computas

Distributed Computing 2.0

Filip van Laenen
fvl@computas.com
@filipvanlaenen

About Computas

- Approx. 180 employees
- Located in Oslo area
- 100% employee owned
- Core business: BPM, Case Management Systems
  - Also: Integration, IAM, Web Portals, Semantic Web, ...
- Technologies: .Net and Java

And we're hiring...

Agenda

- From Distributed Computing 1.0 to Distributed Computing 2.0
- Introduction to SHA1CRK
- Example message protocol
- Distributed database choices and issues
- Deciding on the next task
- Composing the final result
- Incorrect partial results
- Malicious clients
- More information
- Questions and comments

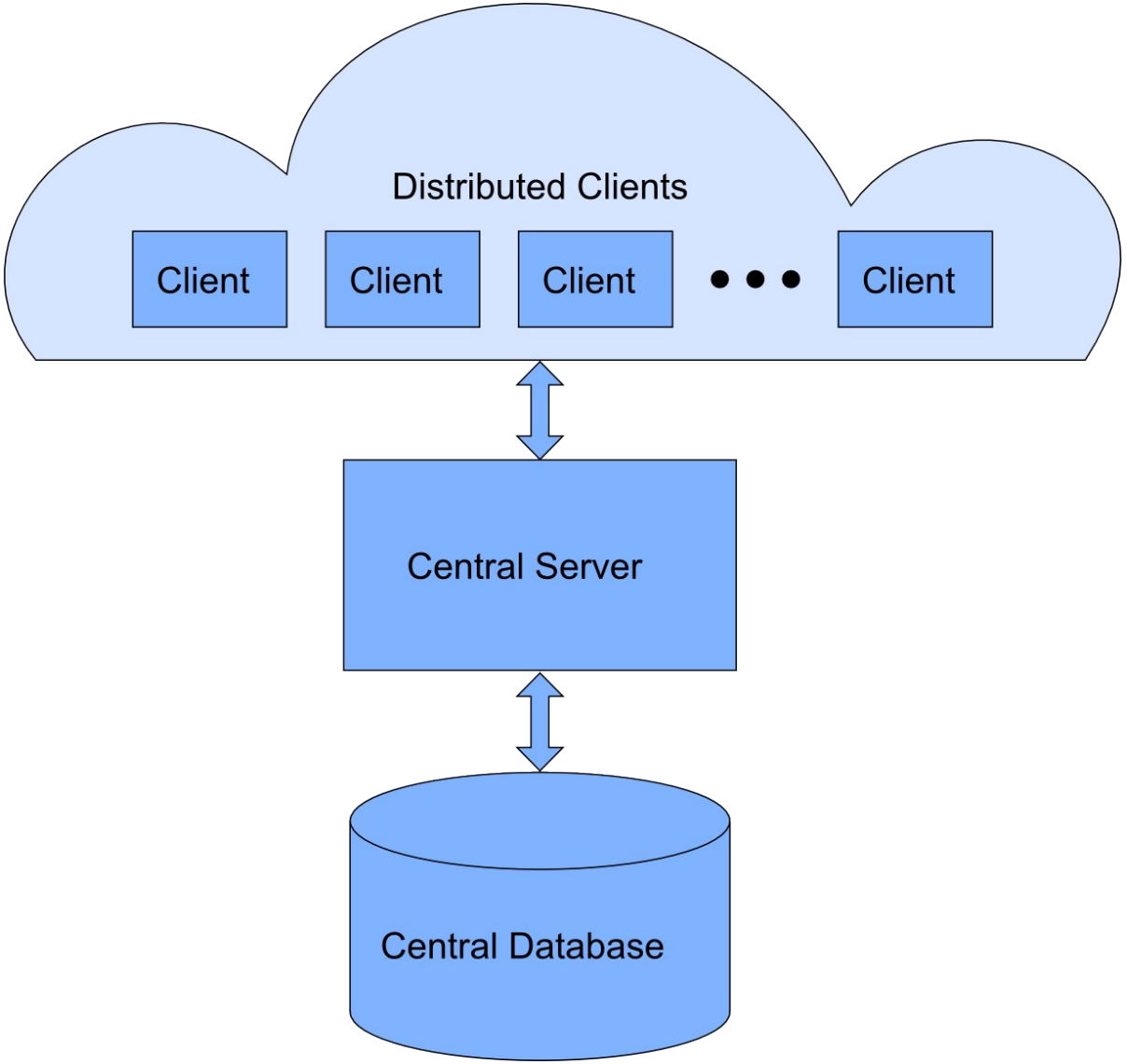
Distributed Computing (1)

- Definition
  - Distributed computing
  - Parallell computing
  - Grid computing
  - Volunteer computing
- Typical areas
  - Data mining
  - Combinatorial problems
  - Prime numbers
  - Cryptology

