

Web Application Security

Richard A. Kemmerer

Reliable Software Group

Computer Science Department

University of California

Santa Barbara, CA 93106, USA

<http://www.cs.ucsb.edu/~rsg/>

Why are web applications important?

- Web has become a ubiquitous application delivery medium
- Easy to develop, deploy, and access web applications

Pervasive: deployed by virtually all companies, institutions, and organizations

Critical: access/manage sensitive information (e.g., health records, financial information, personal data)

Open: widely accessible through firewalls

Dynamic: change frequently

Web application attacks

- Web-related security flaws represent a substantial fraction of all reported security flaws
 - Cross-site scripting
 - Buffer overflows
 - SQL injection
 - Command execution
 - Exploitation of weak cookies
 - Sniffing of unencrypted sensitive information
 - Bypassing client-side validation
 - Misuse of hidden form fields

Result: web-based applications are vulnerable and have become a popular attack target

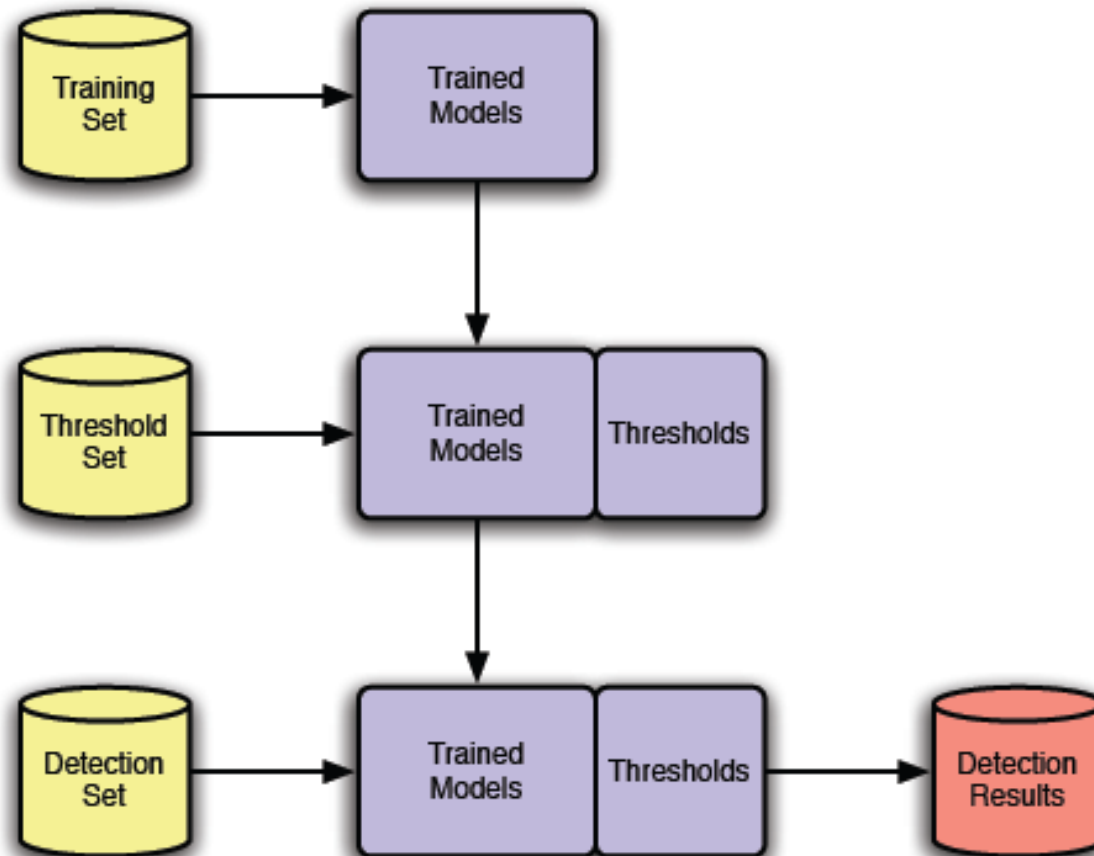
Intrusion detection systems (IDS)

- Examine a time-ordered stream of events from a set of domains
 - Network (e.g., packets, streams)
 - Host (e.g., system calls, audit records)
 - Application (e.g., syslog, HTTP access logs)
- Look for event sequences that represent attacks
 - Misuse detection
 - Use signatures to characterize “bad” events
 - Reliably low false positive rates
 - Poor generalization
 - Anomaly detection
 - Use models to characterize “good” events
 - Traditionally have higher false positive rates
 - Sensitive to novel, unforeseen attacks

