How Extended Unix Tools Can Measure the Changing Security Posture of Power-Control Networks

Gabriel A. Weaver, Edmond Rogers, Rakesh Bobba, Sean W. Smith
Dartmouth College, TCIPG Center

TCIPG Seminar
1/4/13

Practitioners identify and categorize meaningful structures within a variety of data sources in order to evaluate security.

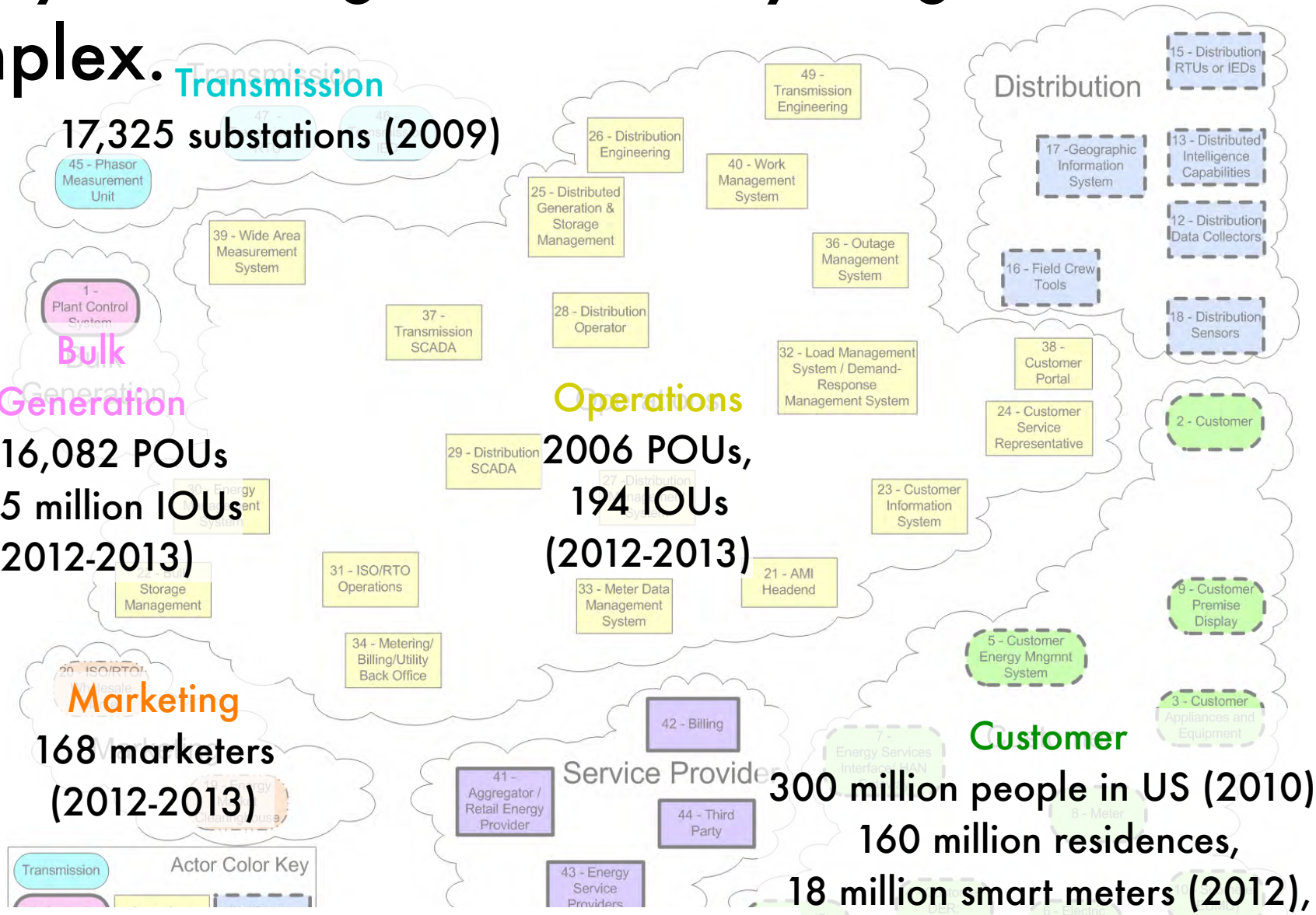
Our research interprets many of these structures (lines, interface blocks) as languages.

We built tools to process and analyze text with respect to those languages.

Just as programmers use high-level languages to program more efficiently...

So can practitioners use high-level languages to audit and maintain power-control networks.

Today's smart grid is already large and complex.



Transmission

17,325 substations (2009)

Bulk Generation

1416,082 POUs
~1.5 million IOUs
(2012-2013)

Operations

2006 POUs,
194 IOUs
(2012-2013)

Marketing

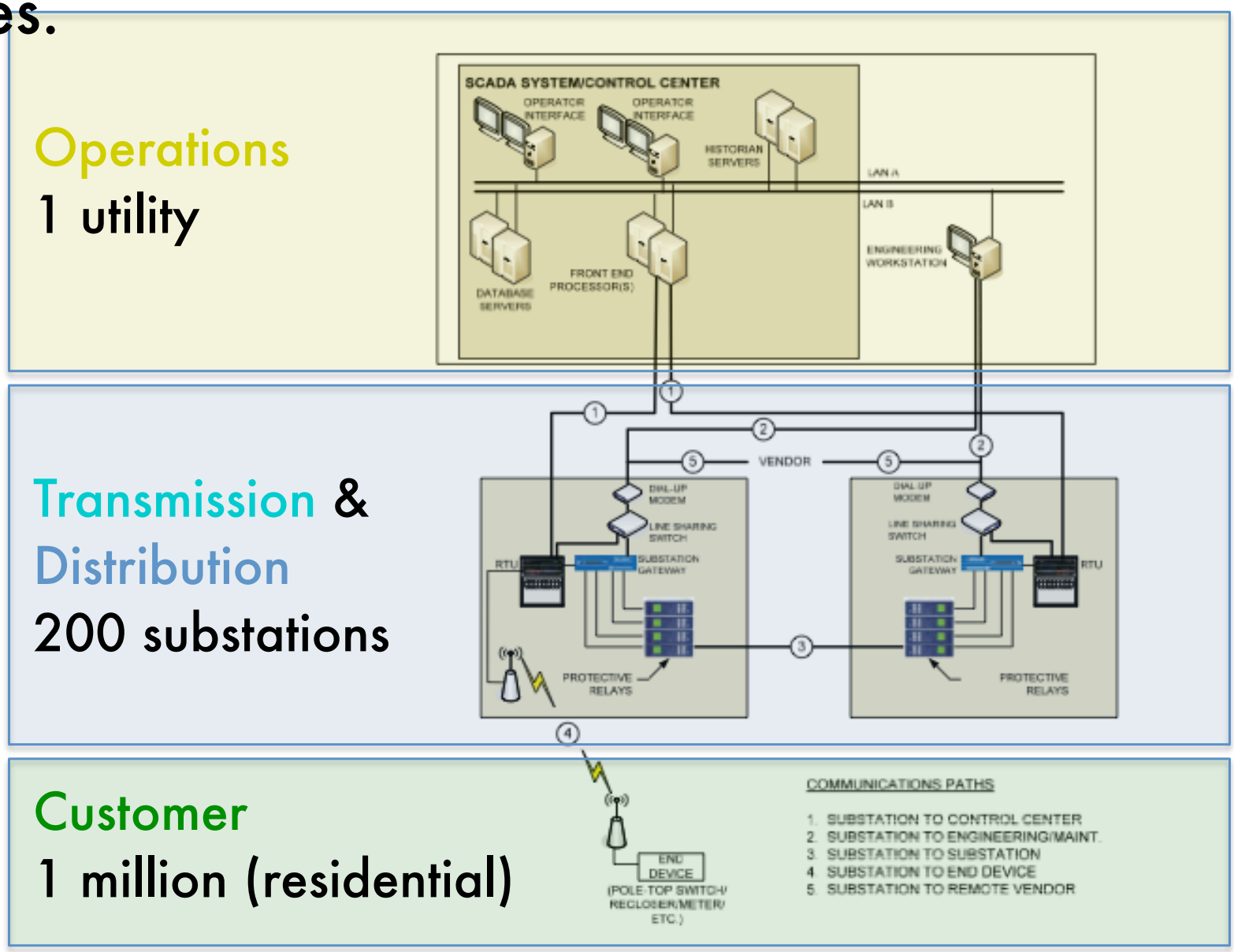
168 marketers
(2012-2013)

Customer

300 million people in US (2010),
160 million residences,
18 million smart meters (2012),
250 million registered cars (2010)

[NIST Smart Grid Program Overview, 2012]
[APPA 2012-2013 Annual Directory & Statistical Report, 2012]

Substation communications at one utility involve many devices.



[INL National SCADA Test Bed Substation Automation Evaluation Report, 2009]

In the Electrical Power Grid, security policies and related artifacts are expressed in a variety of forms.

	Device	Data Type
Operations (1 IOU)	SCADA/Corporate Network	Cisco IOS, Juniper, IEC 61850, CIM
	Data Historian	C37.118
	Operator Interface	Windows Registries, logs
	Engineering Workstation	Windows Registries, logs
Transmission/ Distribution (200 substations)	RTU/Substation Gateway	DNP3, IEC 61850, RADIUS
	Engineering Workstation	Windows Registries
	Substation LAN	Cisco IOS, SCL IED, GOOSE, CIM
	PMU/Relays	C37.118, SCL IED, GOOSE, CIM
Customer (1 million)	Meters	RTU, DNP3
	Electric Cars	Green Button (ESPI XML)
	Appliances	Green Button (ESPI XML)

NERC CIP requires utilities to manage this data via baseline configuration development and change control.

Relevant Provisions

CIP 003-4: Change control and configuration management

CIP 010-1: Baseline configuration development and comparison

CIP 005-4: Update network documentation within 30 days of a change.

Practical Considerations

1. Audits currently consume 30 man days per day of audit.
2. Audits cost large IOUs from hundreds of thousands to millions of dollars.
3. Utilities are currently on a 3 year audit cycle, but FERC would like annual audits.
4. Fines for noncompliance are enough to "bankrupt small nation states." [Conversations with Edmond Rogers, 2012]

High-level research barriers prevent cheaper, more consistent audit.

We need "common terms and measures specific to each energy subsector available to baseline security posture in operational settings."

1. "Try to provide actionable and timely information of security posture from vast quantities of disparate data from a variety of sources and levels of granularity" [Roadmap to Achieve Energy-Delivery Systems Cybersecurity, 2011].

2. "New measurement methods and models are needed to sense, control, and optimize the grid's new operational paradigm." [NIST Smart Grid Program Overview, 2012]

3. Need to develop cybersecurity solutions that are (a) robust to changes in technology and (b) develop capabilities that might be applicable elsewhere. [DOE Cybersecurity Information Exchange, Samara N. Moore, 2012]

We view these barriers as symptoms of three core limitations of textual analysis.

	Description	Baseline configuration and change control in the Power Grid
Tools Gap Problem	There is a gap between practitioner tools and security policy languages.	A wide variety of disparate data for devices on grid, but no common framework.
Granularity of Reference Problem	Practitioners cannot process policy at multiple levels of abstraction.	Many smart-grid formats (SCL, GOOSE, CIM, ESPI-XML) have hierarchical object models.
Discovery Needs Problem	Practitioners need to measure security policy and how it evolves.	Practitioners need to measure how device configurations change and baseline security policy.

Outline

1. Motivation
2. Theoretical Toolbox
3. XUTools Capabilities
   i. Baseline Configuration Development
   ii. Change Control
4. Ongoing Research
5. Conclusions

Outline

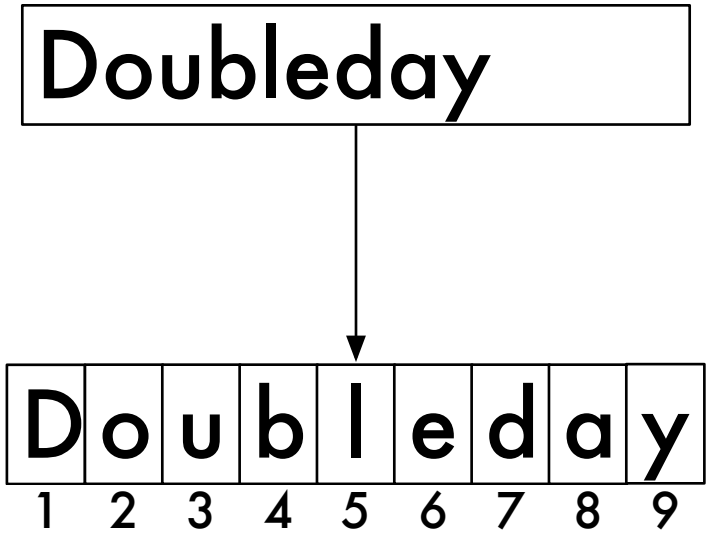
1. Motivation
2. Theoretical Toolbox
3. XUTools Capabilities
   i. Baseline Configuration Development
   ii. Change Control
4. Ongoing Research
5. Conclusions

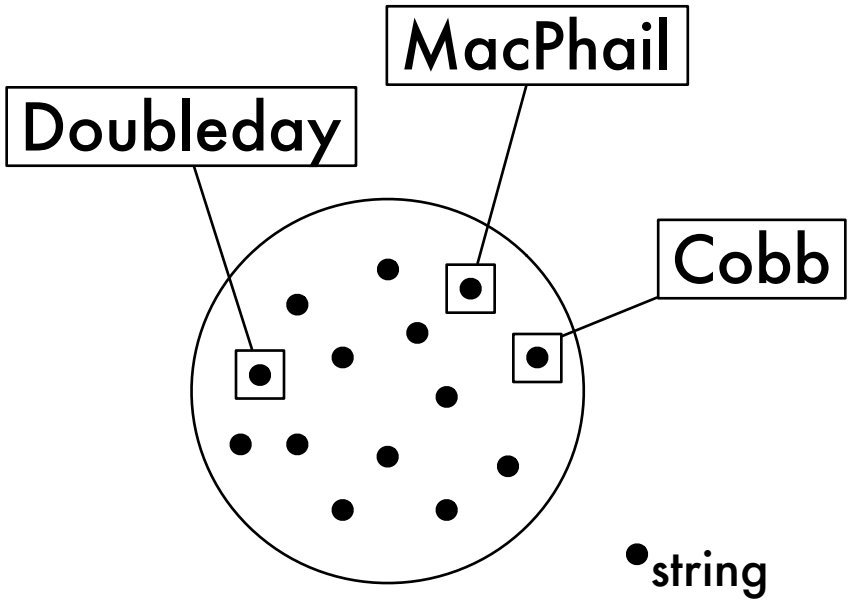
We can reduce audit cost by formalizing security policy analyses involved in baseline configuration development and change control.