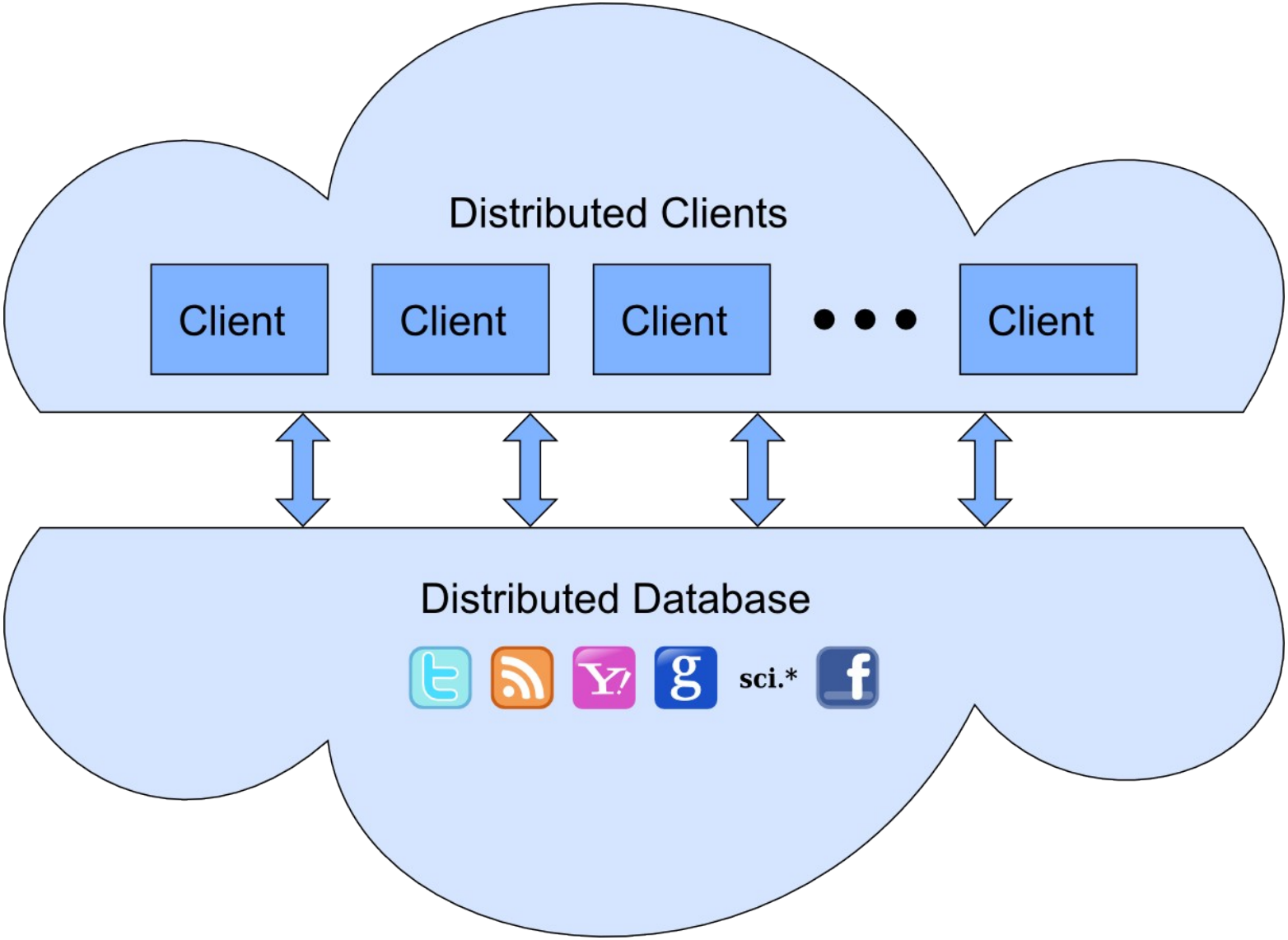
Distributed Computing (2)

- Typical solution strategies
  - Exhaustive searches
  - Brute force attacks
  - “Educated force” attacks
- Example projects
  - SETI@Home
  - Folding@Home
  - Seventeen or Bust, GIMPS
  - MD5CRK
  - And many more on BOINC