Learning-based detection

- Generate models of normal behavior using a history of attack-free samples
 - Models associated with features in a particular domain
 - Models limited in scope to samples observed during training period
- Potential to capture “tighter,” installation-specific behaviors
 - Use limited to subset of software functionality
 - Patterns of user behavior

Anomaly detection phases



Web-based anomaly detection

- Examines web requests sent from clients to server
 - Application to be executed
 - Application parameters (attribute names and values)

Example

`GET /cgi-bin/show.cgi?sID=12345&file=images/foo.png`

- Applies statistical models to each attribute of each application in two phases
 - Learning phase:*
 - Builds profiles of normal behavior for each application parameter
 - Detection phase:*
 - Detects deviations from learned profile

Web-based anomaly detection

- Examines web requests sent from clients to server
 - Application to be executed
 - Application parameters (attribute names and values)

Example

GET /**cgi-bin/show.cgi**?sID=12345&file=images/foo.png

- Applies statistical models to each attribute of each application in two phases
 - Learning phase:*
 - Builds profiles of normal behavior for each application parameter
 - Detection phase:*
 - Detects deviations from learned profile

Web-based anomaly detection

- Examines web requests sent from clients to server
 - Application to be executed
 - Application parameters (**attribute names** and values)

Example

GET /cgi-bin/show.cgi?**sID**=12345&**file**=images/foo.png

- Applies statistical models to each attribute of each application in two phases
 - Learning phase:*
 - Builds profiles of normal behavior for each application parameter
 - Detection phase:*
 - Detects deviations from learned profile

Web-based anomaly detection

- Examines web requests sent from clients to server
 - Application to be executed
 - Application parameters (attribute names and **values**)

Example

GET /cgi-bin/show.cgi?sID=12345&file=images/foo.png

- Applies statistical models to each attribute of each application in two phases
 - Learning phase:*
 - Builds profiles of normal behavior for each application parameter
 - Detection phase:*
 - Detects deviations from learned profile

Detection models used

- Length
- Character distribution
- Structural inference
- Token set
- Frequency
- Interval
- Invocation order

Length model

Observation

Many string lengths are either fixed in size or vary over a small range

- Model attempts to approximate actual (unknown) distribution of string lengths
- Weak bound results in significant tolerance to variations

