CS161, Spring 2008
March 12th

John Bethencourt

Private Information Retrieval

Logistics

- Dawn travelling today
- So guest lecture by me
- Optional topic
  - Not covered by midterm, final, or homeworks
  - But hopefully fun and intriguing
- More advanced, modern cryptography
  - Like ecash, search on encrypted data
  - Techniques for things other than just message privacy, authentication, etc.

Motivation: Searching for Information

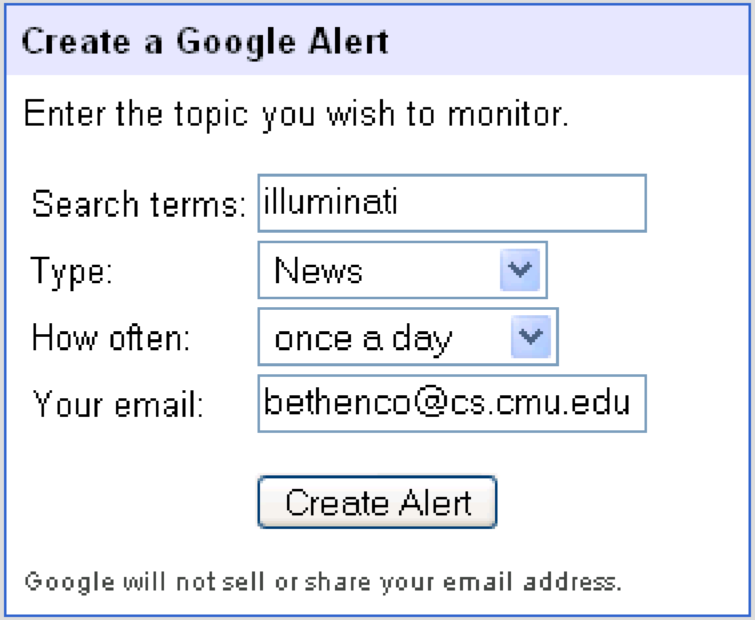
- Too much info online to download
  - WWW pages
  - Message boards
  - Web email
- So we search
  - User specifies search criteria
  - Normally textual keywords
  - Server returns relevant content

Motivation: Searching Privately

- What if keywords are secret?
  - Personal privacy
    (example query: “lice removal”)
  - Commercial interests
    (“takeover bid”)
  - Legal issues
    (“how to grow marijuana”)

- We want *private* searches
  - Client gives server encrypted query
  - Server runs search algorithm, returns data
  - Client recovers matching content
  - Server does not know what client searched for

Example: Google News Alerts



Create a Google Alert

Enter the topic you wish to monitor.

Search terms: illuminati

Type: News

How often: once a day

Your email: bethenco@cs.cmu.edu

Google will not sell or share your email address.

- Google News
  - Google continuously crawls 4,500 news sources
  - Estimated 135,000 news articles each day
  - Alerts service
    - User registers search keywords
    - Matching articles emailed as they are discovered

Example:
Private Google News Alerts

- What about a *private* alerts service?
  - User registers encrypted search keywords
  - Periodically receives matching articles

- Google remains oblivious
  - Doesn't know search keywords
  - Doesn't know which articles you got

Google doesn't know what I want, but they can still give it to me!