First, we must understand the languages that practitioners use to express and analyze security policies.

Therefore, we begin with the definition of a language.

What is a language?

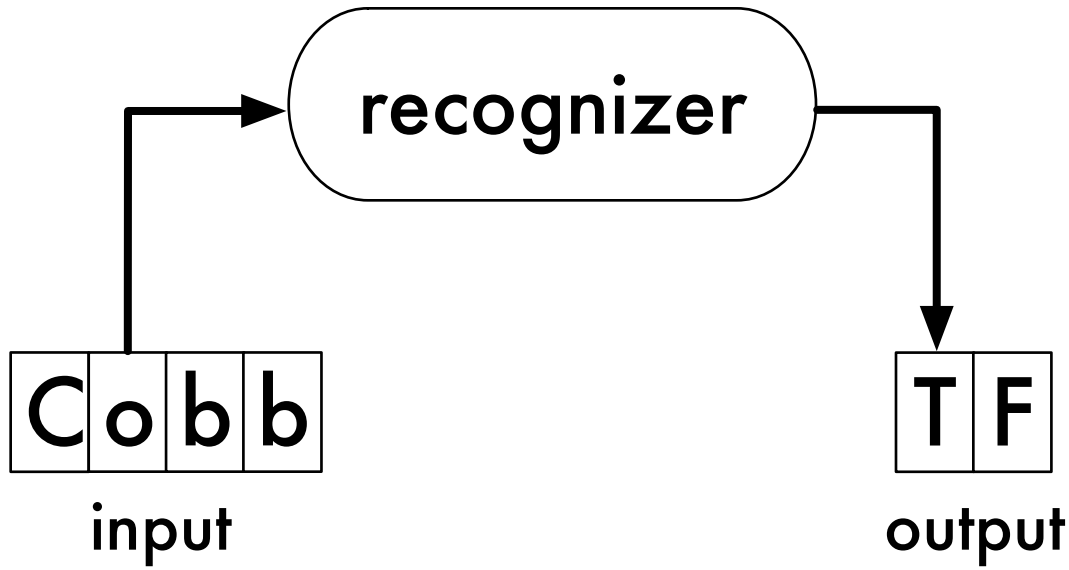




A string is a sequence of symbols taken from some alphabet.

A language is an unordered collection of unique strings.

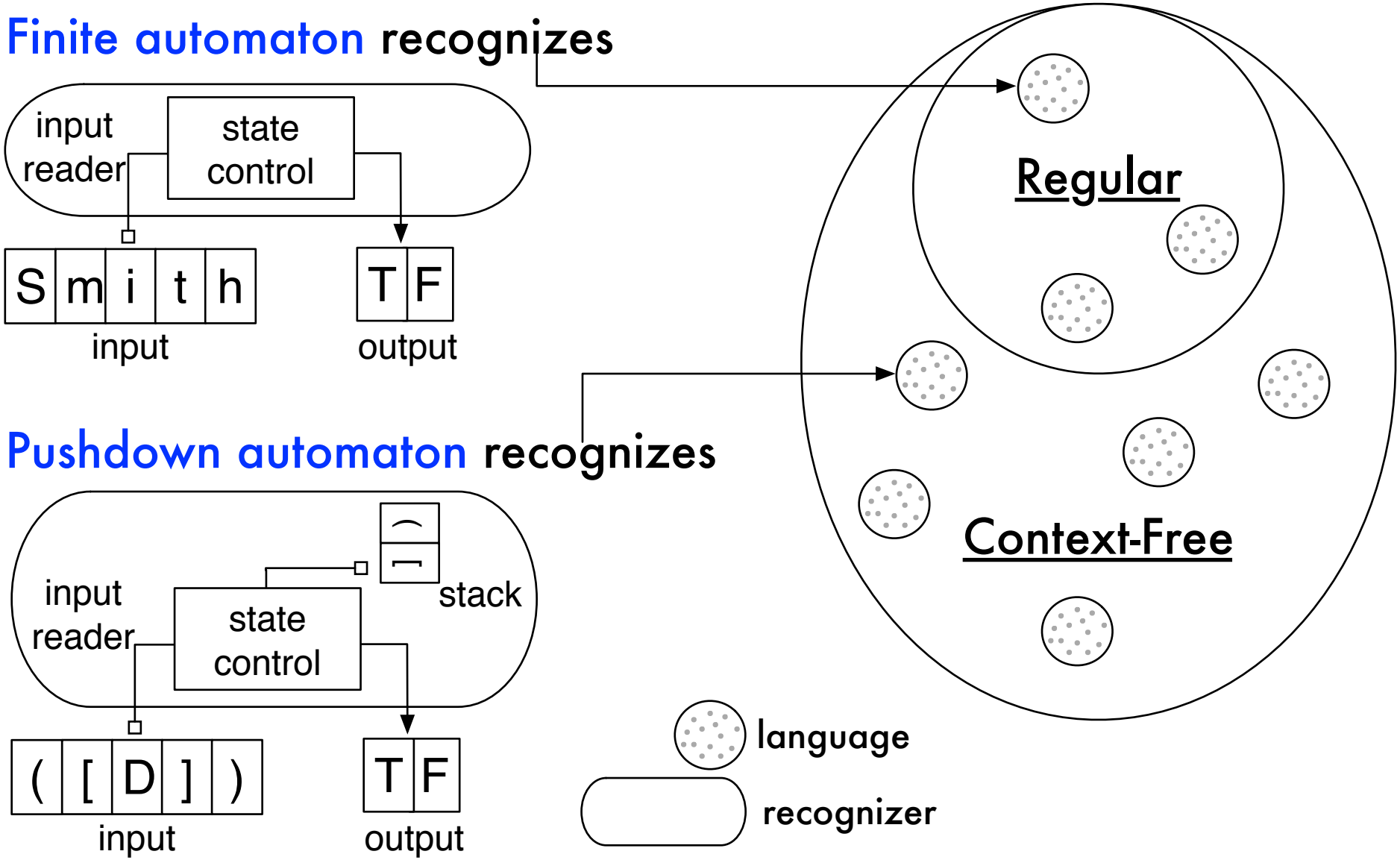
How do we determine whether a language contains a given string?



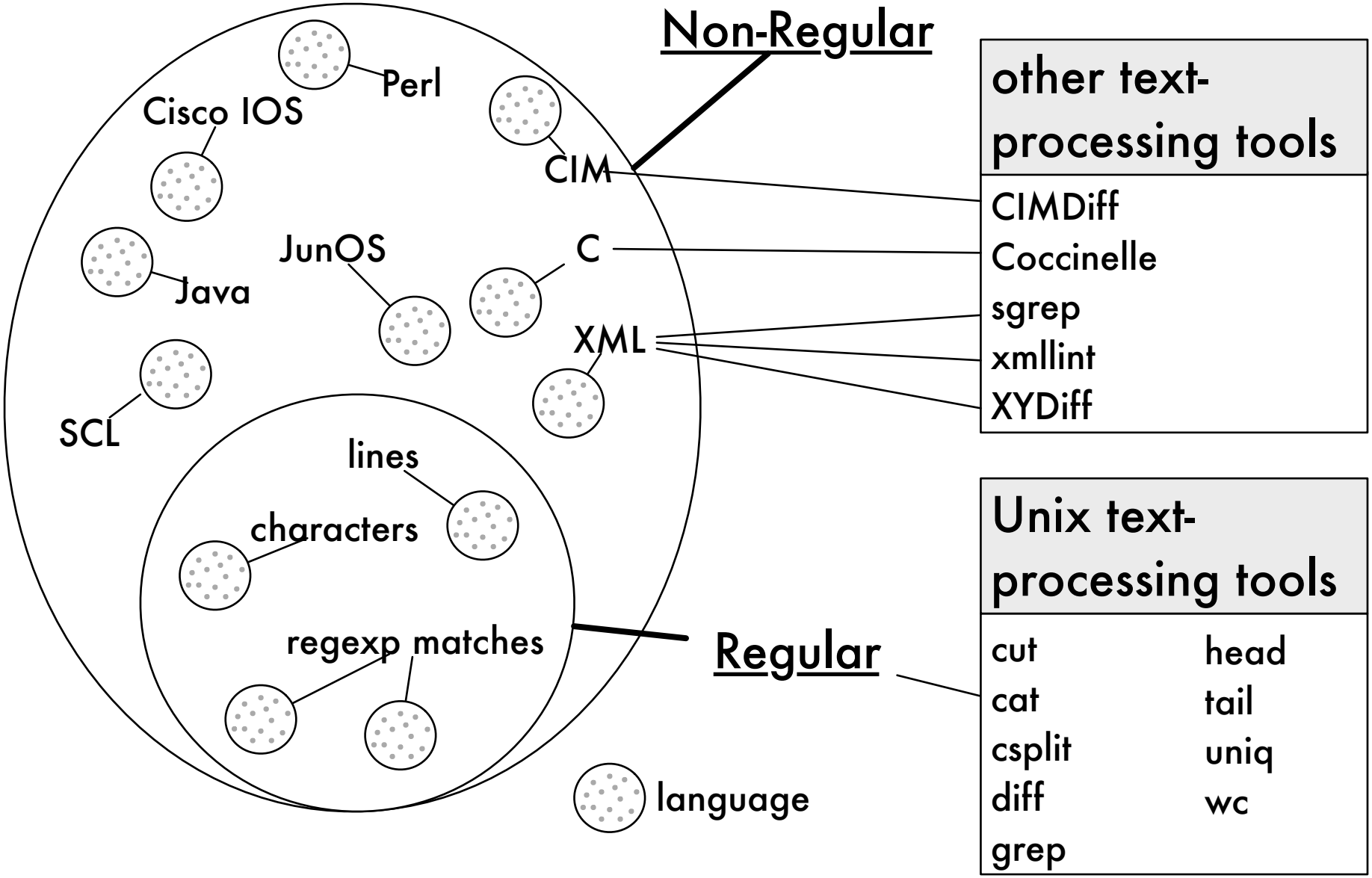
A recognizer for a language is computational machine that outputs TRUE if an input string is in the language.

Language Theory and The Tools Gap Problem

Language theory categorizes languages into different classes based upon recognizer complexity.

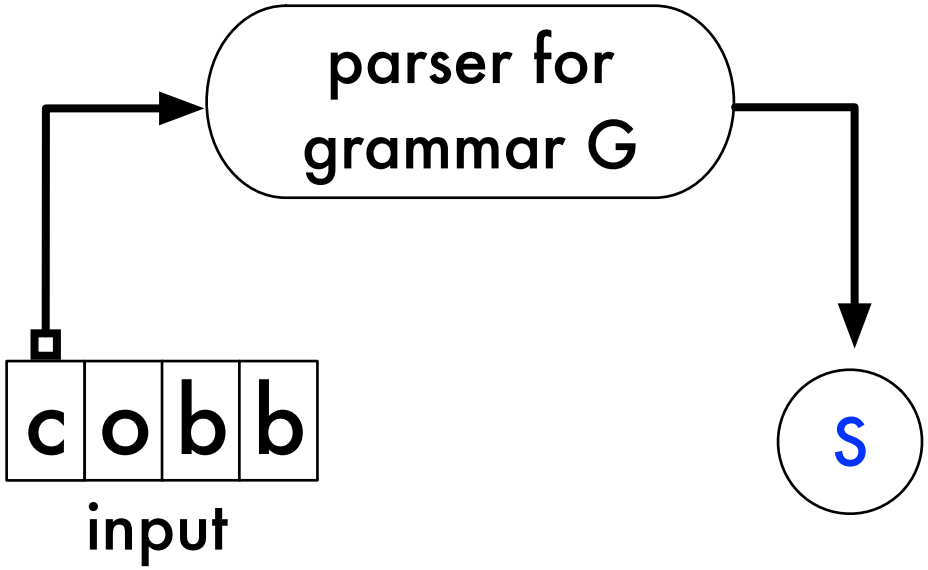


Language theory gives us a framework to understand the Tools Gap Problem.



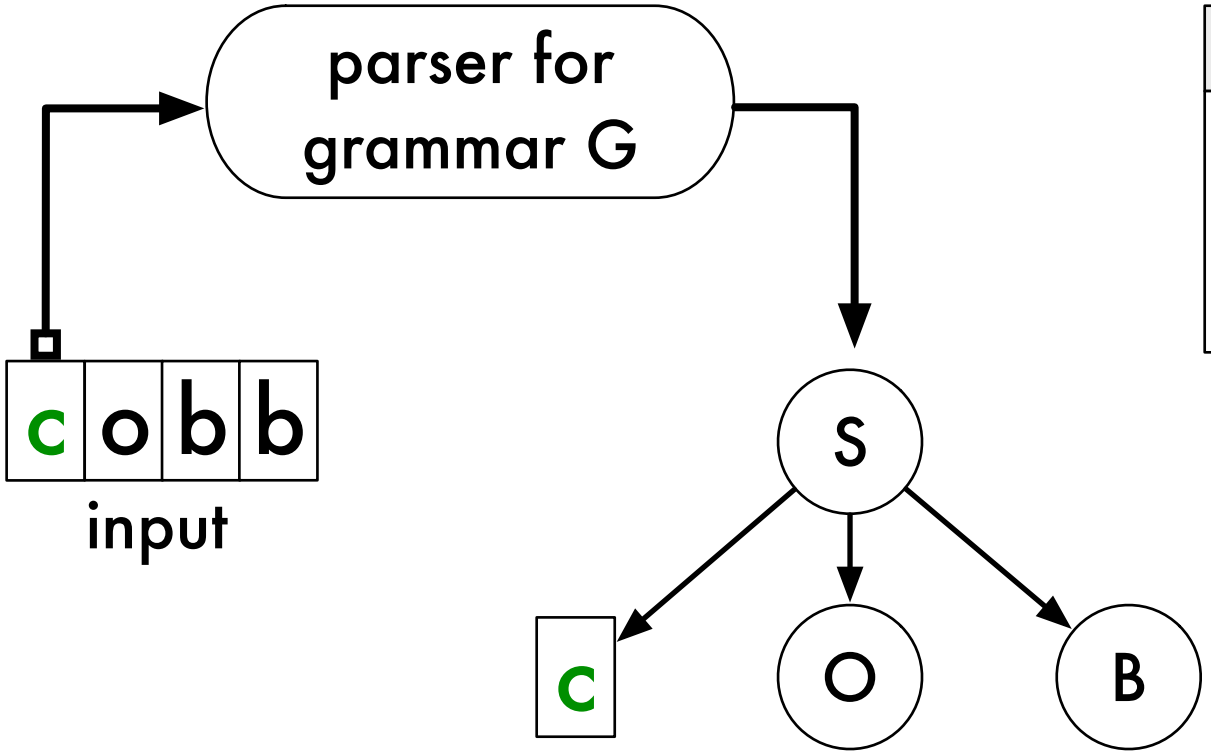
Parsing and the Granularity of Reference Problem

What is parsing?