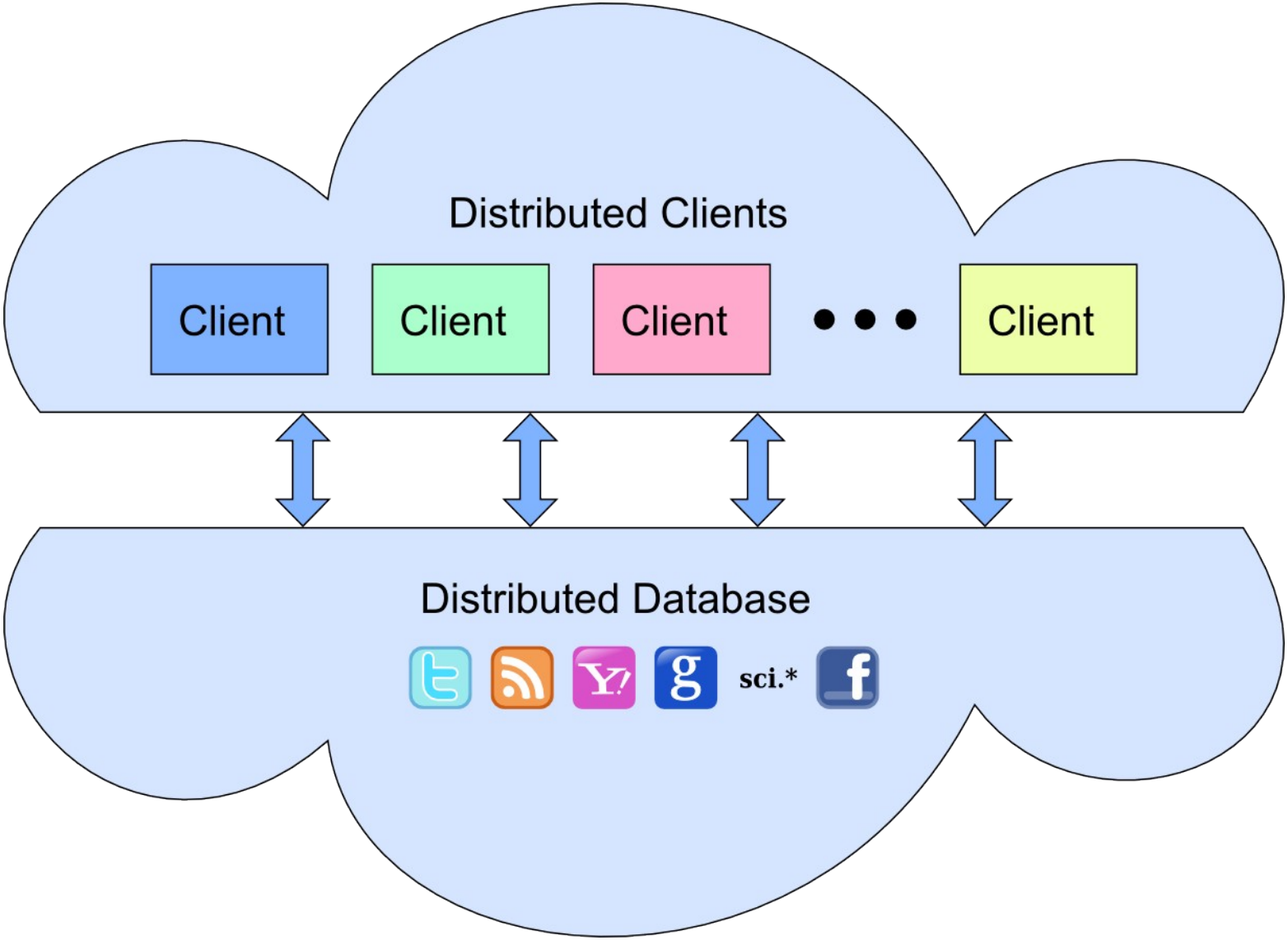
Traditional Architecture



Distributed Computing 2.0



Client Specialization



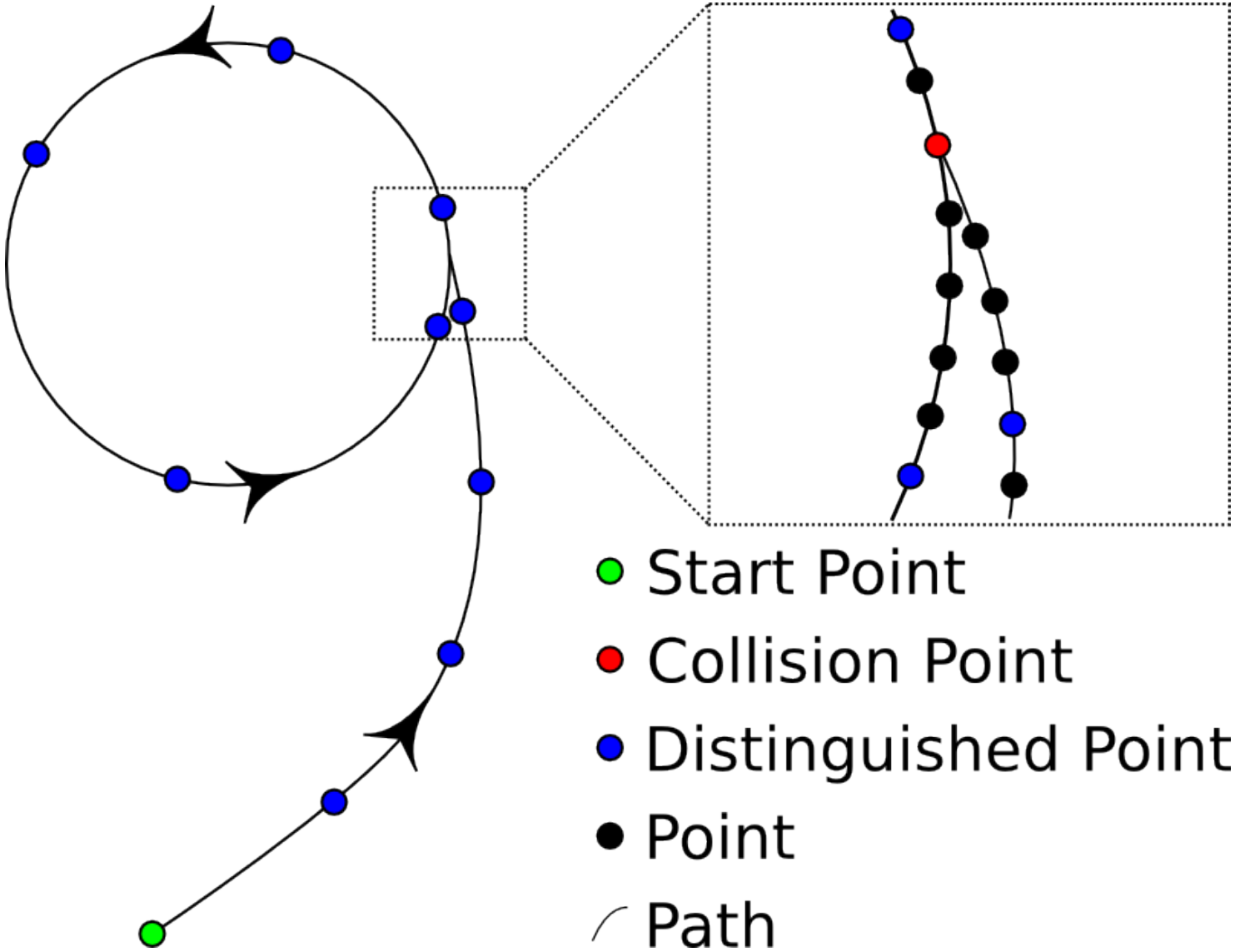
Introduction to SHA1CRK

- Predecessors:
  - MD5CRK
  - SHA-1 Collision Search Graz
- SHA-1: 160 bit cryptographic hash function
  - MD5: 128 bit cryptographic hash function
- Goal: Find a collision
- Message channel: Twitter
- Task decision: Locally
- Strategy: Pollard's ρ collision search

