Telcordia Technologies

Performance from Experience

VIRUS ATTACKS – SOME COMPUTATIONAL TOOLS

URL: http://www.argreenhouse.com/bios/rafail/index.shtml

June 10th 2002

Rafail Ostrovsky
Telcordia Technologies
Rafail@research.telcordia.com
(973) 829-4079

Algorithmic and Modeling Issues

- How do you "detect" a virus? (the issue of modeling and algorithms)
- How do you function "under attack" (again, modeling, algorithms)
- General "tools" category, when dealing with malicious adversaries
- General paradigm: in cryptography we worry about privacy, since the less adversary knows the easier it is to defend against such an adversary!

Telcordia Technologies

Performance from Experience

Authentication

- Private key: collision resistance: three variants
  - Universal hash
  - Universal one-way hash
  - Collision resistance (once vs. many)

  [Q: can we compute any of this using DNA?]

- Public-key: what is the right definition? (existential attack)

Telcordia Technologies

Performance from Experience

How do you “boost” confidence that your software is not infected?

3 POSSIBLE ASNWERS:

- Authentication starting from a smart-card
- Assume that most of your machines are “OK”
- Assume that most of your machines are OK most of the time (and reboot)

Telcordia Technologies

Performance from Experience

Secure computation (under attack)

- Issue of PRIVACY vs SECURITY
- Honest-but-curios
- Dishonest
- Honest-looking
- General theorems [Yao, GMW, BGW, CCD, RB, CO]

Telcordia Technologies

Performance from Experience

Mobile adversary (proactive security)

- Faults move, but allow erasure
- How do you maintain a secret? (Shamir's secret sharing – proactivization tool)
- From polynomial interpolation to general multiparty computation – the need to compute on "shares".

Telcordia Technologies

Performance from Experience

Software protection (assuming smart cards)

- Formal model: what are the issues?
- Authentication and time-stamping: not sufficient!
- Hiding the access pattern: the problem and statement of the general result.

Telcordia Technologies

Performance from Experience

Private Information Retrieval

Remote database access

— Applications:
audit applications, searching for sensitive data in a public database, the ability to hide info that we are searching for.

— Database has N items, user wants to get j'th item
without the database knowing what j is.

— How do you retrieve info with small communication?

solution: design of a new "encryption" function

Telcordia Technologies
Performance from Experience

Content-sensitive “fingerprints”

Random “fingerprint” of a document (hash function)

- Compute a short “fingerprint” so that related documents have “related” fingerprints. (we have a patent on this.)
- Can be viewed as dimension reduction from high to low dimension that “preserves” relative distances

(DEMO: take document, modify, see changes in the fingerprint)

Telcordia Technologies

Performance from Experience

Searching on Encrypted data

Consider having remote "mail" spool file:

- User publishes PK, keeps SK.
  Different senders send mail encrypted with PK, which get stored on a remote Database (spool file)

- later, user wants to search for mails containing keywords but without revealing any other info to the spool file.

- We can also do it in the private-key setting (think of **internet**-based private memory, which you can update, read and write into, but no one else can touch.)

Telcordia Technologies
Performance from Experience

Nearest neighbor search

Searching for Approximate NN

- Given N points in high-dimensional space, first process them in some way. (different dimensions may indicate different incomparable attributes of the problem)
- Given any new "query" point, quickly find nearest neighbor among original N points. (easy in small dimensions, hard in in high dimensions)
- Applications include finding "related" events, find "outliers" searching for "related" events when searching for terrorist activity.

Telcordia Technologies

Performance from Experience

Clustering in High Dimensions

— "Classical" problems in clustering:
Given N points in high-dimensional space, form K clusters s.t.

min-sum clustering, (minimizing intra-cluster distance)
k-clustering, (minimizing the sum of distances to the center)
diameter clustering (minimizing diameter with approx. equal point split)

— Applications: data mining topics and searching for "outliers" and "topics" in streams of data.

Telcordia Technologies
Performance from Experience

Searching streams of data for “patterns”

- Given streams of mail msgs, and examples of “suspicious” mails, try to find mail messages that should be inspected more closely
- Goal: not too many false positives but do not miss too many important messages
- Can be cast as a “learning” problem, which in turn can be formulated as a searching for a nearest-neighbor in for certain metric.

Telcordia Technologies

Performance from Experience

Stronger (than usual) notion of encryption security:

- Standard notion of security: "can not see inside"
- Consider (as a toy example) the case of sending "internet orders" to buy and sell stocks by a trading house.
- Attacks of the type "get me the same as he does" do not contradict standard notion of security!
- Need security that says that can not created "related" cipher text!

Telcordia Technologies

Performance from Experience

Advanced Encryption properties (cont.)

- Need security that says that can not created "related" cipher text!
- Deniable encryption (application to stegonography)
- Self-referential security
- Forward security

Telcordia Technologies

Performance from Experience

Efficient database consistency

- How to "commit" to several copies of the database s.t. you can prove consistency.
- Key problem: how do you convince someone that the answer is NO???
- The problem is nontrivial even for single key-value pairs, becomes more challenging for relational DB.

Telcordia Technologies

Performance from Experience

Hiding command-and-control centers

- Directive issued from one of the nodes, you want to HIDE which node is a leader
- Both external and internal attacks
- Another issue: hiding for the Database access WHICH data is critical – how do you do it? (notice that encryption is not enough!) Must "relocate" data without too much overhead.

Telcordia Technologies
Performance from Experience

Efficient Zero-Knowledge proofs

- Convince that you "know" a secret without revealing what the secret is.

  TOY EXMAPLE: how do you convince someone that you have magic ability to count leafs on a tree?

  more practical issues: interactive passwords, forcing "honest" behavior among cheaters

- Can be done in the non-interactive setting as well.

Telcordia Technologies

Performance from Experience

Databases Discovery/Reconciliation

- Given N points;
  - "smudge and mix them" (I.e. create siblings)
  - Try to recover the "pairs"
- Applications:
  - How do you find many "related" entries in multiple databases or in streaming data
  - How do you find "many" relations

Telcordia Technologies

Performance from Experience

Conclusions (1): Maintain Distributing Trust in Data and Computation

- "Maintain distributed secret, (such as private key) where only a group of people jointly can "sign" or "decrypt" the value
- Proactive security: even if some machines are broken into and then rebooted on a *repeated* bases, the secret must sill remain hidden.
- distributed computation: ability to compute joint secret (such a winner in an auction) without revealing any additional info.

Telcordia Technologies

Performance from Experience

Conclusions (2):

- Many computational tools, some may be useful for in biological settings
- Security is defined only in terms of attacks
- If you think you need any of this machinery, I'll be happy to discuss further:

rafail@research.telcordia.com

Telcordia Technologies

Performance from Experience