Distributed Overlay Construction to Support Policy-based Access Control

Bong Jun Ko*, Starsky W Ho, Kang-won Lee
(IBM T. J. Watson Research)
Sid Chau (University of Cambridge)
A Tale of Two: “Once upon a time …”
A Tale of Two: Building an information sharing network

- **Rules of information sharing**
  - “Everyone is permitted to access the information of his/her country.”
  - “Generals are permitted to access the information of the other country.”
  - “Soldiers are not permitted to access the information of the other country.”

- **Trust**
  - Authorized personnel will not send the information to unauthorized personnel, but will send and forward the information to other authorized personnel over multi-hop paths.

- **Mission:** “**Construct a network of secure links that can be used to pass the information around, without violating the policies.**”

- **Mission#2:** “**Don’t create too many links. They are expensive.**”
A Tale of Two: Building an information sharing network

- Rules of information sharing
  - “Everyone is permitted to access the information of his/her country.”
  - “Generals are permitted to access the information of the other country.”
  - “Soldiers are not permitted to access the information of the other country.”

- Trust
  - Authorized personnel will not send the information to unauthorized personnel, but will send and forward the information to other authorized personnel over multi-hop paths.

- Mission: “Construct a network of secure links that can be used to pass the information around, without violating the policies.”

- Mission#2: “Don’t create too many links. They are expensive.”
Node grouping by access control policy

Access-control policy statement to node grouping

“If $f_k(r) = true$ and $g_k(n) = true$, then permit $n$’s access to $r$.”

$\rightarrow N_k = \{ n \mid g_k(n) = true \}$ : access-control group $k$

If \{origin(data)=Oblivia\} AND
{ {country(person)=Oblivia} OR {rank(person)=General} },
then
Permit the person’s access to the data

If \{origin(data)=Memorabilia\} AND
{ {country(person)=Memorabilia} OR {rank(person)=General} },
then
Permit the person’s access to the data
Node grouping by access control policy (modern example)

Access-control policy statement to node grouping

“If $f_k(r) = true$ and $g_k(n) = true$, then permit $n$’s access to $r$”.

$\rightarrow N_k = \{n \mid g_k(n) = true\} : \text{access-control group } k$

- If organization($n$) == organization($s$), then permit $n$’s access to data from $s$.
- If roles($n$) includes dataType($s$), permit $n$’s access to data from $s$.

Permit n’s access to data from $s$:
- If organization($s$) == A and organization($n$) == A,
- If organization($s$) == B and organization($n$) == B,
- If dataType($s$) == AUDIO and AUDIO $\in$ roles($n$),
- If dataType($s$) == VIDEO and VIDEO $\in$ roles($n$), or
- If dataType($s$) == VIBR and VIBR $\in$ roles($n$)
System architecture

Access controlled groups

Shared overlay network

Physical Networks
Policy-compatible Overlay Construction (PoCO) Problem

- **Given**
  - Set of nodes: \( N = \{1, 2, \ldots, n\} \)
  - Set of groups of nodes: \( \Omega = \{N_1, N_2, \ldots N_K\} \), \( N_k \) is a subset of \( N \)

- A graph \( G=(N,E) \) is **policy-compatible** w.r.t. \( \Omega \) if, for each group \( N_k \), the subgraph \( G_k = (N_k, E_k) \) in \( G \) is a connected graph.

- Find a policy-compatible graph \( G^*=(N,E^*) \) such that \( |E^*| \leq |E| \) for all policy-compatible graphs \( G = (N,E) \).

- NP-Complete problem

**Q: Which of the followings is NOT policy-compatible?**

(a) ![Diagram (a)](image1)
(b) ![Diagram (b)](image2)
(c) ![Diagram (c)](image3)
(d) ![Diagram (d)](image4)
Centralized Overlay Construction

- **Group-connectedness**
  - $C_i = \{(u,v) | u$ and $v$ are not connected in $N_i, u \in N_i, v \in N_i\}$
  - $C = \Sigma_i |C_i|

- **Greedy addition of links**
  - At each step, a link that decreases $C$ the most is selected (until $C = 0$).

Start: $C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}$, $C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}$

Step1: $C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}$, $C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}$

Step2: $C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}$, $C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}$

Step3: $C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}$, $C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}$

$N_1 = \{1,2,3\}$, $N_2 = \{2,3,4\}$
Distributed Overlay Construction

- Each node adds a link that maximizes group-connected pairs in its respective groups.
- Each node also deletes *redundant* links in its groups.
  - A link \((u,v)\) is redundant if deleting it does not render \(u\) and \(v\) disconnected in their group(s).
- Each node exchanges *its local link information* with other nodes that belong to the same group via distributed protocols.

\[N_1 = \{1,2,3\},\]
\[N_2 = \{2,3,4\}\]

Start:
\[C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}, \quad C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}\]

Node 1:
\[C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}, \quad C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}\]

Node 4:
\[C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}, \quad C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}\]

Node 1:
\[C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}, \quad C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}\]

Node 3:
\[C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}, \quad C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}\]

\[C_1 = \{\{1,2\}, \{2,3\}, \{1,3\}\}, \quad C_2 = \{\{2,3\}, \{3,4\}, \{2,4\}\}\]
Distributed Overlay Construction: Convergence

Starting from an arbitrary graph $G_0$, the evolving sequence of the graphs $(G_0, G_1, G_2, \cdots)$ generated by the distributed process converges to a stable graph $G = (N, E)$ which is compatible w.r.t. the policy grouping in finite number of steps.
Performance (Empirical Data)

- Conference program committee membership data
- 482 individuals in 32 conference PCs in 2006.
- Policy: “Data can be shared among members in the same PC”.

* G-MST (Generalized Minimum Spanning Tree): Similar to Reverse-delete algorithm for MST, but deletes redundant edges
Performance (Synthetic Input)

- $|N|$: # of nodes (10 ~ 50)
- $|\Omega|$: # of groups (10, 20)
- $p$: Probability that a node belongs to a group
Performance (Synthetic Input)

- \(|N|\): # of nodes (10 ~ 50)
- \(|\Omega|\): # of groups (10, 20)
- \(p\): Probability that a node belongs to a group

![Graph 1](image1.png)

|\(N| = 50, |\Omega| = 10

![Graph 2](image2.png)

|\(\Omega| = 20, p = 0.5
Performance: Dynamics of distributed algorithm

|N| = 50

|N| = 40

|N| = 30

|N| = 20

|N| = 10

Total # of link additions/deletion by distributed algorithm
Extension for link cost optimization

- Link cost function: $c: N \times N \rightarrow R^+$
  - Accounts for node distance, trust level, etc.

- Construct overlay that minimizes the total cost: $c(E) = \sum_{(u,v) \in E} c(u,v)$
Conclusion

- A new problem of building a network constrained by access control policy.

- Build a shared overlay used across multiple access-control groups.

- Optimize over the number of links in the shared overlay
  - Greedy heuristics (centralized and distributed) work pretty well.

- Open problems
  - Partial knowledge in the distributed construction
    - Nodes may not have complete knowledge of group topology.
  - Risk-aware overlay
    - Some nodes are more trustworthy than others.
  - Resilient, efficient, and robust overlay
    - Small diameter
    - Dynamics: Node join/leave; policy updates
    - Compromised nodes; nodes under attack