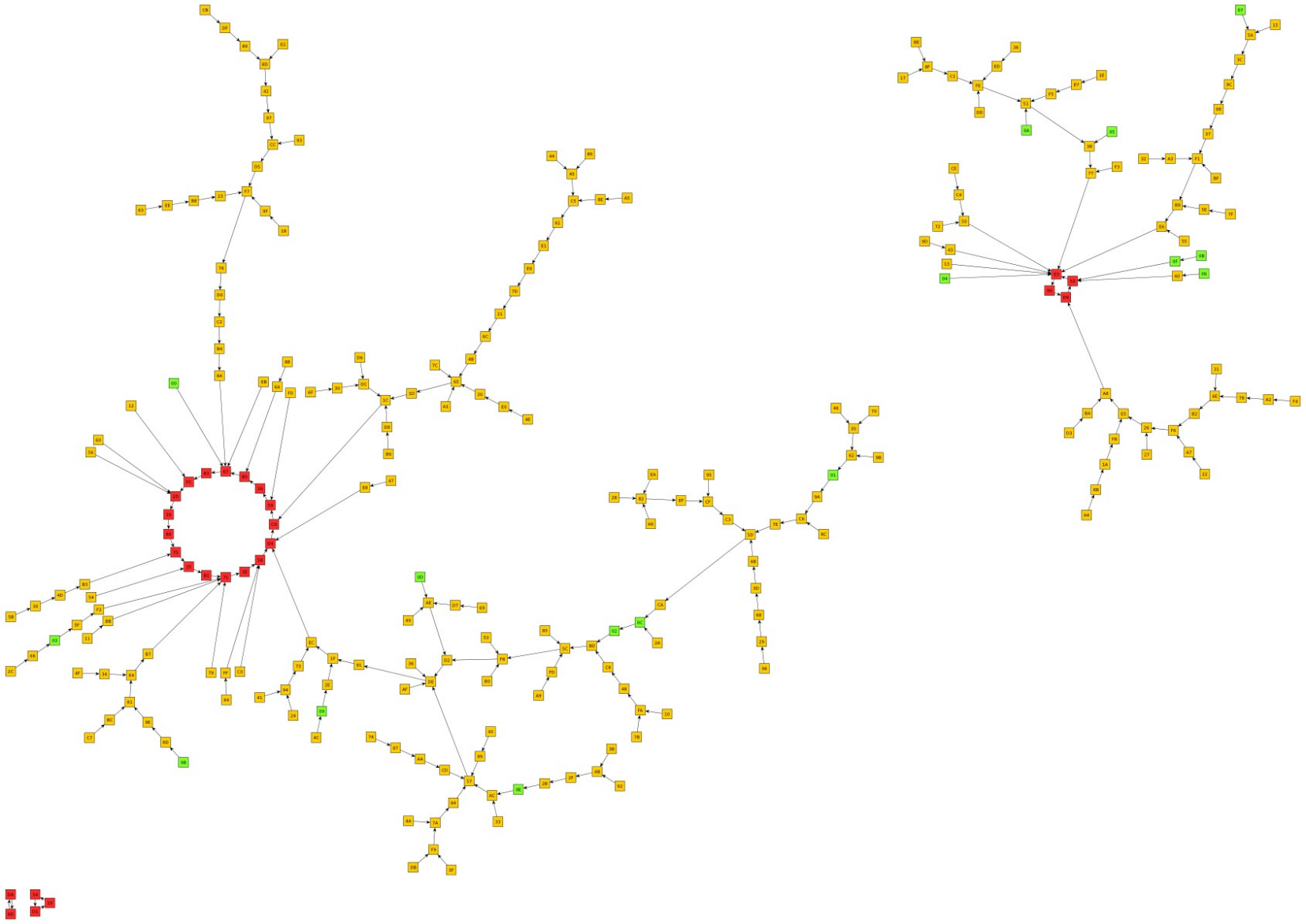
Cryptographic hash functions

- Definition:
  - Takes an arbitrary block of data and returns a fixed-size bit string, such that an accidental or intentional change to the data will change the hash value.
- Used in digital signatures
- Ideal cryptographic hash function properties:
  - Easy to compute hash value of a given message
  - Infeasible to find a message for a given hash value
  - Infeasible to modify message without changing its hash value
  - Infeasible to find two different messages with same hash value