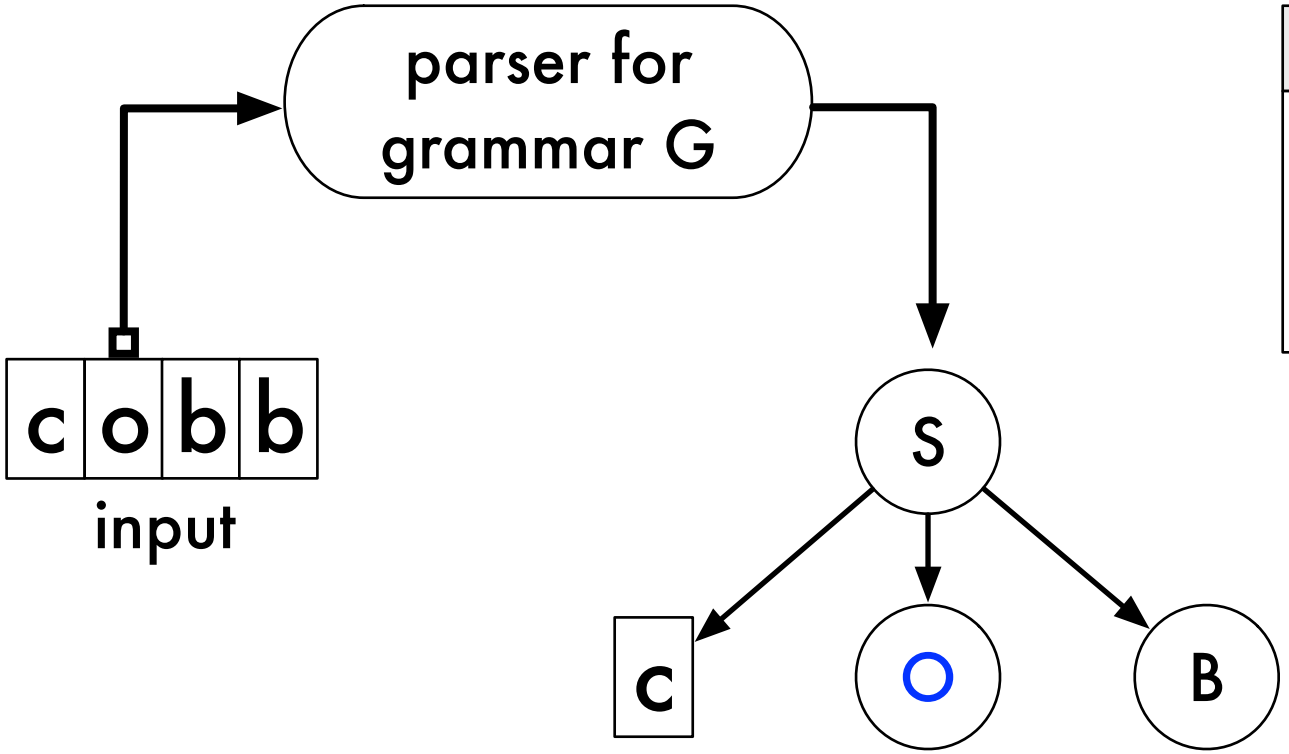
grammar G		
S	→	cOB
O	→	o | oO
B	→	bB | b

What is parsing?



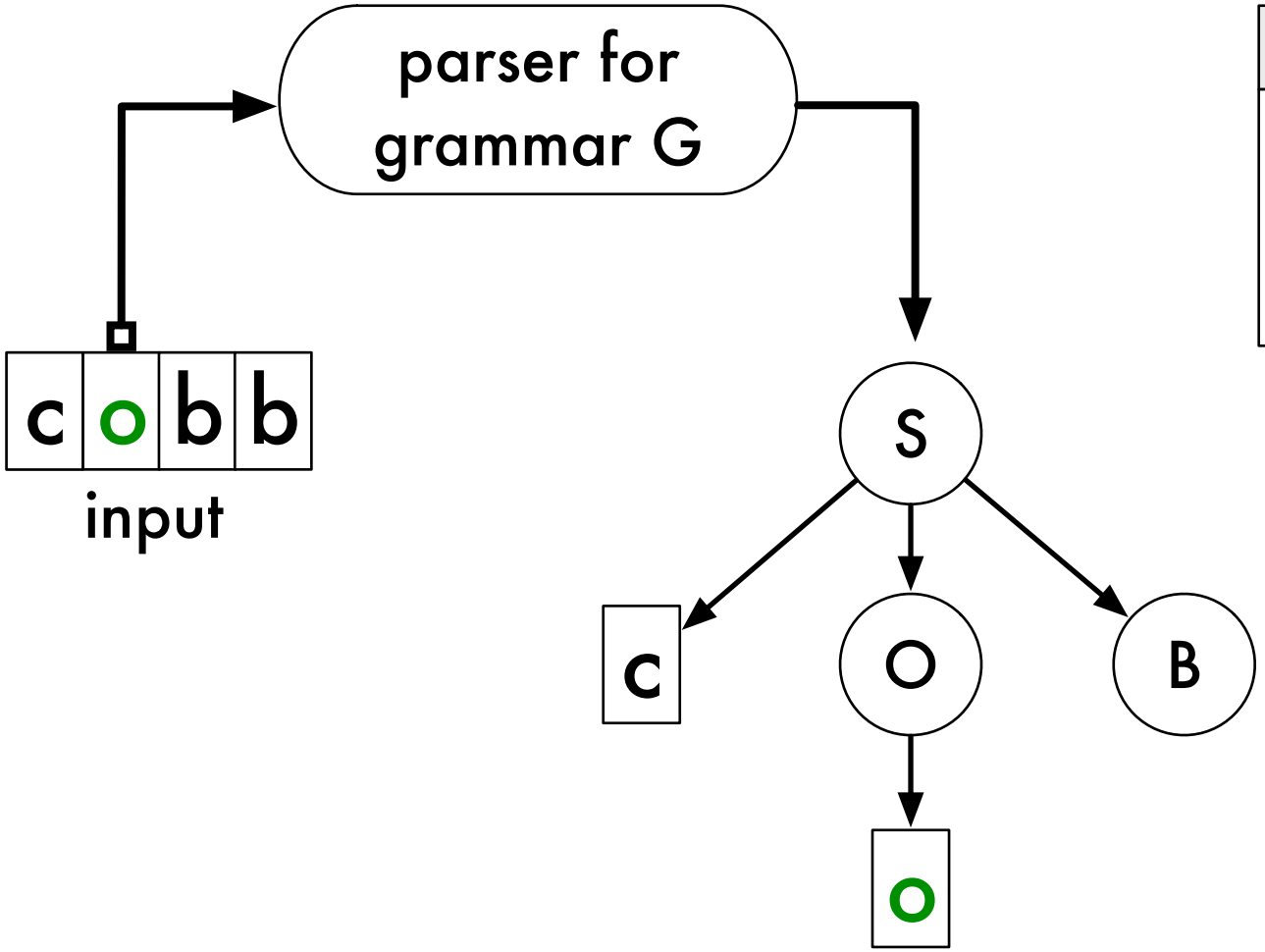
grammar G		
S	→	cOB
O	→	o | oO
B	→	bB | b

What is parsing?



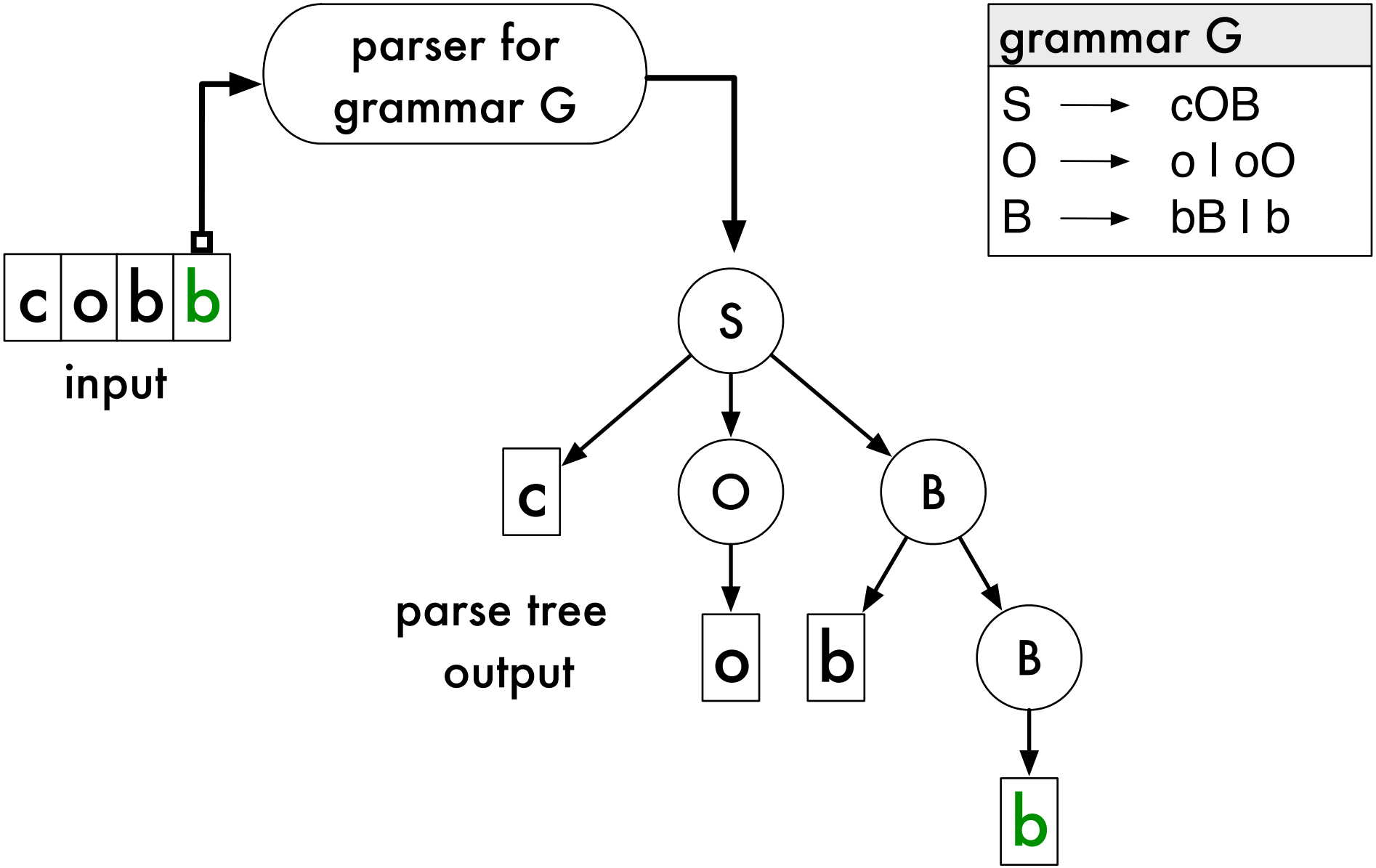
grammar G		
S	→	cOB
O	→	o | oO
B	→	bB | b

What is parsing?



grammar G		
S	→	cOB
O	→	o | oO
B	→	bB | b

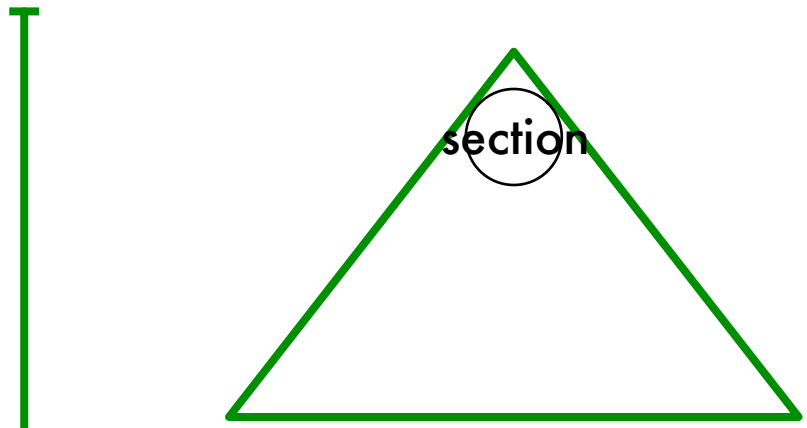
What is parsing?