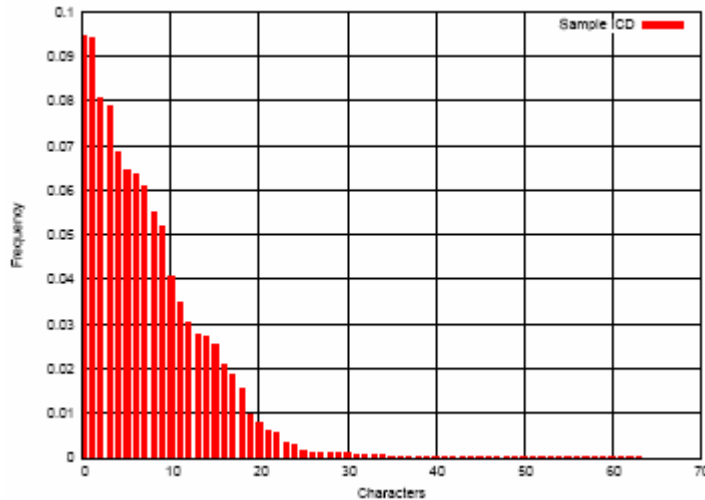
Chebyshev inequality

$$p(l) = p(|x - \mu| > |l - \mu|) = \frac{\sigma^2}{(l - \mu)^2}$$

String character distribution model

Observation

Many strings take values that have similar character distributions

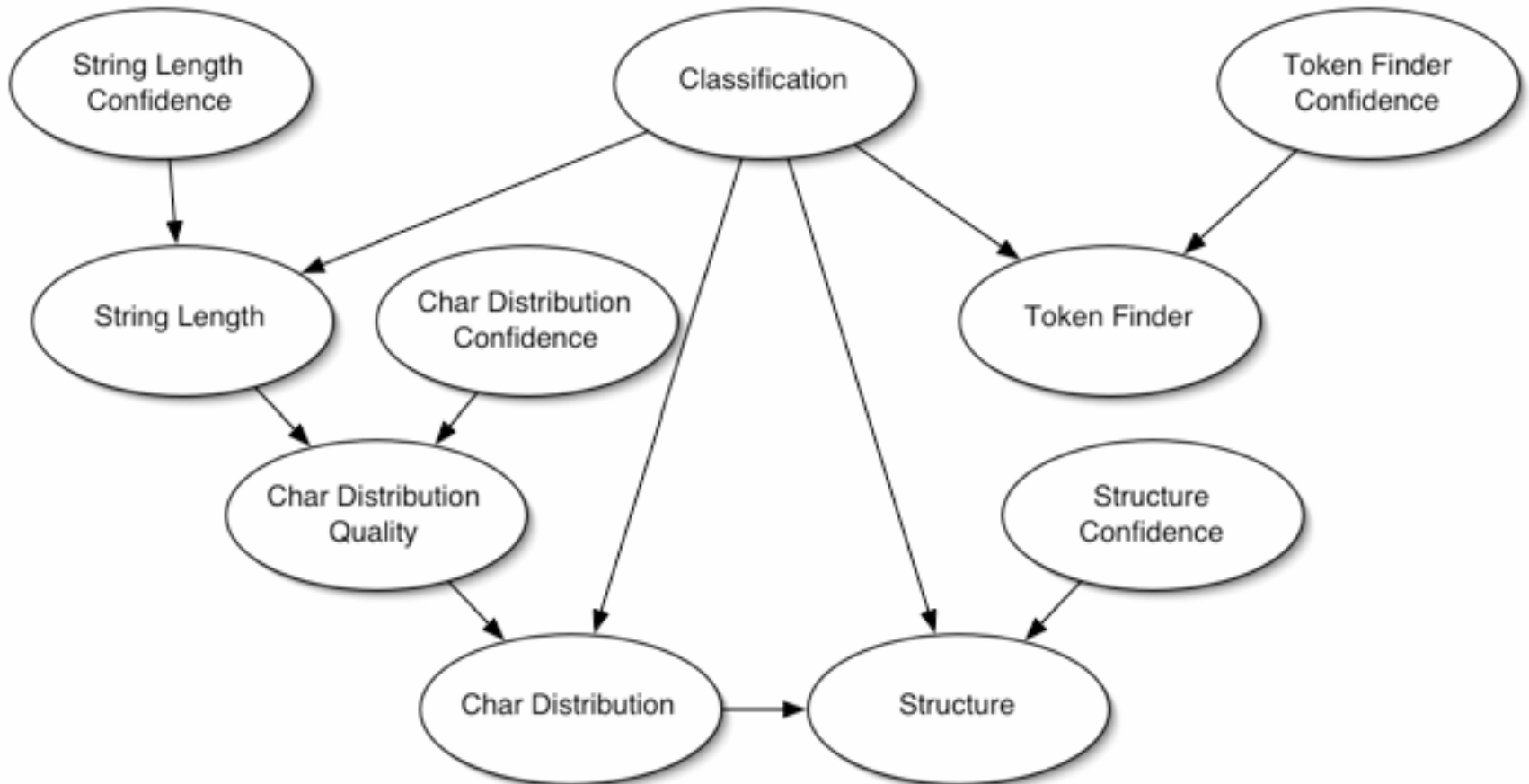


- Model creates idealized character distribution (ICD) for strings observed during the training phase
- Anomaly score calculated using variant of Pearson Chi-squared test

Anomaly score aggregation

- Multiple models per event imply a need for score combination
- The simple approach (weighted summation) cannot represent dependencies between models
 - When one feature is anomalous, another feature may be expected to be anomalous as well
 - An anomalous feature might indicate that the quality of another model output increases
- Dependencies can be represented using Bayesian decision networks