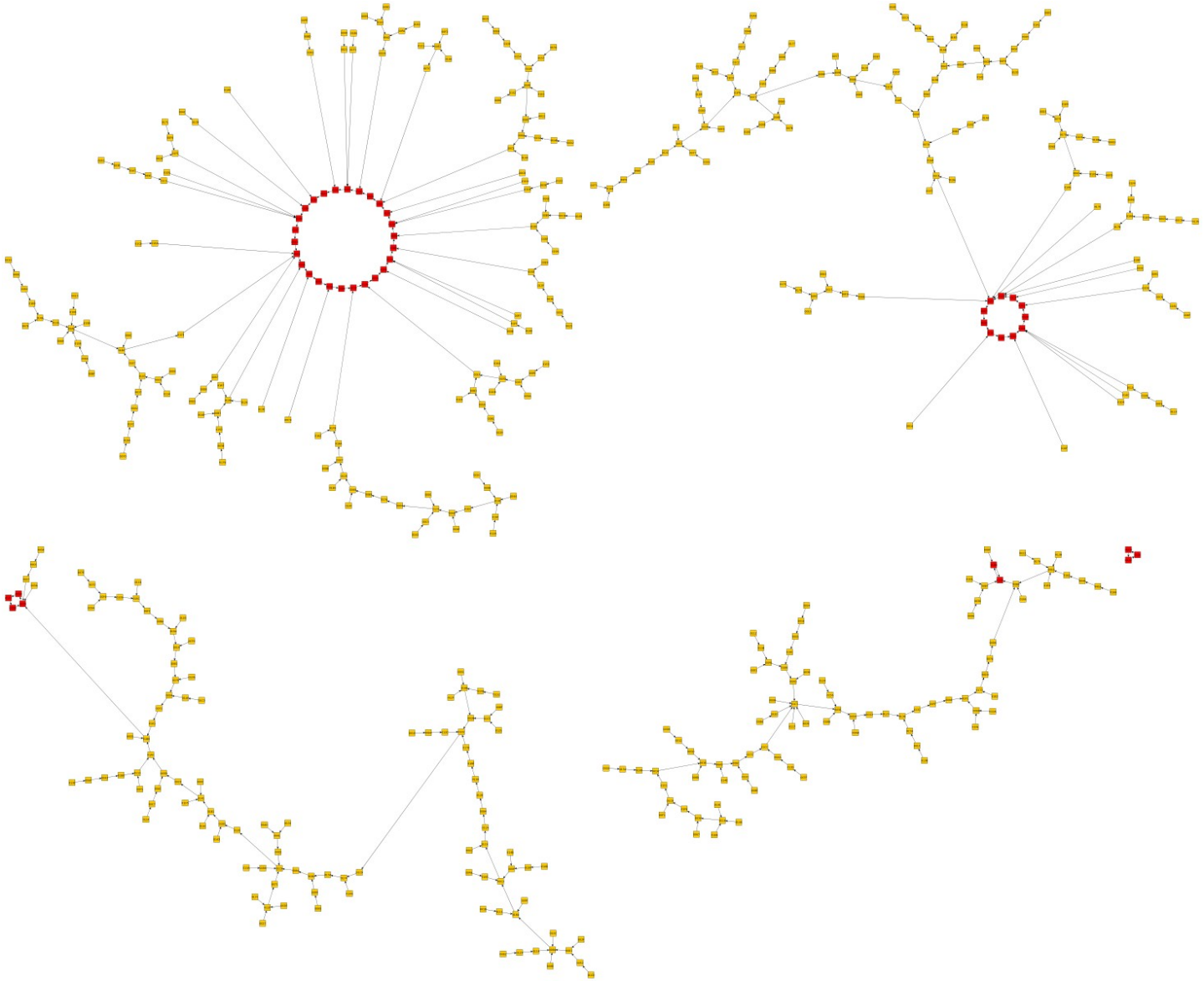
Pollard's Rho Collision Search



Reduced SHA-1 Map (8 Bits)



Reduced SHA-1 Map (9 Bits)



Expected Time to Find a Collision (1)

computas

- Search space:
  - Hash length: 160 bits
  - Therefore: max. $2^{160} = 1.46 \times 10^{46}$ possible hashes
- Number of hashes/distinguished points for a collision:

Probability	Hashes	Distinguished Points
10^{-6}	1.7×10^{21}	2.6×10^{16}
0.1%	5.4×10^{22}	8.3×10^{17}
1%	1.7×10^{23}	2.6×10^{18}
10%	5.5×10^{23}	8.5×10^{18}
25%	9.2×10^{23}	1.4×10^{19}
50%	1.4×10^{24}	2.2×10^{19}
75%	2.0×10^{24}	3.1×10^{19}

Expected Time to Find a Collision (2)

computas

- Storage space:
  - Single distinguished point: 160 bits = 20 bytes
  - Segment: 2 × 20 bytes = 40 bytes
- Total storage space required for a collision:

Probability	Hashes	Distinguished Points
10^{-6}	58 ZB	910 PB
0.1%	1.8 YB	29 EB
1%	5.8 YB	91 EB
10%	19 YB	294 EB
25%	31 YB	485 EB
50%	48 YB	754 EB
75%	68 YB	1.1 ZB

Expected Time to Find a Collision (3)

computas

- My velocity:
  - 400,000 hashes/second (at 100% CPU, both cores)
  - 3,000 distinguished points/year (24/365)
- Number of computers to complete in 1 year:

Probability	2011	2020	2030	2040
10^{-6}	8.9×10^{12}	8.7×10^{9}	8.5×10^{6}	8,272
0.1%	2.8×10^{14}	2.7×10^{11}	2.7×10^{8}	262,000
1%	8.9×10^{14}	8.7×10^{11}	8.5×10^{8}	829,000
10%	2.9×10^{15}	2.8×10^{12}	2.7×10^{9}	2.7×10^{6}
25%	4.8×10^{15}	4.7×10^{12}	4.5×10^{9}	4.4×10^{6}
50%	7.4×10^{15}	7.2×10^{12}	7.1×10^{9}	6.9×10^{6}
75%	1.0×10^{16}	1.0×10^{13}	1.0×10^{10}	9.7×10^{6}

Expected Time to Find a Collision (4)

computas

- SHA-2:
  - Two block sizes: 256 and 512 bits
  - Truncated versions: 224 and 384 bits
- Number of computers to find a collision in 1 year:

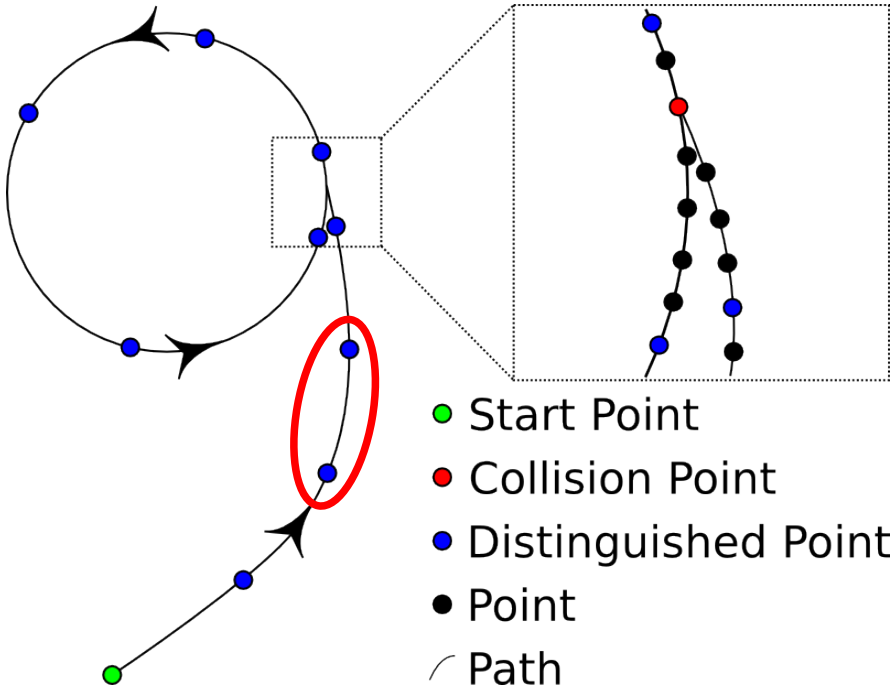
Probability	2011		2040	
10^{-6}	2.5×10^{27}	8.5×10^{65}	2.3×10^{18}	7.9×10^{56}
0.1%	7.9×10^{28}	2.7×10^{67}	7.4×10^{19}	2.5×10^{58}
1%	2.5×10^{29}	8.5×10^{67}	2.3×10^{20}	7.9×10^{58}
10%	8.1×10^{29}	2.8×10^{68}	7.6×10^{20}	2.6×10^{59}
25%	1.3×10^{30}	4.6×10^{68}	1.2×10^{21}	4.2×10^{59}
50%	2.1×10^{30}	7.1×10^{68}	1.9×10^{21}	6.6×10^{59}
75%	2.9×10^{30}	1.0×10^{69}	2.7×10^{21}	9.3×10^{59}

Cryptanalysis of SHA-1

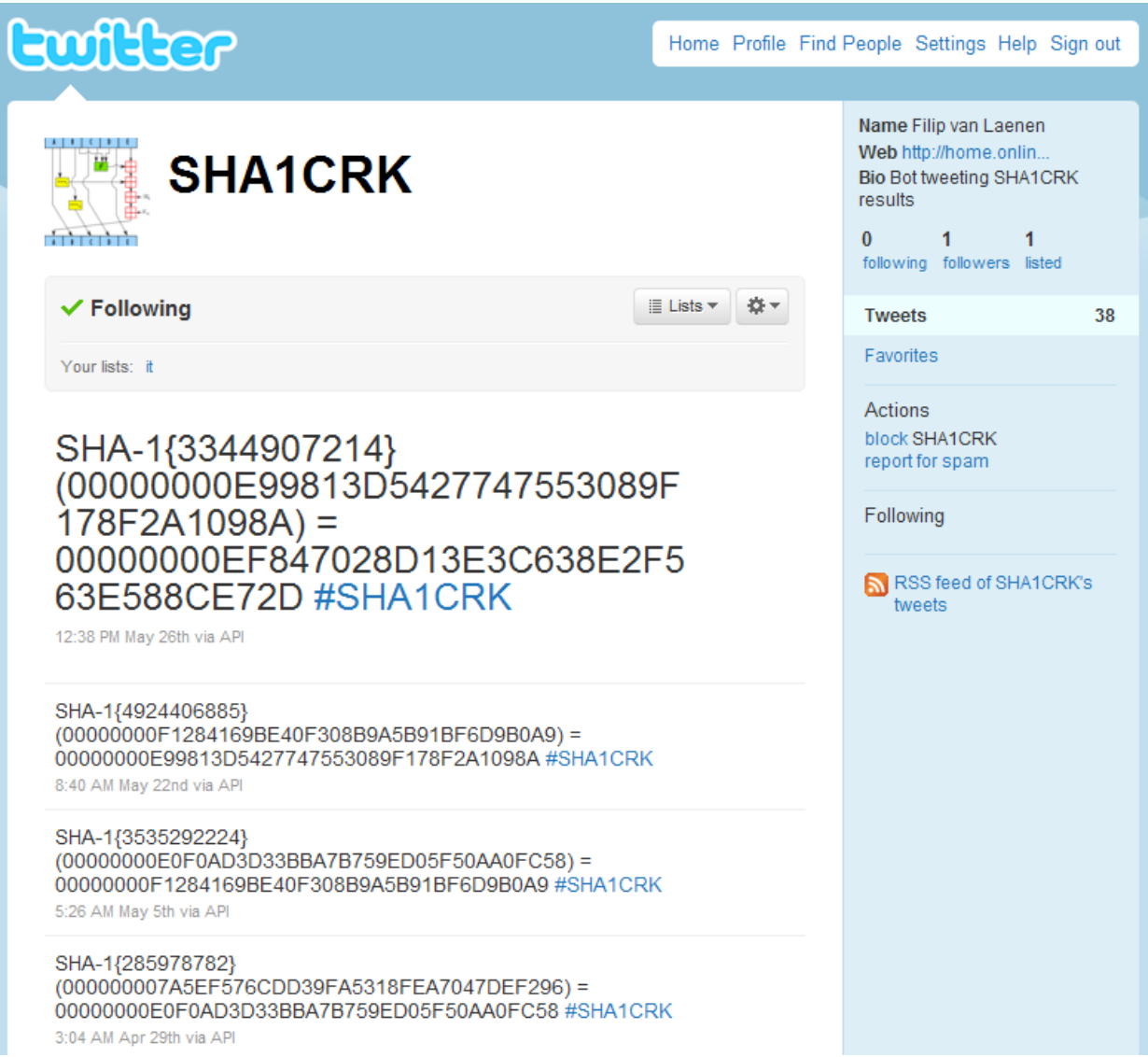
- Brute force attack: 2^{80} hashes
- "Educated force" attacks described:
  - Wang, Yin and Yu (2005): 2^{69} hashes
  - Wang, Yao and Yao (2005): 2^{63} hashes
  - McDonald, Hawkes and Pieprzyk (2009): 2^{52} hashes (unconfirmed)
- However: no implementations available yet
  - Can the "educated force" attacks be distributed?