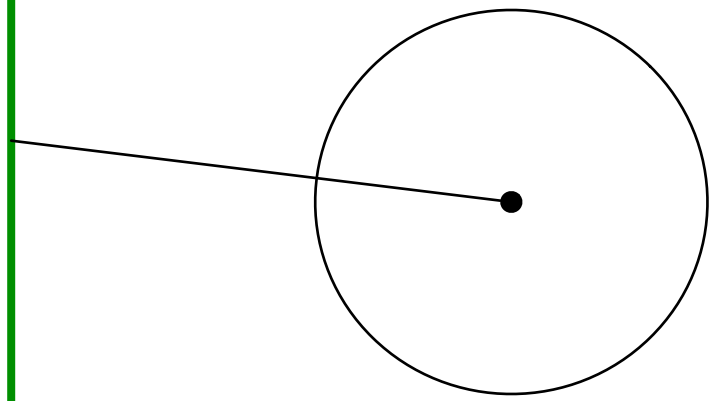
Parse trees give us a formalism for the Granularity of Reference Problem

Analysts' policy language (RFC 3647)

6 TECHNICAL SECURITY CONTROLS
The requirements for technical security measures of a CA or RA are determined by the types of services offered. The precise level of security...
6.1 KEY PAIR GENERATION AND INSTALLATION
6.1.1 KEY PAIR GENERATION
Key pairs for the Grid-CA are generated on a dedicated IT system unequipped with networking capability or directly within a Hardware Security Module (HSM).
6.1.1.1 HSM REQUIREMENTS
The keys are stored only on external data storage media and ...
6.1.2 PRIVATE KEY DELIVERY TO SUBSCRIBER
No cryptographic key pairs are generated for subscribers



Parse Tree under TEI-XML Grammar

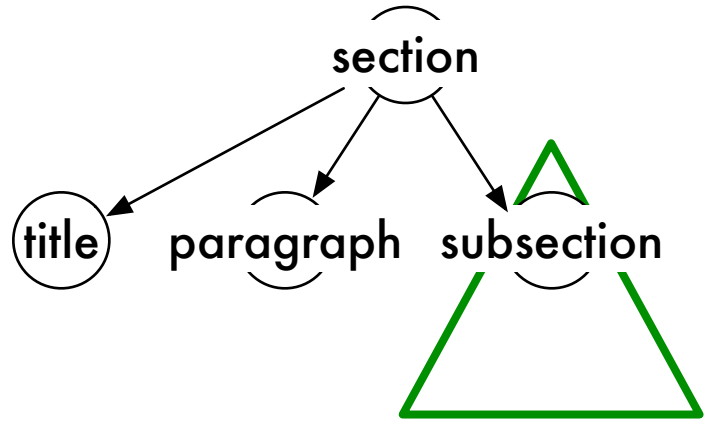


Language of Sections

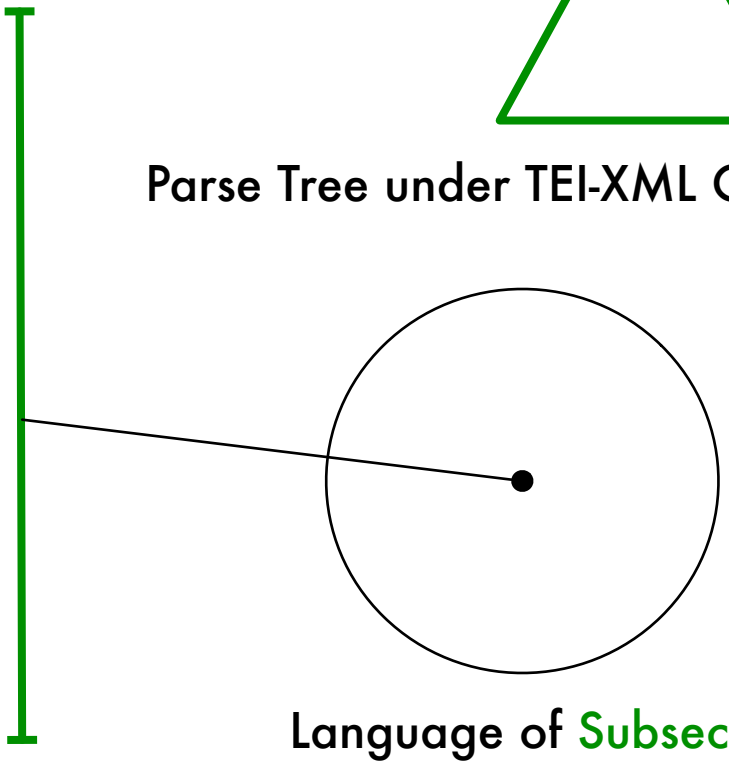
Parse trees give us a formalism for the Granularity of Reference Problem

Analysts' policy language (RFC 3647)

6 TECHNICAL SECURITY CONTROLS
The requirements for technical security measures of a CA or RA are determined by the types of services offered. The precise level of security...
6.1 KEY PAIR GENERATION AND INSTALLATION
6.1.1 KEY PAIR GENERATION
Key pairs for the Grid-CA are generated on a dedicated IT system unequipped with networking capability or directly within a Hardware Security Module (HSM).
6.1.1.1 HSM REQUIREMENTS
The keys are stored only on external data storage media and ...
6.1.2 PRIVATE KEY DELIVERY TO SUBSCRIBER
No cryptographic key pairs are generated for subscribers



Parse Tree under TEI-XML Grammar

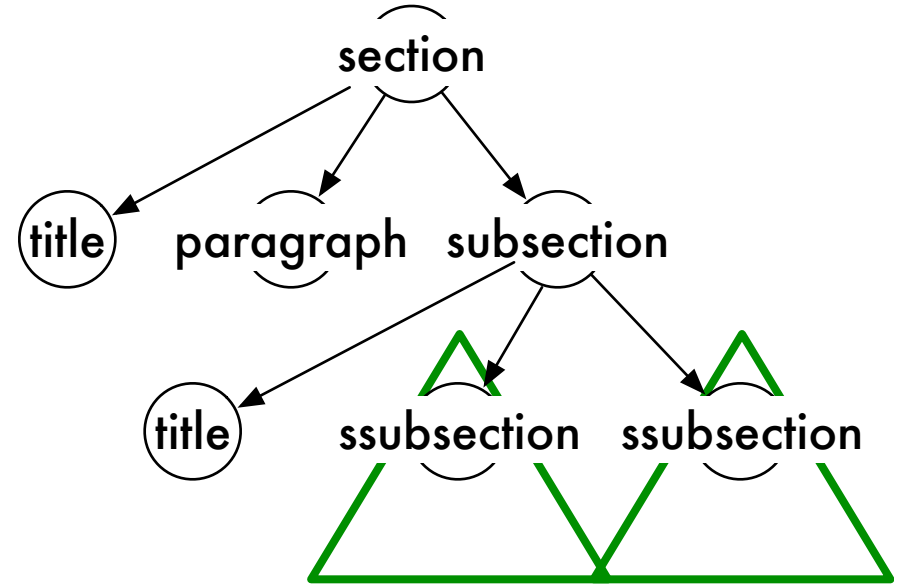


Language of Subsections

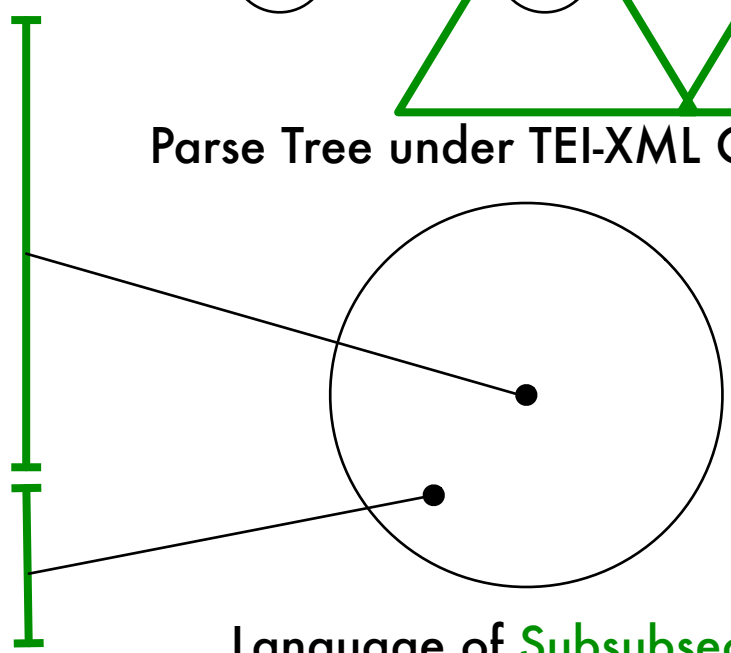
Parse trees give us a formalism for the Granularity of Reference Problem

Analysts' policy language (RFC 3647)

6 TECHNICAL SECURITY CONTROLS
The requirements for technical security measures of a CA or RA are determined by the types of services offered. The precise level of security...
6.1 KEY PAIR GENERATION AND INSTALLATION
6.1.1 KEY PAIR GENERATION
Key pairs for the Grid-CA are generated on a dedicated IT system unequipped with networking capability or directly within a Hardware Security Module (HSM).
6.1.1.1 HSM REQUIREMENTS
The keys are stored only on external data storage media and ...
6.1.2 PRIVATE KEY DELIVERY TO SUBSCRIBER
No cryptographic key pairs are generated for subscribers



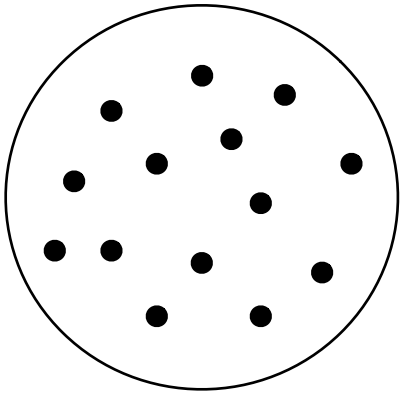
Parse Tree under TEI-XML Grammar



Language of Subsubsections

Discrete Mathematics and the Policy Discovery Needs Problem

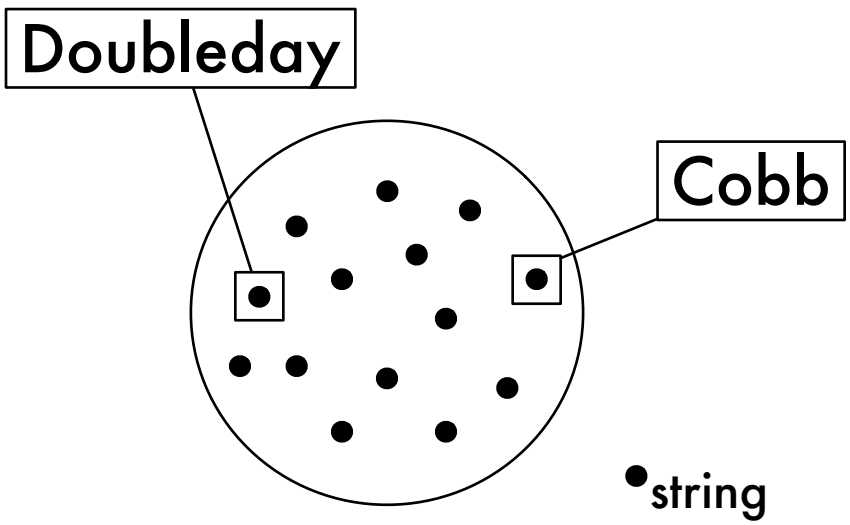
What is a datatype?



operations
distance equality ...

A datatype is a set paired with operations on elements in that set.

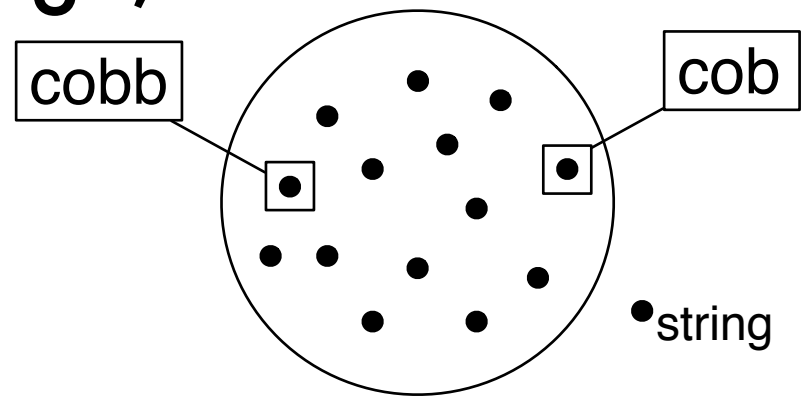
Datatypes give us a formalism for the Policy Discovery Needs Problem



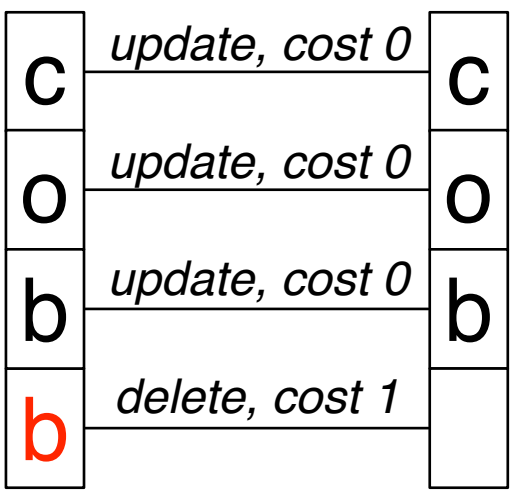
operations
distance equality parsing

1. We view a corpus, a collection of texts, as a datatype.
2. A corpus datatype consists of a language and operations upon that language.

