CS experts in 3 years
  - House panel plans cyber security training for members and staff (cyber flu shots – Oct. 27, Oct.30, 2009)

Questions

- My contact: shambhu@buffalo.edu
- Center Website: http://www.cse.buffalo.edu/caeiae

Video Presentation

- Social Engineering Attack
  - Interview with Kevin Mitnick (Courtesy: CBS 60 Minutes Segment) – 20 minutes
- Cyber Security Awareness
  - NYS CSCIC (Courtesy: William Pelgrin, Director) – 25 minutes