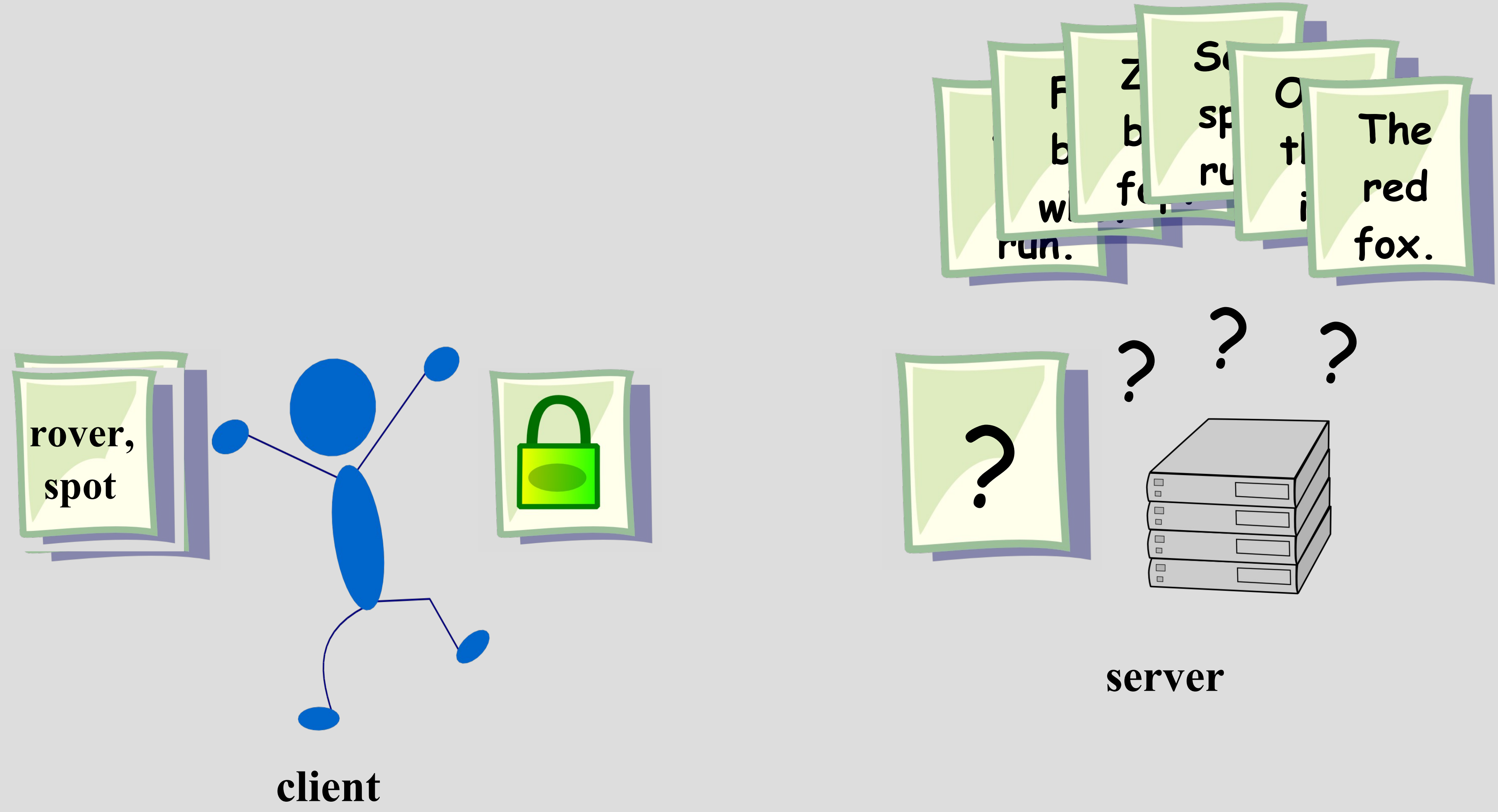
Private Information Retrieval

- Techniques we'll cover today enable such things
  - Could actually build private Google news alerts system
  - Other applications also possible, including malicious ones! (later in lecture)
- Private information retrieval (PIR)
  - Group of related cryptographic building blocks
  - Lets us build things like this

Private Information Retrieval

- Setting
  - Server has database, filesystem, or collection of documents
  - Client has keyword, name, index or some other method of specifying one they want to download
- PIR scheme
  - Protocol between client and server
  - At end, client has desired file
  - Server doesn't know which one that was

Private Information Retrieval



Cryptographic Background

- We are going to show (partly) how to build a scheme like this

- But first, a little background material
  - Paillier (“pie – yay”) cryptosystem
  - Homomorphic encryption
  - Probabilistic encryption

Building a PIR Scheme

- Now we can build a simple PIR scheme
- Setting
  - Files to be retrieved are text documents
  - Client's query is list of keywords
  - Client should get all files which include one or more of the keywords
- Simplifying assumptions
  - Public, global dictionary of possible search words
  - Only one word in query
  - Only one files matches
  - Each file fits in one Paillier plaintext (e.g., 1024 bits)