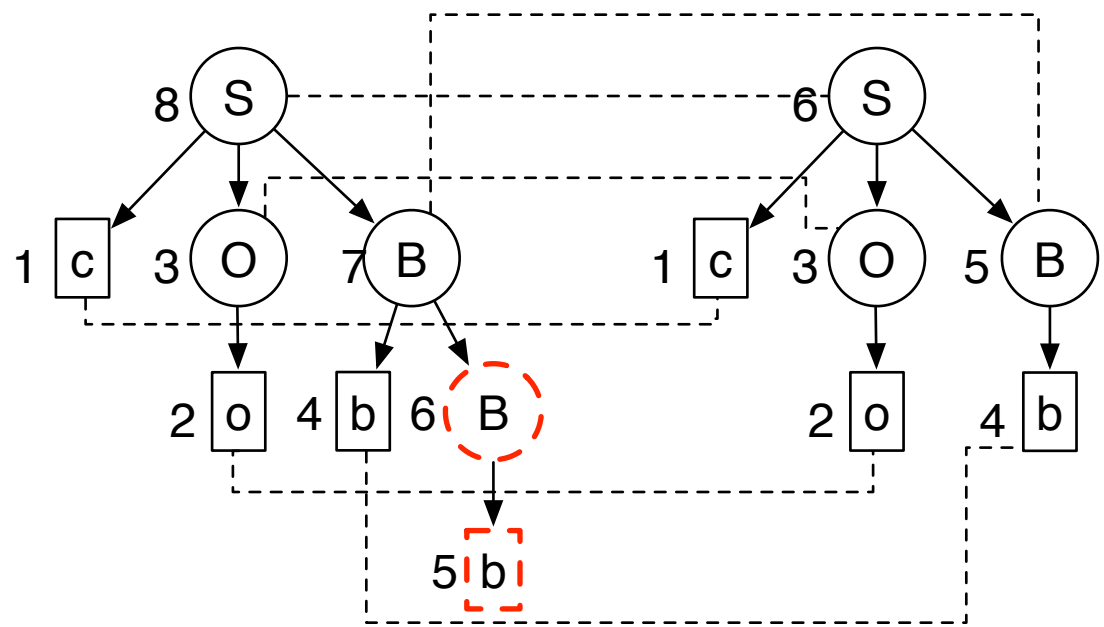
For a context-free language, we use two notions of distance.



string_edit_distance(cobb, cob) = 1



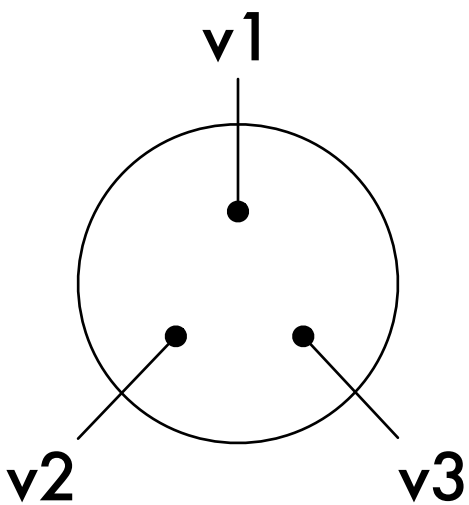
tree_edit_distance(T_{cob} , T_{cobb}) = 2



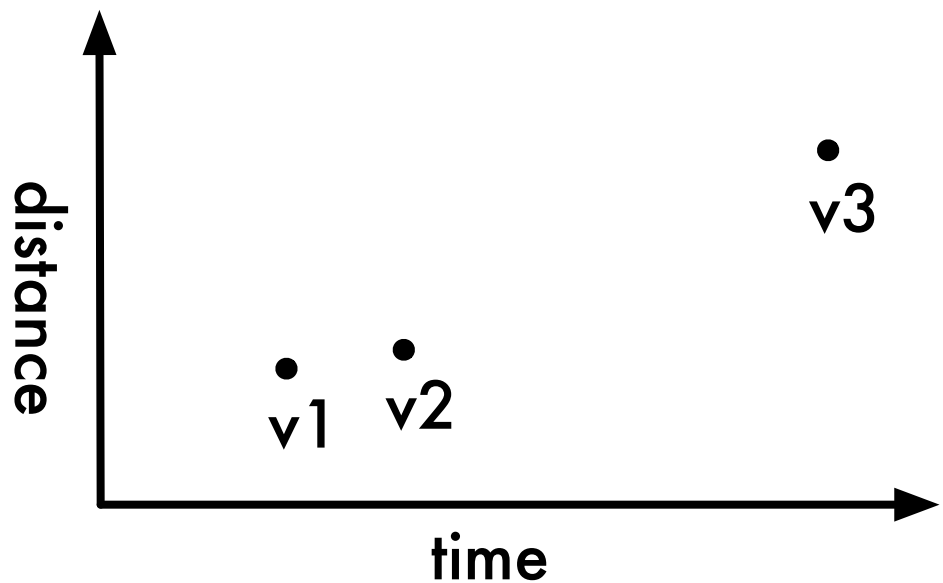
unmapped nodes 5 and 6 are deleted

Distance metrics let us measure trends in how high-level language constructs evolve.

<u>CERN Certificate Policies</u>



<u>Evolution of CERN Certificate Policies</u>



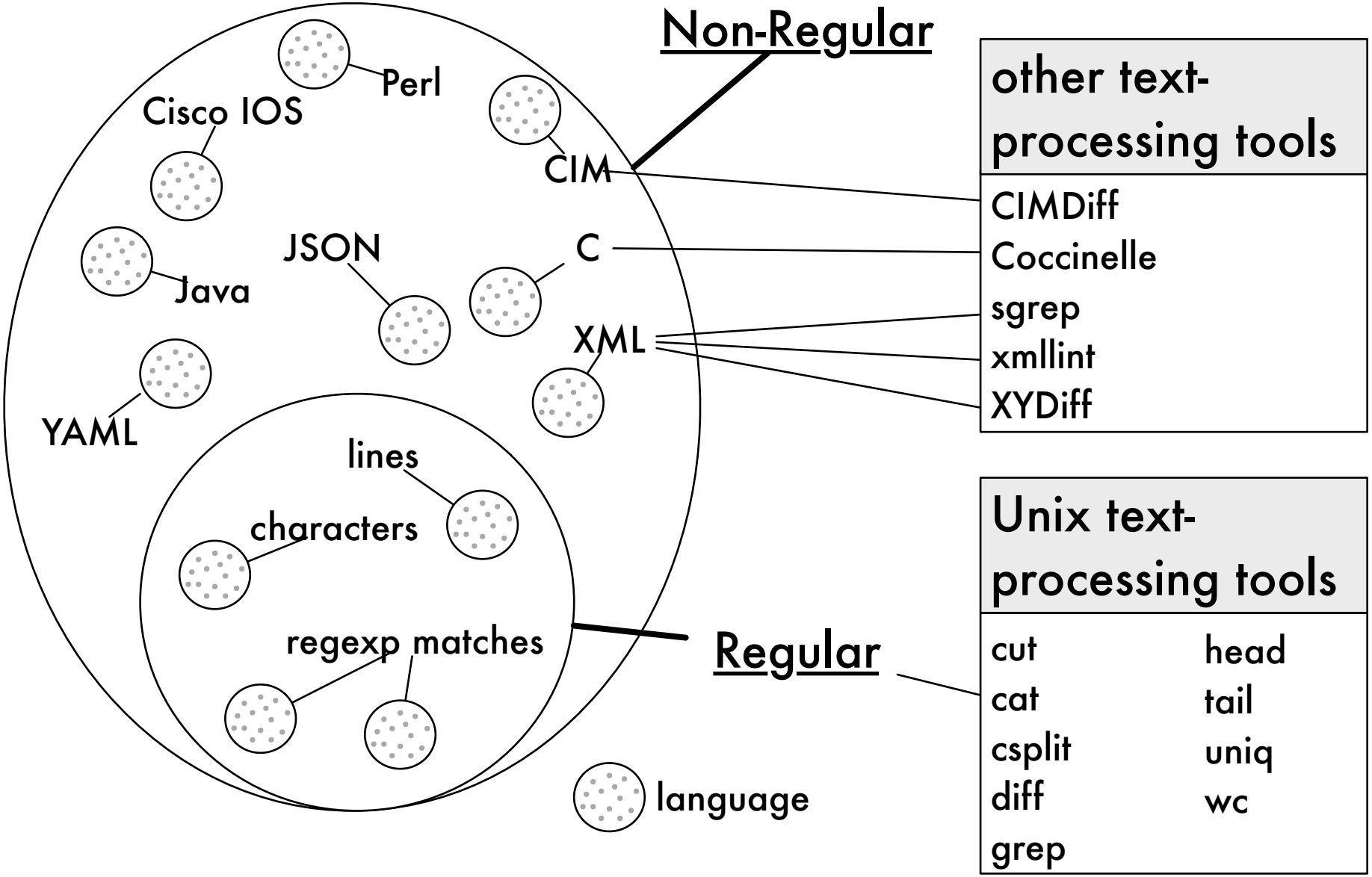
Outline

1. Motivation
2. Theoretical Toolbox
3. XUTools Capabilities
   i. Baseline Configuration Development
   ii. Change Control
4. Ongoing Research
5. Conclusions

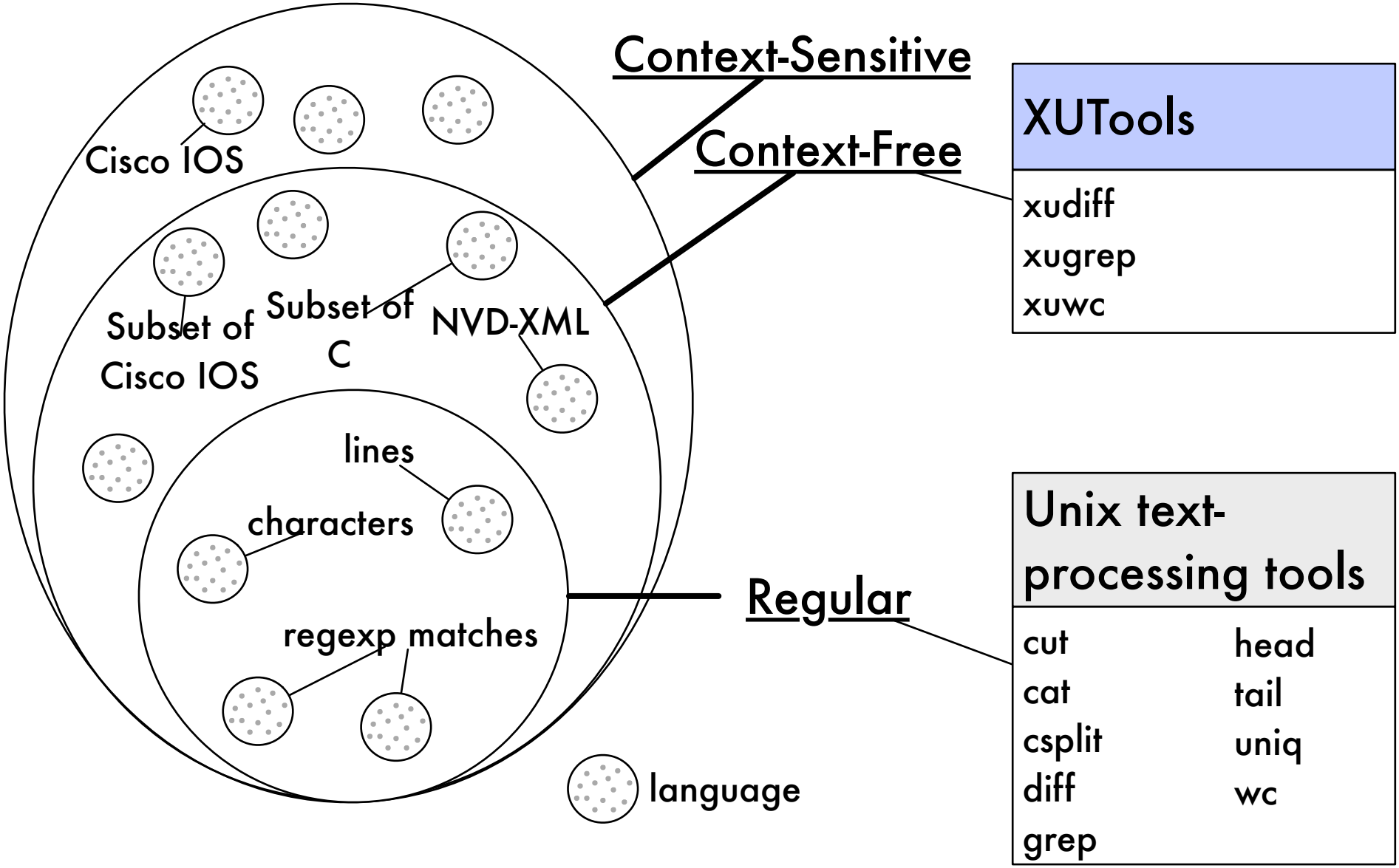
XUTools Capabilities

1. Our extended Unix tools apply the theoretical tools to address core limitations of textual analysis.
2. As a result power-control network security audit becomes more consistent and efficient.

Traditional Unix tools operate on regular languages that don't recognize arbitrary hierarchical structure.



We built extended Unix tools (XUTools) to operate on languages with arbitrary hierarchical structure.