Bayesian network example

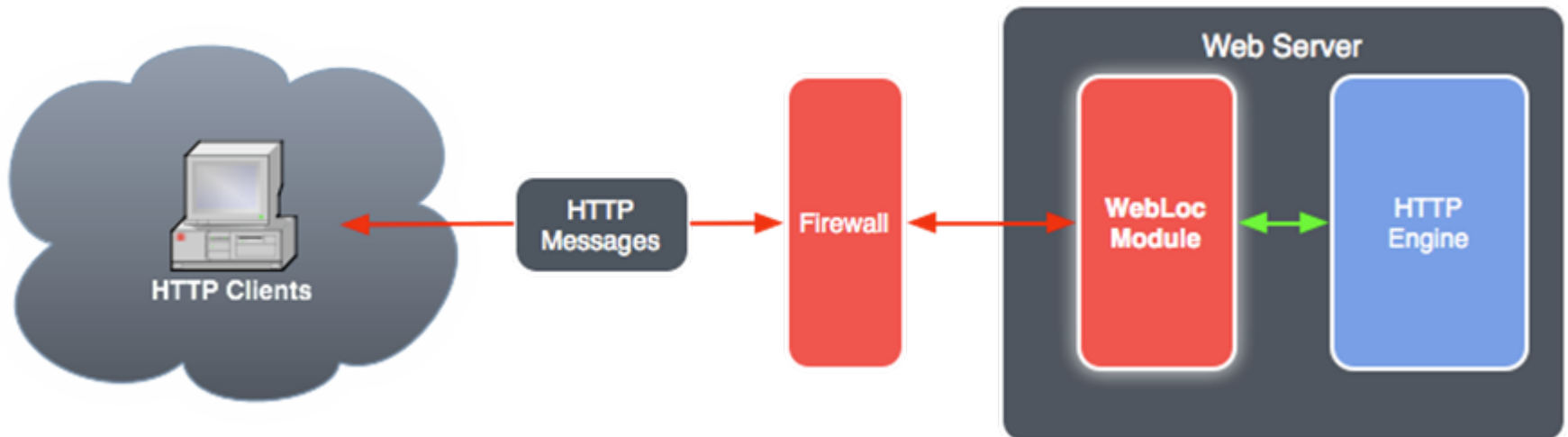


Technology transfer to WebLoc

- Web application security module that ISVs can integrate with their web-based offerings
- Detects and reports/blocks attacks that are
 - Generic
 - Specific to each web application deployment
- Sophisticated site-specific detection algorithms
 - Customizes detection to particular configurations and sites using multi-model learning
 - Responds to changes in the web site and performs necessary retraining
- Aggressive reduction of the impact of false positives
 - Reduces false positives by capturing inter-model dependencies using Bayesian analysis
 - Intelligent reporting using attack aggregation to reduce the cost of false positives
- Requires minimal expertise to deploy, configure, and maintain

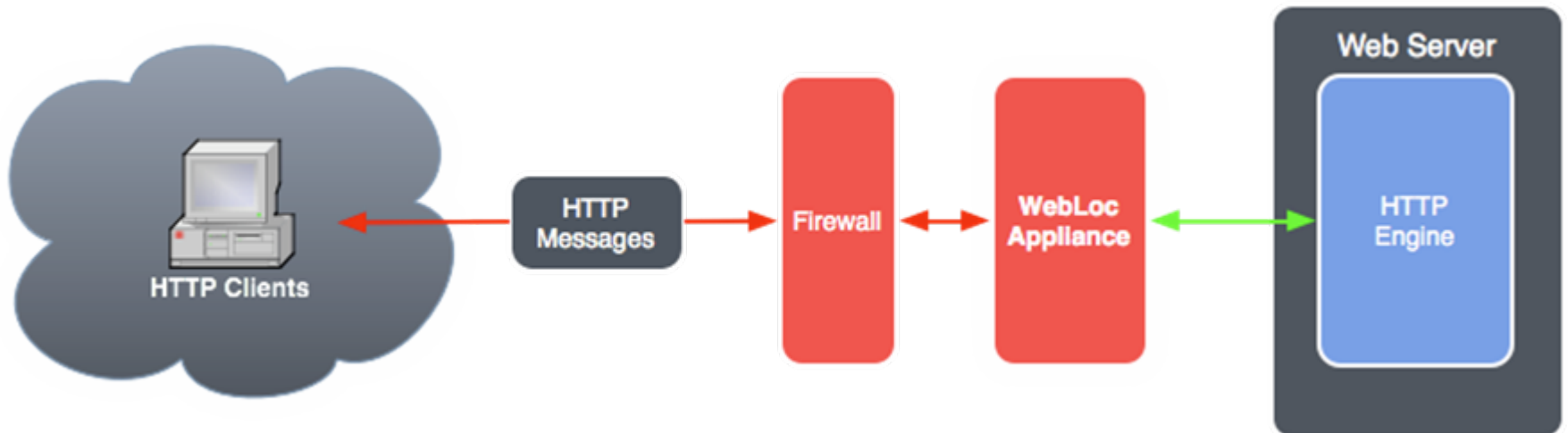
Architecture

- Interposed between clients and web application
- Inspects HTTP requests and responses
- Option to report or block



Architecture

- Interposed between clients and web application
- Inspects HTTP requests and responses
- Option to report or block



Questions?