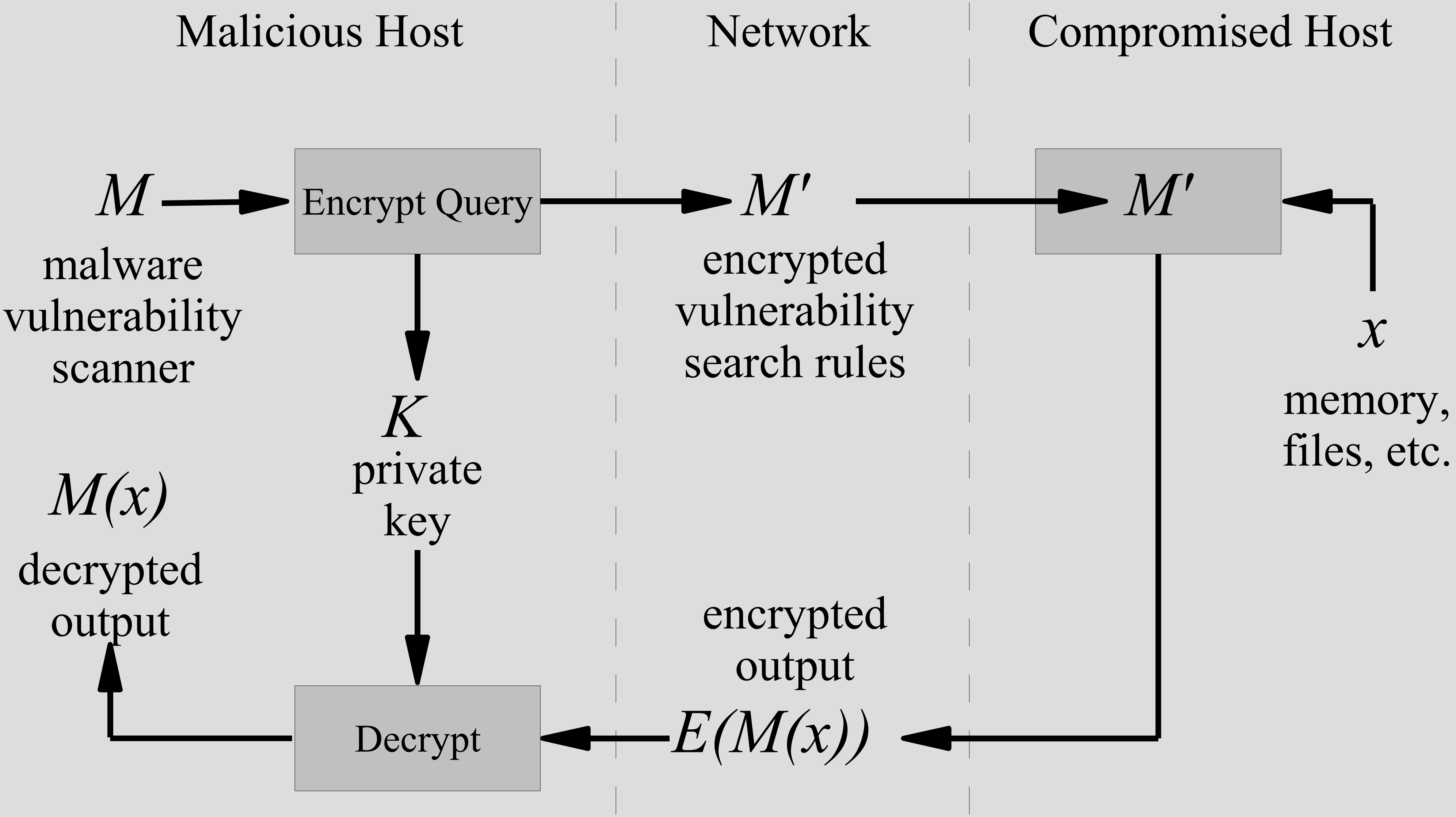
Generalizing the Scheme

- Easy tweaks
  - Handling longer, but fixed length files (just do them blockwise in parallel)
  - Handling more than one search keyword
  - Getting rid of the global dictionary
- Most important and difficult
  - Handling multiple matching files
  - Need to keep them from clobbering each other in c_{buf}
  - Multiple ways of doing so with various efficiency
- Trickier tweaks
  - Variable length files
  - More complex queries (“foo” AND “bar”)

Using this for Evil ...

- Private information retrieval
  - Invented for privacy purposes
  - However, malicious uses also possible
- What if malware uses PIR?
  - Malware author is client
  - Compromised host is server
  - Malware can be stealthy while retrieving data, hide what it is looking for
- Example application: host based 0-day vulnerability scanner

Example Malicious Application



More on Malicious Applications

- Run of the mill Internet miscreants use malware to steal credit cards, run DDOS attacks, etc.
  - Question: Do they care to hide what their malware is retrieving from your machine?
  - Answer: Probably not.
- However, another class of attackers may find this much more appealing ...
  - An anecdote may illustrate this

Trojangate

- In 2005, there was a **massive** scale industrial espionage scandal
  - Three top Israeli telecom corporations responsible for infiltrating competitors systems with trojans
  - Trojans introduced through email attachments and hand delivered on CD's through careful social engineering attacks
- Sought commercially sensitive information
  - Used lists of keywords to trigger keystroke logging, screen capture
  - Searched for and retrieved sensitive documents
  - 10's of thousands of documents exfiltrated to over 100 receiving servers

Trojangate

- Trojans very stealthy
  - Specifically written for espionage, never seen in wild before
  - Kept low profile and were not discovered for a year and a half
- End results
  - Large economic fallout with stock losses, etc.
  - Top executives arrested
  - A possible attempted homicide

Targeted Malware

- These techniques become most interesting in contexts such as these
  - Malware seeking to retrieve specific confidential information
  - Small portion of malware and attacks overall, but perhaps more interesting
- Information sought may cast suspicion on malware originator
  - E.g., everyone wants credit card numbers
  - But how many people want a document detailing the five year strategy of a specific corporation?

Targeted Malware

- Relatively small threat, but widely considered to be growing

- SANS Institute: “Top 10 Security Menaces for 2008”
  1. Browser Vulnerabilities
  2. More advanced botnets
  3. **Espionage efforts by well resourced organizations**
  4. ...

Summary

- Private information retrieval
  - Lets a client retrieve specific information from a server without revealing search criteria
  - Mainly useful for privacy, but subversive applications also possible

- See how it's actually done:
  - Open source PIR toolkit available under GPL
  - http://acsc.csl.sri.com/privss/
  - Or google for “privss”

- Questions?

More Logistic Notes

- Today was last lecture on cryptography, rest of semester on systems security topics
- Monday
  - Midterm review (Rusty and Todd)
  - HW3 back
- Wednesday
  - In class midterm
  - Closed book, closed notes
- Spring break!
  - Faulty testing program for project out by end of break
- Wednesday, April 2
  - Milestone 1 due