Secure Distributed Systems

CompSci 661 / 461

This video:

- John Douceur's "Sybil Attack" paper.
- His proof that there is no reasonable defense.

Sybil Attacks

Sybil Attacks

Wednesday, January 31, 2018 1:12 PM

appear as many identities in a distributed system

2002 paper by John Douceur

Main idea:

It's always possible for one entity to appear as many identities in a distributed system

- except when regulated by a centralized server (and even then...)

How do we identify identities or entities typically?

- IP addresses, cryptographic keys, email addresses mobile phone numbers, tuition payments.

Douceur's paper is an impossibility result

- his goal is to make the fewest (weakest) assumptions possible so that the result applies to as many scenarios as possible.

Things that don't work as a consequence of Douceur

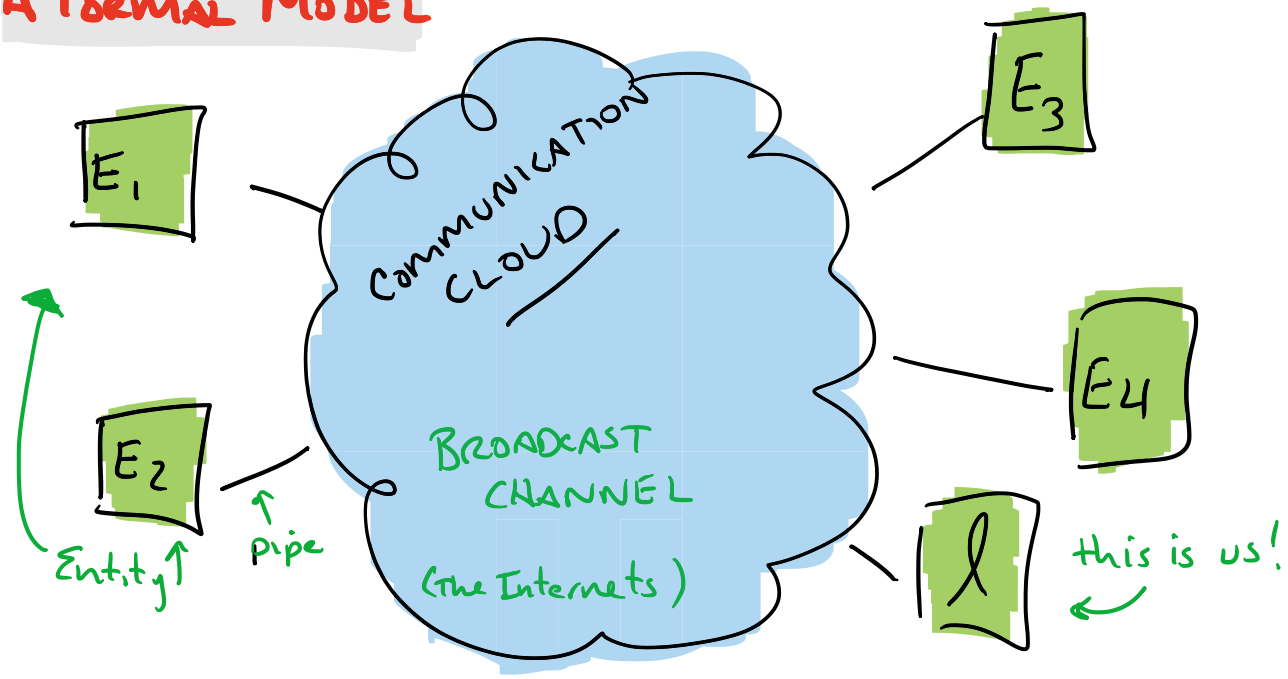
- reputation systems
- Internet voting

The only answer is a centralized trusted authoritative server

We can see this in practice

- you have one UMass email/account
- what did that authority do?
  - ask for money?
  - a signature?

A Formal Model



Set of entities E; pipes; and a cloud

① E is composed of two sets $E : C \cup F$

- C "correct" $c \in C$ (honest)
- F "faulty" $f \in F$ (deceptive)

② Entities communicate via messages

- all messages are broadcast

③ All messages are received within a bounded delay

- delivery is guaranteed
- but order of delivery at each is not

Assumptions

What are the consequences of these assumptions?

- they are very general
  - applies to all kinds of networks and systems.
- the attack is not due to a Denial of Service on a centralized (or distributed) authority
  - That is, the proofs assumes DoS is not possible for the attacker, and yet she succeeds.