SHA1CRK Message Protocol (1)

- Segment calculated:
  - **SHA-1{<*int*>}(<*hex(40)*>) = <*hex(40)*> #SHA1CRK**

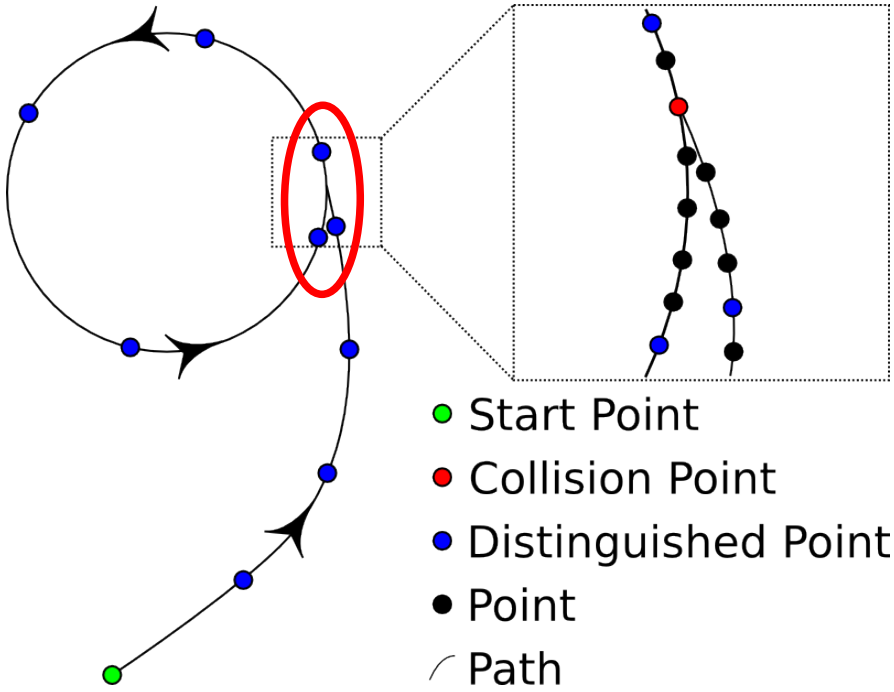


SHA1CRK Message Protocol (2)



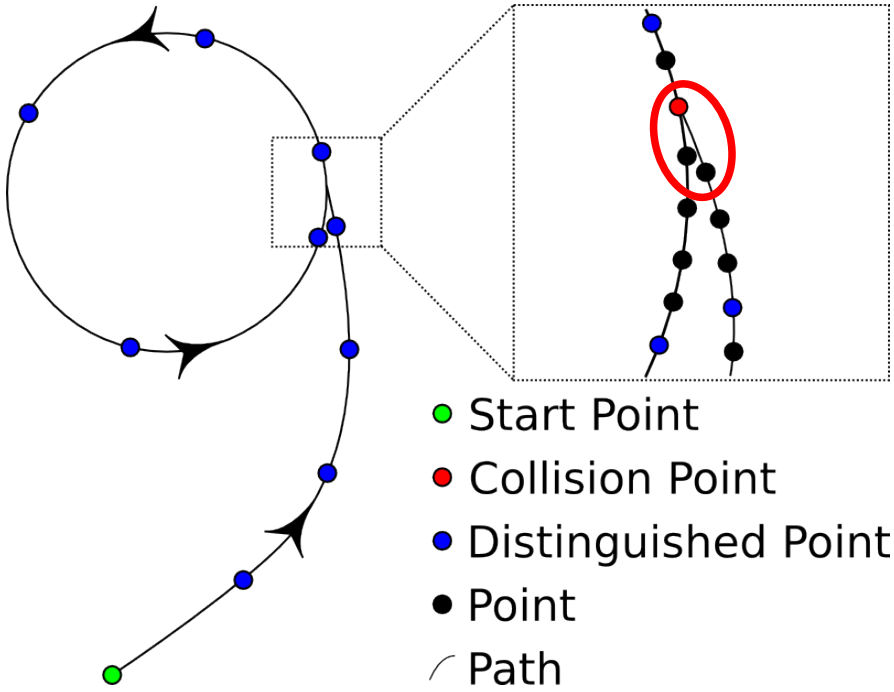
SHA1CRK Message Protocol (3)

- Two segments collide:
  - **SHA-1{<*int*>}(<*hex(40)*>) = SHA-1{<*int*>}(<*hex(40)*>) #SHA1CRK**



SHA1CRK Message Protocol (4)