Capabilities for Baseline Configuration

i. Inventory security primitives
ii. Identify important security primitives

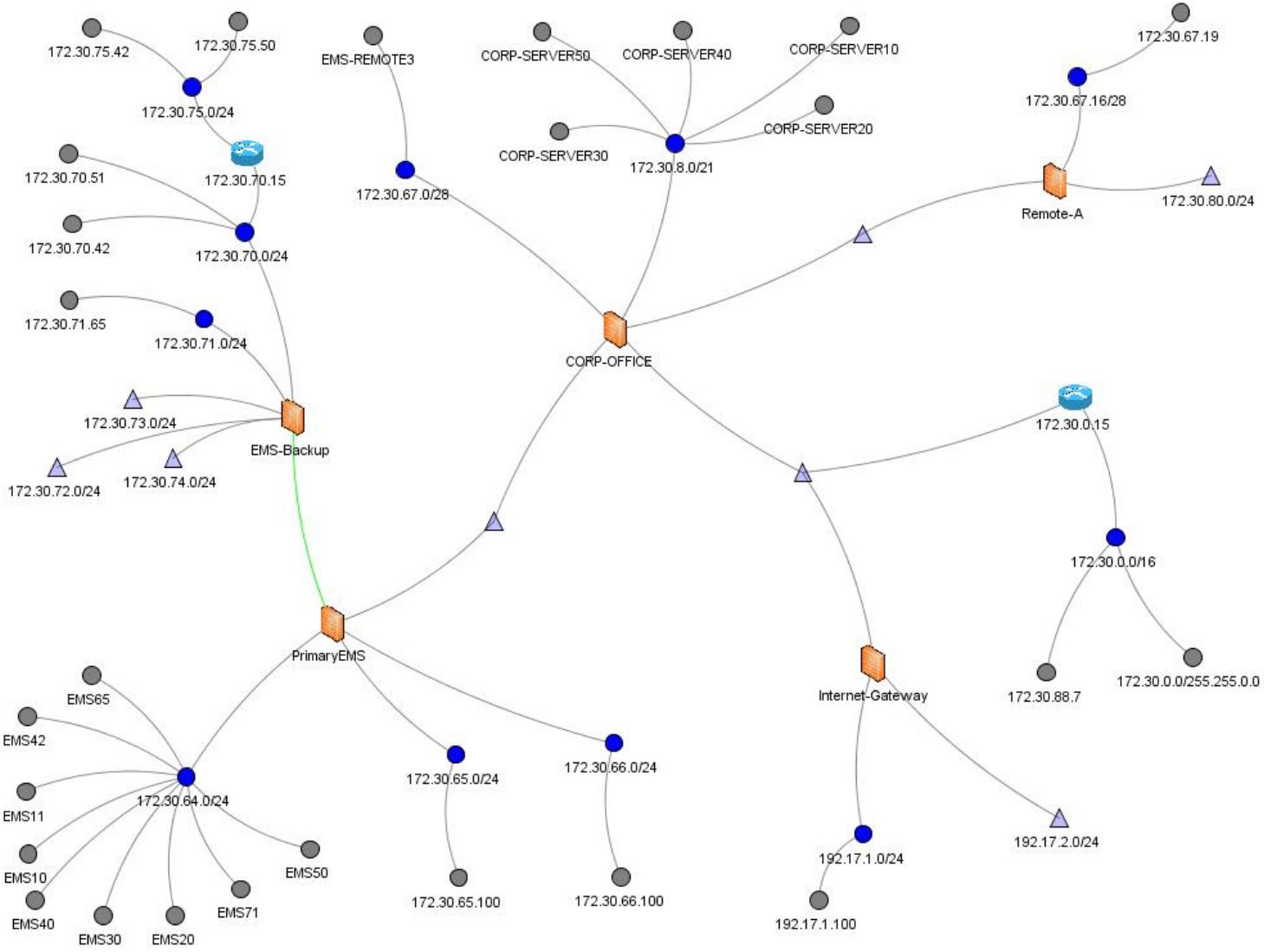
Baseline Configuration Use Cases

Network Configuration	Catalog the roles—groupings of users, devices, and protocols—defined on access-control policy. More roles make firewalls harder to manage, more prone to misconfiguration [Benson 2009].
Windows Machines	Use the Windows Registry as a representation of a baseline configuration [IEEE PECI 2012]. 80% of the machines (not including embedded systems) on the power grid are Windows Machines [Edmond Rogers 2012].

Baseline Configuration – Network Devices

Examples	1. Catalog the roles—groupings of users, devices, and protocols—defined on access-control policy. 2. Identify the most *important* interfaces or roles on a network device.
Current Approach	There are not many tools available to help practitioners in the realm of baseline configuration. Many utilities use spreadsheets to manually document baseline configurations of systems (scalability issues, can't keep up) Manual documentation is error-prone, inconsistent, and does not scale.

Demonstration

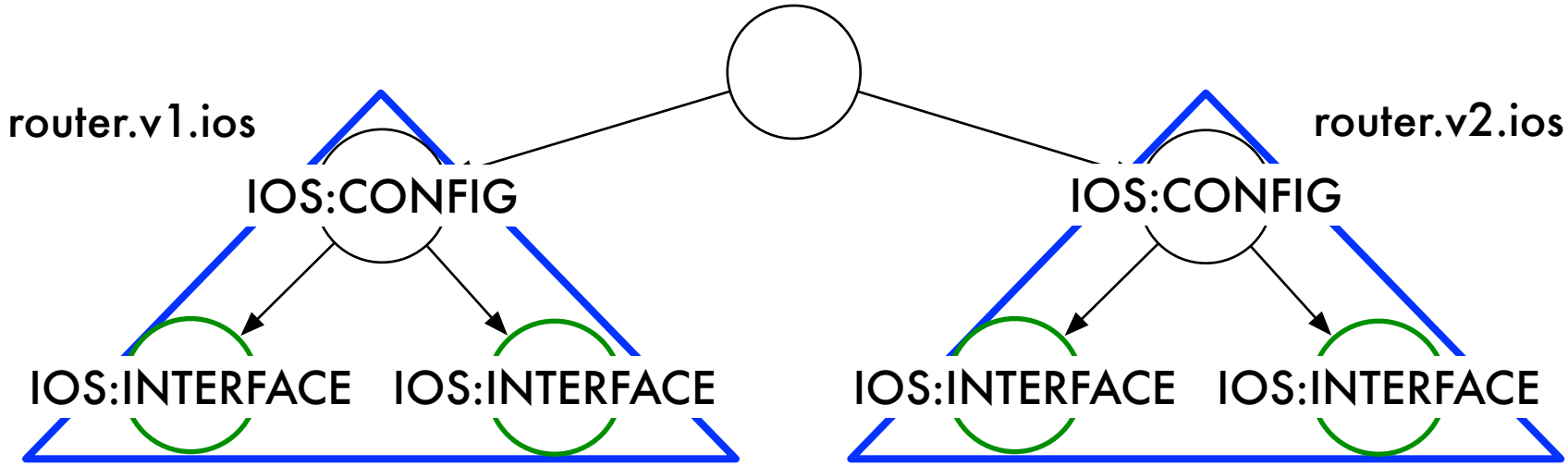


xuwc design

xuwc usage

```
xuwc [--count=<language_name>]
     [--context=<language_name>]
     <xupath> <input_file>+
```

```
xuwc //ios:interface router.v1.ios router.v2.ios
```



By default, xuwc counts the number of language construct occurrences per file.

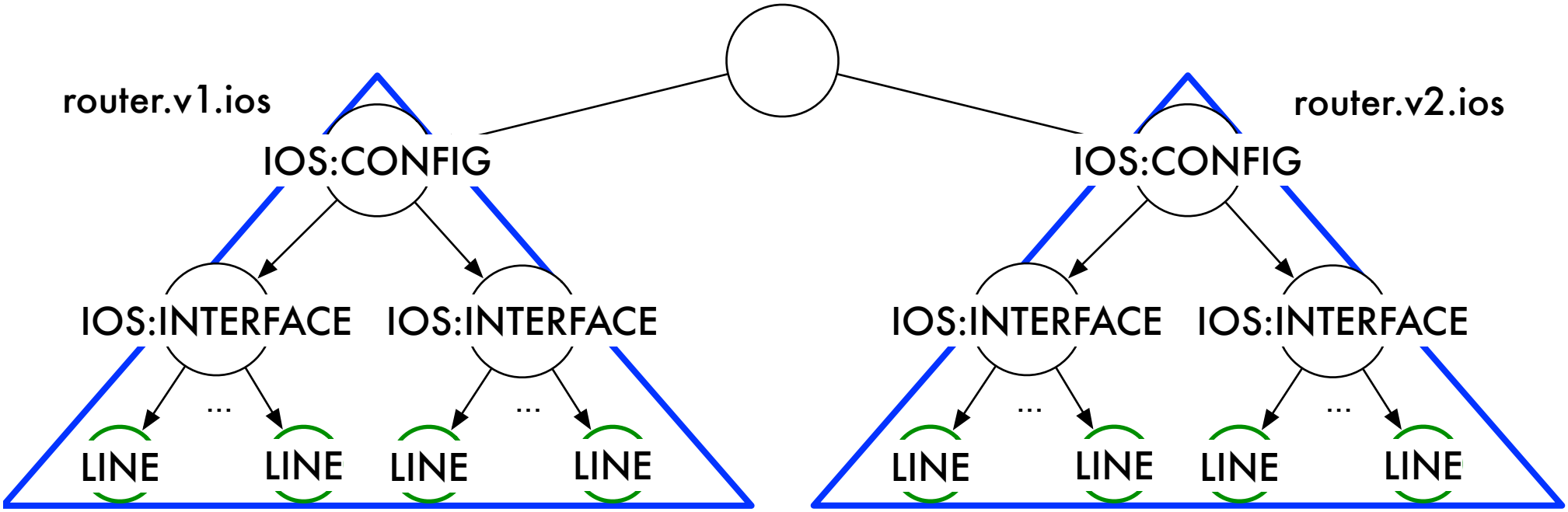
2 interfaces in router.v1.ios
2 interfaces in router.v2.ios

xuwc design

xuwc usage

```
xuwc [--count=<language_name>]
     [--context=<language_name>]
     <xupath> <input_file>+
```

```
xuwc //ios:interface/builtin:line router.v1.ios router.v2.ios
```



By default, xuwc counts the number of language construct occurrences per file.

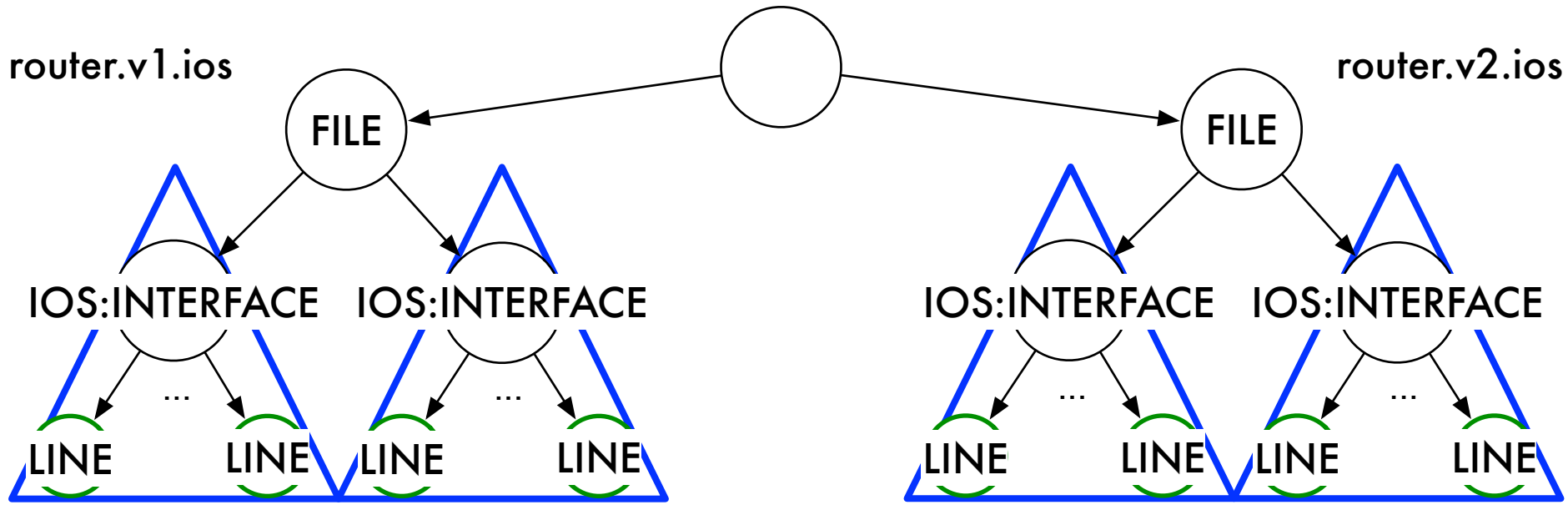
16 lines in router.v1.ios
16 lines in router.v2.ios

xuwc design

xuwc usage

```
xuwc [--count=<language_name>]
     [--context=<language_name>]
     <xupath> <input_file>+
```

```
xuwc --context=ios:interface //ios:interface/builtin:line
  router.v1.ios router.v2.ios
```



The context flag tells xuwc to count number of language construct occurrences per container language construct.

7 lines in <router.v1, Loopback0>
9 lines in <router.v1, GigabitEth>
7 lines in <router.v2, Loopback0>
9 lines in <router.v2, GigabitEth>

Capabilities for Change Control

i. Changelog generation
ii. Measure trends

Change Control Use Cases

Network Configuration	Compare network configuration files in terms of high-level language constructs rather than lines.
Windows Machines	Compare a baseline Windows registry configuration against a production machine's registry [IEEE PECI 2012]. 80% of the machines (not including embedded systems) on the power grid are Windows Machines [Edmond Rogers 2012].

Change Control – Network Devices

Examples	1. Summarize how network configuration devices have changed over time relative to hierarchical network configuration languages. 2. View trends in how the network has changed over time at a variety of levels of granularity.
Current Approach	RANCID gives practitioners a line-level view of changes between consecutive versions of configuration files. ChangeGear and Remedy are ticketing systems that rely upon manual documentation that may be error prone or become outdated. Splunk and search-based approaches are not context-free and rely upon regular expressions to identify key/value pairs.

Demonstration: Change logs at different levels of abstraction via XUDiff.