More Assumptions:

- Each node has a set of computational resources
  - enough CPU to perform cryptography
  - but not enough CPU to break the crypto

  (i.e., the crypto works as advertised and is not the problem that allows the Sybil attack)
- Therefore, any solution that involves crypto that you can think of can be deployed
  - and yet Sybils still happen.

VOCABULARY

IDENTITY: (what is observed) an abstract representation that persists across messages.

ENTITY: (the truth; can't see this) each entity e presents an identity i to the cloud.

We say that local identity l **ACCEPTS** i when e presents itself successfully to l.

CORRECT ENTITIES - present one legitimate identity to the cloud

FAULTY ENTITIES - present one legitimate; and one counterfeit identity to the cloud.

Problem Statement:

Is there a method we can use by which only legitimate identities are accepted?

RESULTS

BASED on four lemmas; Here's a summary:

In general, there are 3 sources of information about identities:

1. a trusted, central agency/server;
2. itself;
3. other (untrusted) entities via their identities.

In the absence of option 1, I either

A. directly validates

OR

B. accepts the recommendation of others.

A FOR DIRECT VALIDATION

- even when resource constrained, a member of F can create a constant number of identities
- Unless all members of a set S are forced to validate all entities simultaneously, then the number of identities available to F is unbounded.

(B) FOR INDIRECT VALIDATION

- A sufficiently large set F can counterfeit an unbounded set of identities;
- UNLESS all entities validate at once; otherwise a single $f \in F$ can create a constant number of fradulent identities.

Here's the more formal proof:

LEMMA 1: DIRECT VALIDATION

IF ρ is the ratio of the resources of a faulty entity f to the resources of a minimally capable entity.

THEN f can present itself as $g = \lfloor \rho \rfloor$ identities to local identity l.

PROOF: It's almost direct. Let f have at least $g \cdot r$ resources, where r is the minimum. Therefore, f can present g identities to l. QED

ρ is the greek letter rho.

Is there an upper bound for an attacker? yes!
and it depends on what is the bounded resource.

① communication: $\mathcal{A}$ accepts only replies within an interval (i.e. network delay or bandwith is limited).

② storage: the challenge to f is to store a large amount of data; $\mathcal{A}$ queries a random portion of it.

③ processing: a crypto puzzle

Implicitly all three involve time.

Douceur suggests this puzzle for ③

given y, find x and z

such that

$$LSB_n(hash(x|y|3)) = 0$$

or more simply $hash(x|y) < target$

hey! that's bitcoin!

Lemma 2: Indirect Validation

IF ℓ accepts entities that are not validated simultaneously,

THEN $\mathcal{S}$ can present an arbitrarily large number of distinct identities to ℓ.

Proof: (again, this is pretty direct.)
Each identity is presented in serial freeing up resources to present the next identity. QED

LEMMA 3: INDIRECT VALIDATION

(recommendations of others: i_1 may VOUCH for i_2)

IF ℓ accepts any identity VOUCHED for by q accepted identities,

THEN a set F can present an arbitrarily large number of distinct identities to ℓ,

IF EITHER $|F| \geq q$

OR

the collective resources of F is equal to $q + |F|$ minimally capable entities.

PROOF: Let r_F be the total resources of set F.

Let r_K be the resources available to each $f_K \in F$

in other words

$$r_F = \sum_{\forall K} r_K$$

Let r_m be the minimally capable bar.

Then

$$q + |F| \leq \frac{r_F}{r_m} = \sum_{f_K \in F} \left(\frac{r_K}{r_m} \right) < \sum_{f_K \in F} \left\lfloor \frac{r_K}{r_m} \right\rfloor + |F|$$

the amount of resources attacker is faking; she has more than that

F has at least that many resources

Since $r_F = \sum_{\forall K} r_K$

There is an easier way to say all that, right?

if $\frac{r_F}{r_m} > q$, then F can be q. QED

Lemma 4

IF the set C do not coordinate time intervals during which they accept identities

and if ℓ accepts any entity vouched for by q accepted identities

THEN even a single, minimally capable faulty entity f can present

$q = \lfloor \frac{|C|}{q} \rfloor$ distinct identities to ℓ.

Proof: Let C be our correct set.

Partition C into subsets, each of size q or larger:

$$C = C_1 \cup C_2 \cup C_3 \cup \cdots \cup C_k$$

1. f then presents identity i_k to each C_k during time interval T_k.
2. After each interval, there are q identities from C that can vouch for C_k.
3. Each i_k is presented to ℓ, and we find that each has q identities to vouch for it.

QED