- Collision found:
  - **SHA-1(<*hex(40)*>) = SHA-1(<*hex(40)*>) #SHA1CRK**



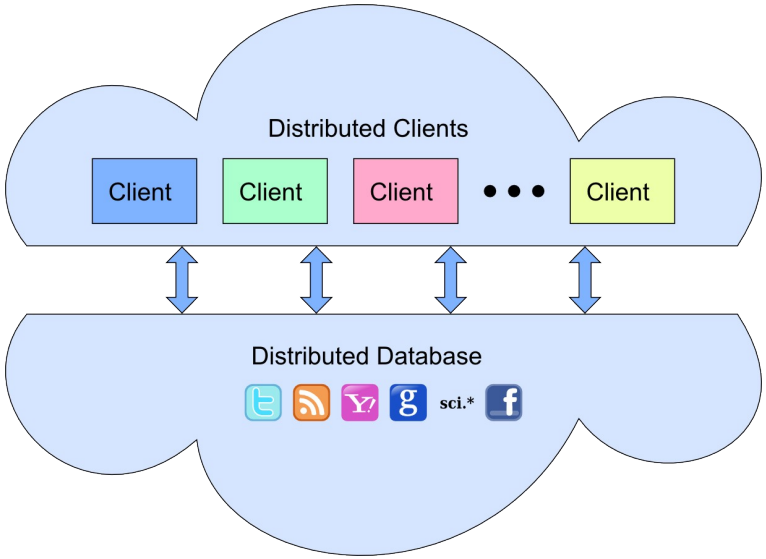
SHA1CRK Message Protocol Extensions

- Hamming distance between end-points of two segments:
  - **SHA-1{<*int*>}(<*hex(40)*>) ≈ SHA-1{<*int*>}(<*hex(40)*>), d=<*int*> #SHA1CRK**
- Additional information from partial results
- Administrative information:
  - Client upgrades
  - New project members
- Statistics

Distributed Database Options

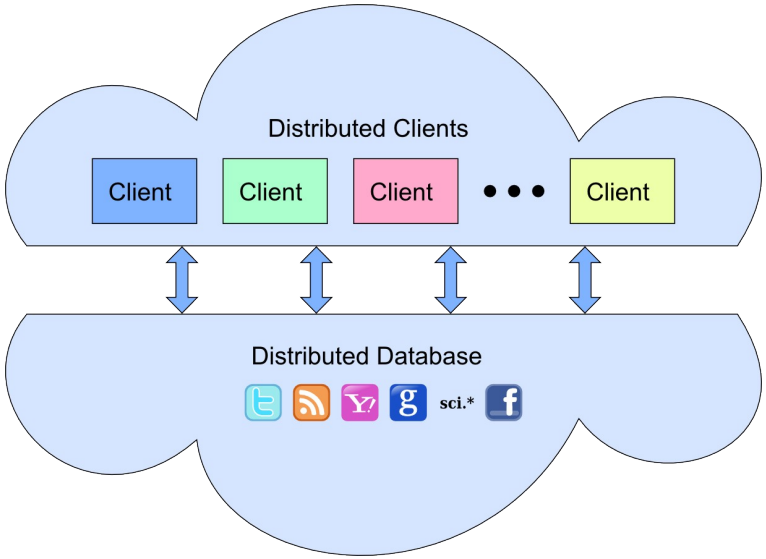
- Twitter
- RSS
  - Blogs
  - File based
- Yahoo groups
- Google groups
- Usenet groups
- Facebook?
- Flickr?

- Combination of more than one?



Distributed Database Issues

- Bandwidth
- Searchability
  - Breadth and depth
  - External searchability
- Peer discovery
- Availability
- Reliability
- Legal issues



The Next Task

- Client decision
  - Task collisions must be handled
  - Good if large space and all tasks have the same value
- Public announcements/reservation
  - Must still handle collisions
  - Preferable when tasks have different values
- Authoritative clients
  - Bottleneck
  - Small space and/or tasks have different values

Composing the Final Result

- How do you know you're done if you can't see all partial results?
  - Problem for Twitter, RSS, ...
    - But can be solved if necessary
  - Not a problem in Yahoo groups, Google groups, Usenet groups
- Specialized client
- May or may not be important
  - Is the final result unique?

Incorrect Partial Results

- How much damage can be done?
  - Incorrect end-point
    - Collision may be missed
    - False collision warning
  - Incorrect starting point
    - Collision may be missed
- Consequences depend on problem
- Solution:
  - Independent verification
  - Independent recalculation
  - Correcting messages

SHA1CRK Contradiction Messages

- Contradiction, disagreement, correction?
- Two segments do not collide:
  - **SHA-1{<*int*>}(<*hex(40)*>) ≠ SHA-1{<*int*>}(<*hex(40)*>) #SHA1CRK**
- Miscalculation of a segment:
  - **SHA-1{<*int*>}(<*hex(40)*>) ≠ <*hex(40)*> #SHA1CRK**

Malicious Clients

- Malice implies malicious intent
- Naming and shaming
  - **SHA-1{*<int>*}(*<hex(40)>*) ≠ *<hex(40)>* @*<ID>* #SHA1CRK**
- Each client must decide for itself to trust other clients
  - Or trust other clients to make the decision
  - Reputation messages

More Information

- fvl@computas.com or f.a.vanlaenen@ieee.org
- @filipvanlaenen
- @sha1crk
- http://filipvanlaenen.net/sha1crk/index.html

Questions?

Contact:

Computas AS
Lysaker Torg 45, pb 482
1327 Lysaker
Norway

Tel +47-67 83 10 00
Fax +47-67 83 10 01
Org.nr: NO 986 352 325 MVA
www.computas.com