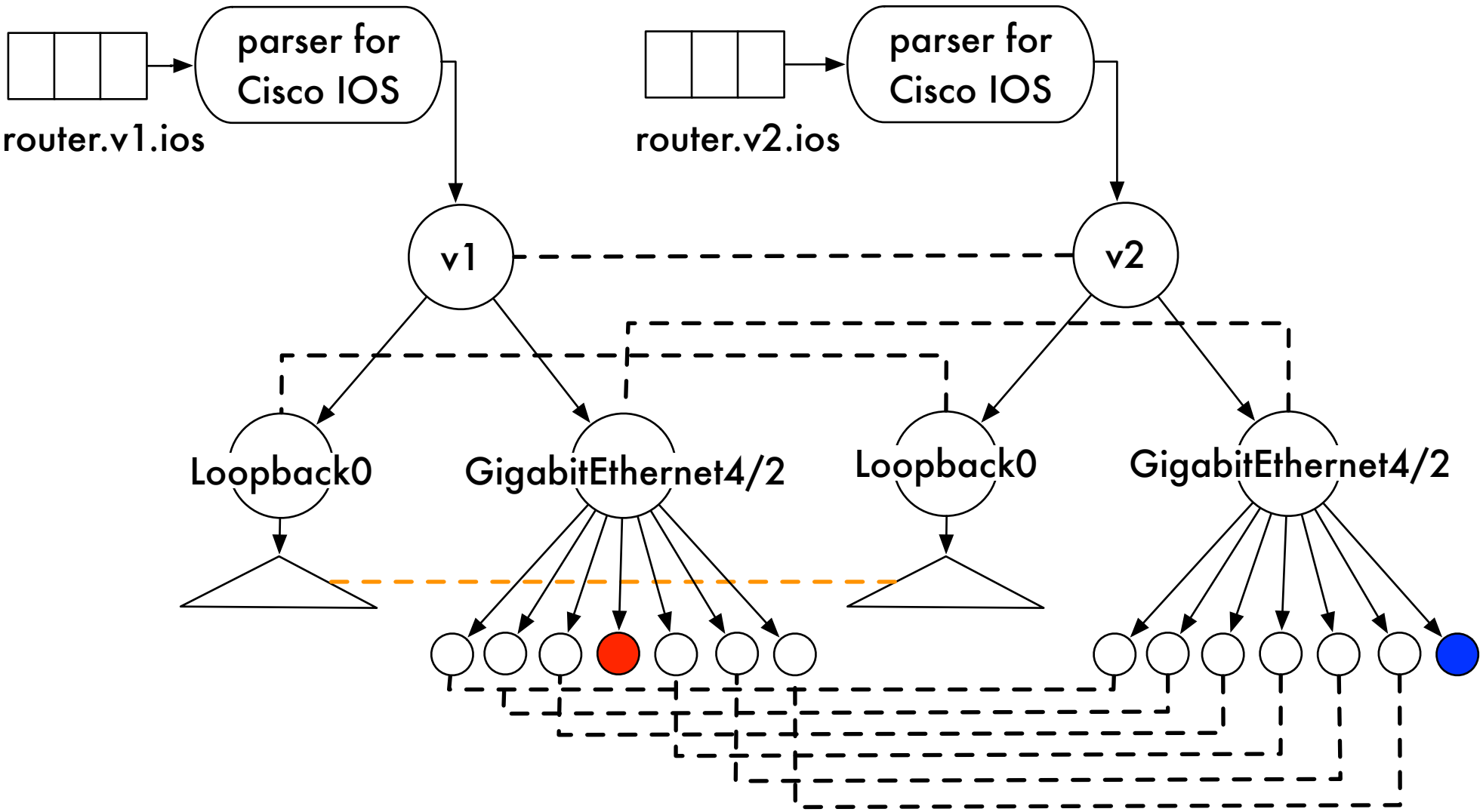
router.v1.example	router.v2.example
interface Loopback0 description really cool description ip address 333.444.1.185 255.255.255.255 no ip unreachables ip pim sparse-dense-mode crypto map azalea ! interface GigabitEthernet4/2 description Core Network ip address 444.555.2.543 255.255.255.240 ip access-group outbound_filter in ip access-group inbound_filter out no ip redirects no ip unreachables no ip proxy-arp !	interface Loopback0 description really cool description ip address 333.444.1.581 255.255.255.255 no ip unreachables ip pim sparse-dense-mode crypto map daffodil ! interface GigabitEthernet4/2 description Core Network ip address 444.555.2.543 255.255.255.240 ip access-group outbound_filter in no ip redirects no ip unreachables no ip proxy-arp ip flow ingress !

xudiff design

xudiff usage

```
xudiff [--cost=<cost_function>]
       <xupath> <input_file1> <input_file2>
```

```
xudiff.py //ios:config router.v1.ios router.v2.ios
```

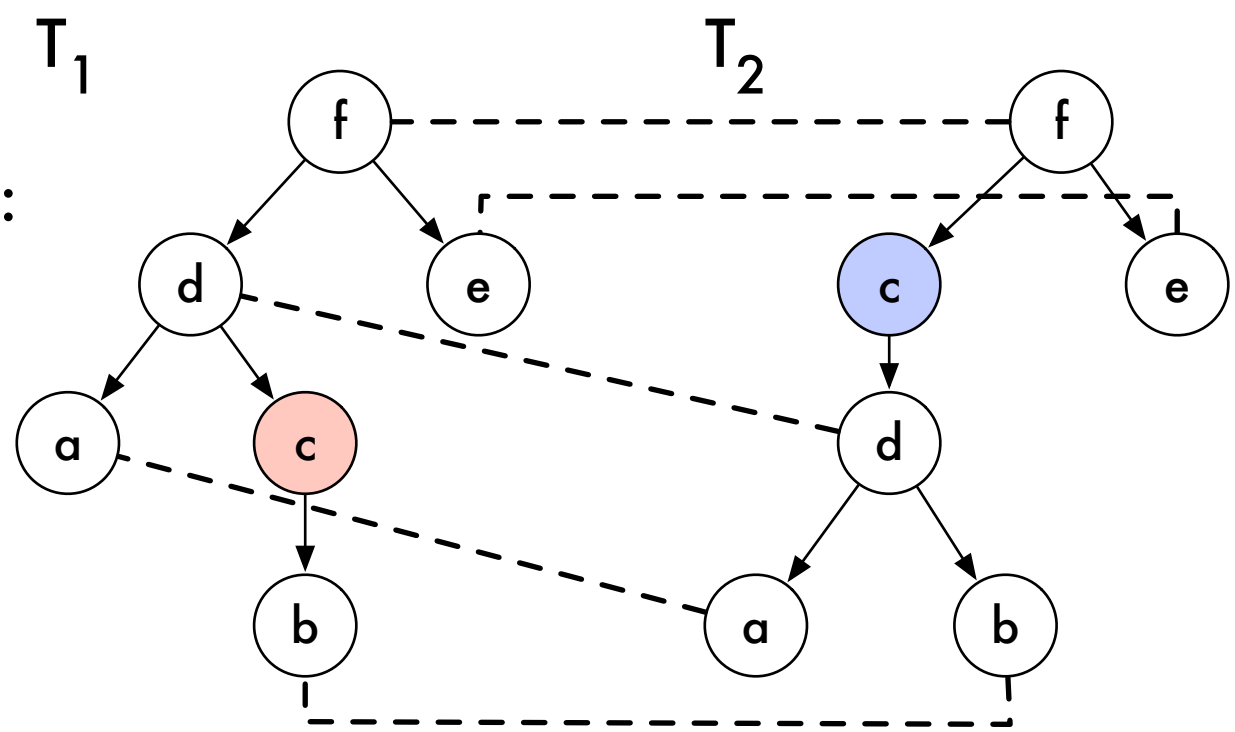


XUDiff uses the Zhang and Shasha Tree Edit Distance Algorithm.

Computes a Mapping (M) between nodes in tree 1 (T_1) and tree 2 (T_2).

M has three properties:

1. one-to-one
2. sibling order preserved
3. ancestor order preserved



T_1 and T_2 must be ordered, labeled trees: sibling order is significant and the nodes have labels (e.g. 'a')

[Zhang and Shasha, 1989]

Outline

1. Motivation
2. Theoretical Toolbox
3. XUTools Capabilities
   i. Baseline Configuration Development
   ii. Change Control
4. Ongoing Research
5. Conclusions

Project Activity in December 2012

12.5.12 Smart Grid Cybersecurity Info. Exchange
12.13.12 USENIX LISA 2012
12.17.12 Ph.D Thesis Defense
12.24.12 XUTools User Group
Juniper OS parser contributed (Egor)

Conclusions

Our XUTools Operate on Hierarchical Object Models.

Format/ Protocol	Ops.	Trans.	Cust.	Description	Hierarchical Data Model
CIM	X	X		Common Information Model	**Yes**
Cisco IOS	X	X		Network Configurations	**Yes**
C37.118	X	X		PMU Data	
DNP3	X	X	X		
ESPI-XML			X	Green Button Energy Usage	**Yes**
ICCP	X	X			
IEC 60870			X		
IEC 61850	X	X	X	GOOSE/SCL (IED)	**Yes**
Modbus	X	X	X		
Windows Registries	X	X		Windows Configuration	**Yes**

Many possible future directions for XUTools...

Format/ Protocol	Ops.	Trans.	Cust.	XUTools Application
CIM	X	X		Process CIM. Define 'equivalence classes' between CIM and 61850 structures. Use this to compute a common communications interface in substations.
Cisco IOS	X	X		Extend analyses for network configuration management.
C37.118	X	X		Parse C37.118 from different IEDs.
DNP3	X	X	X	Define equivalence classes between IEC 61850 and DNP3 structures. Use this to compute a common communications interface in substations.
IEC 61850	X	X	X	Compare IED Configuration Descriptions (ICD) as defined by IEC 61850.
Windows Registries	X	X		Develop and compare baseline configurations for Windows machines via the Windows Registry. Useful for NERC CIP 010-1 [PECI 2012]

Our current XUTools capabilities are useful for both auditor and administrator.

	capability	provision	vulnerability
baseline configuration	Inventory security primitives.	CIP 010: Baseline configuration development	NISTIR 7.2.18: Secure and validate field device settings.
	Identify important security primitives.		
change control	Changelog generation	CIP 004-5: Update network documentation within 30 days of a change.	NISTIR 6.2.2.5: Inadequate change and configuration management. NISTIR 6.2.3.1: Inadequate periodic security audits.
	Change trends	CIP 003-4: Change control and configuration management.	

Thank You!
Questions?

www.xutools.net

We have a mailing list!
gweave01@cs.dartmouth.edu

Practitioners may compute baseline network configuration relative to a language of security primitives.

How many security primitives are defined on a network device?

security primitive	Cisco IOS implementation
interface	interface
role	object group

xugrep design

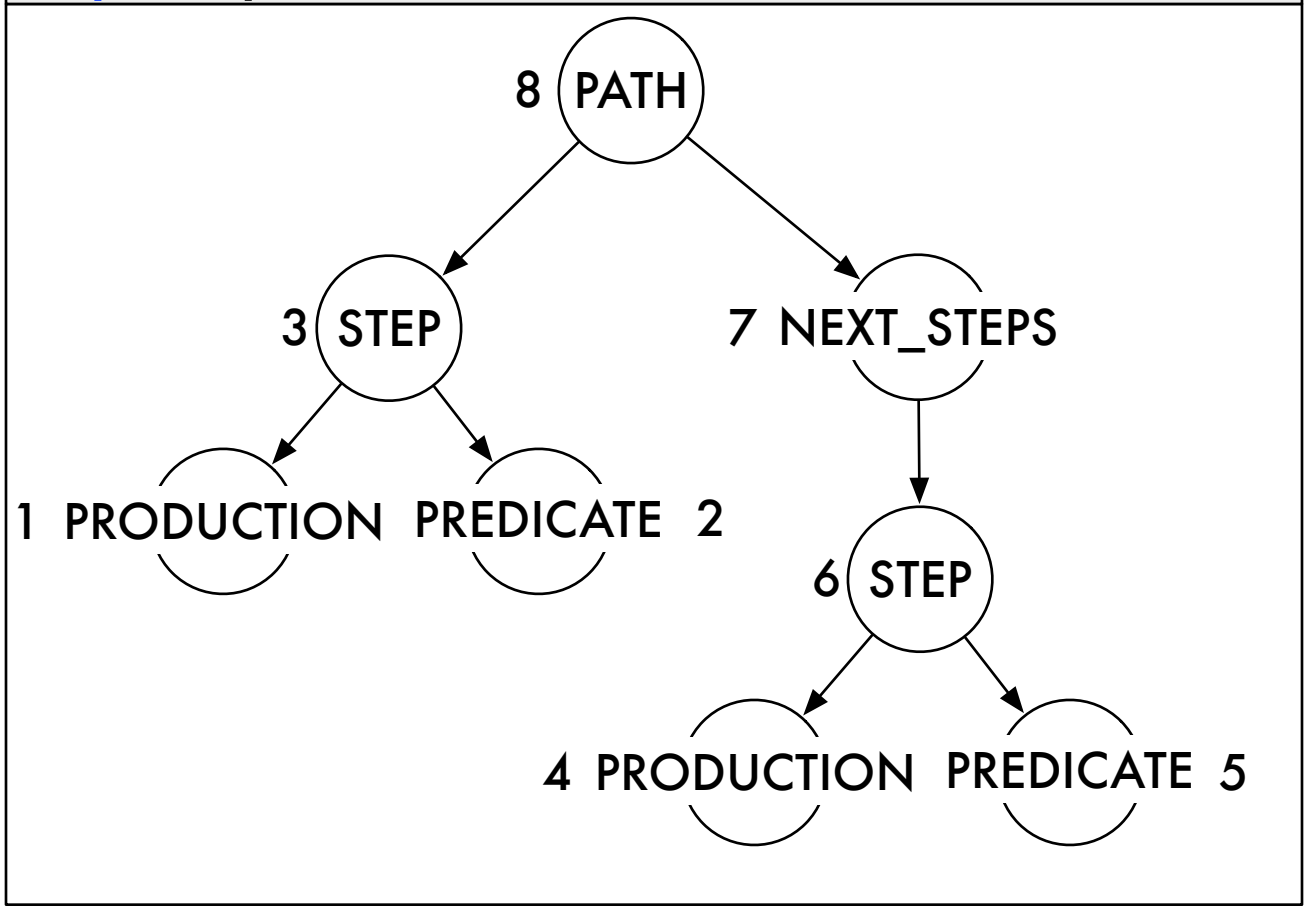
xugrep arguments

xugrep usage

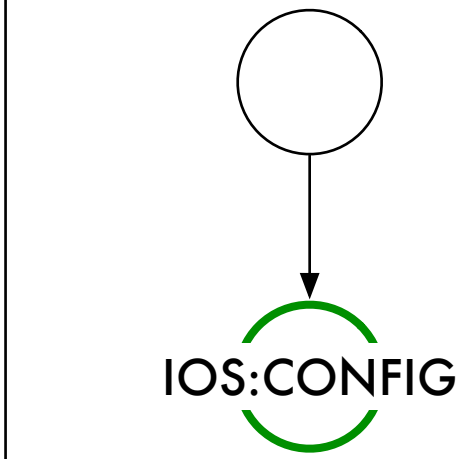
```
xugrep [-1] <xupath> <input_file>+
```

```
xugrep //ios:interface/builtin:line router.v1.ios
```

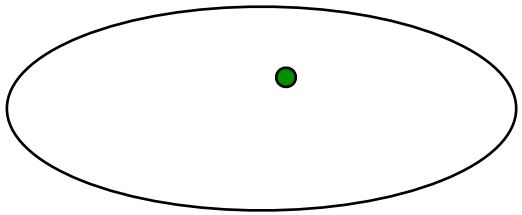
xupath parse tree



xupath query tree



current corpus



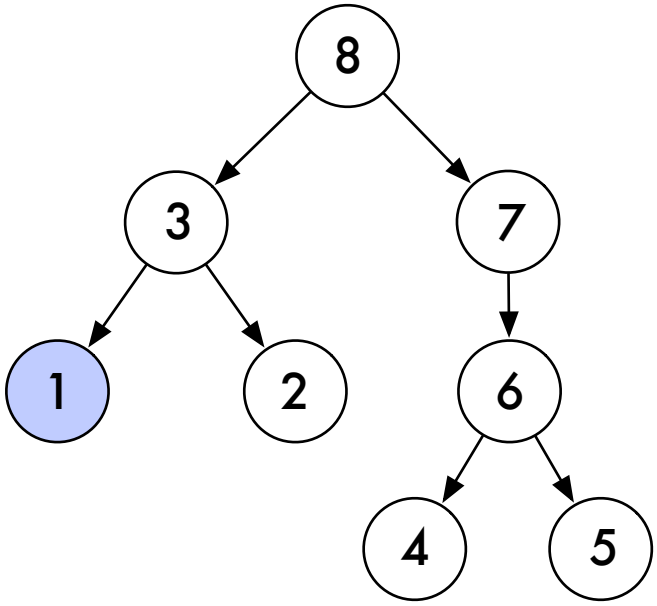
xugrep design

xugrep usage

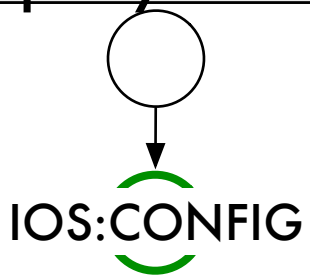
```
xugrep [-l] <xupath> <input_file>+
```

```
xugrep //ios:interface/builtin:line router.v1.ios
```

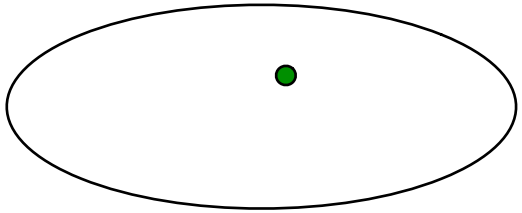
xupath parse tree



xupath query tree



current corpus



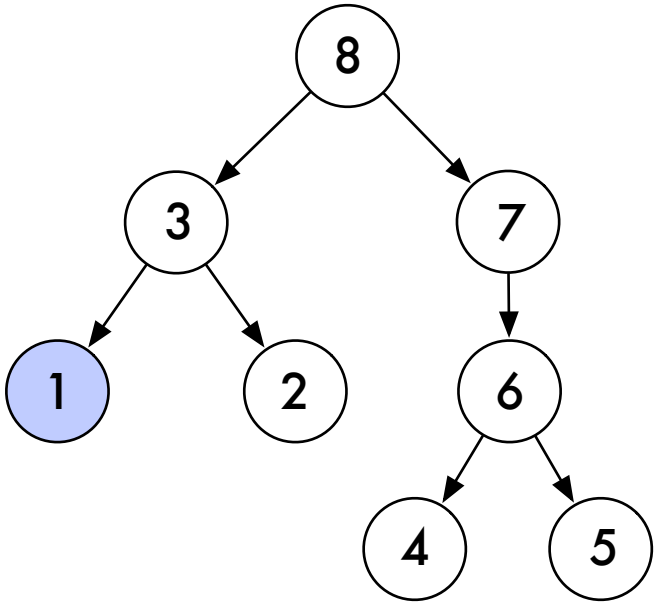
xugrep design

xugrep usage

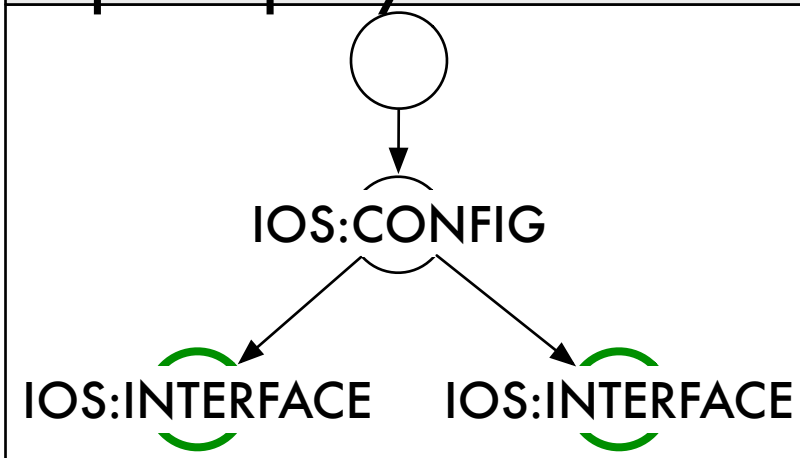
```
xugrep [-1] <xupath> <input_file>+
```

```
xugrep //ios:interface/builtin:line router.v1.ios
```

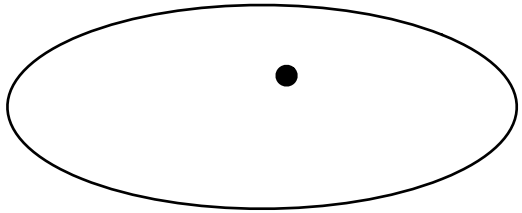
xupath parse tree



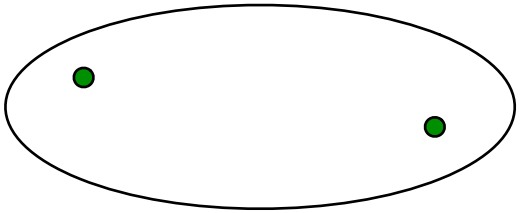
xupath query tree



old corpus



current corpus



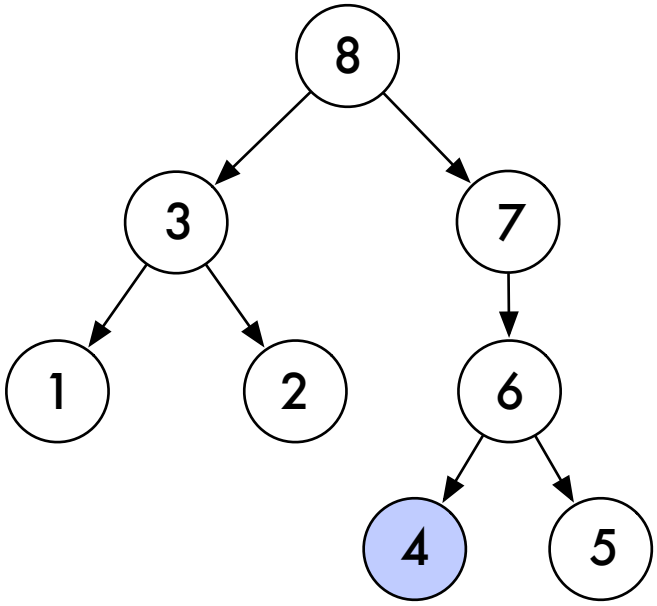
xugrep design

xugrep usage

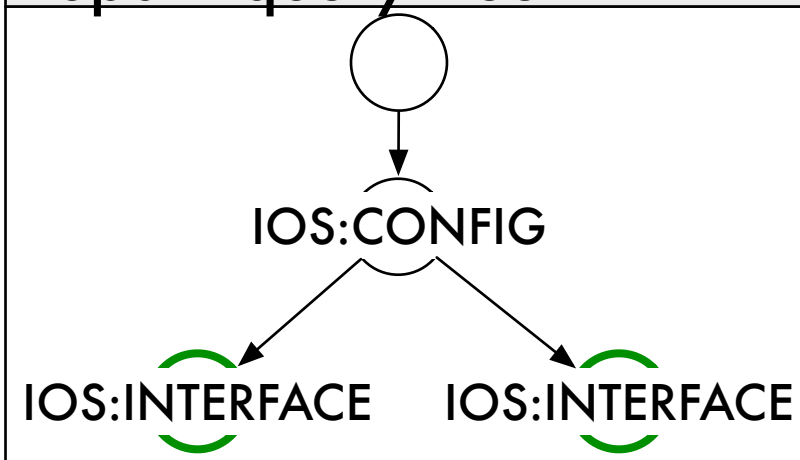
```
xugrep [-1] <xupath> <input_file>+
```

```
xugrep //ios:interface/builtin:line router.v1.ios
```

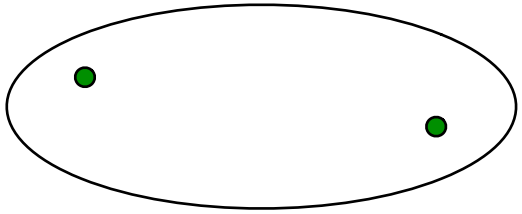
xupath parse tree



xupath query tree



current corpus



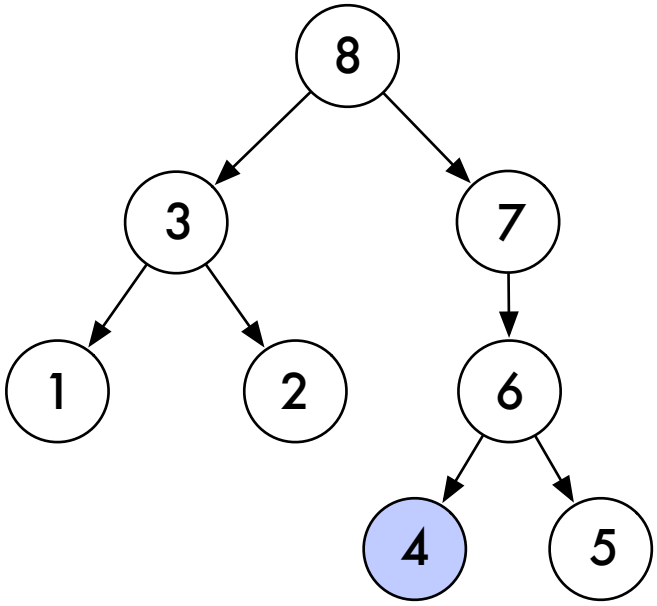
xugrep design

xugrep usage

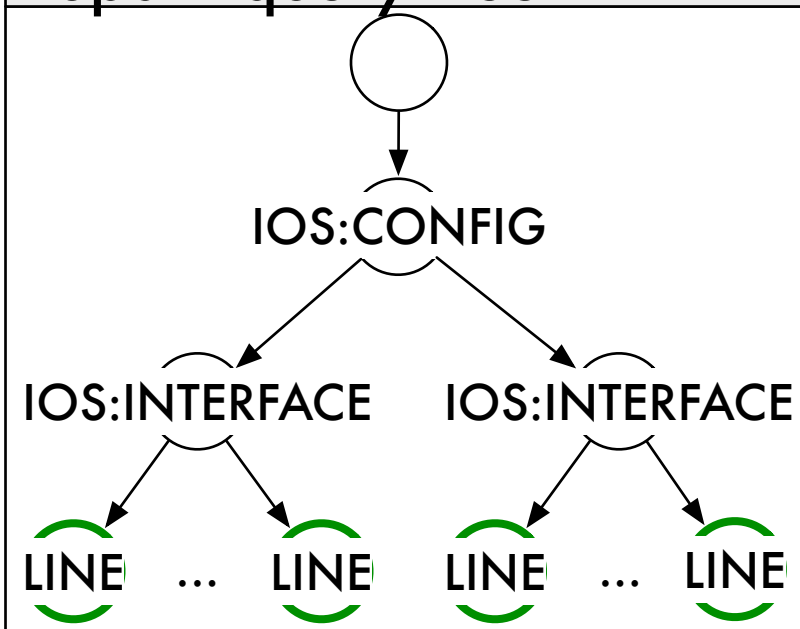
```
xugrep [-1] <xupath> <input_file>+
```

```
xugrep //ios:interface/builtin:line router.v1.ios
```

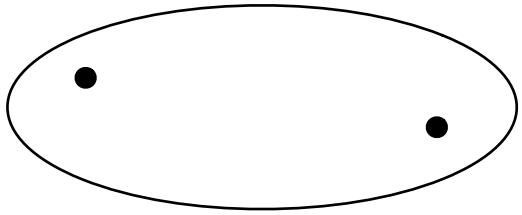
xupath parse tree



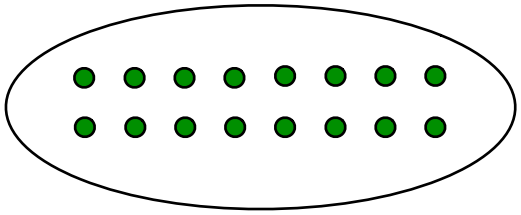
xupath query tree



old corpus



current corpus



xuwc design

Passage	Title	Description
6.2.2.2	Inadequate Security Policy	Security policies must be well practiced and monitored
6.2.2.5	Inadequate Change and Configuration Management	Examples include failing to document changes, misconfigurations
6.2.3.1	Inadequate Periodic Security Audits	Audits should not rely exclusively on interviews with sys admins
6.3.1.10	Logging and Auditing Vulnerability	Log forging/injection (CWE-117)

Passage	Title	Description
7.2.18	Securing and Validating Field Device Settings	Ensure settings remain the same as intended in the config. mgmt. process

CIP Passage	Title	Description
003-4	Change Control and Configuration Management	Must be able to “identify, control, and document” meaningful changes
004-5	Update network documentation	Update documentation within 60 days of a change to the network.
010-1	Baseline Configuration development and comparison	Must be able to compare configurations against baseline configurations.